行政院國家科學委員會專題研究計畫 成果報告

模糊測度 Choquet 積分應用於教育測驗分析之研究(I)
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 97-2410-H-468-014-
執 行 期 間 ：97 年 08 月 01 日至 98 年 07 月 31 日
執 行 單 位 ：亞洲大學生物科技學系

計 畫 主 持 人：劉湘川
共 同 主 持 人：郭伯臣
計畫參與人員： 學士級-專任助理人員：林莞惠

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 98 年 10 月 30 日

行政院國家科學委員會補助研究計畫

模糊測度 Choquet 積分應用於教育測驗分析之研究(I)

成果報告(精簡版)

（2008/08/01~2009/07/31）

計畫類別：　■個別型計畫　□整合型計畫
計畫編號：　NSC 97-2413-H-468-014-
執行期間：　97 年 08 月 01 日 至 98 年 07 月 31 日

計畫主持人：　劉湘川 亞洲大學生物資訊學系暨心理學系
共同主持人：　郭伯臣 國立臺中教育大學教育測驗統計研究所

計畫參與人員：
專任助理人員:
林莞惠　靜宜大學應用數學系統計資訊組
兼任助理人員:
劉育隆　亞洲大學資訊工程學系博士班
杜雨潔　國立台中教育大學教育測驗統計研究所博士班
林士勛　國立台中教育大學教育測驗統計研究所

成果報告類型（依經費核定清單規定繳交）：精簡報告

報告附件： 出席國際學術會議心得報告及發表論文各一份

處理方式：本研究計畫涉及專利或其他智慧財產權，兩年後可公開查詢

執行單位：亞洲大學

中　華　民　國　九 十 八 年 七 月 三十一 日

壹、前言

現有之教育測驗理論，包含試題反應理論，均建構在可加性機率測度及魯貝格積分(Lebesgue Integral)理論基礎上，機率測度理論是一種古典測度理論，其基本理論規定任何隨機事件之機率為其機率密度之可加測度，這樣的規定最大的好處在於計算簡便，但「機率之可加性運算」顯然必須滿足；「不同機率密度間無交互作用之基本假設」，然而在諸多實際應用上並不能完全適用，因而有其他不同之非可加性測度應運而生，例如可能性測度(Possibility measure)、似真性測度(Plausibility measure) 、信任性測度(belief measure)、必然性測度(Necessity measure)等，事實上，上述四種非可加性測度及眾所周知之機率測度都是單調性測度之特例，由於單調性測度之單調性條件甚多，同時確定並不容易，有其不明確性，故菅野道夫(Sugeno,)於 1974 年在提出λ測度之同時，首先將單調性測度稱為模糊測度，繼而，依循菅野道夫之說法，王正元與喬治．克里爾(Zhenyuan Wang & George J. Klir) 於 1992 年出版第一本有關單調性測度之書籍，將其命名為「模糊測度理論(Fuzzy Measure Theory)」，該模糊測度理論是古典測度理論之推廣理論，然而單調性測度發展之初，只討論明確數而非模糊數，事實上不宜稱為模糊測度，特別是目前單調性測度之發展，已由明確數擴張至模糊數了，則關於模糊數之單調性測度可稱為模糊化單調性測度(Fuzzified monotone measure)，若將單調性測度仍稱為模糊測度，則關於模糊數之模糊測度就有模糊不清之議，加之非單調性測度也已被學界引進，故而王正元與喬治．克里爾於 2009 年出版之模糊測度理論擴充版已更名為「廣義測度理論(Generalized Measure Theory)」，由於目前學界所熟悉且容易溝通之名詞，仍稱之為模糊測度，故本研究計畫之單調性測度亦稱為模糊測度，另外單調性測度必須配合單調性積分才能竟其工，換言之、魯貝格積分必須相應擴張為單調性積分。此外，對應於單調性測度稱為模糊測度，則單調性積分也常被稱為模糊積分。第一個提出改進魯貝格積分之單調性積分者，應是義大利數學家 魏塔利(Giusseppe Vitali, 1925,1997), 他於 1925 以義大利文發表，延遲至 1997 年才被翻譯成英文方為人知，其後被法國數學家薛奎爾(Gustave Choquet)重新發現於 1954 年再度提出，經二十餘年之澎勃發展，學界已習慣稱之為 Choquet 積分，故本研究計畫亦稱為 Choquet 積分。雖然菅野道夫(Sugeno,)於 1974 年也提出新的模糊積分，稱為 Sugeno 積分，與其提出之λ測度廣為工程、管理等學界應用，但 Sugeno 積分既不及 Choquet 積分之靈敏，且非魯貝格積分之推廣積分方法，故暫時未列入本研究計畫之內容。Sugeno(1974)將模糊測度分為次可加測度，可加測度、及超可加測度，三種，主要因為其所提出之λ測度隨λ值而異，只有該三種可能，而目前學界也以為模糊測度只有該三種分類，本人於 2006 年首先指出，實務所須，單調性混合模糊測度是不可忽略的，另指出 Sugeno 之λ測度與 Zadeh(1978)之可能性測度均為單值模糊測度，適用性有限，有必要發展具有上述四種類別之多值模糊測度，並於 2007 年起，陸續提出系列具上述改良性值之多值模糊測度族。除可提供工程、管理等學界應用外，並希望能兼顧理論與應用之進一步發展，且能轉化應用於教育測驗領域。

貳、研究目的與方法

本研究計畫為三年期研究計畫之第一年計畫；「模糊測度 Choquet 積分應用於教育測驗分析之研究(I)」本年度計畫主要在探討「應用作者所提供之多種新模糊測度 Choquet 積分法來建立更具預測效力之教育測驗預測模式，並發展應用系統程式」

其中包含：

一、多值模糊測度數學理論之探究：

1. 作者提出 L 測度，並探究其基本數理性質
2. 作者提出改進之完備 L 測度，並探究其基本數理性質
3. 作者提出新測度 δ 測度，並探究其基本數理性質
4. 作者提出基於 L 測度與 δ 測度之組合多值模糊測度，並探究其基本數理性質
5. 作者等提出分組資料之 C 測度，並探究其基本性質
6. 作者提出 γ 模糊密度(γ 支撐)，並驗證其優於 C 支撐與 V 支撐

二、電腦分析系統程式設計

1. 基於 γ 支撐之 L 測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計
2. 基於 γ 支撐之完備 L 測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計
3. 基於 γ 支撐之 δ 測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計
4. 基於 γ 支撐之 L(δ)組合多值模糊測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計
5. 基於 γ 支撐之 C 測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計

三、兩組教育測驗評量預測實證資料

1.苗栗某中學 60 位學生以其國中數學、理化，生物，及地球科學之畢業成績預測其國中基本能力測驗之自然科成績

2. 臺中某國民小學 128 位學生之上臂三頭肌、上臂二頭肌、肩胛下、腸棘上等四處的皮脂厚度推估出來的體脂肪率預測體脂肪計體脂肪率。

四、以預測量均方誤差為比較準則，進行上述資料之各種預測模式之 k 折交互驗證法

(K-Folds Cross Validation Method) 比較研究

預測模式列示於下：

1. 複迴歸預測模式
2. 脊迴歸預測模式

3. 基於 γ 支撐之 Zadeh P 測度 Choquet 積分迴歸預測模式

4. 基於 γ 支撐之 Sugeno λ 測度 Choquet 積分迴歸預測模式

5. 基於 γ 支撐之 L 測度 Choquet 積分迴歸預測模式

6. 基於 γ 支撐之完備 L 測度 Choquet 積分迴歸預測模式

7. 基於 γ 支撐之 δ 測度 Choquet 積分迴歸預測模式

8. 基於 γ 支撐之 L(δ)組合多值模糊測度 Choquet 積分迴歸預測模式

9. 基於 γ 支撐之 C 測度 Choquet 積分迴歸預測模式

叁、研究成果與發表論文目錄

一、. 提出 L 測度之重要數理性質及其 Choquet 積分迴歸預測模式

(一) 提出 L 測度之重要數理性質如下：

1. L 測度滿足有界性能與單調性是以模糊測度

2. L 測度是決定係數 L 在定義域 $[0,\infty)$ 上之連續遞增函數

3. L 測度為多值模糊測度，$L\in[0,\infty)$，不同之決定係數 L 值決定了不同之模糊測度，換言之，L 測度有無限多模糊測度解，且其公式解具封閉型式

4. L=0 時，L 測度恰好為 Zadeh 之 P 測度

5. L 測度可為混合模糊測度、次可加模糊測度，及超可加模糊測度

(二)完成基於 γ 支撐之 L 測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計

(三) 發表 EI 級論文兩篇如下驗證了 L 測度 Choquet 積分迴歸預測模式優於複迴歸預測模式、脊迴歸預測模式、基於 γ 支撐之 Zadeh P 測度 Choquet 積分迴歸預測模式、及基於 γ 支撐之 Sugeno 之 λ 測度 Choquet 積分迴歸預測模式。

(見附漸次出席國際會議發表論文及心得報告)

1. .Hsiang-Chuan Liu, Yu-Chieh Tu, Wen-Chih Lin, and Chin-Chun Chen (2008). Choquet integral regression model based on L-Measure and γ-Support. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* (Hong Kong, 30-31, Aug. 2008.) Volume: 2, pp.777-782. ISBN: 978-1-4244-2238-8. **(EI paper)**

2. Hsiang-Chuan Liu, Yu-Du Jheng, Guey-Shya Chen and Bai-Cheng Jeng. (2008)

Choquet Integral Logistic Regression Algorithms Based on L-Measure and γ-Support. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* (Hong Kong, 30-31, Aug. 2008.) .Volume: 2, pp.771-776. ISBN: 978-1-4244-2238-8. INSPEC Accession Number: 10299006. **(EI paper)**

二、. 提出完備 L 測度之重要數理性質及其 Choquet 積分迴歸預測模式

(一) 提出完備 L 測度之重要數理性質如下：

1. 在既定之模糊密度條件下，提出最大模糊測度；B 測度，及完備測度定義，並指出 Sugeno 之 λ 測度、Zadeh 之 P 測度、及 L 測度均未包含 B 測度，換言之，λ 測度、 P 測度、及 L 測度均非完備測度。
2. 完備 L 測度滿足有界性能與單調性是以模糊測度
3. 完備 L 測度是決定係數 L 在定義域 $[0,\infty)$ 上之連續遞增函數
4. 完備 L 測度為多值模糊測度，$L\in[0,\infty)$，不同之決定係數 L 值決定了不同之模糊測度，換言之，L 測度有無限多模糊測度解，且其公式解具封閉型式
5. 完備 L=0 時，完備 L 測度恰好為最小測度；Zadeh 之 P 測度
6. $L\to\infty$ 時，完備 L 測度恰好為最大測度；B 測度
7. 完備 L 測度可為混合模糊測度、次可加模糊測度，及超可加模糊測度

(二)完成基於 γ 支撐之完備 L 測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計

(三) 發表 EI 級論文同時被刊登於專書如下，驗證了完備 L 測度 Choquet 積分迴歸預測模式優於複迴歸預測模式、脊迴歸預測模式、Zadeh P 測度 Choquet 積分迴歸預測模式、Sugeno 之 λ 測度 Choquet 積分迴歸預測模式、及 L 測度 Choquet 積分迴歸預測模式。

1. Hsiang-Chuan Liu, “A theoretical approach to the completed L-fuzzy measure”, Conference Proceedings of 2009 International Institute of Applied Statistics Studies (2009IIASS), July 24-28 2009.Qindao, China, pp. 1121-1124, 2009. ISBN:978-0-9806057-4-7. (EI paper)

2. Hsiang-Chuan Liu (2009). “A theoretical approach to the completed L-fuzzy measure”, *Quantitative Analysis Techology and Related Engineering Applications*, pp. 1121-1124, 2009, AUSSINO ACADEMIC PUBLISH HOUSE Sydney Australia, ZHU Koulai & Henry ZHANG, ISBN:978-0-9806057-4-7.

三、. 提出 δ測度之重要數理性質及其 Choquet 積分迴歸預測模式

(一) 提出δ測度之重要數理性質如下：

1. δ測度滿足有界性能與單調性是以模糊測度

2. δ測度是決定係數δ在定義域$[-1,1]$上之連續遞增函數

3. δ測度為多值模糊測度，$\delta \in [-1,1]$，不同之決定係數δ值決定了不同之模糊測度，換言之，δ測度有無限多模糊測度解，且其公式解具封閉型式

4. δ=-1 時，δ測度恰好為 Zadeh 之 P 測度

5. δ=0 時，δ測度恰好為可加測度，當模糊密度之和為 1 時，Sugeno 之λ測度即可加測度，此時δ測度恰好亦為λ測度，並指出 L 測度及完備 L 測度均未包含可加測度。

6. $-1 \le \delta < 0$ 時，δ測度為次可加測度，$0 < \delta \le 1$ 時，δ測度為超可加測度

7. δ測度不可能為混合模糊測度及完備測度。

(二)完成基於γ支撐之δ測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計

(三) 發表 EI 級期刊論文如下，驗證了δ測度 Choquet 積分迴歸預測模式優於複迴歸預測模式、脊迴歸預測模式、Zadeh P 測度 Choquet 積分迴歸預測模式、及 Sugeno 之λ測度 Choquet 積分迴歸預測模式。

Hsiang-Chuan Liu, Der-Bang Wu, Yu-Du Jheng and Tian-Wei Sheu (2009). “Theory of Multivalent Delta-Fuzzy Measures and its Application”, WSEAS TRANSACTION ON INTERNATIONAL SCIENCE AND APPLICATION ,Vol. 6, No. 6 1061-1070, June 2009. ISSN: 1790-0832. (EI Journal)

四、. 提出基於 L 測度與δ測度之組合多值模糊測度之重要數理性質及其 Choquet 積分迴歸預測模式

(一) 提出基於 L 測度與δ測度之組合多值模糊測度；L(δ)測度之重要數理性質如下：

1. L(δ)測度滿足有界性能與單調性是以模糊測度

2. L(δ)測度是決定係數 L 在定義域$[-1,\infty)$上之連續遞增函數

3. L(δ)測度為多值模糊測度，$L \in [-1,\infty)$，不同之決定係數 L 值決定了不同之模糊測度，換言之，L(δ)測度有無限多模糊測度解，且其公式解具封閉型

式，並指示δ測度之無限多模糊測度解，遠少於 L 測度，及 L(δ)測度之無限多模糊測度解，而 L(δ)測度之無限多模糊測度解亦多於與 L 測度之無限多模糊測度解

4. L=-1 時，L(δ)測度恰好為 Zadeh 之 P 測度
5. L=0 時，L(δ)測度恰好為可加測度，當模糊密度之和為 1 時，Sugeno 之λ測度即可加測度，此時 L(δ)測度恰好亦為λ測度。
6. $-1 \le L < 0$ 時，δ測度為次可加測度，$0 < L < \infty$ 時，L(δ)測度為超可加測度
7. L(δ)測度不可能為混合模糊測度及完備測度。

(二)完成基於γ支撐之 L(δ)測度 Choquet 積分迴歸預測模式之電腦分析系統程式設計

(三) 發表 EI 級期刊論文如下，驗證了 L(δ)測度 Choquet 積分迴歸預測模式優於複迴歸預測模式、脊迴歸預測模式、Zadeh P 測度 Choquet 積分迴歸預測模式、Sugeno 之λ測度 Choquet 積分迴歸預測模式、L 測度 Choquet 積分迴歸預測模式及δ測度 Choquet 積分迴歸預測模式。

Hsiang-Chuan Liu, Chin-Chun Chen, Der-Bang Wu, and Tian-Wei Sheu (2009). "Theory and Application of the Composed Fuzzy Measure of L-Measure and Delta-Measures", WSEAS TRANSACTION ON INTERNATIONAL SCIENCE AND CONTRAL , Issue 8. Vol. 4, pp. 359-368, Augest 2009. ISSN: 1991-8763. (EI Journal)

五、. 提出基於 C 測度與δ測度之組合多值模糊測度之重要數理性質及其 Choquet 積分迴歸預測模式

(一) 提出基於 C 測度之重要性質如下：

1. 基於複雜度之 C 測度滿足有界性能與單調性是以模糊測度
2. C 測度適合於分組資料之模糊測度度即可加測度，此時 L(δ)測度恰好亦為λ測度。

(二)完成了 C 測度 Choquet 積分預測模式之電腦分析系統程式設計

(三) 發表 SCI 級期刊論文如下，驗證了 C 測度 Choquet 積分預測模式優於複迴歸預測模式、脊迴歸預測模式、Zadeh P 測度 Choquet 積分迴歸預測模式、Sugeno 之λ測度 Choquet 積分迴歸預測模式、

Jiunn-I Shieh, Hsin-Hong Wu, Hsiang-Chuan Liu., (2009). Applying complexity-based Choquet integral to evaluate students' performance. *Expert Systems with Applications, 36* (2009), 5100-5106. **ISSN:** 0957-4174. (SCI, impact factor; 2.596)

六、 提出基於 Pearson 相關係數之γ模糊密度(γ支撐)，並驗證其優於 C 支撐與 V 支撐發表論文(同一之論文) 如下

1. .Hsiang-Chuan Liu, Yu-Chieh Tu, Wen-Chih Lin, and Chin-Chun Chen (2008). Choquet integral regression model based on L-Measure and γ-Support. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* (Hong Kong, 30-31, Aug. 2008.) Volume: 2, pp.777-782. ISBN: 978-1-4244-2238-8. **(EI paper**)
2. Hsiang-Chuan Liu, Yu-Du Jheng, Guey-Shya Chen and Bai-Cheng Jeng. (2008) Choquet Integral Logistic Regression Algorithms Based on L-Measure and γ-Support. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* (Hong Kong, 30-31, Aug. 2008.) .Volume: 2, pp.771-776. ISBN: 978-1-4244-2238-8. INSPEC Accession Number: 10299006. **(EI paper**)

四 、結論

本研究計畫第一年度經數理分析之探討，提出一種有效之模糊密度；γ支撐，四種改善之多值模糊測度及一種分組資料可用之模糊測度，同時完成了基於γ支撐之上述各種模糊測度之 Choquet 積分迴歸模式，包含複迴歸預測模式及脊迴歸預測模式之電腦分析系統程式設計、進行兩組教育測驗資料之五折交互驗證比較研究，各種模糊測度之 Choquet 積分迴歸模式，均獲得有效之成果，並發表了 1 篇 SCI 期刊論文，2 篇 EI 期刊論文，及 3 篇 EI 研討會論文。

五、附錄：發表論文

Expert Systems with Applications 36 (2009) 5100–5106

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Applying a complexity-based Choquet integral to evaluate students' performance

Jiunn-I Shieh [a,*], Hsin-Hung Wu [b], Hsiang-Chuan Liu [c]

[a] Department of Information Science and Applications, Asia University, No. 500 Lioufeng Road, Wufeng Shiang, Taichung 413, Taiwan
[b] Department of Business Administration, National Changhua University of Education, Taiwan
[c] Institute of Bioinformatics, Asia University, Taiwan

ARTICLE INFO

Keywords:
Fuzzy measure
Discrete Choquet integral
Entropy
Complexity

ABSTRACT

The weighted arithmetic mean and the regression methods are the most often used operators to aggregate criteria in decision making problems with the assumption that there are no interactions among criteria. When interactions among criteria exist, the discrete Choquet integral is proved to be an adequate aggregation operator by further taking into accounts the interactions. In this study, we propose a complexity-based method to construct fuzzy measures needed by the discrete Choquet integral and a real data set is analyzed. The advantage of the complexity-based method is that no population probability is to be estimated such that the error of estimating the population probability is reduced. Four methods, including weighted arithmetic method, regression-based method, the discrete Choquet integral with the entropy-based method, and our proposed discrete Choquet integral with the complexity-based method, are used in this study to evaluate the students' performance based on a Basic Competence Test. The results show that the students' overall performance evaluated by our proposed discrete Choquet integral with the complexity-based method is the best among the four methods when the interactions among criteria exist.

1. Introduction

The most often used operator to aggregate criteria in decision making problems is the classical weighted arithmetic mean (Fishburn, 1970). In many practical applications the decision criteria present some interaction. However, the problem of modeling such an interaction remains a difficult question, which is often overlooked (Domingo & Torra, 2002). The reason is that practitioners are lack of suitable tools to deal with the interactions such that criteria are assumed to be independent and exhaustive. This comes primarily from the absence of a precise definition of interactions as well as the complexity and difficulty of identifying the interaction phenomena among criteria. It is known that the mutual independence among the criteria is a necessary condition for aggregation operator to be additive. That is, if some criteria are preferentially dependent with the others, then no additive aggregation operator can model the preferences of the decision maker (Domingo & Torra, 2002).

The weighted arithmetic mean and regression method are unable to overcome the undesirable phenomenon of dependence. In contrast, the Choquet integral takes into account the interactions among criteria. In addition, there is a key issue unsolved in the application of fuzzy integral with the determination of density values to decide the fuzzy measures in the fusion process. In this study, entropy-based method and our proposed complexity-based method to construct the fuzzy measures in the discrete Choquet integral are discussed.

This paper is outlined as follows: Section 2 reviews weighting methods, fuzzy measures, and discrete Choquet integrals with two different constructs in fuzzy measures. A procedure of using Choquet integral is provided in Section 3. A case study of applying the weighted arithmetic mean method, regression method, Choquet integral with the entropy-based method, and our proposed Choquet integral with the complexity-based method is performed in Section 4 to analyze the students' overall performance on Basic Competence Test when the interactions exist. Finally, conclusions are summarized in Section 5.

2. Weighting methods, fuzzy measures, and discrete Choquet integral

The classical weighted arithmetic mean method is the most commonly used operator to aggregate criteria in decision making problems without further considering the interactions among criteria. The regression method is to maximize the linear relation among the criteria without further taking into considering the interactions among criteria. On the contrary, the discrete Choquet integral is proved to be an adequate aggregation operator that extends the weighted arithmetic mean method by taking into

* Corresponding author. Tel.: +886 4 2332 3456x1832; fax: +886 4 2331 6699.
E-mail address: jishieh@yahoo.com.tw (Jiunn-I Shieh).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.06.003

consideration the interactions among criteria. The philosophy of the Choquet integral was first introduced in capacity theory (Choquet, 1953) and used as a (fuzzy) integral with respect to a fuzzy measure proposed by Höhle (1982) and then rediscovered later by Murofushi and Sugeno (1989, 1991).

Choquet integral is defined to integrate functions with respect to the fuzzy measures (Murofushi & Sugeno, 1989). Fuzzy integrals are very useful for global evaluation models but the number of parameters of fuzzy measures is large. The definitions of fuzzy measures and Choquet integrals are as follows (Murofushi & Sugeno, 1989):

Definition 1. Let N be a finite set of criteria. A discrete fuzzy measure on N is a set function v: $2^N \to [0,1]$ which satisfies the following axioms:

(i) $v(\phi) = 0$, $v(N) = 1$ (boundary conditions);
(ii) $A \subseteq B$ implies $v(A) \leqslant v(B)$ (monotonicity) for $A, B \in 2N$.

For each subset of criteria $S \subseteq N$, $v(S)$ can be interpreted as the weight of the coalition S.

Definition 2. Let v be a fuzzy measure on $N = \{1, 2, \ldots, n\}$. The discrete Choquet integral of function x: $N \to R$ with respect to v is defined by $C_v(x) = \sum_{i=1}^{n} x_{(i)}[v(A_{(i)}) - v(A_{(i+1)})]$, where $(\cdot)$ indicates a permutation on N such that $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$. Also $A_{(i)} = \{(i), \ldots, (n)\}$, and $A_{(n+1)} = \phi$. For instance, if $x_1 \leqslant x_3 \leqslant x_2$, then rank x_1, x_2, x_3 from the smallest one to the largest one. The result is $x_{(1)} = x_1$, $x_{(2)} = x_3$, $x_{(3)} = x_2$. Finally, $C_v(x_1, x_2, x_3) = x_1 * [v(\{(1), (2), (3)\})] + (x_3 - x_1) * [v\{(2), (3)\}] + (x_2 - x_3) * [v(\{(3)\})]$.

The discrete Choquet integral takes into account the interaction by means of the fuzzy measure v. If the criteria are independent, the fuzzy measure is additive. Then, the discrete Choquet integral coincide with the weighted arithmetic mean method. That is, $C_v(x) = \sum_{i=1}^{n} v(\{i\}) * x_i$, $x \in R^n$. For example, there are five students and three courses (D_1, D_2, and D_3). Assume the raw data and a fuzzy measure v on each subset are in Tables 1 and 2, respectively. In Table 2, (0,0,0), (1,0,0), (0,1,0), (1,1,0), (0,0,1), (1,0,1), (0,1,1), and (1,1,1) represent empty set, $\{D_1\}$, $\{D_2\}$, $\{D_1,D_2\}$, $\{D_3\}$, $\{D_1,D_3\}$, $\{D_2,D_3\}$, and $\{D_1,D_2,D_3\}$, respectively. For the first student, the raw scores are 70, 81, and 75. First, rank the scores from the smallest to the largest, i.e., 70, 75, and 81. Then, the overall performance evaluated by Choquet integral is computed by $70 * v(\{D_1,D_2,D_3\}) + (75 - 70) * v(\{D_2,D_3\}) + (81 - 75) * v(\{D_2\}) = 70 * 1 + 5 * 0.5 + 6 * 0.1667 = 73.5002$. By the same philosophy, the overall performance values of the second, third, fourth, and fifth students evaluated by Choquet integral are 77.6667, 74.6667, 81.0002, and 77.8334, respectively.

To evaluate a discrete Choquet integral, we need a fuzzy measure first. How to find a suitable fuzzy measure becomes an issue. To be a fuzzy measure, the measure needs to satisfy the axioms of the fuzzy measure. We note that entropy measure and complexity-based measure are qualified to be fuzzy measures. The former one is proposed by Kojadinovic (2004) and the latter one is proposed in our study.

To measure the uncertainty of a random variable, the concept of entropy was introduced (Shannon, 1948). The basic idea is that an item with large entropy in its ratings is more important in a user's interest than an item with small entropy. Based on this idea, an entropy-based method is in the following (Yu, Wen, Xu, & Ester, 2001): Given a discrete random variable A, let p^A be the probability of A, then define entropy of A to be $h(A) = -\sum p^A \log_2 p^A$, where $p^A > 0$. With the similar formula, let B be a discrete random vector which contains at least two discrete random variables, then generalize this idea to a random vector and call p^B be the joint probability and $h(B)$ the joint entropy. By using the idea of joint entropy to calculate the entropy of the subsets of criteria of N, define the fuzzy measure v_1 as the following: $v_1(S) = \frac{h(S)}{h(N)}$ for all $S \subseteq N$ (Kojadinovic, 2004). By using the idea of entropy, we need to decide the number of level to be used to classify the raw data into the level of the score for each criterion. For example, let the number of level to be used be 2 and S contain only two random variables X_1 and X_2. In additional, assume the raw data are in Table 3.

The raw data in Table 3 can be classified into Table 4 by histogram equalization of "hist.m" program of Matlab 7.0 for each random variable. To generate the complete information of fuzzy measure v_1, first to compute $h(N)$. A joint pattern (1,2) means that $X_1 = 1$ and $X_2 = 2$, and (2,2) means that $X_1 = 2$ and $X_2 = 2$. There are 3 of pattern (1,2) and 2 of pattern (2,2). Thus, $p^S(X_1 = 1, X_2 = 2) = 3/5 = 0.6$, and $p^S(X_1 = 2, X_2 = 2) = 2/5 = 0.4$. Therefore, $h(N) = -0.6 * \log_2(0.6) - 0.4 * \log_2(0.4) = 0.9710$. Next, $h(S)$ is computed when $S = X_1$ and X_2, i.e., $h(X_1)$ and $h(X_2)$. In this case, there are 3 pattern "1" and 2 pattern "2" in X_1. From Table 4, $p^{X_1}(X_1 = 1) = 3/5 = 0.6$, $p^{X_1}(X_1 = 2) = 2/5 = 0.4$, and $h(X_1) = -0.6 * \log_2(0.6) - 0.4 * \log_2(0.4) = 0.9710$. In contrast to X_1, there are 5 pattern "2" in X_2. From Table 4, $p^{X_2}(X_2 = 1) = 0/5 = 0$, $p^{X_2}(X_2 = 2) = 5/5 = 1$,

Table 1
Example of the raw data used to demonstrate computation of the overall performance by Choquet integral

Student	D_1	D_2	D_3
1	70	81	75
2	70	85	86
3	65	85	84
4	75	91	85
5	75	80	82

Table 2
A fuzzy measure used to demonstrate computation of the overall performance by Choquet integral

D_1	D_2	D_3	Fuzzy measure v
0	0	0	0
1	0	0	0.1667
0	1	0	0.1667
1	1	0	0.5
0	0	1	0.1667
1	0	1	0.5
0	1	1	0.5
1	1	1	1

Table 3
Example of the raw data used to construct fuzzy measures based on entropy and complexity methods

Student	X_1	X_2
1	70	81
2	70	85
3	65	85
4	75	91
5	75	80

Table 4
The level of the score for each criterion classified from the raw data in Table 3 when the number of level is two

Student	X_1	X_2
1	1	2
2	1	2
3	1	2
4	2	2
5	2	2

and $h(X_2) = -1 * \log_2(1) = 0$. By $v_1(S) = \frac{h(S)}{h(N)}$ for all $S \subseteq N$, the fuzzy measure v_1 is completely defined by the following Table 5. Although our example is to compute the fuzzy measure of a random vector with two discrete random variables, the entropy method is also easy to compute the fuzzy measure of a random vector with more than two discrete random variables. However, the entropy-based weighting scheme might take the risk to estimate the probability for each criterion. If the sample size is small, it often makes a larger error to estimate the population probability. Under such circumstances, we propose a complexity method to improve the prediction.

The basic concept of complexity is that the more substructures in a system, the more complex the system. This concept is in agreement with our intuitive understanding that it is the connectedness of the system elements that matters more. Thus, the more connected the system, the higher the number of substructures in it. Then, it is a good reason to count how many substructures in a structure (Bonchev & Rouvray, 2003). The complexity C of a discrete random variable X is defined to be the function which counts the number of distinct patterns in X. The complexity C of n discrete random variables $X_1, X_2, \ldots, X_n$ is defined as the function which counts the number of distinct patterns in joint pattern of $X_1,X_2,\ldots,X_n$. For a finite number of random variables $X_1,X_2,\ldots,X_n$, the complexity is finite. Thus, $C(X_1,X_2,\ldots,X_n)$ always can be normalized to be 1. Moreover, it is very natural to defined $C(\phi)$ to be zero, where ϕ is an empty set. By using the idea of complexity C to calculate the complexity of the subsets of criteria of N, define C_1 as the following: $C_1(S) = \frac{C(S)}{C(N)}$ for all $S \subseteq N$. It is easy to check that C_1 has property of monotonicity. That is, $X \subseteq Y$ implies $C_1(X) \leqslant C_1(Y)$ for $X, Y \in 2^N$. In addition, $C_1(\phi) = 0$. By the definition of fuzzy measure, C_1 is a fuzzy measure.

Let the number of level to be 2 and S contain only two random variables X_1 and X_2. By using the raw data from Table 3, the raw data can be classified by histogram equalization of "hist.m" program of Matlab 7.0 for each random variable, as shown in Table 4. To generate the complete information of fuzzy measure v_1, compute $C(N)$. From Table 4, there are two different joint patterns, i.e., (1,2) and (2,2). Thus, the complexity of N is 2. Next, $C(S)$ is computed when $S = X_1$ and X_2. That is, compute $C(X_1)$ and $C(X_2)$. There are two different patterns in X_1. Then, $C(X_1) = 2$. Moreover, there are only 1 pattern in X_2, i.e., $C(X_2) = 1$. By $C_1(S) = \frac{C(S)}{C(N)}$ for all $S \subseteq N$, the fuzzy measure C_1 is completely defined by the following Table 6. Although our example is to compute the fuzzy measure of a random vector with two discrete random variables, the complexity method is also quite easy to compute the fuzzy measure of a random vector with more than two discrete random variables.

Table 5
A fuzzy measure constructed by the entropy method

X_1	X_2	Fuzzy measure v_1
0	0	0/0.9710 = 0
1	0	0.9710/0.9170 = 1
0	1	0/0.9710 = 0
1	1	0.9710/0.9710 = 1

Table 6
A fuzzy measure constructed by the complexity method

X_1	X_2	Fuzzy measure C_1
0	0	0/1 = 0
1	0	2/2 = 1
0	1	1/2 = 0.5
1	1	2/2 = 1

In this study, four methods, including classical weighted arithmetic mean method, regression-based method, the Choquet integral with the entropy method and our proposed Choquet integral with the complexity-based method, are applied in a case study of a Basic Competence Test to evaluate the students' performance.

3. A procedure of using the discrete Choquet integral

A five-step procedure of applying the Choquet integral based on Calvo, Kolesarova, Komornikova, and Mesiar (2001) is as follows:

Step 1. Decide the range of level to be used to classify the raw data into the level of the score for each criterion in our study by Scott's rule and Sturge's formula. Assume that m is the number of the level of scores and $m = 2, 3, 4, 5, 6, 7, 8, 9$ are the range in our study. Then, transform the scores of the raw data into the level of the scores for each item when $m = 2, 3, 4, 5, 6, 7, 8, 9$.
Step 2: Check the mutual interaction and the strength of interaction among criteria. First, calculate the Chi-square divergence between a pair of criteria, and use statistical test to determine if there is any mutual interaction among the criteria for each $m = 2, 3, 4, 5, 6, 7, 8, 9$. For the analysis of correlation, we chose Cramer's coefficients to determine if there is strong mutual interaction among criteria. Compute Cramer's coefficients for each $m = 2, 3, 4, 5, 6, 7, 8, 9$. Note that if there is no interaction among criteria, we expect that the accuracy of the Choquet integral method is as well as that of weighted arithmetic mean method.
Step 3. For each m make the following calculations: (1) use credit hours to get the weight for each course; (2) use regression method to get the weight for each course; (3) by using the results from Step 2, compute fuzzy measures based on entropy and joint entropy for each subset of all courses. Then, the importance for each subset is resolved; (4) use the results from Step 2, compute fuzzy measures based on the complexity for each subset of all courses. Thus, the importance for each subset is available.
Step 4: Calculate the weighted arithmetic mean and regression methods among all courses from the raw data. Later, transform the results into the level of the scores for each course when $m = 2, 3, 4, 5, 6, 7, 8, 9$. Use the Choquet integral with the entropy method and the complexity-based method to compute overall performance values discussed in Step 3 for each $m = 2, 3, 4, 5, 6, 7, 8, 9$. Finally, transform the results into the level of the scores for each $m = 2, 3, 4, 5, 6, 7, 8, 9$.
Step 5: Calculate the accuracy for each method for each $m = 2, 3, 4, 5, 6, 7, 8, 9$.

4. A case study

A data set comes from a class with 45 students in a junior high school, and each student took three courses (namely physics and chemistry, biology, and geoscience) for natural science. The credit hours for these three courses are 16, 4, and 4, respectively. The maximum score for each course is 100 points. Later, all students took a Basic Competence Test for all junior high school students. The maximum and minimum scores of the Basic Competence Test are 60 and 1, respectively. To simplify the notations, physics and chemistry, biology, and geoscience are denoted by C_1, C_2, C_3, while the score of the Basic Competence Test is denoted by Obj. The detailed information is depicted in Table 7.

The first step is to use two rules to help decide the range of the number of level. One is Scott rule with the formula of $m = \frac{R \cdot n^{1/3}}{3.49 \cdot \sigma}$, where R is full range, m is the number of the level, σ is the standard deviation of the distribution, and n is the available sample of size

Table 7
The detailed information in the case study

Student	C_1	C_2	C_3	Obj	Student	C_1	C_2	C_3	Obj
1	70	81	75	41	26	78	80	76	37
2	70	85	86	42	27	88	84	80	35
3	65	85	84	33	28	55	65	60	5
4	75	91	85	25	29	78	85	75	27
5	75	80	82	27	30	72	84	78	47
6	68	75	76	33	31	64	76	70	27
7	70	77	72	35	32	60	70	65	20
8	80	78	70	31	33	69	80	70	35
9	83	81	85	50	34	66	78	66	17
10	75	79	83	31	35	62	70	66	13
11	62	74	68	35	36	61	72	65	28
12	68	74	80	30	37	68	74	71	11
13	77	85	81	37	38	53	65	59	9
14	66	76	74	29	39	67	70	64	36
15	78	88	83	31	40	59	65	68	16
16	57	67	62	15	41	74	82	75	49
17	56	70	63	12	42	58	66	62	15
18	68	80	74	31	43	76	74	78	38
19	53	66	58	21	44	84	81	78	37
20	65	81	73	32	45	76	72	74	35
21	62	76	69	12					
22	67	75	71	22					
23	74	71	68	28					
24	61	69	65	28					
25	64	70	67	24					

Table 8
The results of Cramer's correlation coefficients

	C_1	C_2	C_3	C_1	C_2	C_3
	$m = 2$			$m = 3$		
C_1	1	0.5307	0.5737	1	0.5437	0.5284
C_2	0.5307	1	0.7441	0.5437	1	0.6765
C_3	0.5737	0.7441	1	0.5284	0.6765	1
$p < 0.01$				$p < 0.01$		
	$m = 4$			$m = 5$		
C_1	1	0.5097	0.5744	1	0.5131	0.5885
C_2	0.5097	1	0.6026	0.5131	1	0.5583
C_3	0.5744	0.6026	1	0.5885	0.5583	1
$p < 0.01$				$p < 0.01$		
	$m = 6$			$m = 7$		
C_1	1	0.4848	0.5537	1	0.4991	0.537
C_2	0.4848	1	0.6212	0.4991	1	0.5821
C_3	0.5537	0.6212	1	0.537	0.5821	1
$p < 0.01$				$p < 0.01$		
	$m = 8$			$m = 9$		
C_1	1	0.515	0.5329	1	0.5164	0.5375
C_2	0.515	1	0.5336	0.5164	1	0.5049*
C_3	0.5329	0.5336	1	0.5375	0.5049*	1
$p < 0.01$						

* $p > 0.01$.

(Scott, 1979). In practice, σ is replaced by the estimated standard deviation, s. In our study, the sample of size n is 45. From the raw data, $R = 35$, 26, 28, and 45 for each item and $s = 8.4887$, 6.7182, 7.6480, and 10.7691, respectively. By the above formula, m would be 4.2021, 3.9443, 3.7313, and 4.2587, respectively. Thus, $m = 4$ or 5 are possible candidates. The other one is the Sturge's formula: $m = 1 + 3.3 ^* \log_{10}(n)$ (Scott, 1992). From the latter formula, m is 6.4556. Thus, $m = 6$ or 7 are possible candidates. In this study, set $m = 2, 3, 4, 5, 6, 7, 8, 9$ for extending m values around the possible candidates by two levels. That is, $m = 2, 3, 4, 5, 6, 7, 8, 9$.

The second step is to check whether there exist mutual interactions at significant level of 0.01 among courses and observe the strength of mutual interactions among courses. First, use the results in Step 1 and "crosstab.m" program of Matlab 7.0 to compute the corresponding p-values and Chi-square divergence between a pair of criteria for each $m = 2, 3, 4, 5, 6, 7, 8, 9$. Later, compute Cramer's correlation coefficient by using Chi-square values by the following formula: $G = \sqrt{\frac{\chi^2}{n*L}}$, where $n = 45$ and $L = m - 1$ for each $m = 2, 3, 4, 5, 6, 7, 8, 9$. From p-values under $m = 2, 3, 4, 5, 6, 7, 8, 9$, summarized in Table 8, clearly there exist mutual interactions at significant level of 0.01 among courses when $m = 2, 3, 4, 5, 6, 7, 8$ except $m = 9$. From Cramer's correlation coefficient in Table 8, we know the strength of mutual interactions among courses is stronger. Thus, we expect the accuracy of the Choquet integral method is better than those of weighted arithmetic mean and regression methods when $m = 2, 3, 4, 5, 6, 7, 8$.

The third step is to calculate the importance for each course by weighted arithmetic mean and regression methods, and the results are summarized in Table 9. From Table 9, C_1 (physics and chemistry) has the highest importance than C_2 (biology) and C_3 (geoscience) by the weighted arithmetic mean method, i.e., $C_1 > C_2 = C_3$. In contrast to the weighted arithmetic mean method, the regression method shows different importance as follows: $C_1 > C_3 > C_2$. That is, it suggests that the class needs to put more efforts on geoscience to improve the score on the Basic Competence Test. For the evaluation of the Choquet integral with the entropy-based and the complexity-based methods, calculate the importance for each subset generated by all courses for $m = 2, 3, 4, 5, 6, 7, 8, 9$. The numerical figures of fuzzy measures for each subset are computed by Matlab and provided in Table 10. From Table 10, the importance of complexity-based method is larger than that of entropy-based method for each subset of all criteria. This means that the importance of entropy-based method is underestimated. The reasons may come from the error of estimating a population probability by a small sample of size 45.

The fourth step is to compute the overall performance of students by the four methods. For each student, the overall performance and the score of the Basic Competence Test are transformed into the level of the scores for each item, as shown in Table 11, where $M1$, $M2$, $M3$, and $M4$ represent the weighted arithmetic mean method, the regression method, the Choquet integral with the entropy-based method, and the Choquet integral with the complexity-based method, respectively. The different numerical figures in the Choquet integral column depicted in Table 11 have different meanings. The higher the value of Choquet integral is, the better it is. Finally, the fifth step is to compare the predictions of different methods under different m, depicted in Table 12, where higher value means better accuracy. Obviously, the Choquet integral with the complexity-based method has the best accuracy among the four methods. The reasons may be that to estimate a population by the sample probability is worsen when m is greater than 4. It is worth to note that the regression method has better accuracy than the weighted arithmetic mean method since the regression method is to minimize the error without the assumption of mutual interaction among courses.

Table 9
Weights for each course by the weighted arithmetic mean and regression methods

Weighted arithmetic mean method			Regression method			
C_1	C_2	C_3	β	C_1	C_2	C_3
16/24	4/24	4/24	45.8959	0.5062	0.1650	0.3732

Table 10
Entropy-based and complexity-based fuzzy measures with $m = 2, 3, 4, 5, 6, 7, 8, 9$

Entropy-based				Complexity-based			
C_1	C_2	C_3	Fuzzy measure	C_1	C_2	C_3	Fuzzy measure
m = 2							
0	0	0	0	0	0	0	0
1	0	0	0.1667	1	0	0	0.2857
0	1	0	0.1667	0	1	0	0.2857
1	1	0	0.5	1	1	0	0.5714
0	0	1	0.1667	0	0	1	0.2857
1	0	1	0.5	1	0	1	0.5714
0	1	1	0.5	0	1	1	0.5714
1	1	1	1	1	1	1	1
m = 3							
0	0	0	0	0	0	0	0
1	0	0	0.1667	1	0	0	0.2308
0	1	0	0.1667	0	1	0	0.2308
1	1	0	0.5	1	1	0	0.5385
0	0	1	0.1667	0	0	1	0.2308
1	0	1	0.5	1	0	1	0.5385
0	1	1	0.5	0	1	1	0.5385
1	1	1	1	1	1	1	1
m = 4							
0	0	0	0	0	0	0	0
1	0	0	0.1579	1	0	0	0.2
0	1	0	0.1579	0	1	0	0.2
1	1	0	0.579	1	1	0	0.6
0	0	1	0.1579	0	0	1	0.2
1	0	1	0.579	1	0	1	0.6
0	1	1	0.5263	0	1	1	0.55
1	1	1	1	1	1	1	1
m = 5							
0	0	0	0	0	0	0	0
1	0	0	0.1667	1	0	0	0.2
0	1	0	0.1667	0	1	0	0.2
1	1	0	0.5	1	1	0	0.52
0	0	1	0.1667	0	0	1	0.2
1	0	1	0.5417	1	0	1	0.56
0	1	1	0.5	0	1	1	0.52
1	1	1	1	1	1	1	1
m = 6							
0	0	0	0	0	0	0	0
1	0	0	0.1786	1	0	0	0.2069
0	1	0	0.1786	0	1	0	0.2069
1	1	0	0.6786	1	1	0	0.6897
0	0	1	0.1786	0	0	1	0.2069
1	0	1	0.6429	1	0	1	0.6552
0	1	1	0.6072	0	1	1	0.6207
1	1	1	1	1	1	1	1
m = 7							
0	0	0	0	0	0	0	0
1	0	0	0.1936	1	0	0	0.2188
0	1	0	0.1936	0	1	0	0.2188
1	1	0	0.7097	1	1	0	0.7188
0	0	1	0.1936	0	0	1	0.2188
1	0	1	0.7097	1	0	1	0.7188
0	1	1	0.6774	0	1	1	0.6875
1	1	1	1	1	1	1	1
m = 8							
0	0	0	0	0	0	0	0
1	0	0	0.1945	1	0	0	0.2162
0	1	0	0.1945	0	1	0	0.2162
1	1	0	0.6667	1	1	0	0.6757
0	0	1	0.1945	0	0	1	0.2162
1	0	1	0.6667	1	0	1	0.6757
0	1	1	0.7222	0	1	1	0.7297
1	1	1	1	1	1	1	1
m = 9							
0	0	0	0	0	0	0	0
1	0	0	0.2286	1	0	0	0.25
0	1	0	0.2286	0	1	0	0.25
1	1	0	0.7143	1	1	0	0.7222
0	0	1	0.2286	0	0	1	0.25
1	0	1	0.8286	1	0	1	0.8333
0	1	1	0.8	0	1	1	0.8056
1	1	1	1	1	1	1	1

Table 11
The results of overall performance evaluated by four methods and the scores of the Basic Competence Test are transformed into the level of the scores with $m = 2, 3, 4, 5, 6, 7, 8, 9$

Student	M1	M2	M3	M4	Obj	Student	M1	M2	M3	M4	Obj
m = 2											
1	2	2	2	2	2	26	2	2	2	2	2
2	2	2	2	2	2	27	2	2	2	2	2
3	2	2	2	2	2	28	1	1	1	1	1
4	1	2	2	2	1	29	1	2	2	2	1
5	1	2	2	2	1	30	2	2	2	2	2
6	2	2	2	2	2	31	1	1	1	1	1
7	2	2	2	2	2	32	1	1	1	1	1
8	2	2	2	2	2	33	2	2	2	2	2
9	2	2	2	2	2	34	1	1	1	1	1
10	2	2	2	2	2	35	1	1	1	1	1
11	2	1	1	1	2	36	2	1	1	1	2
12	2	2	2	2	2	37	1	1	1	1	1
13	2	2	2	2	2	38	1	1	1	1	1
14	2	1	2	2	2	39	2	1	1	1	2
15	2	2	2	2	2	40	1	1	1	1	1
16	1	1	1	1	1	41	2	2	2	2	2
17	1	1	1	1	1	42	1	1	1	1	1
18	2	2	2	2	2	43	2	2	2	2	2
19	1	1	1	1	1	44	2	2	2	2	2
20	2	1	2	2	2	45	2	2	2	2	2
21	1	1	1	1	1						
22	1	1	1	1	1						
23	2	2	1	1	2						
24	2	1	1	1	2						
25	1	1	1	1	1						
m = 3											
1	2	2	2	2	3	26	3	3	3	3	3
2	2	3	3	3	3	27	3	3	3	3	2
3	2	2	3	3	2	28	1	1	1	1	1
4	3	3	3	3	2	29	3	3	3	3	2
5	3	3	3	3	2	30	2	3	3	3	3
6	2	2	2	2	2	31	2	2	2	2	2
7	2	2	2	2	2	32	1	1	1	1	1
8	3	3	3	3	2	33	2	2	2	2	2
9	3	3	3	3	3	34	2	2	2	2	1
10	3	3	3	3	2	35	1	1	1	1	1
11	1	1	2	2	2	36	1	1	1	1	2
12	2	2	2	2	2	37	2	2	2	2	1
13	3	3	3	3	3	38	1	1	1	1	1
14	2	2	2	2	2	39	2	2	1	2	3
15	3	3	3	3	2	40	1	1	1	1	1
16	1	1	1	1	1	41	2	3	3	3	3
17	1	1	1	1	1	42	1	1	1	1	1
18	2	2	2	2	2	43	3	3	2	3	3
19	1	1	1	1	2	44	3	3	3	3	3
20	2	2	2	2	2	45	2	2	2	2	2
21	1	2	2	2	1						
22	2	2	2	2	2						
23	2	2	2	2	2						
24	1	1	1	1	2						
25	1	1	1	2	2						
m = 4											
1	3	3	3	3	4	26	3	4	3	4	3
2	3	4	4	4	4	27	4	4	4	4	3
3	3	3	4	3	3	28	1	1	1	1	1
4	4	4	4	4	2	29	4	4	4	4	2
5	3	4	4	4	2	30	3	3	4	3	4
6	2	3	3	3	3	31	2	2	2	2	2
7	3	3	3	3	3	32	1	1	1	1	2
8	3	3	3	3	3	33	3	3	3	3	3

Table 11 (*continued*)

Student	M1	M2	M3	M4	Obj	Student	M1	M2	M3	M4	Obj
9	4	4	4	4	4	34	2	2	2	2	2
10	3	4	4	4	3	35	2	2	2	2	1
11	2	2	2	2	3	36	2	2	2	2	3
12	2	3	3	3	3	37	2	2	2	2	1
13	4	4	4	4	3	38	1	1	1	1	1
14	2	2	3	3	3	39	2	2	2	2	3
15	4	4	4	4	3	40	1	1	1	1	1
16	1	1	1	1	1	41	3	3	3	3	4
17	1	1	1	1	1	42	1	1	1	1	1
18	2	3	3	3	3	43	3	3	3	3	3
19	1	1	1	1	2	44	4	4	4	4	3
20	2	2	3	3	3	45	3	3	3	3	3
21	2	2	2	2	1						
22	2	2	2	2	2						
23	3	3	2	3	3						
24	1	1	1	2	3						
25	2	2	2	2	2						
m = 5											
1	4	4	4	4	4	26	4	4	4	4	4
2	5	4	5	5	5	27	4	5	5	5	4
3	4	4	4	4	4	28	1	1	1	1	1
4	3	5	5	5	3	29	3	4	5	5	3
5	3	4	4	5	3	30	5	4	4	4	5
6	4	3	3	3	4	31	3	3	3	3	3
7	4	3	3	3	4	32	2	2	2	2	2
8	3	4	4	4	3	33	4	3	3	3	4
9	5	5	5	5	5	34	2	2	3	3	2
10	3	4	4	5	3	35	1	2	2	2	1
11	4	2	2	2	4	36	3	2	2	2	3
12	3	3	4	3	3	37	1	3	3	3	1
13	4	5	5	5	4	38	1	1	1	1	1
14	3	3	3	3	3	39	4	2	2	2	4
15	3	5	5	5	3	40	2	2	1	2	2
16	2	1	1	1	2	41	5	4	4	4	5
17	1	1	1	1	1	42	2	1	1	1	2
18	3	3	3	3	3	43	4	4	4	4	4
19	2	1	1	1	2	44	4	5	5	5	4
20	3	3	3	3	3	45	4	4	3	4	4
21	1	2	3	2	1						
22	2	3	3	3	2						
23	3	3	3	3	3						
24	3	2	2	2	3						
25	3	2	2	2	3						
m = 6											
1	5	4	4	4	5	26	5	5	5	5	5
2	5	5	6	6	5	27	4	6	6	6	4
3	4	4	5	5	4	28	1	1	1	1	1
4	3	6	6	6	3	29	3	5	5	5	3
5	3	5	5	5	3	30	6	5	5	5	6
6	4	4	4	4	4	31	3	3	3	3	3
7	4	4	4	4	4	32	2	2	2	2	2
8	4	5	5	5	4	33	4	4	4	4	4
9	6	6	6	6	6	34	2	3	3	3	2
10	4	5	5	5	4	35	2	2	2	2	2
11	4	2	3	3	4	36	4	2	2	2	4
12	4	4	4	4	4	37	1	3	3	3	1
13	5	5	6	6	5	38	1	1	1	1	1
14	4	3	3	4	4	39	5	3	2	3	5
15	4	6	6	6	4	40	2	2	2	2	2
16	2	1	1	1	2	41	6	5	5	5	6
17	1	1	2	1	1	42	2	1	1	1	2
18	4	4	4	4	4	43	5	5	4	5	5
19	3	1	1	1	3	44	5	6	5	6	5
20	4	3	4	4	4	45	4	4	4	4	4
21	1	3	3	3	1						
22	3	3	3	3	3						
23	4	4	3	4	4						
24	4	2	2	2	4						
25	3	2	2	3	3						
m = 7											
1	4	5	5	5	6	26	6	6	5	6	5
2	5	6	7	7	6	27	7	7	7	7	5
3	4	5	6	6	5	28	1	1	1	1	1

Table 11 (*continued*)

Student	M1	M2	M3	M4	Obj	Student	M1	M2	M3	M4	Obj
4	6	7	7	7	4	29	6	6	6	6	4
5	5	6	6	6	4	30	5	5	6	6	7
6	4	4	4	5	5	31	3	3	4	4	4
7	4	4	4	4	5	32	2	2	2	2	3
8	6	5	5	5	5	33	4	4	4	4	5
9	7	7	7	7	7	34	3	3	4	3	2
10	5	6	6	6	5	35	2	2	2	3	2
11	3	3	3	3	5	36	2	2	2	2	4
12	4	5	5	5	4	37	4	4	3	4	1
13	6	6	6	7	5	38	1	1	1	1	1
14	4	4	4	4	4	39	3	3	2	3	5
15	6	7	7	7	5	40	2	2	2	2	2
16	1	1	1	1	2	41	5	5	5	5	7
17	1	1	2	2	2	42	1	1	1	1	2
18	4	4	5	5	5	43	5	5	5	5	6
19	1	1	1	1	3	44	7	7	6	7	5
20	4	4	5	4	5	45	5	5	4	5	5
21	3	3	3	3	2						
22	4	4	4	4	3						
23	4	4	3	4	4						
24	2	2	2	2	4						
25	3	3	2	3	3						
m = 8											
1	5	5	6	6	7	26	6	7	6	7	6
2	6	7	7	8	7	27	8	8	8	8	6
3	5	6	7	7	5	28	1	1	1	1	1
4	7	8	8	8	4	29	7	7	7	7	4
5	6	7	6	7	4	30	6	6	6	7	8
6	4	5	5	5	5	31	3	4	4	4	4
7	5	5	5	5	6	32	2	2	2	3	3
8	6	6	6	6	5	33	5	5	5	5	6
9	8	8	7	8	8	34	4	4	4	4	3
10	6	7	7	7	5	35	3	3	2	3	2
11	3	3	3	3	6	36	3	3	3	3	5
12	4	5	5	5	5	37	4	4	4	4	2
13	7	7	7	8	6	38	1	1	1	1	1
14	4	4	4	5	5	39	3	3	3	3	6
15	7	8	8	8	5	40	2	2	2	2	2
16	1	2	1	2	2	41	6	6	6	6	8
17	1	2	2	2	2	42	2	2	1	2	2
18	4	5	5	5	5	43	6	6	5	6	6
19	1	1	1	1	3	44	8	8	7	8	6
20	4	4	5	5	5	45	6	6	5	5	6
21	3	3	4	4	2						
22	4	4	4	4	4						
23	5	5	4	4	5						
24	2	2	2	3	5						
25	3	3	3	3	4						
m = 9											
1	5	6	6	6	8	26	7	7	6	7	7
2	6	7	8	9	8	27	9	9	9	9	6
3	5	6	8	8	6	28	1	1	1	1	1
4	8	8	9	9	4	29	7	7	7	8	5
5	7	7	7	8	5	30	6	7	7	7	9
6	5	5	5	6	6	31	4	4	4	4	5
7	5	5	5	5	6	32	2	3	2	3	3
8	7	7	6	7	6	33	5	5	5	5	6
9	9	9	8	9	9	34	4	4	4	4	3
10	7	8	7	8	6	35	3	3	3	3	2
11	3	3	4	4	6	36	3	3	3	3	5
12	5	6	6	6	5	37	5	5	4	5	2
13	7	8	8	8	7	38	1	1	1	1	1
14	4	5	5	5	5	39	4	4	3	3	7
15	8	8	9	9	6	40	2	3	2	2	3
16	2	2	1	2	2	41	6	7	6	7	9
17	2	2	2	2	2	42	2	2	1	2	2
18	5	5	6	6	6	43	7	7	6	7	7
19	1	1	1	1	4	44	8	8	8	8	7
20	4	5	6	6	6	45	6	6	5	6	6
21	3	4	4	4	2						
22	4	5	4	5	4						
23	5	5	4	5	5						
24	3	3	2	3	5						
25	3	3	3	3	4						

Table 12
The accuracy for each method with $m = 2, 3, 4, 5, 6, 7, 8, 9$

m	Weighted arithmetic mean	Regression	Choquet integral based on entropy	Choquet integral based on complexity
2	0.7333	0.8	0.8222	0.8222
3	0.5778	0.6222	0.6	0.6444
4	0.4444	0.4889	0.5333	0.5333
5	0.3111	0.4	0.3333	0.4
6	0.3778	0.4667	0.3778	0.4667
7	0.2444	0.2667	0.2667	0.2889
8	0.2444	0.3556	0.2889	0.3556
9	0.2889	0.3111	0.2	0.3333

5. Conclusions

A case study of applying weighted arithmetic mean method, regression method, the Choquet integral with the entropy-based method, and the proposed Choquet integral with the complexity-based method is presented in this study to evaluate the overall performance of students in a junior high school based on a Basic Competence Test. The weighted arithmetic mean and regression methods assume there is no any interaction between courses, whereas the Choquet integral methods can be used to deal with interactions among courses. The advantage of the proposed Choquet integral with the complexity-based method is that no population probability is to be estimated such that the error of estimating the population probability is reduced. Typically, in order to accurately estimate the population probabilities, the sample of size should be large enough.

In our study, the statistical tests show that there exists interaction between any two courses which result in the best performance of our proposed method consistently for different m values. The poor performance of the Choquet integral with the entropy-based method might result from the smaller sample of size. Note that our sample of size is 45. Thus, the error of estimating a population probability based on a small sample of size might be larger. Finally, the proposed Choquet integral with the complexity-based method is suitable to deal with particularly small sample of sizes.

References

Bonchev, D. G., & Rouvray, D. H. (2003). *Complexity: Introduction and fundamentals*. Taylor & Francis, CRC Press.

Calvo, T., Kolesarova, A., Komornikova, M., & Mesiar, R. (2001). Aggregation operators: Properties, classes and construction methods. *Fuzziness and Soft Computing, 97*, 1–103.

Choquet, G. (1953). Theory of Capacities. *Annales de l'Institut Fourier, 5*, 131–295.

Domingo, J., & Torra, V. (2002). Aggregation techniques for statistical confidentiality. In R. Mesiar, T. Calvo, & G. Mayor (Eds.), *Aggregation operators: New trends and applications* (pp. 260–271). Berlin, Germany: Physica-Verlag, Springer.

Fishburn, P. C. (1970). *Utility theory for decision making*. New York: John Wiley.

Höhle, U. (1982). Integration with respect to fuzzy measures. In *Proceedings of the IFAC symposium on theory and application of digital control* (pp. 35–37). New Delhi, India.

Kojadinovic, I. (2004). Estimation of the weights of interacting criteria from the set of profiles by means of information-theoretic functionals. *European Journal of Operational Research, 155*, 741–751.

Murofushi, T., & Sugeno, M. (1989). An interpretation of fuzzy measure and the Choquet integral as an integral with respect to a fuzzy measure. *Fuzzy Sets and Systems, 29*, 201–227.

Murofushi, T., & Sugeno, M. (1991). A theory of fuzzy measures – representations, the Choquet integral, and null sets. *Journal of Mathematical Analysis and Applications, 159*(2), 532–549.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika, 66*, 605–610.

Scott, D. W. (1992). *Multivariate density estimation: Theory practice, and visualization*. New York: John Wiley.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423. 623–656.

Yu, K., Wen, Z., Xu, X., & Ester, M. (2001). Feature weighting and instance selection for collaborative filtering. In *Proceedings of 2nd international workshop on management of information on the web – web data and text mining (MIV'01)*.

Theory of Multivalent Delta-Fuzzy Measures and its Application

Hsiang-Chuan Liu, Der-Bang Wu, Yu-Du Jheng, and Tian-Wei Sheu

Abstract**—**The well known fuzzy measures, Lambda-measure and P-measure, have only one formulaic solution, the former is not a closed form, and the latter is not sensitive enough. In this paper, a novel fuzzy measure, called Delta-measure, is proposed. This new measure proves to be a multivalent fuzzy measure which provides infinitely many solutions to closed form, and it can be considered as an extension of the above two measures. In other words, the above two fuzzy measures can be treated as the special cases of Delta-measure. For evaluating the Choquet integral regression models with our proposed fuzzy measure and other different ones, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted. The performances of Choquet integral regression models with fuzzy measure based on Delta-measure, Lambda-measure and P-measure, respectively, a ridge regression model and a multiple linear regression model are compared. Experimental result shows that the Choquet integral regression models with respect to Delta-measure based on Gamma-support outperforms other forecasting models.

Keywords**—**Lambda-measure, P-measure, Delta-measure, Gamma-support, Choquet integral regression model.

I. Introduction

When there are interactions among independent variables, traditional multiple linear regression models do not perform well enough. The traditional improved methods exploited ridge regression models [1]. In this paper, we suggest using the Choquet integral regression models [5,6,7,8,9,10] based on some single or compounded fuzzy measures [2,3,4, 12,13] to improve this situation. The well-known fuzzy measures, λ-measure [2,3] and P-measure [4], have only one formulaic solution of fuzzy measure, the former is not a closed form, and the latter is not sensitive enough. In this paper, we proposed a new fuzzy measure, δ-measure, which offers infinitely many solutions to a fuzzy measure with closed form and without changing the given singleton measure, and thereby, we can obtain an improved Choquet integral regression model with respect to this new fuzzy measure.

This paper is organized as follows: The multiple linear regression and ridge regression [1] are introduced in section II; two well known fuzzy measure, λ-measure [2] and P-measure [4], are introduced in section III; our new measure, δ-measure, is introduced in section IV; the fuzzy support, γ-support [7] is described in section V; the Choquet integral regression model [6],[7],[8] based on fuzzy measures are described in section VI; experiment and result are described in section VII; and final section is for conclusions and future works.

Manuscript received March 20, 2009: Revised version received May 20, 2009. This work was supported in part by the National Science Council grant of Taiwan Government (NSC 97-2410-H-468-014).

H. C. Liu is with the Department of Bioinformatics, Asia University, Taichung, 41345 Taiwan (e-mail: lhc@asia.edu.tw).

D. B. Wu, is with National Taichung University, Taichung, Taiwan 40306. He is now with the Graduate Institute of Educational Measurement and Department of Mathematics Education, National Taichung University, Taichung, 40306 Taiwan. (phone: 886-93673-1088; fax: 886-42218-3500; e-mail:wudb@hotmail.com).

Y. D. Jheng with National Taichung University, Taichung, Taiwan 40306. He is now with the Graduate Institute of Educational Measurement, National Taichung University, Taichung, 40306 Taiwan. (e-mail: doora0622@yahoo.com.tw).

T. W. Sheu with National Taichung University, Taichung, Taiwan 40306. He is now with the Graduate Institute of Educational Measurement, National Taichung University, Taichung, 40306 Taiwan. (e-mail:sheu@mail.ntcu.edu.tw).

II. The Multiple Linear Regression, Ridge Regression

Let $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{\varepsilon} \sim \mathrm{N}\left(\underline{0}, \sigma^2 I_n\right)$ be a multiple linear model, $\hat{\underline{\beta}} = \left(X'X\right)^{-1} X'Y$ be the estimated regression coefficient vector, and $\hat{\underline{\beta}}_k = \left(X'X + kI_n\right)^{-1} X'Y$ be the estimated ridge regression coefficient vector, Kenard and Baldwin [1] suggested

$$\hat{k} = \frac{n\hat{\sigma}^2}{\hat{\underline{\beta}}'\hat{\underline{\beta}}} \tag{1}$$

III. Fuzzy Measures

The two well known fuzzy measures, the λ-measure proposed by Sugeno in 1974, and P-measure proposed by Zadah in 1978, are concise introduced as follows.

A. Axioms of Fuzzy Measures [2, 3, 4]

A fuzzy measure μ on a finite set X is a set function $\mu : 2^X \to [0,1]$ satisfying the following axioms:

1) $\mu(\phi) = 0$, $\mu(X) = 1$ (boundary conditions) (2)

2) $A \subseteq B \Rightarrow \mu(A) \le \mu(B)$ (monotonicity) (3)

B. *Singleton Measures* [2, 6, 7]

A singleton measure of a fuzzy measure μ on a finite set X is a function $s : X \to [0,1]$ satisfying:

$$s(x) = \mu(\{x\}),\ x \in X \tag{4}$$

$s(x)$ is called the fuzzy density of singleton x.

For given singleton measures s, a λ-measure, g_λ, is a fuzzy measure on a finite set X, satisfying:

$$\begin{aligned} &A, B \in 2^X,\ A \cap B = \phi,\ A \cup B \neq X \\ &\Rightarrow g_\lambda(A \cup B) \\ &\quad = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B) \end{aligned} \tag{5}$$

$$\prod_{i=1}^{n}\left[1 + \lambda s(x_i)\right] = \lambda + 1 > 0,\ s(x_i) = g_\lambda(\{x_i\}) \tag{6}$$

Note that once the singleton measure is known, we can obtain the values of λ uniquely by using the previous polynomial equation. In other words, λ-measure has a unique solution without closed form. Moreover, for given singleton measures s, *If* $\sum_{x\in X} s(x) = 1$ *then* $g_\lambda(A) = \sum_{x \in A} s(x)$, in other word, if $\sum_{x\in X} s(x) = 1$ then λ-measure is just the additive measure

C. *P-measure [4]*

For given singleton measures s, a P-measure, g_P, is a fuzzy measure on a finite set X, satisfying:

$$\begin{aligned} &\forall A \in 2^X \\ &\Rightarrow g_P(A) = \max_{x\in A}\{s(x)\} = \max_{x \in A}\{g_P(\{x\})\} \end{aligned} \tag{7}$$

Note that for any subset of X, A, P-measure considers only the maximum value and will lead to insensitivity.

IV. A NEW METHOD - DELTA-MEASURES

A. *Definition of δ-measure*

For given singleton measure s, a δ-measure, g_δ, is a fuzzy measure on a finite set X, $|X| = n$, satisfying:

1) $\delta \in [-1,1],\ \sum_{x\in X} s(x) = 1$

2) $g_\delta(\phi) = 0, g_\delta(X) = 1$

3) $\forall A \subset X, A \neq X \Rightarrow$

$$g_\delta(A) = \left[1 + \delta \max_{x\in A} s(x)\right] \frac{(1+\delta)\sum_{x\in A} s(x)}{1 + \delta \sum_{x\in A} s(x)} - \delta \max_{x \in A} s(x) \tag{8}$$

B. *Important Properties of δ-measure*

To prove that δ-measure is a fuzzy measure, we need to prove the following theorem 1 firstly.

Theorem 1

For given singleton measure s,

If $A \subseteq B \subseteq X$ then

$$\sum_{x\in B} s(x) - \sum_{x\in A} s(x) \geq \max_{x\in B}\{s(x)\} - \max_{x\in A}\{s(x)\} \geq 0 \tag{9}$$

[Proof]

Let $B = A \cup C = A \cup \{x_1, x_2, ..., x_n\}, C = \{x_1, x_2, ..., x_n\}$

If $\max_{x\in B}\{s(x)\} = \max_{x\in A}\{s(x)\}$, then

$\sum_{x\in B} s(x) - \sum_{x\in A} s(x) \geq 0 = \max_{x\in B}\{s(x)\} - \max_{x\in A}\{s(x)\}$,

its true,

now suppose that $\max_{x\in B}\{s(x)\} > \max_{x\in A}\{s(x)\}$

(I) If n=1, let $B = A \cup C = A \cup \{x_1\}$, then

(i)

$$\begin{aligned} &\sum_{x\in B} s(x) = \sum_{x\in A} s(x) + \sum_{x\in C} s(x) = \sum_{x\in A} s(x) + s(x_1) \\ &\Rightarrow s(x_1) = \sum_{x\in B} s(x) - \sum_{x\in A} s(x) \geq 0 \end{aligned}$$

(ii) Since

$$\max_{x\in B}\{s(x)\} = \max\left\{\max_{x\in A}\{s(x)\}, s(x_1)\right\} > \max_{x\in A}\{s(x)\}$$

$$\Rightarrow s(x_1) = \max_{x\in B}\{s(x)\} \geq \max_{x\in B}\{s(x)\} - \max_{x\in A}\{s(x)\}$$

(iii) from (i) and (ii), we can obtain

$$\sum_{x\in B} s(x) - \sum_{x\in A} s(x) \geq \max_{x\in B}\{s(x)\} - \max_{x\in A}\{s(x)\}$$

(II) If n=k , let $B = A \cup C = A \cup \{x_1, x_2, ..., x_k\}$
satisfying

$$\sum_{x \in B} s(x) - \sum_{x \in A} s(x) \geq \max_{x \in B}\{s(x)\} - \max_{x \in A}\{s(x)\} \quad (10)$$

To prove that if n=k+1 ,

$B' = A \cup C = A \cup \{x_1, x_2, ..., x_k, x_{k+1}\} = B \cup \{x_{k+1}\}$

Satisfying

$$\sum_{x \in B'} s(x) - \sum_{x \in A} s(x) \geq \max_{x \in B'}\{s(x)\} - \max_{x \in A}\{s(x)\} \quad (11)$$

Since $B' = B \cup \{x_{k+1}\}$, and $s(x_{k+1}) \leq \max_{x \in B'}\{s(x)\}$

(i) if $\max_{x \in B}\{s(x)\} = \max_{x \in B'}\{s(x)\}$, then

$$\sum_{x \in B'} s(x) - \sum_{x \in A} s(x) = s(x_{k+1}) + \sum_{x \in B} s(x) - \sum_{x \in A} s(x)$$
$$\geq s(x_{k+1}) + \max_{x \in B}\{s(x)\} - \max_{x \in A}\{s(x)\}$$
$$= s(x_{k+1}) + \max_{x \in B'}\{s(x)\} - \max_{x \in A}\{s(x)\}$$
$$\geq \max_{x \in B'}\{s(x)\} - \max_{x \in A}\{s(x)\}$$
$$\Rightarrow \sum_{x \in B'} s(x) - \sum_{x \in A} s(x)$$
$$\geq \max_{x \in B'}\{s(x)\} - \max_{x \in A}\{s(x)\}$$

(ii) Now suppose that $\max_{x \in B}\{s(x)\} < \max_{x \in B'}\{s(x)\}$,

then $s(x_{k+1}) = \max_{x \in B'}\{s(x)\}$, and

$$\sum_{x \in B'} s(x) - \sum_{x \in A} s(x)$$
$$= s(x_{k+1}) + \sum_{x \in B} s(x) - \sum_{x \in A} s(x)$$
$$\geq s(x_{k+1}) = \max_{x \in B'}\{s(x)\}$$
$$\geq \max_{x \in B'}\{s(x)\} - \max_{x \in A}\{s(x)\}$$
$$\Rightarrow \sum_{x \in B'} s(x) - \sum_{x \in A} s(x)$$
$$\geq \max_{x \in B'}\{s(x)\} - \max_{x \in A}\{s(x)\}$$

(III) By mathematical induction, from (I) and (II), the proof is completed.

Theorem 2

For given singleton measure s, $\forall \delta \in [-1,1]$, δ-measure is a fuzzy measure.

[Proof];

(I) To prove the boundary conditions; $0 \leq g_\delta(A) \leq 1$

(i) if $\delta = -1$ it is trivial

(ii) if $\delta > -1$

$$0 \leq \sum_{x \in A} s(x) \leq 1 \Rightarrow \quad 1 + \delta \sum_{x \in A} s(x) > 0$$

Since

$$g_\delta(A)\left[1 + \delta \sum_{x \in A} s(x)\right] =$$
$$\left[1 + \delta \max_{x \in A}\{s(x)\}\right]\left[(1+\delta)\sum_{x \in A} s(x)\right]$$
$$-\left[\delta \max_{x \in A}\{s(x)\}\right]\left[1 + \delta \sum_{x \in A} s(x)\right]$$

and

$$\delta > -1,$$
$$\left[1 + \delta \sum_{x \in A} s(x)\right] \geq \left[1 + \delta \max_{x \in A}\{s(x)\}\right] > 0,$$
$$0 < \max_{x \in A}\{s(x)\} \leq \sum_{x \in A} s(x) \leq 1$$

Hence

$$g_\delta(A)\left[1 + \delta \sum_{x \in A} s(x)\right]$$
$$= (1+\delta)\sum_{x \in A} s(x) + \delta \max_{x \in A}\{s(x)\} \sum_{x \in A} s(x)$$
$$- \delta \max_{x \in A}\{s(x)\}$$

$$if\ -1 \leq \delta \leq 0,\ then$$

$$g_\delta(A)\left[1 + \delta \sum_{x \in A} s(x)\right] = (1+\delta)\sum_{x \in A} s(x) \quad (12)$$
$$+(-\delta)\max_{x \in A}\{s(x)\}\left[1 - \sum_{x \in A} s(x)\right] \geq 0$$

$$if\ \delta > 0,\ then$$

$$g_\delta(A)\left[1 + \delta \sum_{x \in A} s(x)\right]$$
$$= \left(1 + \delta \max_{x \in A}\{s(x)\}\right)\sum_{x \in A} s(x) +$$
$$\delta\left[\sum_{x \in A} s(x) - \max_{x \in A}\{s(x)\}\right] \geq 0$$

Therefore

$$g_\delta(A)\left[1 + \delta \sum_{x \in A} s(x)\right] \geq 0 \text{ , and } \quad g_\delta(A) \geq 0$$

(13)

(iii)

$$\left[1+\delta \max_{x\in A}\{s(x)\}\right](1+\delta)\sum_{x\in A} s(x)-$$

$$\delta \max_{x\in A}\{s(x)\}\left[1+\delta\sum_{x\in A} s(x)\right]-\left[1+\delta\sum_{x\in A} s(x)\right]$$

$$=\left[1+\delta \max_{x\in A}\{s(x)\}\right]\left[\sum_{x\in A} s(x)-1\right]\le 0$$

$$\Rightarrow\left[1+\delta \max_{x\in A}\{s(x)\}\right](1+\delta)\sum_{x\in A} s(x)-$$

$$\delta \max_{x\in A}\{s(x)\}\left[1+\delta\sum_{x\in A} s(x)\right]$$

$$\le\left[1+\delta\sum_{x\in A} s(x)\right]$$

$$\Rightarrow\left[1+\delta \max_{x\in A}\{s(x)\}\right]\frac{(1+\delta)\sum_{x\in A} s(x)}{\left[1+\sum_{x\in A} s(x)\right]}-\delta \max_{x\in A}\{s(x)\}\le 1$$

Therefore $0\le g_\delta(A)\le 1, \forall A\subset X$

(II) To prove the monotonicity;

$$A\subset B\subset X\Rightarrow g_\delta(A)\le g_\delta(B)$$

(i) Let $\mathrm{g_P}(A)=\max_{x\in A}\{s(x)\}, \mathrm{g}_\sigma(A)=\sum_{x\in A} s(x)$

$A\subset B\subset X$ Let

$\Rightarrow \mathrm{g_P}(A)\le \mathrm{g_P}(B),\ \mathrm{g}_\sigma(A)\le \mathrm{g}_\sigma(B)$

$\mathrm{g_P}(B)=\mathrm{g_P}(A)+c\le \mathrm{g}_\sigma(B)=\mathrm{g}_\sigma(A)+d$

From theorem 1 we know that $0\le c\le d\le 1$, then

$$-\left[\left[1+\delta g_P(A)\right]\frac{(1+\delta)g_\sigma(A)}{\left[1+\delta g_\sigma(A)\right]}-\delta g_P(A)\right]$$

$$=\left[1+\delta \mathrm{g}_P(A)\right]\frac{(1+\delta)g_\sigma(B)}{1+\delta g_\sigma(B)}$$

$$+\delta c\frac{(1+\delta)g_\sigma(B)}{\left[1+\delta g_\sigma(B)\right]}-\delta c$$

$$-\left[\left(1+\delta \mathrm{g}_P(A)\right)\frac{(1+\delta)\mathrm{g}_\sigma(A)}{1+\delta \mathrm{g}_\sigma(A)}\right] \tag{14}$$

$$\Rightarrow D'=\left[1+\delta g_\sigma(B)\right]D$$

$$=\left[1+\delta \mathrm{g}_P(A)\right](1+\delta)g_\sigma(B)+\delta c(1+\delta)g_\sigma(B)$$

$$-\delta c\left[1+\delta g_\sigma(B)\right]$$

$$-\left[1+\delta g_\sigma(B)\right]\left[\left(1+\delta \mathrm{g}_P(A)\right)\frac{(1+\delta)\mathrm{g}_\sigma(A)}{1+\delta \mathrm{g}_\sigma(A)}\right]$$

$$=\left[1+\delta\left[\mathrm{g}_P(A)\right]\right](1+\delta)[d]$$

$$+\left[1+\delta\left[\mathrm{g}_\sigma(A)\right]\right]\left[\delta c\left[\mathrm{g}_\sigma(B)-1\right]\right] \tag{15}$$

$if\ \delta\in[-1,0]\ then\ \left[1+\delta\left[\mathrm{g}_P(A)\right]\right]\ge 0,$

$(1+\delta)\ge 0,$

$\left[1+\delta\left[\mathrm{g}_\sigma(A)\right]\right]\ge 0,$

$\left[\delta c\left[\mathrm{g}_\sigma(B)-1\right]\right]\ge 0,$

$where\left[\mathrm{g}_\sigma(B)\right]\in[0,1],$

$d>0\quad then\ D'\ge 0,$

$and\ g_\delta(A)\le g_\delta(B)$

$if\ 0\le\delta\le 1, since\ d\ge c\ge 0$

$$\Rightarrow\left[1+\delta\left[\mathrm{g}_P(A)\right]\right](1+\delta)[d]$$

$$+\left[1+\delta\left[\mathrm{g}_\sigma(A)\right]\right]\left[\delta c\left[\mathrm{g}_\sigma(B)-1\right]\right]$$

$$\ge\left[1+\delta\left[\mathrm{g}_P(A)\right]\right](1+\delta)[c]$$

$$+\left[1+\delta\left[\mathrm{g}_\sigma(A)\right]\right]\left[\delta c\left[\mathrm{g}_\sigma(B)-1\right]\right]\ge 0$$

where $(1+\delta)\ge\left[1+\delta\left[\mathrm{g}_\sigma(A)\right]\right]\ge 0, and$

$\left[1+\delta\left[\mathrm{g}_P(A)\right]\right]\ge\left[\delta\left[1\text{-}\mathrm{g}_\sigma(B)\right]\right]\ge 0$

$then\ A\subset B\Rightarrow \mathrm{g}_\delta(A)\le \mathrm{g}_\delta(B)$

Theorem 3

(i) δ-measure is increasing function on δ

(ii) if $\delta=-1$ then δ-measure is just the P-measure

(iii) if $\delta=0$ then δ-measure is just the additive measure

(iv) if $-1<\delta<0$ then δ-measure is a sub-additive measure

(v) if $0<\delta<1$ then δ-measure is a supper-additive measure

[Proof];

(i) δ-measure is increasing function onδ

Let $-1\le\delta_1<\delta_2\le 1$ to prove that for each

$$A\subset X\Rightarrow g_{\delta_1}(A)\le g_{\delta_2}(A)$$

Let $f(\delta)=g_\delta(A)$

$$=\left[1+\delta g_P(A)\right]\frac{(1+\delta)g_\sigma(A)}{1+\delta g_\sigma(A)}-\delta g_P(A) \tag{16}$$

Since $1-g_\sigma(A)\ge 0,\ g_\sigma(A)\ge g_P(A)$

Then

$$= \ f'(\delta)=\frac{\left[1-g_\sigma(A)\right]\left[g_\sigma(A)-g_P(A)\right]}{\left[1+\delta g_\sigma(A)\right]^2}\ge 0 \quad (17)$$

$$f''(\delta)=\frac{-2g_\sigma(A)\left[1-g_\sigma(A)\right]\left[g_\sigma(A)-g_P(A)\right]}{\left[1+\delta g_\sigma(A)\right]^3}<0 \quad (18)$$

Therefore δ-measure is a concaved downward and increasing function on δ.

(ii) and (iii) are trivial

(iv) If $-1<\delta<0$, since δ-measure is increasing function on δ then $\forall A\subset X \Rightarrow g_\delta(A)\le g_0(A)=\sum_{x\in A}s(x)$,in other word, δ-measure is sub-additive

(v) If $0<\delta<1$, since δ-measure is increasing function on δ then $\forall A\subset X \Rightarrow g_\delta(A)\ge g_0(A)=\sum_{x\in A}s(x)$,in other word, δ-measure is supper-additive.

Theorem 4

If $\sum_{x\in X}s(x)=1$ and $\delta=0$ then δ-measure is just the λ-measure

Theorem 5

P -measure, additive measure and λ-measure are the special cases of δ-measure

V. Γ - SUPPORT [7]

For given singleton measure s of a fuzzy measure μ on a finite set X, if $\sum_{x\in X}s(x)=1$, then s is called a fuzzy support measure of μ, or a fuzzy support of μ, or a support of μ. Two kinds of fuzzy supports are introduced as below.

Let μ be a fuzzy measure on a finite set $X=\{x_1,x_2,...,x_n\}$, y_i be global response of subject i and $f_i(x_j)$ be the evaluation of subject i for singleton x_j, satisfying:

$$0<f_i(x_j)<1,\ i=1,2,...,N,\ j=1,2,...,n$$

$$\gamma(x_j)=\frac{1+r\left(f(x_j)\right)}{\sum_{k=1}^{n}\left[1+r\left(f(x_k)\right)\right]},\ j=1,2,...,n \quad (19)$$

where $r\left(f(x_j)\right)=\frac{S_{y,x_j}}{S_y S_{x_j}}$ (20)

$$S_y^2=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N}y_i\right)^2 \quad (21)$$

$$S_{x_j}^2=\frac{1}{N}\sum_{i=1}^{n}\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N}f_i(x_j)\right]^2 \quad (22)$$

$$S_{y,x_j}=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N}y_i\right)\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N}f_i(x_j)\right] \quad (23)$$

satisfying $0\le\gamma(x_j)\le 1$ and $\sum_{j=1}^{n}\gamma(x_j)=1$ (24)

then the function $\gamma: X\to[0,1]$ satisfying $\mu(\{x\})=\gamma(x)$, $\forall x\in X$ is a fuzzy support of μ, called γ-support of μ.

VI. CHOQUET INTEGRAL REGRESSION MODELS

A. *Choquet Integral [3, 5, 9, 10]*

Let μ be a fuzzy measure on a finite set X. The Choquet integral of $f_i: X\to R_+$ with respect to μ for individual i is denoted by

$$\int_C f_i d\mu=\sum_{j=1}^{n}\left[f_i\left(x_{(j)}\right)-f_i\left(x_{(j-1)}\right)\right]\mu\left(A_{(j)}^i\right)\ ,i=1,2,...,N \quad (25)$$

where $f_i\left(x_{(0)}\right)=0$, $f_i\left(x_{(j)}\right)$ indicates that the indices have been permuted so that

$$0\le f_i\left(x_{(1)}\right)\le f_i\left(x_{(2)}\right)\le ...\le f_i\left(x_{(n)}\right) \quad (26)$$

$$A_{(j)}=\left\{x_{(j)},x_{(j+1)},...,x_{(n)}\right\} \quad (27)$$

B. *Choquet Integral Regression Models [6 - 12]*

Let $y_1,y_2,...,y_N$ be global evaluations of N objects and $f_1(x_j),f_2(x_j),...,f_N(x_j),\ j=1,2,...,n$, be their evaluations of x_j, where $f_i: X\to R_+$, $i=1,2,...,N$.

Let μ be a fuzzy measure, $\alpha,\beta\in R$,

$$y_i=\alpha+\beta\int_C f_i dg_\mu+e_i\ ,\ e_i\sim N(0,\sigma^2)\ ,\ i=1,2,...,N \quad (28)$$

$$\left(\hat{\alpha},\hat{\beta}\right)=\arg\min_{\alpha,\beta}\left[\sum_{i=1}^{N}\left(y_i-\alpha-\beta\int_C f_i dg_\mu\right)^2\right] \quad (29)$$

then $\hat{y}_i=\hat{\alpha}+\hat{\beta}\int f_i dg_\mu,\ i=1,2,...,N$ is called the Choquet integral regression equation of μ, where

$$\hat{\beta}=S_{yf}/S_{ff} \quad (30)$$

$$\hat{\alpha}=\frac{1}{N}\sum_{i=1}^{N}y_i-\hat{\beta}\frac{1}{N}\sum_{i=1}^{N}\int f_i dg_\mu \quad (31)$$

$$S_{yf}=\frac{\sum_{i=1}^{N}\left[y_i-\frac{1}{N}\sum_{i=1}^{N}y_i\right]\left[\int f_i dg_{\mu^*}-\frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]}{N-1} \quad (32)$$

$$S_{ff}=\frac{\sum_{i=1}^{N}\left[\int f_i dg_{\mu^*}-\frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]^2}{N-1} \quad (33)$$

VII. Experiment and Result

A. Education Data

The total scores of 60 students from a junior high school in Taiwan are used for this research. The examinations of four courses, physics and chemistry, biology, geoscience and mathematics, are used as independent variables, the score of the Basic Competence Test of junior high school is used as a dependent variable.

The data of all variables listed in Table III is applied to evaluate the performances of four Choquet integral regression models with P-measure, λ-measure and δ-measure based on γ-support respectively, a ridge regression model, and a multiple linear regression model by using 5-fold cross validation method to compute the mean square error (MSE) of the dependent variable. The formula of MSE is

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \quad (34)$$

The singleton measures, γ-support of the P-measure, λ-measure and δ-measure are listed as follows which can be obtained by using the formula (19).

{0.2488, 0.2525, 0.2439, 0.2547}

For any fuzzy measure, μ-measures, once the fuzzy support of the μ-measure is given, all event measures of μ can be found, and then, the Choquet integral based on μ and the Choquet integral regression equation based on μ can also be found by using above corresponding formulae.

The experimental results of five forecasting models are listed in Table I. We find that the Choquet integral regression model with δ-measure based on γ-support outperforms other forecasting regression models.

TABLE I MSE OF REGRESSION MODELS

Regression model		5-fold CV MSE
	measure	
Choquet Integral Regression model	δ	48.7672
	λ	49.1832
	p	53.9582
Ridge regression		59.1329
Multiple linear regression		65.0664

B. Fat Data [3, 5, 9, 10, 11]

In this study, anthropometric dimensions were measured following a standard protocol [11]. High was measured to the nearest 0.1 cm using anthropometers. Body weight was measured to the nearest 0.1 kg at the same time the bioelectric impedance was measured using a body fat analyzer (TBF310; Tanita, Tokyo, Japan) to estimate the percentage of body fat (%fat). Skinfold thicknesses at biceps, triceps, subscapular, and suprailiac of the right side of body were measured with GMP skinfold calipers (Siber Hegener and Co. Ltd, Switzerland). The measurements were performed by one experienced operator that took two repeated measurements at the test site of the same subject. The mean of the two readings from each site was used to calculate body composition.

A real data set with 128 samples from a elementary school in Taiwan including the independent variables, 4 Skinfold determination values, and the dependent variable, the measurements of the BIA of each student listed in Table IV is applied to evaluate the performances of three Choquet integral regression models with P-measure, λ-measure and L-measure based on γ-support respectively, a ridge regression model, and a multiple linear regression model by using 5-fold cross validation method to compute the mean square error (MSE) of the dependent variable.

The singleton measures, γ-support of the P-measure, λ-measure and δ-measure are listed as follows which can be obtained by using the formula (19).

{0.2396, 0.2466, 0.254, 0.2596}

The formulas of MSE is by using 5-fold cross validation method to compute the mean square error (MSE) of the depen dent variable.

For any fuzzy measure, μ-measures, once the fuzzy support of the μ-measure is given, all event measures of μ can be found, and then, the Choquet integral based on μ and the Choquet integral regression equation based on μ can also be found.

The singleton measures, γ-support of the P-measure, λ-measure and L-measure can be obtained by using the formulas (6).

The experimental results of five forecasting models are listed in Table II. We find that the Choquet integral regression model with δ-measure based on γ-support outperforms other forecasting regression models.

TABLE II MSE OF REGRESSION MODELS

Regression model		5-fold CV MSE
	measure	
Choquet Integral Regression model	δ	14.4228
	λ	14.9218
	p	18.3846
Ridge regression		15.7434
Multiple linear regression		16.1122

VIII. CONCLUSION S

In this paper, multivalent fuzzy measure, δ-measure, is proposed. This new measure is proved that it is of closed form with infinitely many solutions, and it can be considered as an extension of the two well known fuzzy measures, λ-measure and P-measure. By using 5-fold cross-validation RMSE, an experiment is conducted for comparing the performances of a multiple linear regression model, a ridge regression model, and the Choquet integral regression model with respect to P-measure, λ-measure, and our proposed δ-measure based on γ-support respectively. The result shows that the Choquet integral regression models with respect to the proposed δ-measure based on γ-support outperforms other forecasting models.

In the future, we will apply the proposed Choquet integral regression model with fuzzy measure based on γ-support to develop multiple classifier system.

REFERENCES

[1] A. E. Hoerl, R. W. Kenard, and K. F. Baldwin, Ridge Regression: Some Simulation, *Communications in Statistics*, vol. 4, No. 2, pp. 105-123, 1975.

[2] Z. Wang, and G. J. Klir, *Fuzzy Measure Theory*, Plenum Press, New York, 1992.

[3] M. Sugeno, *Theory of Fuzzy Integrals and its Applications*, unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[4] L. A. Zadeh, *Fuzzy Sets and Systems*, vol. 1, pp. 3, 1978.

[5] G. Choquet, Theory of Capacities, *Annales de l'Institut Fourier*, vol. 5, pp. 131-295, 1953.

[6] H.-C. Liu, C.-C. Chen, D.-B. Wu, and Y.-D. Jheng, A New Weighting Method for Detecting Outliers in IPA Based on Choquet Integral, *IEEE International conference on Industrial Engineering and Engineering Management* 2007, Singapore, December 2007.

[7] Liu, H.-C. Tu, Y.-C. Chen, C.-C. and Weng, W.-S. The Choquet Integral with Respect to λ-Measure Based on γ-support, *2008 International Conferences on Machine Learning and Cybernetics*, Kunming, China, July 2008.

[8] J.-I Shieh, H.-H. Wu, H.-C. Liu, Applying Complexity-based Choquet Integral to Evaluate Students' Performance, *Expert Systems with Applications*, 36, pp. 5100-5106, 2009.

[9] S. Srivastava, A. Bansal, D. Chopra, G. Goe, Modeling and Control of a Choquet Fuzzy Integral Based Controller on a Real Time System. *WSEAS Transactions on Systems*, vol. 7, no. 5, pp. 1571-1578, July 2006.

[10] H.-C. Liu, Chen, C.-C., Chen, G.-S., & Jheng, Y.-D. (2007). Power-transformed-measure and its Choquet Integral Regression Model, *Proc. of 7th WSEAS International Conference on Applied Computer Science* (ACS'07), pp.101-104, 2007.

[11] J. S. Weiner, and J. A. Lourie, *Practical Human Biology*, Academic Press, London, 1981.

[12] H.-C. Liu, C. C. Chen, D. B. Wu, & Y. D. Jheng, A New Weighting Method for Detecting Outliers in IPA Based on Choquet Integral. *Proceedings of International Conference on Industrial Engineering & Engineering Management* (IEEM2007), pp. 2015-2019, Dec., 2007.

[13] G.S. Chen, Y.-D. Jheng, H.-C. Yao, and H.-C. Liu, Stroke Order Computer-based Assessment with Fuzzy Measure Scoring. *WSEAS Transactions Information Science & Applications*, vol. 2, no. 5, 62-68, 2008.

TABLE III THE DATA SET WITH FOUR COURSES AND SCIENCE SCORES OF THE BCT

No.	C1	C2	C3	C4	BCT	No.	C1	C2	C3	C4	BCT
1	72	66	78	72	19	31	66	68	75	74	25
2	86	80	82	81	35	32	68	70	74	76	40
3	56	63	69	75	21	33	57	65	75	70	24
4	78	86	86	86	33	34	74	70	80	75	35
5	66	72	80	76	23	35	49	60	69	64	13
6	68	74	77	80	28	36	51	60	63	64	18
7	74	86	87	88	44	37	58	64	68	66	32
8	54	56	62	68	7	38	73	78	84	81	39
9	71	74	80	77	26	39	56	56	65	61	6
10	68	70	80	75	33	40	61	62	70	70	25
11	53	56	70	63	22	41	57	60	68	64	23
12	67	70	80	75	35	42	57	64	67	70	26
13	70	66	70	74	13	43	50	52	68	60	7
14	60	65	75	70	23	44	84	80	76	72	49
15	68	68	78	76	35	45	62	66	76	71	22
16	58	66	76	71	37	46	70	74	78	82	32
17	61	66	72	78	33	47	69	70	80	75	26
18	68	68	80	74	26	48	63	74	74	74	42
19	56	66	76	71	21	49	66	78	80	82	39
20	59	62	70	78	29	50	67	70	80	75	31
21	62	64	76	70	36	51	56	65	75	70	23
22	71	72	78	75	26	52	50	54	66	60	18
23	74	63	69	75	12	53	71	75	85	80	41
24	59	70	80	76	37	54	74	77	80	85	26
25	75	75	85	80	39	55	71	72	76	80	31
26	73	78	84	81	24	56	60	65	75	70	21
27	62	68	72	74	29	57	59	57	70	68	17
28	77	74	80	76	42	58	50	56	65	68	13
29	63	60	68	69	17	59	72	76	80	78	38
30	56	61	75	68	22	60	81	76	78	80	33

C1 : physics and chemistry
C2 : biology
C3 : geoscience
C4 : mathematics
BCT : Basic Competence Test of nature science

Table IV Measurements of BIA and four skinfold determinations of percent body fat

No	BIA	biceps	triceps	Sub-scapular	Sup-railiac	No	BIA	biceps	triceps	Sub-scapular	Sup-railiac
1	11.8	15.8	16.2	27.6	16.3	65	22.8	27.4	32.2	33.6	24.5
2	17.8	20.4	12.8	35.6	23.0	66	5.4	14.4	8.0	17.2	17.5
3	6.2	10.4	9.4	12.6	8.2	67	8.6	9.6	11.4	14.4	14.8
4	7.0	10.2	5.6	10.6	13.6	68	7.4	9.6	8.0	10.2	13.4
5	5.2	10.8	12.0	13.8	11.8	69	20.2	29.2	28.2	40.0	23.9
6	20.2	24.6	39.8	40.0	34.7	70	28.4	30.6	38.6	39.8	39.5
7	14.8	18.8	22.4	24.8	18.6	71	6.8	10.4	8.8	14.6	15.2
8	12.4	18.2	20.2	27.6	18.4	72	11.0	12.4	10.4	14.4	11.7
9	17.0	20.6	21.2	27.8	21.6	73	20.8	30.8	38.8	40.0	37.7
10	16.6	22.6	18.8	32.6	16.6	74	25.0	27.6	40.0	40.0	29.6
11	10.2	13.4	8.4	12.8	15.1	75	4.8	6.2	5.6	7.0	9.7
12	21.0	26.7	36.2	40.0	23.9	76	8.4	7.0	8.0	11.8	15.6
13	10.0	13.4	17.4	17.8	18.1	77	11.0	14.0	10.0	12.4	17.7
14	9.8	13.8	11.6	19.2	15.5	78	10.6	11.6	7.0	14.4	12.9
15	8.6	10.8	11.4	14.4	8.5	79	12.4	16.8	17.0	25.8	14.9
16	21.8	25.6	36.6	36.8	31.7	80	11.4	14.0	20.8	26.4	25.1
17	25.2	29.4	30.2	30.6	31.2	81	12.8	15.8	20.2	19.8	16.4
18	10.2	11.0	10.2	14.8	12.2	82	12.0	15.6	16.0	22.8	17.3
19	10.4	15.8	13.0	25.6	16.6	83	13.0	20.4	24.0	27.0	26.2
20	17.6	22.6	23.2	34.6	23.2	84	14.2	15.4	22.4	22.8	18.9
21	12.6	12.4	14.2	16.0	14.2	85	11.0	16.4	14.2	15.8	17.9
22	9.0	11.2	9.4	11.8	9.3	86	22.4	29.8	35.0	36.2	28.5
23	12.2	19.2	17.4	27.8	19.7	87	6.4	7.6	8.6	11.4	12.1
24	4.6	7.0	8.8	11.2	7.2	88	6.8	10.6	9.6	14.6	12.2
25	6.4	8.8	11.0	12.6	10.8	89	16.2	18.4	27.2	27.4	24.9
26	23.8	29.0	37.0	35.0	30.7	90	22.4	26.8	25.4	33.4	30.4
27	8.4	15.8	17.8	23.0	21.6	91	9.6	11.2	10.4	18.0	11.7
28	12.2	16.6	16.4	20.6	18.7	92	10.8	17.2	24.0	24.8	21.9
29	7.2	12.8	8.6	18.6	15.2	93	13.0	16.2	12.4	18.4	14.2
30	21.4	31.2	31.4	39.4	28.6	94	5.6	12.4	11.4	15.6	14.5
31	18.2	23.0	40.0	40.0	28.2	95	19.4	25.0	36.2	39.0	29.9
32	9.2	12.6	40.0	17.8	16.0	96	14.4	22.4	29.8	35.0	24.8
33	10.2	18.8	17.8	20.8	18.4	97	25.4	29.4	37.0	40.0	24.6
34	19.2	24.4	35.2	35.0	34.1	98	9.4	11.2	11.4	12.4	8.9
35	6.8	12.0	8.0	14.4	16.1	99	17.4	22.6	19.4	31.6	22.7
36	16.8	20.8	25.6	27.8	20.7	100	24.0	30.8	40.0	40.0	29.4
37	35.8	38.6	40.0	40.0	30.1	101	3.8	6.0	6.4	6.8	10.8
38	10.0	11.6	10.4	18.6	8.3	102	11.0	19.4	11.6	18.4	13.7
39	5.4	12.2	12.4	21.4	19.2	103	22.6	24.4	40.0	40.0	33.3
40	11.2	18.0	23.6	30.8	22.1	104	9.2	10.0	11.0	19.2	13.4
41	5.4	11.2	6.8	11.6	11.9	105	18.2	19.0	31.0	29.4	24.5
42	7.6	8.4	9.4	13.6	12.8	106	6.8	12.4	14.0	17.8	14.1
43	6.6	9.8	9.6	12.0	9.3	107	7.4	11.6	10.0	16.0	11.0
44	32.4	37.2	40.0	40.0	18.2	108	9.2	10.6	12.4	14.4	12.7
45	7.8	14.0	11.0	17.8	31.3	109	29.4	23.6	39.8	40.0	37.4
46	17.8	26.6	34.2	40.0	22.8	110	6.8	7.8	9.8	12.8	10.9
47	22.0	27.8	38.2	39.4	23.6	111	12.4	14.6	15.8	19.8	16.9
48	14.4	15.8	18.8	23.8	14.9	112	8.2	9.8	9.2	16.0	14.5
49	15.8	18.4	21.4	24.0	24.9	113	16.4	20.8	25.2	30.4	24.9
50	7.4	12.8	10.2	17.0	14.3	114	9.4	11.4	12.0	21.8	14.3
51	16.2	29.0	21.6	29.8	24.1	115	16.4	22.4	33.2	36.8	25.1
52	6.0	7.4	7.6	9.8	8.6	116	7.0	11.4	13.8	17.4	11.2

53	12.2	15.4	16.2	18.8	17.8	117	10.4	12.6	14.8	23.8	18.0
54	11.6	12.0	9.8	13.0	8.9	118	5.6	8.2	10.2	8.6	7.7
55	17.8	22.6	38.0	31.0	24.8	119	10.8	11.8	17.8	21.2	19.9
56	13.2	16.8	18.6	23.4	20.7	120	9.6	15.8	14.4	19.4	18.6
57	4.4	7.2	8.2	9.8	14.3	121	5.0	6.8	7.4	9.4	6.0
58	16.2	21.8	28.2	32.6	27.2	122	9.8	12.2	12.4	15.4	13.5
59	11.4	19.4	28.8	32.8	22.3	123	13.8	18.0	16.4	21.0	19.3
60	11.2	13.0	18.8	22.6	21.9	124	8.8	12.8	9.8	11.8	13.3
61	8.6	11.4	7.2	10.2	7.5	125	15.8	21.0	35.4	39.8	27.3
61	20.4	26.2	31.0	32.8	25.8	126	10.8	16.6	15.6	23.2	16.5
63	7.0	8.8	11.6	9.4	12.0	127	9.0	10.6	10.0	16.8	11.9
64	14.6	17.4	12.8	16.8	14.7	128	8.8	12.4	10.0	10.8	11.3

Theory and Application of the Composed Fuzzy Measure of L-Measure and Delta-Measures

Hsiang-Chuan Liu, Chin-Chun Chen, Der-Bang Wu, and Tian-Wei Sheu

Abstract—The well known fuzzy measures, λ-measure and P-measure, have only one formulaic solution. Two multivalent fuzzy measures with infinitely many solutions were proposed by our previous works, called L-measure and δ-measure, but the former do not include the additive measure as the latter and the latter has not so many measure solutions as the former. Due to the above drawbacks, in this paper, an improved fuzzy measure composed of above both, denoted L_δ -measure, is proposed. For evaluating the Choquet integral regression models with our proposed fuzzy measure and other different ones, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted. The performances of Choquet integral regression models with fuzzy measure based L_δ -measure, L-measure, δ-measure, λ-measure, and P-measure, respectively, a ridge regression model, and a multiple linear regression model are compared. Experimental result shows that the Choquet integral regression models with respect to extensional L-measure based on γ-support outperforms others forecasting models.

Keywords—Lambda-measure, P-measure, Delta-measure, Gamma-support, composed fuzzy measure, Choquet integral regression model.

I. INTRODUCTION

When there are interactions among independent variables, traditional multiple linear regression models do not perform well enough. The traditional improved methods exploited ridge regression models [1]. In this paper, we suggest using the Choquet integral regression models [7-15] based on some single or compounded fuzzy measures [2-5, 7-15] to improve this situation. The well-known fuzzy measures, λ-measure [2-4] and P-measure [5] have only one formulaic solution of fuzzy measure, the former is not a closed form, and the latter is not sensitive enough. Two multivalent fuzzy measures with infinitely many solutions were proposed by our previous works, called L-measure [7-9] and δ-measure [10,11], but L-measure do not include the additive measure and δ-measure has not so many measure solutions as L-measure. Due to the above drawbacks, in this paper, an improved fuzzy measure composed of above two multivalent fuzzy measures, denoted L_δ -measure, is proposed. This improved multivalent fuzzy measure is not only including the additive measure, but also having the same infinitely many measure solutions as L-measure. For evaluating the Choquet integral regression models with our proposed fuzzy measure and other different ones, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted. The performances of Choquet integral regression models with fuzzy measure based L_δ -measure, L-measure, δ-measure, λ-measure, and P-measure, respectively, a ridge regression model, and a multiple linear regression model are compared.

This paper is organized as follows: The multiple linear regression and ridge regression [1] are introduced in section II; two well known fuzzy measure, λ-measure [2] and P-measure [5], are introduced in section III; our new measure, δ-measure, is introduced in section IV; the fuzzy support, γ-support [7] is described in section V; the Choquet integral regression model [6-8] based on fuzzy measures are described in section VI; experiment and result are described in section VII; and final section is for conclusions and future works.

Manuscript received May 20, 2009: Revised version received June 28, 2009. This work was supported in part by the National Science Council grant of Taiwan Government (NSC 97-2410-H-468-014).

H. C. Liu, is with the Department of Bioinformatics, Asia University, Taichung, 41345, Taiwan (e-mail: lhc@asia.edu.tw).

C. C. Chen, is with National Taichung University, Taichung, 40306, Taiwan. He is now with Department of General Education, Min-Hwei College, Tainan 736 , Taiwan. (e-mail: ccchen@mail.mhchcm.edu.tw).

D. B. Wu, is with National Taichung University, Taichung, 40306, Taiwan. He is now with the Graduate Institute of Educational Measurement and Department of Mathematics Education, National Taichung University, Taichung, 40306 Taiwan. (phone: 886-93673-1088; fax: 886-42218-3500; e-mail:wudb@hotmail.com).

T. W. Sheu, is with National Taichung University, Taichung, Taiwan 40306. He is now with the Graduate Institute of Educational Measurement, National Taichung University, Taichung, 40306 Taiwan. (e-mail: sheu@mail.ntcu.ed u.tw).

II. THE MULTIPLE LINEAR REGRESSION, RIDGE REGRESSION

Let $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{\varepsilon} \sim \mathrm{N}\left(\underline{0}, \sigma^2 I_n\right)$ be a multiple linear model, $\underline{\hat{\beta}} = \left(X'X\right)^{-1} X'Y$ be the estimated regression coefficient vector, and $\underline{\hat{\beta}}_k = \left(X'X + kI_n\right)^{-1} X'Y$ be the estimated ridge regression coefficient vector, Hoerl, Kenard and Baldwin [1] suggested

$$\hat{k} = \frac{n\hat{\sigma}^2}{\underline{\hat{\beta}}'\underline{\hat{\beta}}} \tag{1}$$

III. FUZZY MEASURES

The two well known fuzzy measures, the λ-measure proposed by Sugeno in 1974, and P-measure proposed by Zadah in 1978, are concisely introduced as follows.

A. Axioms of Fuzzy Measures

Definition 1 fuzzy measure [2-4]

A fuzzy measure μ on a finite set X is a set function $\mu: 2^X \to [0,1]$ satisfying the following axioms:

1) $\mu(\phi)=0$, $\mu(X)=1$ (boundary conditions) (2)

2) $A \subseteq B \Rightarrow \mu(A) \le \mu(B)$ (monotonicity) (3)

B. Singleton Measures

Definition 2 singleton measure [2-7]

A singleton measure of a fuzzy measure μ on a finite set X is a function $s: X \to [0,1]$ satisfying:

$$s(x)=\mu(\{x\}),\ x\in X \tag{4}$$

$s(x)$ is called the fuzzy density of singleton x.

C. λ-measure

Definition 3 λ-measure [3]

For a given singleton measures s, λ-measure, g_λ, is a fuzzy measure on a finite set X, satisfying:

$$\begin{aligned} & A,B\in 2^X,\ A\cap B=\phi,\ A\cup B\neq X \\ & \Rightarrow g_\lambda(A\cup B) \\ & \quad = g_\lambda(A)+g_\lambda(B)+\lambda g_\lambda(A)g_\lambda(B) \end{aligned} \tag{5}$$

$$\prod_{i=1}^{n}\left[1+\lambda s(x_i)\right]=\lambda+1>0,\ s(x_i)=g_\lambda(\{x_i\}) \tag{6}$$

Where the real number, λ, is also called the determine coefficient of λ-measure.

Note that once the singleton measure is known, we can obtain the values of λ uniquely by using the previous polynomial equation. In other words, λ-measure has a unique solution without closed form. Moreover, for given singleton measures s, $If \sum_{x\in X} s(x)=1\ then\ g_\lambda(A)=\sum_{x\in A} s(x)$, in other word, if $\sum_{x\in X} s(x)=1$ then λ-measure is just the additive measure

D. P-measure

Definition 4 P-measure [5]

For given a singleton measures s, P-measure, g_P, is a fuzzy measure on a finite set X, satisfying:

$$\begin{aligned} & {}^\forall A\in 2^X \\ & \Rightarrow g_P(A)=\max_{x\in A}\{s(x)\}=\max_{x\in A}\{g_P(\{x\})\} \end{aligned} \tag{7}$$

Note that for any subset of X, A, P-measure considers only the maximum value and will lead to insensitivity.

E. L-measure

Definition 5 L-measure [7-9]

For given a singleton measure $s(x)$, L-measure, g_L, is a fuzzy measure on a finite set X, $|X|=n$, satisfying:

1) $L\in[0,\infty)$ (8)

2) ${}^\forall A\subset X, n-|A|+(|A|-1)L>0 \Rightarrow$

$$g_L(A)=\max_{x\in A}\left[s(x)\right]+\frac{(|A|-1)L\sum_{x\in A}s(x)\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\left[n-|A|+(|A|-1)L\right]\sum_{x\in X}s(x)} \tag{9}$$

Where the real number, L, is also called the determine coefficient of L-measure.

Theorem 1 [7-9]

(i) for each $L\in[0,\infty)$, L-measure is a fuzzy measure, in other words, L-measure has infinitely many solutions of fuzzy measures, for each $L\in[0,\infty)$.

(ii) $L\in[0,\infty)$, L-measure is an increasing function on real number L.

(iii) if $L=0$ then L-measure is just the P-measure

F. δ-measure

Definition 6 δ-measure [10,11]

For given singleton measure $s(x)$, a δ-measure, g_δ, is a fuzzy measure on a finite set X, $|X|=n$, satisfying:

1) $\delta\in[-1,1],\ \sum_{x\in X}s(x)=1$ (10)

2) $g_\delta(\phi)=0, g_\delta(X)=1$ (11)

3) $\forall A\subset X, A\neq X \Rightarrow$

$$g_\delta(A)=\left[1+\delta\max_{x\in A}s(x)\right]\frac{(1+\delta)\sum_{x\in A}s(x)}{1+\delta\sum_{x\in A}s(x)}-\delta\max_{x\in A}s(x) \tag{12}$$

Where the real number, δ, is also called the determine coefficient of δ-measure.

Theorem 2 [11]

(i) $\delta\in[-1,1]$,δ-measure is an increasing function on δ

(ii) if $\delta=-1$, then δ-measure is just the P-measure

(iii) if $\delta=0$, then δ-measure is just the additive measure

(iv) if $-1<\delta<0$, then δ-measure is a sub-additive measure

(v) if $0<\delta<1$, then δ-measure is a supper-additive measure.

(vi) If $\sum_{x\in X} s(x)=1$ and $\delta=0$ then δ-measure is just the λ-measure

(vii) P -measure, additive measure and λ-measure are the special cases of δ-measure

IV. Comparison between two fuzzy measures

Definition 7 $\mu_1-measure\le\mu_2-measure$, $\mu_2-measure\ge\mu_1-measure$ [8,9]

For any given fuzzy density function, $s(x)$, on a finite set, X, If μ_1 and μ_2 are two fuzzy measures, satisfying $g_{\mu_1}(A)\le g_{\mu_2}(A), \forall A\subset X$, then we say that μ_1-measure is not larger than μ_2-measure, or μ_2-measure is not smaller than μ_1-measure, denoted as $\mu_1-measure\le\mu_2-masure$,or $\mu_2-measure\ge\mu_1-measure$

Theorem 3 [8,9]

For any given fuzzy density function, $s(x)$, on a finite set, X, P-measure is not larger than any other fuzzy measure, μ , that is $P-measure\le\mu-measure$

V. Composed Measure of L- Method and-Delta-Measures

A. *Definition of Generalized L-measure*

Definition 8 Generalized L-measure

For given singleton measure $s(x)$, a generalized L-measure based on a fuzzy measure, μ , L_μ, is a fuzzy measure on a finite set X, $|X|=n$, satisfying:

1) $L\in[0,\infty)$ (13)

2) $\forall A\subset X, n-|A|+(|A|-1)L>0\Rightarrow$

$$g_{L_\mu}(A)=\max_{x\in A}\left[s(x)\right]+\frac{\left[(|A|-1)L\right]\mu(A)\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\left[n-|A|+(|A|-1)L\right]\mu(X)} \tag{14}$$

Where the real number, L, is also called the determine coefficient of L_μ-measure.

Theorem 4

(i) For each $L\in[0,\infty)$, L_μ-measure is a fuzzy measure, In other words, L_μ-measure has infinite many fuzzy measures with determine coefficient L, $L\in[0,\infty)$.

(ii) $L\in[0,\infty)$ L_μ-measure is an increasing function on L,

(iii) if $L=0$ then L_μ-measure is just the μ-measure

(iv) if μ-measure is the P-measure then L_μ-measure is just the L-measure

(v) for each $L\in[0,\infty)$,

P-measure ≤ L-measure ≤ L_μ-measure

Proof.

(i) the boundary conditions are trivial, Now to prove the monotonicity.

Let $\forall A,B\in 2^X$, $A\subset B$ to prove $g_{L_\mu}(A)\le g_{L_\mu}(B)$ (15)

If $\max_{x\in A}\left[s(x)\right]=\max_{x\in B}\left[s(x)\right]$,

$$\text{since }\frac{(|B|-1)L\mu(B)}{\left[n-|B|+(|B|-1)L\right]}\ge\frac{(|A|-1)L\mu(A)}{\left[n-|A|+(|A|-1)L\right]} \tag{16}$$

We can obtain $g_{L_\mu}(B)\ge g_{L_\mu}(A)$

If $\max_{x\in B}\left[s(x)\right]=\max_{x\in A}\left[s(x)\right]+a, a>0$ (17)

$$g_{L_\mu}(B)-g_{L_\mu}(A)=a\left[1-\frac{(|B|-1)L\mu(B)\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\left[n-|B|+(|B|-1)L\right]\mu(X)}\right]$$
$$+\left[\frac{(|B|-1)L\mu(B)}{\left[n-|B|+(|B|-1)L\right]}-\frac{(|A|-1)L\mu(A)}{\left[n-|A|+(|A|-1)L\right]}\right]\frac{\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\mu(X)} \tag{18}$$

$$\text{Since }\quad 1-\frac{(|B|-1)L\mu(B)\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\left[n-|B|+(|B|-1)L\right]\mu(X)}\ge 0 \tag{19}$$

$$\text{and}\qquad \frac{\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\mu(X)}\ge 0 \tag{20}$$

We can also obtain that $g_{L_\mu}(B)\ge g_{L_\mu}(A)$, therefore L_μ-measure is a fuzzy measure.

(ii)

$$f(L)=g_{L_\mu}(A)=\max_{x\in A}\left[s(x)\right]+\frac{\left[(|A|-1)L\right]\mu(A)\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\left[n-|A|+(|A|-1)L\right]\mu(X)}$$
$$\Rightarrow f'(L)=\frac{(|A|-1)\mu(A)\left[1-\max_{x\in A}\left[s(x)\right]\right]\left[n-|A|\right]}{\mu(X)\left[n-|A|+(|A|-1)L\right]^2}\ge 0 \tag{21}$$

Hence L_μ-measure is an increasing function on L.

(iii), (iv) and (v) are trivial.

B. *Definition of L_δ-measure*

Definition 9 L_δ-measure

For given singleton measure $s(x)$, the composed measure of L-measure and δ-measure , denoted L_δ-measure, g_{L_δ}, is a fuzzy measure on a finite set X, $|X|=n$, satisfying:

1) $L\in[-1,\infty), \sum_{x\in X}s(x)=1$ (22)

2) $g_{L_\delta}(\phi)=0, g_{L_\delta}(X)=1$ (23)

3) $\forall A \subset X \Rightarrow$

$$g_{L_\delta}(A)=\begin{cases}\max\limits_{x\in A} s(x) & \text{if } L=-1\\ \dfrac{(1+L)\sum\limits_{x\in A} s(x)\left[1+L\max\limits_{x\in A} s(x)\right]}{1+L\sum\limits_{x\in A} s(x)}-L\max\limits_{x\in A} s(x) & \text{if } L\in(-1,0]\\ \dfrac{L(|A|-1)\sum\limits_{x\in A} s(x)\left[1-\sum\limits_{x\in A} s(x)\right]}{\left[n-|A|+L(|A|-1)\right]\sum\limits_{x\in X} s(x)}+\sum\limits_{x\in A} s(x) & \text{if } L\in(0,\infty)\end{cases} \quad (24)$$

C. *Important Properties of L_δ-measure*

Theorem 5 Important Properties of L_δ-measure

(i) $L\in[-1,\infty)$, L_δ-measure is a fuzzy measure family

(ii) $L\in[-1,\infty)$, L_δ-measure is an increasing function on L

(iii) if $L=-1$ then L_δ-measure is just the P-measure

(iv) if $L=0$ then L_δ-measure is just the additive measure

(v) if $-1<L<0$ then L_δ-measure is a sub-additive measure

(vi) if $0<L<\infty$ then L_δ-measure is a supper-additive measure

(vii) If $\sum\limits_{x\in X} s(x)=1$ and $L=0$ then L_δ-measure is just the λ-measure

(viii) P -measure, additive measure and λ-measure are the special cases of L_δ-measure

Proof.

(i) if $L\in[-1,0)$, then L_δ-measure is a special case of δ-measure, since δ-measure is a fuzzy measure, then L_δ-measure is also a fuzzy measure.

if $L\in[0,\infty)$, then L_δ-measure is a special case of generalized L-measure based on the additive measure, since any generalized L-measure is also a fuzzy measure, then L_δ-measure is also a fuzzy measure.

Therefore, for each $L\in[-1,\infty)$, L_δ-measure is a fuzzy measure.

(ii) if $L\in[-1,0)$, then L_δ-measure is a special case of δ-measure, since δ-measure is an increasing function with upper bound, additive measure , then L_δ-measure is also an increasing function with upper bound, additive measure.

if $L\in[0,\infty)$, then L_δ-measure is a special case of generalized L-measure based on the additive measure, since generalized L-measure based on the additive measure is also an increasing function with lower bound, additive measure, then L_δ-measure is also an increasing function with lower bound, additive measure.

Therefore, for each $L\in[-1,\infty)$, L_δ-measure is also an increasing function on L'

(iii), (iv), (v), (vi), (vii) and (viii) are trivial.

VI. Γ- SUPPORT

Definition 10: γ-support [7]

For given singleton measure s of a fuzzy measure μ on a finite set X, if $\sum\limits_{x\in X} s(x)=1$, then s is called a fuzzy support measure of μ, or a fuzzy support of μ, or a support of μ. One of fuzzy supports is introduced as below.

Let μ be a fuzzy measure on a finite set $X=\{x_1,x_2,...,x_n\}$, y_i be global response of subject i and $f_i(x_j)$ be the evaluation of subject i for singleton x_j, satisfying:

$$0<f_i(x_j)<1,\ i=1,2,...,N,\ j=1,2,...,n \quad (25)$$

$$\gamma(x_j)=\frac{1+r\left(f(x_j)\right)}{\sum\limits_{k=1}^{n}\left[1+r\left(f(x_k)\right)\right]},\ j=1,2,...,n \quad (26)$$

$$\text{where } r\left(f(x_j)\right)=\frac{S_{y,x_j}}{S_y S_{x_j}} \quad (27)$$

$$S_y^2=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N} y_i\right)^2 \quad (28)$$

$$S_{x_j}^2=\frac{1}{N}\sum_{i=1}^{n}\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N} f_i(x_j)\right]^2 \quad (29)$$

$$S_{y,x_j}=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N}y_i\right)\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N}f_i(x_j)\right] \quad (30)$$

$$\text{satisfying } 0\le\gamma(x_j)\le 1 \text{ and } \sum_{j=1}^{n}\gamma(x_j)=1 \quad (31)$$

then the function $\gamma: X\to[0,1]$ satisfying $\mu(\{x\})=\gamma(x)$, $\forall x\in X$ is a fuzzy support of μ, called γ-support of μ.

VII. CHOQUET INTEGRAL REGRESSION MODELS

A. *Choquet Integral*

Definition 11 Choquet Integral [2-6]

Let μ be a fuzzy measure on a finite set X. The Choquet integral of $f_i: X\to R_+$ with respect to μ for individual i is denoted by

$$\int_C f_i d\mu=\sum_{j=1}^{n}\left[f_i(x_{(j)})-f_i(x_{(j-1)})\right]\mu(A_{(j)}^i),\ i=1,2,...,N \quad (32)$$

where $f_i\left(x_{(0)}\right)=0$, $f_i\left(x_{(j)}\right)$ indicates that the indices have been permuted so that

$$0 \le f_i\left(x_{(1)}\right) \le f_i\left(x_{(2)}\right) \le \ldots \le f_i\left(x_{(n)}\right) \tag{33}$$

$$A_{(j)} = \left\{x_{(j)}, x_{(j+1)}, \ldots, x_{(n)}\right\} \tag{34}$$

B. *Choquet Integral Regression Models*

Definition 12 Choquet Integral Regression Models [7-15]
Let $y_1, y_2, \ldots, y_N$ be global evaluations of N objects and $f_1\left(x_j\right), f_2\left(x_j\right), \ldots, f_N\left(x_j\right),\ j=1,2,\ldots,n$, be their evaluations of x_j, where $f_i : X \to R_+$, $i=1,2,\ldots,N$.

Let μ be a fuzzy measure, $\alpha, \beta \in R$,

$$y_i = \alpha + \beta \int_C f_i dg_\mu + e_i\ ,\ e_i \sim N\left(0, \sigma^2\right)\ ,\ i=1,2,\ldots,N \tag{35}$$

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg\min_{\alpha,\beta}\left[\sum_{i=1}^{N}\left(y_i - \alpha - \beta \int_C f_i dg_\mu\right)^2\right] \tag{36}$$

then $\hat{y}_i = \hat{\alpha} + \hat{\beta}\int f_i dg_\mu,\ i=1,2,\ldots,N$ is called the Choquet integral regression equation of μ, where

$$\hat{\beta} = S_{yf} / S_{ff} \tag{37}$$

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{\beta}\frac{1}{N}\sum_{i=1}^{N}\int f_i dg_\mu \tag{38}$$

$$S_{yf} = \frac{\sum_{i=1}^{N}\left[y_i - \frac{1}{N}\sum_{i=1}^{N} y_i\right]\left[\int f_i dg_{\mu^*} - \frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]}{N-1} \tag{39}$$

$$S_{ff} = \frac{\sum_{i=1}^{N}\left[\int f_i dg_{\mu^*} - \frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]^2}{N-1} \tag{40}$$

VIII. Experiment and Result

A. *Education Data*

The total scores of 60 students from a junior high school in Taiwan are used for this research [9-13]. The examinations of four courses, physics and chemistry, biology, geoscience and mathematics, are used as independent variables, the score of the Basic Competence Test of junior high school is used as a dependent variable.

The data of all variables listed in Table III is applied to evaluate the performances of five Choquet integral regression models with P-measure, λ-measure and δ-measure, L-measure measure and L_δ-measure based on γ-support respectively, a ridge regression model, and a multiple linear regression model by using 5-fold cross validation method to compute the mean square error (MSE) of the dependent variable. The formula of MSE is

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{41}$$

The singleton measures, γ-support of the P-measure, λ-measure, δ-measure, L-measure and L_δ-measure are listed as follows which can be obtained by using the formula (26).

$$\{0.2488,\ 0.2525,\ 0.2439,\ 0.2547\} \tag{42}$$

For any fuzzy measure, μ-measures, once the fuzzy support of the μ-measure is given, all event measures of μ can be found, and then, the Choquet integral based on μ and the Choquet integral regression equation based on μ can also be found by using above corresponding formulae.

The experimental results of seven forecasting models are listed in Table I. We find that the Choquet integral regression model with L_δ-measure based on γ-support outperforms other forecasting regression models.

TABLE I MSE OF REGRESSION MODELS

Regression model	measure	5-fold CV MSE
Choquet Integral Regression model	L_δ	47.5722
	L	48.4610
	δ	48.7672
	λ	49.1832
	p	53.9582
Ridge regression		59.1329
Multiple linear regression		65.0664

B. *Fat Data*

In this study, anthropometric dimensions were measured following a standard protocol [11, 16]. High was measured to the nearest 0.1 cm using anthropometers. Body weight was measured to the nearest 0.1 kg at the same time the bioelectric impedance was measured using a body fat analyzer (TBF310; Tanita, Tokyo, Japan) to estimate the percentage of body fat (%fat). Skinfold thicknesses at biceps, triceps, subscapular, and suprailiac of the right side of body were measured with GMP skinfold calipers (Siber Hegener and Co. Ltd, Switzerland). The measurements were performed by one experienced operator that took two repeated measurements at the test site of the same subject. The mean of the two readings from each site was used to calculate body composition.

A real data set with 128 samples from a elementary school in Taiwan including the independent variables, 4 Skinfold determination values, and the dependent variable, the measurements of the BIA of each student listed in Table IV is applied to evaluate the performances of three Choquet integral regression models with P-measure, λ-measure, δ-measure, L-measure and L_δ-measure based on γ-support respectively, a ridge regression model, and a multiple linear regression model by using 5-fold cross validation method to compute the mean square error (MSE) of the dependent variable.

The singleton measures, γ-support of the P-measure, λ-measure, δ-measure, L-measure and L_δ-measure are listed as follows which can be obtained by using the formula (26).

$$\{0.2396, 0.2466, 0.254, 0.2596\} \tag{43}$$

The formulas of MSE is by using 5-fold cross validation method to compute the mean square error (MSE) of the depen dent variable.

TABLE II MSE OF REGRESSION MODELS

Regression model	measure	5-fold CV MSE
Choquet Integral Regression model	L_δ	13.7136
	L	14.2344
	δ	14.4228
	λ	14.9218
	p	18.3846
Ridge regression		15.7434
Multiple linear regression		16.1122

The experimental results of seven forecasting models are listed in Table II. We also find that the Choquet integral regression model with L_δ -measure based on γ-support outperforms other forecasting regression models.

IX. CONCLUSION

In this paper, a multivalent composed fuzzy measure of L-measure and δ-measure, called L_δ -measure, is proposed. This new measure is proved that it is of closed form with infinitely many solutions, and it can be considered as an extension of the two well known fuzzy measures, λ-measure and P-measure. Furthermore, this improved multivalent fuzzy measure is not only including the additive measure, but also having the same infinitely many measure solutions as L-measure. By using 5-fold cross-validation MSE, two experiments are conducted for comparing the performances of a multiple linear regression model, a ridge regression model, and the Choquet integral regression model with respect to P-measure, λ-measure, δ-measure, L-measure and our proposed L_δ -measure -measure based on γ-support respectively. The result shows that the Choquet integral regression models with respect to the proposed L_δ -measure based on γ-support outperforms other forecasting models.

In the future, we will apply the proposed Choquet integral regression model with the new fuzzy measure based on γ-support to develop multiple classifier system.

ACKNOWLEDGMENT

This paper is partially supported by the grant of National Science Council of Taiwan Government (NSC 97-2410--H-468-014).

REFERENCES

[1] A. E. Hoerl, R. W. Kenard, and K. F. Baldwin, Ridge Regression: Some Simulation, *Communications in Statistics*, vol. 4, No. 2, pp. 105-123, 1975.
[2] Z. Wang, and G. J. Klir, *Fuzzy Measure Theory*, Plenum Press, New York, 1992.
[3] Z. Wang, and G. J. Klir, *Generalized Measure Theory*, Springer Press, New York, 2009.
[4] M. Sugeno, *Theory of Fuzzy Integrals and its Applications*, unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.
[5] L. A. Zadeh, *Fuzzy Sets and Systems*, vol. 1, pp. 3, 1978.
[6] G. Choquet, Theory of Capacities, *Annales de l'Institut Fourier*, vol. 5, pp. 131-295, 1953.
[7] H.-C. Liu, Y.-C Tu,. C.-C. Chen, and W.-S. Weng, "The Choquet Integral with Respect to λ-Measure Based on γ-support", *2008 International Conferences on Machine Learning and Cybernetics*, Kunming, China, July 2008.
[8] H.-C. Liu, "Type 2 L-fuzzy measure and its Choquet integral regression model" *2009 Cross-Strait Conferences on Information Science and Information Technology*, Ilan, Taiwan 22-24 June, 2009.(in Chinese)
[9] H.-C. Liu, "A theoretical approach to the completed L-fuzzy measure", 2009 International Institute of Applied Statistics Studies (2009IIASS) 2nd conference, Qindao, China, July 24-29, 2009.
[10] H.-C. Liu, D.-B. Wu, Y-D Jheng, and T.-W. Sheu, "Delta-Fuzzy Measures and its Application", *Proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational Science*, pp. 289-293, May 2009.
[11] H.-C. Liu, D.-B. Wu, Y-D Jheng, and T.-W. Sheu, "Theory of Multivalent Delta-Fuzzy Measures and its Application", *WSEAS Transactions on Information Science and Application*, vol. 6, no. 6, pp. 1061-1070, July 2009.
[12] J.-I Shieh, H.-H. Wu, H.-C. Liu, Applying Complexity-based Choquet Integral to Evaluate Students' Performance, *Expert Systems with Applications*, 36, pp. 5100-5106, 2009.
[13] S. Srivastava, A. Bansal, D. Chopra, G. Goe, Modeling and Control of a Choquet Fuzzy Integral Based Controller on a Real Time System. *WSEAS Transactions on Systems*, vol. 7, no. 5, pp. 1571-1578, July 2006.

[14] H.-C. Liu, Chen, C.-C., Chen, G.-S., & Jheng, Y.-D. (2007). Power-transformed-measure and its Choquet Integral Regression Model, *Proc. of 7th WSEAS International Conference on Applied Computer Science* (ACS'07), pp.101-104, 2007.
[15] G.S. Chen, Y.-D. Jheng, H.-C. Yao, and H.-C. Liu, Stroke Order Computer-based Assessment with Fuzzy Measure Scoring. *WSEAS Transactions Information Science & Applications*, vol. 2, no. 5, 62-68, 2008.
[16] J. S. Weiner, and J. A. Lourie, *Practical Human Biology*, Academic Press, London, 1981.

Hsiang-Chuan Liu received the Ph.D. degree in Statistics from National Taiwan University, Taiwan. He is a professor at the Department of Bioinformatics, Asia University, Taiwan since August 2001 and also a honored professor at the Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taiwan. He was the President of National Taichung University, Taiwan from 1993 to 2000. Dr. Liu is a member of IEEE since 2007. He has funded research and published articles in the areas of Biostatistics, Bioinformatics, Fuzzy Theory, Educational Measurement, and E-Learning.

Chin-Chun Chen is an Associate Professor and the President of Min-Hwei College, Taiwan. He is now also a doctoral student of Graduate Institute of Educational Measurement and Statistics, Taichung University since 2006. His studies focus in Fuzzy Measure, Grey System, etc.

Der-bang Wu, Ph.D. (M'07) became a Member (M) of **NAUN** in 2007. He was born in Yun-Lin County, Taiwan on July, 1958. He earned his first doctoral degree (Ph.D.) from Department of Mathematics, Graduate School of University of Northern Colorado, Greeley, Colorado, USA in 1994. Also, in order to earn his second doctoral degree, he is a doctoral student of Graduate Institute of Educational Measurement and Statistics, Taichung University since 2007. His studies focus in Fuzzy c-mean, Fuzzy Measure, Grey System, Rough set, Educational Measurement, and e-Learning, etc.

Wu is an Associate Professor of Graduate Institute of Mathematics, Education, Taichung University, Taichung, Taiwan. He was a teacher of Elementary School for 4 years (1979-1983) and work in Taichung University for 24 years (1985-2009), mainly engaged in the design and the development of Mathematics Textbooks, Research Instruments, van Hiele Geometrical Thinking, etc.

Dr. Wu is a member of IEEE, Taiwan Association of Mathematics Education, Taiwan Association of Science Education, and International Group of Psychology of Mathematics Education, etc.

Tian-Wei Sheu received the Ph.D. degree in Mathematics from National Osaka University, Japan in 1990. He is the Dean of College of Education and a professor of Graduate Institute of Educational Measurement, National Taichung University, Taichung, 40306 Taiwan. His studies focus in IRT, Educational Measurement, and e-Learning, etc.

TABLE III THE DATA SET WITH FOUR COURSES AND SCIENCE SCORES OF THE BCT

No.	C1	C2	C3	C4	BCT	No.	C1	C2	C3	C4	BCT
1	72	66	78	72	19	31	66	68	75	74	25
2	86	80	82	81	35	32	68	70	74	76	40
3	56	63	69	75	21	33	57	65	75	70	24
4	78	86	86	86	33	34	74	70	80	75	35
5	66	72	80	76	23	35	49	60	69	64	13
6	68	74	77	80	28	36	51	60	63	64	18
7	74	86	87	88	44	37	58	64	68	66	32
8	54	56	62	68	7	38	73	78	84	81	39
9	71	74	80	77	26	39	56	56	65	61	6
10	68	70	80	75	33	40	61	62	70	70	25
11	53	56	70	63	22	41	57	60	68	64	23
12	67	70	80	75	35	42	57	64	67	70	26
13	70	66	70	74	13	43	50	52	68	60	7
14	60	65	75	70	23	44	84	80	76	72	49
15	68	68	78	76	35	45	62	66	76	71	22
16	58	66	76	71	37	46	70	74	78	82	32
17	61	66	72	78	33	47	69	70	80	75	26
18	68	68	80	74	26	48	63	74	74	74	42
19	56	66	76	71	21	49	66	78	80	82	39
20	59	62	70	78	29	50	67	70	80	75	31
21	62	64	76	70	36	51	56	65	75	70	23
22	71	72	78	75	26	52	50	54	66	60	18
23	74	63	69	75	12	53	71	75	85	80	41
24	59	70	80	76	37	54	74	77	80	85	26
25	75	75	85	80	39	55	71	72	76	80	31
26	73	78	84	81	24	56	60	65	75	70	21
27	62	68	72	74	29	57	59	57	70	68	17
28	77	74	80	76	42	58	50	56	65	68	13
29	63	60	68	69	17	59	72	76	80	78	38
30	56	61	75	68	22	60	81	76	78	80	33

C1 : physics and chemistry
C2 : biology
C3 : geoscience
C4 : mathematics
BCT : Basic Competence Test of nature science

Table IV Measurements of BIA and four skinfold determinations of percent body fat

No	BIA	biceps	triceps	Sub-scapular	Sup-railiac	No	BIA	biceps	triceps	Sub-scapular	Sup-railiac
1	11.8	15.8	16.2	27.6	16.3	65	22.8	27.4	32.2	33.6	24.5
2	17.8	20.4	12.8	35.6	23.0	66	5.4	14.4	8.0	17.2	17.5
3	6.2	10.4	9.4	12.6	8.2	67	8.6	9.6	11.4	14.4	14.8
4	7.0	10.2	5.6	10.6	13.6	68	7.4	9.6	8.0	10.2	13.4
5	5.2	10.8	12.0	13.8	11.8	69	20.2	29.2	28.2	40.0	23.9
6	20.2	24.6	39.8	40.0	34.7	70	28.4	30.6	38.6	39.8	39.5
7	14.8	18.8	22.4	24.8	18.6	71	6.8	10.4	8.8	14.6	15.2
8	12.4	18.2	20.2	27.6	18.4	72	11.0	12.4	10.4	14.4	11.7
9	17.0	20.6	21.2	27.8	21.6	73	20.8	30.8	38.8	40.0	37.7
10	16.6	22.6	18.8	32.6	16.6	74	25.0	27.6	40.0	40.0	29.6
11	10.2	13.4	8.4	12.8	15.1	75	4.8	6.2	5.6	7.0	9.7
12	21.0	26.7	36.2	40.0	23.9	76	8.4	7.0	8.0	11.8	15.6
13	10.0	13.4	17.4	17.8	18.1	77	11.0	14.0	10.0	12.4	17.7
14	9.8	13.8	11.6	19.2	15.5	78	10.6	11.6	7.0	14.4	12.9
15	8.6	10.8	11.4	14.4	8.5	79	12.4	16.8	17.0	25.8	14.9
16	21.8	25.6	36.6	36.8	31.7	80	11.4	14.0	20.8	26.4	25.1
17	25.2	29.4	30.2	30.6	31.2	81	12.8	15.8	20.2	19.8	16.4
18	10.2	11.0	10.2	14.8	12.2	82	12.0	15.6	16.0	22.8	17.3
19	10.4	15.8	13.0	25.6	16.6	83	13.0	20.4	24.0	27.0	26.2
20	17.6	22.6	23.2	34.6	23.2	84	14.2	15.4	22.4	22.8	18.9
21	12.6	12.4	14.2	16.0	14.2	85	11.0	16.4	14.2	15.8	17.9
22	9.0	11.2	9.4	11.8	9.3	86	22.4	29.8	35.0	36.2	28.5
23	12.2	19.2	17.4	27.8	19.7	87	6.4	7.6	8.6	11.4	12.1
24	4.6	7.0	8.8	11.2	7.2	88	6.8	10.6	9.6	14.6	12.2
25	6.4	8.8	11.0	12.6	10.8	89	16.2	18.4	27.2	27.4	24.9
26	23.8	29.0	37.0	35.0	30.7	90	22.4	26.8	25.4	33.4	30.4
27	8.4	15.8	17.8	23.0	21.6	91	9.6	11.2	10.4	18.0	11.7
28	12.2	16.6	16.4	20.6	18.7	92	10.8	17.2	24.0	24.8	21.9
29	7.2	12.8	8.6	18.6	15.2	93	13.0	16.2	12.4	18.4	14.2
30	21.4	31.2	31.4	39.4	28.6	94	5.6	12.4	11.4	15.6	14.5
31	18.2	23.0	40.0	40.0	28.2	95	19.4	25.0	36.2	39.0	29.9
32	9.2	12.6	40.0	17.8	16.0	96	14.4	22.4	29.8	35.0	24.8
33	10.2	18.8	17.8	20.8	18.4	97	25.4	29.4	37.0	40.0	24.6
34	19.2	24.4	35.2	35.0	34.1	98	9.4	11.2	11.4	12.4	8.9
35	6.8	12.0	8.0	14.4	16.1	99	17.4	22.6	19.4	31.6	22.7
36	16.8	20.8	25.6	27.8	20.7	100	24.0	30.8	40.0	40.0	29.4
37	35.8	38.6	40.0	40.0	30.1	101	3.8	6.0	6.4	6.8	10.8
38	10.0	11.6	10.4	18.6	8.3	102	11.0	19.4	11.6	18.4	13.7
39	5.4	12.2	12.4	21.4	19.2	103	22.6	24.4	40.0	40.0	33.3
40	11.2	18.0	23.6	30.8	22.1	104	9.2	10.0	11.0	19.2	13.4
41	5.4	11.2	6.8	11.6	11.9	105	18.2	19.0	31.0	29.4	24.5
42	7.6	8.4	9.4	13.6	12.8	106	6.8	12.4	14.0	17.8	14.1
43	6.6	9.8	9.6	12.0	9.3	107	7.4	11.6	10.0	16.0	11.0
44	32.4	37.2	40.0	40.0	18.2	108	9.2	10.6	12.4	14.4	12.7
45	7.8	14.0	11.0	17.8	31.3	109	29.4	23.6	39.8	40.0	37.4
46	17.8	26.6	34.2	40.0	22.8	110	6.8	7.8	9.8	12.8	10.9
47	22.0	27.8	38.2	39.4	23.6	111	12.4	14.6	15.8	19.8	16.9
48	14.4	15.8	18.8	23.8	14.9	112	8.2	9.8	9.2	16.0	14.5
49	15.8	18.4	21.4	24.0	24.9	113	16.4	20.8	25.2	30.4	24.9
50	7.4	12.8	10.2	17.0	14.3	114	9.4	11.4	12.0	21.8	14.3
51	16.2	29.0	21.6	29.8	24.1	115	16.4	22.4	33.2	36.8	25.1
52	6.0	7.4	7.6	9.8	8.6	116	7.0	11.4	13.8	17.4	11.2
53	12.2	15.4	16.2	18.8	17.8	117	10.4	12.6	14.8	23.8	18.0
54	11.6	12.0	9.8	13.0	8.9	118	5.6	8.2	10.2	8.6	7.7
55	17.8	22.6	38.0	31.0	24.8	119	10.8	11.8	17.8	21.2	19.9
56	13.2	16.8	18.6	23.4	20.7	120	9.6	15.8	14.4	19.4	18.6
57	4.4	7.2	8.2	9.8	14.3	121	5.0	6.8	7.4	9.4	6.0

58	16.2	21.8	28.2	32.6	27.2	122	9.8	12.2	12.4	15.4	13.5
59	11.4	19.4	28.8	32.8	22.3	123	13.8	18.0	16.4	21.0	19.3
60	11.2	13.0	18.8	22.6	21.9	124	8.8	12.8	9.8	11.8	13.3
61	8.6	11.4	7.2	10.2	7.5	125	15.8	21.0	35.4	39.8	27.3
61	20.4	26.2	31.0	32.8	25.8	126	10.8	16.6	15.6	23.2	16.5
63	7.0	8.8	11.6	9.4	12.0	127	9.0	10.6	10.0	16.8	11.9
64	14.6	17.4	12.8	16.8	14.7	128	8.8	12.4	10.0	10.8	11.3

A theoretical approach to the completed L-fuzzy measure

Hsiang-Chuan Liu

Department of Bioinformatics, Asia University, Taiwan

Abstract **The well known fuzzy measures, λ-measure and P-measure, have only one formulaic solution. An multivalent fuzzy measure with infinitely many solutions of closed form based on P-measure was proposed by our previous work, called L-measure, but L-measure is not a completed multivalent fuzzy measure, In this paper, A further improved fuzzy measure, called completed L-measure, is proposed. Some important properties of this new multivalent fuzzy measure are also proposed.**

Key words λ-measure , P-measure**,** L-measure, completed L-measure

1 Introduction

When interactions among independent variables exist in forecasting problems, the performance of the multiple linear regression models is not good enough. The traditional improved methods exploited the ridge regression models [1]. Recently, the Choquet integral regression models based on some univalent or multivalent fuzzy measures [2,3,4] were used to improve this situation. The well known fuzzy measures, λ-measure [5] and P-measure [6], have only one formulaic solution of fuzzy measure. A multivalent fuzzy measure with infinitely many solutions of closed form based on P-measure was proposed by our previous work, called L-measure[3], but it is not a completed multivalent fuzzy measure. In this paper, an improved multivalent fuzzy measure, called completed L-measure, is proposed. Some important properties of this new multivalent fuzzy measure are also discussed.

This paper is organized as followings: The basis concepts of fuzzy measures are introduced in section 2; Comparison between two fuzzy measures is introduced in section 3; L-measures is introduced in section 4; completed L –measure and its properties is described in section 5; and final section is for conclusions and the future works.

2 Fuzzy Measures

The well known fuzzy measures, the λ-measure proposed by Sugeno in 1974, and P-measure proposed by Zadah in 1978, are concise introduced as follows

2.1 Definition of fuzzy measures [2, 5]

A fuzzy measure μ on a finite set X is a set function $g_\mu : 2^X \to [0,1]$ satisfying the following axioms: $g_\mu(\phi)=0$, $g_\mu(X)=1$ (boundary conditions) (1)

$$A \subseteq B \Rightarrow g_\mu(A) \le g_\mu(B) \quad \text{(monotonicity)} \tag{2}$$

2.2 Fuzzy density function [2, 6]

A fuzzy density function $s(x)$ of a fuzzy measure μ on a finite set X is a measurable function

$s: X \to [0,1]$ satisfying: $$s(x) = g_\mu(\{x\}),\ x \in X \tag{3}$$

2.3 λ-measure [5]

For each given fuzzy density function $s(x)$, a **λ**-measure on a finite set X is a measurable function,

$g_\lambda: 2^X \to [0,1]$, satisfying:

(i) $A, B \in 2^X$, $A \cap B = \phi,\ A \cup B \neq X \quad \Rightarrow g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$ (4)

(ii) $\prod_{x \in X}\left[1 + \lambda s(x)\right] = \lambda + 1 > 0,\ s(x) = g_\lambda(\{x\})$ (5)

2.4 P-measure [6]

For each given fuzzy density function, $s(x)$, a P-measure on a finite set X is a set function,

$g_P: 2^X \to [0,1]$, satisfying: $\forall A \in 2^X \Rightarrow g_P(A) = \max_{x \in A}\{s(x)\} = \max_{x \in A} g_P(\{x\})$ (6)

3 Comparison between two fuzzy measures [4]

Definition 2 $\mu_1 - measure \leq \mu_2 - masure$, $\mu_2 - measure \geq \mu_1 - measure$

For any given fuzzy density function, $s(x)$, on a finite set, X, If μ_1 and μ_2 are two fuzzy measures, satisfying $g_{\mu_1}(A) \leq g_{\mu_2}(A),\ \forall A \subset X$, then we say that μ_1 -measure is not larger than μ_2 -measure, or μ_2 -measure is not smaller than μ_1 -measure, denoted as $\mu_1 - measure \leq \mu_2 - masure$,or $\mu_2 - measure \geq \mu_1 - measure$

Theorem 1 For any given fuzzy density function, $s(x)$, on a finite set, X, P-measure is not larger than any other fuzzy measure, μ, that is $P - measure \leq \mu - measure$

Proof. For the same given fuzzy density function, $s(x)$, We have $g_P(\{x\}) = g_\mu(\{x\}) = s(x), \forall x \in X$

Let $\forall A \subset X$, if $|A| \leq 1$, it is trivial, now suppose $|A| = k \leq |X|$, and $A = \{x_1, x_2, ..., x_k\} \subset X$

From the monotonicity, we have

$$\begin{aligned} & g_\mu(A) = g_\mu(\{x_1, x_2, ..., x_k\}) \geq g_\mu(\{x_1\}), g_\mu(\{x_2\}), ..., g_\mu(\{x_k\}) \\ & \Rightarrow g_\mu(\{x_1, x_2, ..., x_k\}) \geq \max\left[g_\mu(\{x_1\}), g_\mu(\{x_2\}), ..., g_\mu(\{x_k\})\right] \\ & = \max\left[g_P(\{x_1\}), g_P(\{x_2\}), ..., g_P(\{x_k\})\right] = g_P(\{x_1, x_2, .., x_k\}) = g_P(A) \end{aligned} \tag{7}$$

That is $g_\mu(A) \geq g_P(A)$, the proof is completed

Definition 3 B- measure [4]

For any given fuzzy density function, $s(x)$, on a finite set, X, a B-measure is a set function,

$g_B : 2^X \to [0,1]$, satisfying: $$g_B(A) = \begin{cases} 0 & A = \phi \\ s(x) & A = \{x\}, x \in X \\ 1 & |A| > 1, A \subset X \end{cases} \quad (8)$$

Theorem 2 For any given fuzzy density function, $s(x)$, on a finite set, X, B-measure is not smaller than any fuzzy measure, μ, that is $B - measure \geq \mu - measure$

Proof. It is trivial.

4 L-measures [3, 4]

Definition 4 L-measure

Let $L \in [0, \infty)$, for each given fuzzy density function $s(x)$ on a finite set X, a L-measure is a set function $g_L : 2^X \to [0,1]$ satisfying:

$$^{\forall} A \subset X, |X| - |A| + (|A|-1)L > 0 \Rightarrow g_L(A) = \max_{x \in A}\{s(x)\} + \frac{(|A|-1)L\sum_{x \in A} s(x)\left[1 - \max_{x \in A}\{s(x)\}\right]}{\left[|X| - |A| + (|A|-1)L\right]\sum_{x \in X} s(x)} \quad (9)$$

Theorem 3

(i)L-measure is an increasing continuous function on L.

(ii)If $L = 0$, then L-measure is just the P-measure

(iii)If $L > 0$, then L-measure is not smaller than P-measure

5 Completed L-measures

Definition 5 Completed fuzzy measure

A fuzzy measure is completed, if the P-measure and B-measure are the lower and upper limit fuzzy measures of this measure, respectively.

Theorem 4 L-measure is not a completed fuzzy measure

Proof Since $\lim_{L \to \infty} g_L(A) = \max_{x \in A}\{s(x)\} + \frac{\sum_{x \in A} s(x)}{\sum_{x \in X} s(x)}\left[1 - \max_{x \in A}\{s(x)\}\right] \neq g_B(A)$, the proof is completed.

Definition 6 Completed L- measure, L_C -measure

Let $L \in [0, \infty)$, for each given fuzzy density function $s(x)$ on a finite set X, a completed L-measure or L_C -measure is a set function $g_{L_C} : 2^X \to [0,1]$ satisfying:

$$^{\forall} A \subset X, |X| - |A| + (|A|-1)L > 0 \Rightarrow g_{L_C}(A) = \max_{x \in A}\{s(x)\} + \frac{(|A|-1)L\sum_{x \in A} s(x)\left[1 - \max_{x \in A}\{s(x)\}\right]}{\left[|X| - |A|\right]\sum_{x \in X} s(x) + (|A|-1)L\sum_{x \in A} s(x)} \quad (10)$$

Theorem 5 L_C -measure is a fuzzy measure

Proof:

(I) (To prove the boundary conditions; $0 \leq g_{L_C}(A) \leq 1, \forall A \subset X$)

If $|A| \le 1$, It is trivial. Let $|A| > 1$, since $L_C \in [0,\infty)$ We can obtain

$$0 \le \frac{(|A|-1)L\sum_{x\in A} s(x)}{[|X|-|A|]\sum_{x\in X} s(x)+(|A|-1)L\sum_{x\in A} s(x)} \le 1 \Rightarrow 0 \le \max_{x\in A}\{s(x)\} + \frac{(|A|-1)L\sum_{x\in A} s(x)\left[1-\max_{x\in A}\{s(x)\}\right]}{[|X|-|A|]\sum_{x\in X} s(x)+(|A|-1)L\sum_{x\in A} s(x)} \le 1 \quad (11)$$

Therefore $$0 \le g_{L_C}(A) = \max_{x\in A}\{s(x)\} + \frac{(|A|-1)L\sum_{x\in A} s(x)\left[1-\max_{x\in A}\{s(x)\}\right]}{[|X|-|A|]\sum_{x\in X} s(x)+(|A|-1)L\sum_{x\in A} s(x)} \le 1 \quad (12)$$

That is $0 \le g_{L_C}(A) \le 1, \forall A \subset X, \forall L \in [0,\infty)$

(II) (To prove the monotonicity;)

Let $L \in [0,\infty), A \subset B \subset X$ if $|A| \le 1$, it is trivial.

If $|A| > 1$, let $\max_{x\in B}\{s(x)\} = \max_{x\in A}\{s(x)\} + a,\ where\ 0 \le a \le 1$, (13)

we can obtain

$$g_{L_C}(B) - g_{L_C}(A) = a\left[1 - \frac{(|B|-1)L\sum_{x\in B} s(x)}{[|X|-|B|]\sum_{x\in X} s(x)+(|B|-1)L\sum_{x\in B} s(x)}\right] + L\left[1-\max_{x\in A}\{s(x)\}\right]\left[f(B)-f(A)\right] \ge 0 \quad (14)$$

where

$$\begin{aligned} f(B)-f(A) &= \frac{\sum_{x\in X} s(x)}{D_B D_A}\left[(|B|-1)\sum_{x\in B} s(x)[|X|-|A|] - \left[(|A|-1)\sum_{x\in A} s(x)\right][|X|-|B|]\right] > 0 \\ D_B &= [|X|-|B|]\sum_{x\in X} s(x)+(|B|-1)L\sum_{x\in B} s(x) > 0,\quad D_A = [|X|-|A|]\sum_{x\in X} s(x)+(|A|-1)L\sum_{x\in A} s(x) > 0 \end{aligned} \quad (15)$$

The proof is completed.

Theorem 6 Basic properties of L_C -measure

. For the same given fuzzy density function

(i) L_C -measure is an increasing continuous function on L.

(ii) if $L = 0$ then L_C -measure is just the P-measure.

(iii) P- measure ≤ L -measure ≤ L_C -measure ≤ B- measure

(iv) $\forall A \subset X \Rightarrow \lim_{L\to\infty} g_{L_C}(A) = g_B(A)$

(v) L_C -measure is a completed multivalent fuzzy measure.

Proof;

(i) let $f(L) = g_{L_C}(A)$ and $L \in (0,\infty)$

then $$f'(L) = \frac{[|X|-|A|](|A|-1)\sum_{x\in X} s(x)\sum_{x\in A} s(x)\left[1-\max_{x\in A}\{s(x)\}\right]}{\left[[|X|-|A|]\sum_{x\in X} s(x)+(|A|-1)L\sum_{x\in A} s(x)\right]^2} \ge 0 \quad (20)$$

Therefore L_C -measure is an increasing function on L.

(ii) (iv) and (v) are trivial.

(iii) From Theorem 1 and Theorem 2, we know that

$$\text{P- measure} \le L\text{-measure}, L_C\text{-measure} \le \text{B- measure} \tag{21}$$

Now, to prove that L-measure $\le L_C$-measure

Let $\forall A \subset X$, since $\sum_{x\in A} s(x) \le \sum_{x\in X} s(x)$, then

$$\frac{(|A|-1)L\sum_{x\in A} s(x)\left[1-\max_{x\in A}\{s(x)\}\right]}{\left[|X|-|A|+(|A|-1)L\right]\sum_{x\in X} s(x)} \le \frac{(|A|-1)L\sum_{x\in A} s(x)\left[1-\max_{x\in A}\{s(x)\}\right]}{\left[|X|-|A|\right]\sum_{x\in X} s(x)+(|A|-1)L\sum_{x\in A} s(x)} \tag{22}$$

We can obtain $g_L(A) \le g_{L_C}(A), \quad \forall A \subset X$,

That is L-measure $\le L_C$-measure, the proof is completed.

6 Conclusion

In this paper, an improved multivalent fuzzy measure, completed L-measure, is proposed. Some important properties of this new fuzzy measure are also proposed.

The Choquet integral regression model with the proposed new measure has been practically applied to an educational data. The experiment is not report here because of the lack of space.

In the future, we will apply the Choquet integral regression model with the proposed new fuzzy measure to develop multiple classifier system.

References

[1] A. E. Hoerl, R. W. Kenard, and K. F. Baldwin, Ridge regression: Some simulation, Communications in Statistics, vol. 4, No. 2, pp. 105-123, 1975.

[2] Hsiang-Chuan Liu, The Choquet integral regression model based on ν-measure, Journal of Educational Measurement and Statistics Vol.15, pp1-14, 2006, Taichung, Taiwan. (in Chinese)

[3] H.-C. Liu, Y.-C. Tu, W.-C. Lin, and C. C. Chen, Choquet integral regression model based on L-Measure and γ-Support, Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition. pp.777-782, 2008.

[4] H.-C. Liu, Type-2 L-measure and its Choquet integral regression model, 2009 Cross-Strait Conferences on Information Science and Information Technology, Ilan, Taiwan, 22-24 June, 2009.(in Chinese)

[5] M. Sugeno, Theory of fuzzy integrals and its applications, unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[6] L. A. Zadeh, Fuzzy Sets as a basis for a theory of possibility, Fuzzy Sets and Systems, vol. 1, pp. 3-28, 1978.

Author in brief:

Hsiang-Chuan Liu Ph. D.，Professor, Department of Bioinformatics, Asia University, President of National Taichung University, Taiwan from 1993 to 2000. He has funded research and published articles in the areas of Biostatistics, Bioinformatics and Fuzzy Theory, E-mail: lhc@asia.edu.tw

Acknowledgement:

This paper is partially supported by the National Science Council grant (NSC 97-2410--H-468-014).

.

參加三次國際會議發表 EI 級論文 12 篇心得報告

報告人姓名：劉湘川

會議一、

名稱：2008 International Conference on Wavelet Analysis and Pattern Recognition.

期間及地點：August. 30-31, 2008, Hong Kong, .

發表論文 5 篇題目如下：

1. Hsiang-Chuan Liu, Yu-Chieh Tu, Wen-Chih Lin, and Chin-Chun Chen (2008). Choquet integral regression model based on L-Measure and γ-Support. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* Volume: 2, pp.777-782. ISBN: 978-1-4244-2238-8 INSPEC Accession Number: 10299007. **(EI paper**)

2. Hsiang-Chuan Liu, Yu-Du Jheng, Guey-Shya Chen and Bai-Cheng Jeng. (2008) Choquet Integral Logistic Regression Algorithms Based on L-Measure and γ-Support. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* Volume: 2, pp.771-776. ISBN: 978-1-4244-2238-8. INSPEC Accession Number: 10299006. **(EI paper**)

3. Hsiang-Chuan Liu, Jeng-Ming Yih, Der-Bang Wu, Shin-Wu Liu. (2008). Fuzzy Possibility C-Mean Clustering Algorithms Based on completed Mahalonobis Distances. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* Volume: 1, pp.50-55. ISBN: 978-1-4244-2238-8 INSPEC Accession Number: 10298940. **(EI paper**)

4.. Hsiang-Chuan Liu, Ya-Ching Chiu, Chien-Hsiung Liao and Tung-Sheng Liu. (2008). An improved SVM algorithm based on Normalization and Liu-Transfomation. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition.* Volume: 2, pp. 470-473. ISBN:978-1-4244-2238-8.INSPEC Accession Number: 10288186. **(EI paper**)

5. Hsiang-Chuan Liu, Shin-Wu Liu, Pei-Chun Chang, Wen-Chun Huang and Chien-Hsiung Liao. (2008). A novel classifier for influenza A viruses based on SVM and Logistic regression. *Proceedings of 2008 International Conference on Wavelet Analysis and Pattern Recognition*. Volume: 1, pp.287-291. ISBN: 978-1-4244-2238-8. INSPEC Accession Number: 10288176. **(EI paper**)

會議二、

名稱：Fifth International Conference on Fuzzy Systems and Knowledge Discovery
Fourth International Conference on Natural Computing

期間及地點：October.18-20, 2008, Jinan, China.

發表論文 4 篇題目如下：

6. **Hsiang-Chuan Liu**, Yu-Chieh Tu, Wen-Chun Huang and Chin-Chun Chen. (2008). "The Choquet Integral with Respect to R-Measure Based on γ-Support". *The 5th International Conference on Fuzzy Systems and Knowledge Discovery.* 18-20 October 2008, Jinan, China. Vol. 1, pp. 645-649. ISBN: 978-0-7695-3305-6. INSPEC Accession Number: 10384941. (EI paper).

7. **Hsiang-Chuan Liu**, Der-Bang Wu, Jeng-Ming Yih, and Shin-Wu Liu. (2008). "Fuzzy c-mean algorithm based on complete mahalanobis distances". *The 5th International Conference on Fuzzy Systems and Knowledge Discovery.* 18-20 October 2008, Jinan, China. Vol. 1, pp. 87-91. ISBN: 978-0-7695-3305-6. INSPEC Accession Number: 10384941. (EI paper).

8. **Hsiang-Chuan Liu**, Horng-Jinh Chang, Kuei-Jen Lee, Jiunn-I Shieh,Wen-Chun Huang and Shin-Ming Huang. (2008). "A Novel Classification Algorithm of Thermostable Proteins by Using Hurst Exponent and SVM Classifier" *The 4th International Conference on Natural Computation,.* 18-20 October 2008, Jinan, China. Vol. 5, pp. 24-28. ISBN:978-0-7695-3305-6. (INSPEC Accession Number: 10398844). (EI paper)

9. Pei-Chun Chang, Kuei-Jen Lee, Jiunn-I Shieh, Chung-Hung Li, Jing-Doo Wang and **Hsiang-Chuan Liu**. (2008). "Physiochemical contraints in Influeng A Hemagglatinin". Proceedings of the 2008 Fourth International Conference on Natural Computation, Vol. 5, pp. 85-89. ISBN:978-0-7695-3304-9. (EI paper).

會議三、

名稱：Eighth International Conference on Machine Learning and Sybernetics

期間及地點：July. 12-15, 2009, Baoding, China.

發表論文 3 篇題目如下：

10. **Hsiang-Chuan Liu**, Wei-Sung Chen, Yu-Chieh Tu, Yen-Kuei Yu, "Choquet Integral Regression Model Based on High-Order L-measure" Proceedings of the 2009 International Conference on Machine Learning and Cybernetics (ICMLC 2009). Volume: 6, page(s): 3177-3182, 2009. ISBN: 978-1-4244-3702-3. INSPEC Accession Number: 10845957..(EI paper)

11. Horng-Jinh Chang, **Hsiang-Chuan Liu**, Shang-Wen Tseng and Fengming M. Chang, "A comparison on Choquet integral with respect to different information-based fuzzy measures", Proceedings of the 2009 International Conference on Machine Learning and Cybernetics (ICMLC 2009). Volume: 6, pp. 3161-3166. ISBN: 978-1-4244-3702-3. INSPEC Accession Number: 10845958. (EI paper)

12. Jing-Doo Wang, **Hsiang-Chuan Liu**, Yao-Chug Shi, "A novel approach for evaluating class structure ambiguity", Proceedings of the 2009 International Conference on Machine Learning and Cybernetics (ICMLC 2009). Volume 3 page(s): 1550-1555, 2009. ISBN: 978-1-4244-3702-3. INSPEC Accession Number: 1084570. (EI paper)

與會心得：

1. 感謝國科會補助參加此三次國際會議，各主辦單均很用心在籌劃這次的會議，充分讓人感覺到他們的熱情。
2. 此三次國際會議個人計發表12篇論文。均為EI級論文，其中模糊測度所發展之相關理論模式均以教育測驗資料為應用實例‧ 且在後續研究中將有進一步之發展 。
3. 參加三次國際會議，與世界各國研究與技術開發專業人才進行交流，並從展覽會中收集到一些與研究相關資料，獲益良多，亦看到了他國學者的在會議中互動的方式和技巧，也了解到自己不足及可以改進的地方。
4. 攜回資料名稱及內容：

   會議論文初稿全集 CD，會議議程冊，及未來相關國際會議的 Call for Paper 海報。

Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, 30-31 Aug. 2008

CHOQUET INTEGRAL REGRESSION MODEL BASED ON L-MEASURE AND γ- SUPPORT

HSIANG-CHUAN LIU[1], YU-CHIEH TU[2], WEN-CHIH LIN[3], CHIN-CHUN CHEN[2,4]

[1]Department of Bioinformatics, Asia University, Taiwan
[2]Graduate Institute of Educational Measurement and Statistics, Taichung University, Taiwan
[3]Department of Computer Science and Information Engineering, Asia University, Taiwan
[4]General Education, Min-Hwei College of Health Care Management, Taiwan
E-MAIL:lhc@asia.edu.tw, tuyuchieh@yahoo.com.tw, kbc@mail.ntcu.edu.tw, cc_chen0336@yahoo.com.tw

Abstract:

When the multicollinearity within independent variables occurs in the multiple regression models, its performance will always be poor. Replacing the above models with the ridge regression model is the traditional improved method. In our previous work, we found that, the Choquet integral regression model with R-measure based on the new support, γ-support, proposed by us has the best performance than before. In this study, for finding the further improved model, we replaced R-measure with our new fuzzy measure, L-measure in Choquet integral regression model with the new support, γ-support. For comparing the Choquet integral regression model with P-measure, λ-measure, R-measure and L-measure based on two different fuzzy supports, V-support and γ-support, respectively, the traditional multiple regression model and the ridge regression model, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted. Experimental result shows that the Choquet integral regression model with L-measure based on γ-support has the best performance.

Keywords:

Fuzzy measure; R-measure; L-measure; fuzzy support; γ-support

1. Introduction

When interactions among independent variables exist in forecasting problems, the performance of the multiple linear regression models is poor. The traditional improved methods exploited the ridge regression models [1]. Recently, some Choquet integral regression models based on different fuzzy measures were used by our previous works to further improve this situation [2], [3], [4], [5], [6]

In our previous works [7], [8], we found that if the Choquet integral regression model based on the same fuzzy measure is derived from different fuzzy support, then it may have different performances, in other words, the better performance of a Choquet integral regression model is not only derived from a better fuzzy measure but also first derived from a better fuzzy support. Hence, before we find the better fuzzy measure of a Choquet integral regression model, we need first to find a better fuzzy support of the same fuzzy measure of that Choquet integral regression model. And we found that the Choquet integral regression model with R-measure and λ-measure based on the new support, γ-support, proposed by us has the better than before, and the model with R-measure based on γ-support is better than the model with λ-measure based on γ-support.

In this study, the Choquet integral regression model with P-measure, λ-measure, R-measure and our new fuzzy measure, L-measure based on the V-support and γ-support, respectively, were considered. For comparing the performances of the above different Choquet integral regression models with the multiple regression model and the ridge regression model, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted.

This paper is organized as followings: The multiple linear regression and ridge regression are introduced in section 2, two well known fuzzy measure, P-measure, λ-measure and our R-measure are introduced in section 3, our new measure, L-measure, is introduced in section 4, two kind fuzzy supports: V-support and γ-support are described in section 5. The Choquet integral regression model based on fuzzy measures are described in section 6. Experiment and result are described in section 7, and final section is for conclusions and future works.

2. The multiple linear regression, ridge regression [1]

Let $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{\varepsilon} \sim \mathrm{N}\left(\underline{0}, \sigma^2 I_n\right)$ be a multiple linear model, $\hat{\underline{\beta}} = \left(X'X\right)^{-1} X'Y$ be the estimated regression coefficient vector, and $\hat{\underline{\beta}}_k = (X'X + kI_n)^{-1} X'Y$ be the estimated ridge regression coefficient vector, Kenard and Baldwin [1] suggested

$$\hat{k} = \frac{n\hat{\sigma}^2}{\hat{\underline{\beta}}'\hat{\underline{\beta}}}. \quad (1)$$

3. Fuzzy measures

The well known fuzzy measures, P-measure proposed by Zadah in 1978, and the λ-measure proposed by Sugeno in 1974, are concise introduced as follows.

3.1. Fuzzy measures [9], [10], [11]

A fuzzy measure μ on a finite set X is a set function $\mu : 2^X \to [0,1]$ satisfying the following axioms:

(i) $\mu(\phi) = 0$, $\mu(X) = 1$ (boundary conditions) (2)

(ii) $A \subseteq B \Rightarrow \mu(A) \le \mu(B)$ (monotonicity) (3)

3.2. Singleton measures [4], [5], [6]

A singleton measure of a fuzzy measure μ on a finite set X is a function $s : X \to [0,1]$ satisfying:

$$s(x) = \mu(\{x\}), x \in X \quad (4)$$

$s(x)$ is called the density of singleton x.

3.3. P-measure [12]

For given singleton measures s, a P-measure, g_P, is a fuzzy measure on a finite set X, satisfying:

$$^{\forall}A \in 2^X \Rightarrow g_P(A) = \max_{x \in A} s(x) = \max_{x \in A} g_P(\{x\}) \quad (5)$$

3.4. λ-measure [10], [11]

For given singleton measures s, a λ-measure, g_λ, is a fuzzy measure on a finite set X, satisfying:

(i) $A, B \in 2^X$, $A \cap B = \phi$, $A \cup B \ne X$

$$\Rightarrow g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B) \quad (6)$$

(ii) $\prod_{i=1}^{n} [1 + \lambda s(x_i)] = \lambda + 1 > 0,\ s(x_i) = g_\lambda(\{x_i\})$ (7)

Note that once the singleton measure is known, we can obtain the values of λ uniquely by using the previous polynomial equation. In other words, λ-measure has a unique solution without closed form.

3.5. R-measure [4]

For given singleton measure s, a R-measure, g_R, is a fuzzy measure on a finite set X, $|X| = n$, satisfying:

(i) $R \in [0, \infty)$ (8)

(ii) $\sum_{x \in X} s(x) = \sum_{x \in X} g_R(\{x\}) = 1$ (9)

(iii) $^{\forall}A \subset X, n - |A| + (|A| - 1)R > 0$

$$\Rightarrow g_R(A) = \max_{x \in A}[s(x)] + \frac{(|A|-1)R\sum_{x \in A} s(x)}{[n - |A| + (|A|-1)R]}\left[1 - \max_{x \in X}[s(x)]\right] \quad (10)$$

[Property]

(i) R-measure has infinitely many solutions with closed form.

(ii) When R=0, the R-measure is just a P-measure with closed form.

(iii) g_R is an increasing function of R.

4. L-measure [6]

Liu, Hsiang-Chuan et al get the regular value in the most right end of the definition of R-measure, who replace dynamic value, propose it obviously for being sensitive “L-measure”.

For given singleton measure s, a L-measure, g_L, is a fuzzy measure on a finite set X, $|X| = n$, satisfying:

(i) $L \in [0, \infty)$ (11)

(ii) $\sum_{x \in X} s(x) = \sum_{x \in X} g_L(\{x\}) = 1$ (12)

(iii) $^{\forall}A \subset X, n - |A| + (|A| - 1)L > 0$

$$\Rightarrow g_L(A) = \max_{x \in A}[s(x)] + \frac{(|A|-1)L\sum_{x \in A} s(x)}{[n - |A| + (|A|-1)L]}\left[1 - \max_{x \in A}[s(x)]\right] \quad (13)$$

[Property]

(i) L-measure has infinitely many solutions with closed form.

(ii) When L=0, the L-measure is just a P-measure with closed form.

(iii) g_L is an increasing function of L.

(iv) When the decision coefficient value of L-measure is the same as this of R-measure, the number of measure values for any incident of L-measure always is more than

R-measure.

5. Fuzzy supports

For given singleton measures s of a fuzzy measure μ on a finite set X, if $\sum_{x\in X} s(x)=1$, then s is called a fuzzy support measure of μ, or a fuzzy support of μ, or a support of μ. Two kinds of fuzzy supports are introduced as below.

5.1. V-support [7], [8]

Let μ be a fuzzy measure on a finite set, $X=\{x_1,x_2,...,x_n\}$ be the set of n courses, $f_1(x_j), f_2(x_j),..., f_N(x_j),\ j=1,2,...,n$ be the evaluating scores of subject i for singleton x_j, satisfying:

$$0<f_i(x_j)<1, i=1,2,...,N,\ j=1,2,...,n \tag{14}$$

If $$V(x_j)=\frac{V_{ar}(f(x_j))}{\sum_{k=1}^{n} V_{ar}(f(x_k))},\ j=1,2,...,n, \tag{15}$$

where $$V_{ar}(f(x_j))=\frac{1}{N}\sum_{i=1}^{N}\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N} f_i(x_j)\right]^2 \tag{16}$$

satisfying $$0\le V(x_j)\le 1 \text{ and } \sum_{j=1}^{n} V(x_j)=1 \tag{17}$$

then the function $V: X\to[0,1]$ satisfying $\mu(\{x\})=V(x),\ \forall x\in X$ is a fuzzy support of μ, called V-support of μ.

5.2. γ- support [7], [8]

Let μ be a fuzzy measure on a finite set $X=\{x_1,x_2,...,x_n\}$, y_i be global response of subject i and $f_i(x_j)$ be the evaluation of subject i for singleton x_j, satisfying:

$$0<f_i(x_j)<1, i=1,2,...,N,\ j=1,2,...,n \tag{18}$$

If $$\gamma(x_j)=\frac{1+r(f(x_j))}{\sum_{k=1}^{n}\left[1+r(f(x_k))\right]},\ j=1,2,...,n \tag{19}$$

where $$r(f(x_j))=\frac{S_{y,x_j}}{S_y S_{x_j}} \tag{20}$$

$$S_y^2=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N} y_i\right)^2 \tag{21}$$

$$S_{x_j}^2=\frac{1}{N}\sum_{i=1}^{n}\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N} f_i(x_j)\right]^2 \tag{22}$$

$$S_{y,x_j}=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N} y_i\right)\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N} f_i(x_j)\right] \tag{23}$$

satisfying $$0\le \gamma(x_j)\le 1 \text{ and } \sum_{j=1}^{n}\gamma(x_j)=1 \tag{24}$$

then the function $\gamma: X\to[0,1]$ satisfying $\mu(\{x\})=\gamma(x)$, $\forall x\in X$ is a fuzzy support of μ, called γ-support of μ.

6. Choquet integral regression models

6.1. Choquet integral [4], [11], [12]

Let μ be a fuzzy measure on a finite set X. The Choquet integral of $f_i: X\to R_+$ with respect to μ for individual i is denoted by

$$\int_C f_i d\mu=\sum_{j=1}^{n}\left[f_i(x_{(j)})-f_i(x_{(j-1)})\right]\mu(A_{(j)}^i),\ i=1,2,...,N \tag{25}$$

where $f_i(x_{(0)})=0$, $f_i(x_{(j)})$ indicates that the indices have been permuted so that

$$0\le f_i(x_{(1)})\le f_i(x_{(2)})\le ...\le f_i(x_{(n)}) \tag{26}$$

$$A_{(j)}=\{x_{(j)},x_{(j+1)},...,x_{(n)}\} \tag{27}$$

6.2. Choquet integral regression models [2], [3], [4], [5], [6], [7], [8]

Let $y_1,y_2,...,y_N$ be global evaluations of N objects and $f_1(x_j), f_2(x_j),..., f_N(x_j),\ j=1,2,...,n$, be their evaluations of x_j, where $f_i: X\to R_+,\ i=1,2,...,N$.

Let μ be a fuzzy measure, $\alpha,\beta\in R$,

$$y_i=\alpha+\beta\int_C f_i dg_\mu+e_i,\quad e_i\sim N(0,\sigma^2),\quad i=1,2,...,N$$

(28)

$$\left(\hat{\alpha},\hat{\beta}\right)=\arg\min_{\alpha,\beta}\left[\sum_{i=1}^{N}\left(y_i-\alpha-\beta\int_C f_i dg_\mu\right)^2\right] \quad (29)$$

then $\hat{y}_i=\hat{\alpha}+\hat{\beta}\int f_i dg_\mu,\ i=1,2,\ldots,N$ is called the Choquet integral regression equation of μ, where

$$\hat{\beta}=S_{yf}/S_{ff}$$

$$\hat{\alpha}=\frac{1}{N}\sum_{i=1}^{N}y_i-\hat{\beta}\frac{1}{N}\sum_{i=1}^{N}\int f_i dg_\mu \quad (30)$$

$$S_{hy}=\frac{\sum_{i=1}^{N}\left[y_i-\frac{1}{N}\sum_{i=1}^{N}y_i\right]\left[\int f_i dg_{\mu^*}-\frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]}{N-1}$$

$$S_{hh}=\frac{\sum_{i=1}^{N}\left[\int f_i dg_{\mu^*}-\frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]^2}{N-1} \quad (31)$$

7. Experiment and result

A real data set with 72 samples from a junior high school in Taiwan including the independent variables, examination scores of four courses, and the dependent variable, the score of the Basic Competence Test of junior high school listed in Table 2 is applied to evaluate the performances of three Choquet integral regression models with P-measure, λ-measure, R-measure based and L-measure on V-support, and γ-support respectively, a ridge regression model, and a multiple linear regression model by using 5-fold cross validation method to compute the mean square error (MSE) of the dependent variable. The formulas of MSE is

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \quad (32)$$

For any fuzzy measure, μ-measures, once the fuzzy support of the μ-measure is given, all event measures of μ can be found, and then, the Choquet integral based on μ and the Choquet integral regression equation based on μ can also be found.

The singleton measures, V-support and γ-support of the P-measure, λ-measure, and R-measure can be obtained by using the formulas (15) and (19), respectively.

The experimental results of eight forecasting models are listed in Table I. We can find that the Choquet integral regression model with L-measure outperforms other forecasting regression models.

Table 1 MSE of regression models

Regression model			5-fold CV MSE
	measure	support	
Choquet Integral Regression model	P	V	70.4011
		γ	68.9878
	λ	V	61.0440
		γ	57.5449
	R	V	60.5317
		γ	56.2746
	L	V	59,0159
		γ	56.2711
Ridge regression			63.1253
Multiple linear regression			69.7094

8. Conclusions and future works

When the sub-tests of a composite test are with interaction, the performance of the traditional additive scale method is poor. Non-additive fuzzy measures and fuzzy integral can be applied to improve this situation. In this study, a real data set from a junior high school including the independent variables, test scores of four courses with interaction, and the dependent variable, junior high school graduates' scores of the Basic Competence Test (BCT) are applied to evaluate the performances of the Choquet integral regression model with four well known fuzzy measures, P-measure, λ-measure, R-measure and L-measure based on two different supports, V-support, and γ-support respectively, the traditional multiple linear regression model, and the ridge regression model. Experimental result shows that the following situations:

(i) Choquet integral regression model with L-measure based on γ-support has the best performance.

(ii) Based on the same fuzzy support, not only the γ-support but also the V-support, the Choquet integral regression model with L- measure is better than which with fuzzy measure, λ-measure, P-measure and R-measure.

(iii) The Choquet integral regression model with the same measure, P-measure, λ-measure, R-measure and L-measure, respectively, the performance of which is derived from the γ-support is better than which from the V-support.

(iv) The Choquet integral regression model with λ-measure, R-measure and L-measure based on V-support and γ-support, respectively, are all better than the ridge regression and the multiple regression model.
(v) The Choquet integral regression model with P-measure is not a good model.

In future, we will apply the proposed Choquet integral regression model with the better measure based on the best fuzzy support, γ-support, to develop multiple classifier system.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 96-2413--H-468-001).

References

[1] A. E. Hoerl, R. W. Kenard, and K. F. Baldwin, "Ridge regression: Some simulation", Communications in Statistics, Vol. 4, No. 2, pp. 105-123, 1975.
[2] Hsiang-Chuan Liu, Chin-Chun Chen, Der-Bang Wu, and Yu-Du Jheng, "A new weighting method for detecting outliers in IPA based on Choquet integral", IEEE International conference on Industrial Engineering and Engineering Management 2007, December 2007, Singapore.
[3] Hsiang-Chuan Liu, "The Choquet integral regression model based on r-complete measure", Journal of educational research and development, Vol. 2, No. 4, pp. 87-107, 2006 (in Chinese).
[4] Hsiang-Chuan Liu, Wen-Chih Lin, and Wei-Sheng Weng, "A Choquet Integral Regression Model Based on a New Fuzzy Measure", The 12th International Conference on Fuzzy Theohery & Technology, July 2007, Salt Lake City, Utah, U.
[5] Hsiang-Chuan Liu, Wen-Chih Lin, Kei-Yi Chang, and Wei-Sheng Weng, "A Nonlinear Regression Model Based on Choquet Integral with ε-Measure", 2007 WSEAS International Conferences, Venice, Italy, November 2007.
[6] Hsiang-Chuan Liu, Yu-Du Jheng, Wen-Chih Lin, and Guey-Shya Chen, "A novel fuzzy measure and its Choquet Integral Regression Model", Proceedings of International conference on Machine Learning and Cybernetics 2007, Hong Kong, China, August 2007. Vol. 7, No.3, pp. 1394-1398, ISBN 1-4244-0973X.
[7] Hsiang-Chuan Liu, Yu-Chieh Tu, Chin-Chun Chen, and Wei-Sheng Weng, "The Choquet integral with respect to λ-measure based on γ-support", 2008 International Conferences on Machine Learning and Cybernetics, Kuming, China, July 12-15, 2008 (Accepted).
[8] Hsiang-Chuan Liu, Yu-Chieh Tu, Wen-Chun Huang, and Chin-Chun Chen, "The Choquet integral with respect to R-measure based on γ-support", ICNC'08-FSKD'08, Jinan, China, August 25-27, 2008 (Accepted).
[9] G. Choquet, "Theory of capacities", Annales de l'Institut Fourier, Vol. 5, pp. 131-295, 1953.
[10] M. Sugeno, "Theory of fuzzy integrals and its applications", unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.
[11] Z. Wang, and G. J. Klir, Fuzzy Measure Theory, Plenum Press, New York, 1992.
[12] L. A. Zadeh, "Fuzzy Sets and Systems", Vol. 1, pp. 3, 1978.

Table 2 The data set with four courses and science scores of the BCT

No.	C1	C2	C3	C4	BCT	No.	C1	C2	C3	C4	BCT
1	77	75	79	83	31	31	74	70	80	75	35
2	71	72	78	75	26	32	56	61	75	68	22
3	78	86	86	86	33	33	62	68	72	74	29
4	58	64	68	66	32	34	86	80	82	81	35
5	48	59	65	68	16	35	63	78	88	83	31
6	68	74	77	80	28	36	56	66	76	71	21
7	62	72	84	78	47	37	77	74	80	76	42
8	51	53	65	59	9	38	73	78	84	81	24
9	62	64	76	70	36	39	63	60	68	69	17

10	63	70	81	75	41	40	53	68	80	74	31
11	66	68	75	74	25	41	74	86	87	88	44
12	66	72	80	76	23	42	78	83	81	85	50
13	75	75	85	80	39	43	47	58	66	62	15
14	74	63	69	75	12	44	51	60	63	64	18
15	68	78	85	75	27	45	60	65	75	70	23
16	71	74	80	77	26	46	68	68	80	74	26
17	49	60	69	64	13	47	52	60	70	65	20
18	73	78	84	81	39	48	57	65	75	70	24
19	68	70	74	76	40	49	70	66	70	74	13
20	54	56	62	68	7	50	53	68	74	80	30
21	53	68	74	71	11	51	68	68	78	76	35
22	56	63	69	75	21	52	57	60	68	64	23
23	70	80	78	70	31	53	61	62	70	70	25
24	51	74	82	75	49	54	59	70	80	76	37
25	61	66	72	78	33	55	59	62	70	78	29
26	67	70	80	75	35	56	52	64	76	70	27
27	59	75	80	82	27	57	68	70	80	75	33
28	53	56	70	63	22	58	71	76	74	78	38
29	56	56	65	61	6	59	72	66	78	72	19
30	52	57	67	62	15						

Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, 30-31 Aug. 2008

CHOQUET INTEGRAL LOGISTIC REGRESSION ALGORITHM BASED ON L-MEASURE AND γ-SUPPORT

HSIANG-CHUAN LIU [1], YU-DU JHENG [2], GUEY-SHYA CHEN [2], BAI-CHENG JENG [2]

[1] Department of Bioinformatics, Asia University Taiwan
[2] Graduate Institute of Educational Measurement and Statistics, Taichung University, Taiwan
E-MAIL: lhc@asia.edu.tw, doora0622@yahoo.com.tw, grace@mail.ntcu.edu.tw, jason5907@yahoo.com.tw

Abstract:

Logistic regression algorithm and SVM algorithm are two well-known classification algorithms but when the multi-collinearity between independent variables occurs in above two algorithms, their classifying performance will always be not good. An improved classification algorithm combining the Choquet integral with respect to the λ-measure based on γ-support is proposed by our previous work. In this paper, we replaced the more sensitive fuzzy measure, L-measure with the λ-measure in above improved classification algorithm, and we obtained a further improved algorithm, called Choquet integral logistic regression algorithm based on L-measure and γ-support. For evaluating the performances of the SVM, logistic regression and the Choquet integral logistic regression algorithm with γ-support based on P-measure, λ-measure and L-measure, respectively, a real data experiment by using Leave-one-out Cross-Validation accuracy is conducted. Experimental result shows that our new algorithm has the best performance.

Keywords:

Fuzzy measure; Choquet integral; λ-measure; L-measure; γ-support

1. Introduction

When interactions among independent variables exist in forecasting and classifying problems, the performances of the traditional methods, multiple linear regression algorithms and multiple logistic regression algorithms are always not good. For forecasting problems, recently, some Choquet integral regression algorithms based on different fuzzy measures proposed by our previous works can be used to improve this situation [1], [2], [3], [4], [5], [6]. Therefore, in our previous study, we consider that the proposed Choquet integral regression algorithms may also be used to improve the performance of classification, and an improved classification algorithm combining the Choquet integral with respect to the λ-measure based on γ-support is proposed.

In this paper, we replaced the more sensitive fuzzy measure, L-measure with the λ-measure in above improved classification algorithm, and a further improved algorithm, called Choquet integral logistic regression algorithm based on L-measure and γ-support was obtained. For evaluating the performances of the logistic regression algorithm, a well-known classifying algorithm, Support Vector Machine (SVM), and our new algorithm, Choquet integral logistic regression algorithm with γ-support based on P-measure, λ-measure and L-measure, respectively, a real data experiment by using a Leave-one-out Cross-Validation accuracy is conducted.

This paper is organized as followings: the logistic regression algorithm is introduced in section 2, the SVM algorithm is brief introduced in section 3, fuzzy measures including two well-known measures, P-measure, λ-measure and our new measure, L-measure, are described in section 4, fuzzy support and γ-support are described in section 5. Choquet integral and its regression algorithm are described in section 6. The new algorithm, Choquet integral logistic regression algorithm is introduced in section 7, Experiment and result are described in section 8, and final section is for conclusions and future works.

2. Logistic regression

For no needing to group the original data, our previous study derived the logistic regression algorithm as below by using a pared-down maximal likelihood estimating based on Bernoulli distribution not the binomial distribution.

2.1. Logistic regression model

Let $(x_{i1}, x_{i2}...., x_{in}, y_i)$, $i = 1, 2, ..., N$ be a sample data, satisfying

$$\underline{x}_i = (x_{i1}, x_{i2}, ..., x_{in}) \in R^n,\ y_i \in \{0,1\},$$

978-1-4244-2239-5/08/$25.00 ©2008 IEEE

$$Y_i^{\perp\perp} \sim B(1, p_i), i = 1,2,...,N \quad (1)$$

Logistic regression model is denoted as follows

$$P_i = P(Y_i = 1 \mid x_i) = \frac{1}{1+\exp\left[-\left(\alpha + \underline{\beta}'\underline{x}\right)\right]}, i = 1,2,...,N \quad (2)$$

where $\underline{\beta}' = (\alpha, \beta_1, \beta_2, ..., \beta_n)$ are parameters vector of regression coefficients.

2.2. Logistic regression algorithm

We can obtain the likelihood function and log likelihood function as following equations (3) and (4), respectively.

$$L(p_1, p_2, ..., p_N) = \prod_{i=1,2,...,N} p_i^{y_i} (1-p_i)^{1-y_i} \quad (3)$$

$$\begin{aligned} l &= \log\left[L(p_1, p_2, ..., p_N)\right] \\ &= \sum_{i=1}^{N}\left[y_i \log p_i + (1-y_i)(1-\log p_i)\right] \end{aligned} \quad (4)$$

And we can get

$$\begin{aligned} l = l(\alpha,\beta) &= \sum_{i=1}^{N}\left[y_i \log p_i + (1-y_i)(1-\log p_i)\right] \\ &= -\sum_{i=1}^{N}\left[\log\left(1+exp\left[-\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]\right) + (1-y_i)\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right] \end{aligned}$$

Using Newton-Raphson's iterative algorithm, we can get the estimated regression coefficients of the multiple logistic regression model and the estimated multiple logistic regression equation as follows:

$$\hat{P}_i = \hat{P}(Y_i = 1 \mid x_i) = \frac{1}{1+\exp\left[-\left(\hat{\alpha}+\underline{\hat{\beta}}'\underline{x}\right)\right]} \quad (5)$$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}_{k+1} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}_{k} - \begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \alpha \partial \beta_n} \\ \frac{\partial^2 l}{\partial \beta_1 \partial \alpha} & \frac{\partial^2 l}{\partial \beta_1^2} & \cdots & \frac{\partial^2 l}{\partial \alpha \partial \beta_n} \\ \vdots & \ddots & & \vdots \\ \frac{\partial^2 l}{\partial \beta_n \partial \alpha} & \frac{\partial^2 l}{\partial \beta_n \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_n^2} \end{bmatrix}_{k}^{-1} \begin{bmatrix} \frac{\partial l}{\partial \alpha} \\ \frac{\partial l}{\partial \beta_1} \\ \vdots \\ \frac{\partial l}{\partial \beta_n} \end{bmatrix}_{k} \quad (6)$$

Increment k; until $\left\| \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}_{k+1} - \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}_{k} \right\| < \varepsilon \quad (7)$

where $\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{N}\left[\frac{1}{1+\exp\left[-\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]} - (1-y_i)\right] \quad (8)$

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{N}\left[\frac{1}{1+\exp\left[-\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]} - (1-y_i)\right] x_{ij}, j = 1,2,...,n \quad (9)$$

$$\frac{\partial^2 l}{\partial \alpha^2} = -\sum_{i=1}^{N}\frac{\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2} \quad (10)$$

$$\frac{\partial^2 l}{\partial \beta_j^{\,2}} = -\sum_{i=1}^{N}\frac{x_{ij}^2 \exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2}, j = 1,2,...,n \quad (11)$$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta_j} = \frac{\partial^2 l}{\partial \beta_j \partial \alpha} = -\sum_{i=1}^{N}\frac{x_{ij} \exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2}, j = 1,2,...,n \quad (12)$$

$$\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} = \frac{\partial^2 l}{\partial \beta_k \partial \beta_j} = -\sum_{i=1}^{N}\frac{x_{ij} x_{ik} \exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2}, j,k = 1,2,...,n \quad (13)$$

3. Support vector machine (SVM) [7], [8]

Given the training set of instance-labeled pairs $(\underline{x}_i, y_i), i = 1,2,...,N$, where

$$\underline{x}_i \in R^n, y_i \in \{1,-1\}, i = 1,2,...,N \quad (14)$$

The support vector machine (SVM) algorithm (Boser, Guyon, and Vapnik 1992, Cortes and Vapnik 1995) requires

$$\begin{aligned} &\min_{\underline{w},b,\xi} \frac{1}{2}\underline{w}'\underline{w} + c\sum_{i=1}^{N}\xi_i \\ &\text{subject to } y_i\left(\underline{w}'\phi(\underline{x}_i)+b\right) \ge 1-\xi_i, \\ &\qquad \xi_i \ge 0, \\ &where \quad b,c \in R, \underline{w}, \phi(\underline{x}_i) \in R^m \\ &\qquad \phi : R^n \to R^m \end{aligned} \quad (15)$$

For any testing point $\underline{x}_i \in R^n, y_i \in \{1,-1\}$, we can make an assignment according to the following formula.

$$f(\underline{x}_i) = \text{sign}\left[\underline{w}'\varphi(\underline{x}_i) + b - (1-\xi_i)\right] = \begin{cases} +, & \text{if } y_i = +1 \\ -, & \text{if } y_i = -1 \end{cases} \tag{16}$$

4. Fuzzy measures

The well known fuzzy measures, P-measure proposed by Zadah in 1978, and the λ-measure proposed by Sugeno in 1974, and L-measure proposed by our previous work in 2007 are concisely introduced as follows.

4.1. Fuzzy measures [9], [10], [11]

A fuzzy measure μ on a finite set X is a set function $\mu : 2^X \to [0,1]$ satisfying the following axioms:

(i) $\mu(\phi) = 0$, $\mu(X) = 1$ (boundary conditions) (17)

(ii) $A \subseteq B \Rightarrow \mu(A) \le \mu(B)$ (monotonicity) (18)

4.2. Singleton measures [3], [4], [5], [6]

A singleton measure of a fuzzy measure μ on a finite set X is a function $s : X \to [0,1]$ satisfying:

$$s(x) = \mu(\{x\}),\ x \in X \tag{19}$$

$s(x)$ is called the density of singleton x.

4.3. λ-measure [9], [11]

For given singleton measures s, a λ-measure, g_λ, is a fuzzy measure on a finite set X, satisfying:

(i) $A, B \in 2^X$, $A \cap B = \phi$, $A \cup B \ne X$

$$\Rightarrow g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B) \tag{20}$$

(ii) $\prod_{i=1}^{n}\left[1 + \lambda s(x_i)\right] = \lambda + 1 > 0,\ s(x_i) = g_\lambda(\{x_i\})$ (21)

Note that once the singleton measure is known, we can obtain the values of λ uniquely by using the previous polynomial equation.

4.4. P-measure [10]

For given singleton measures s, a P-measure, g_P, is a fuzzy measure on a finite set X, satisfying:

$$^\forall A \in 2^X \Rightarrow g_P(A) = \max_{x \in A} s(x) = \max_{x \in A} g_P(\{x\}) \tag{22}$$

4.5. L-measure [5]

For given singleton measure s, a L-measure, g_L, is a fuzzy measure on a finite set X, $|X| = n$, satisfying:

(i) $L \in [0, \infty)$ (23)

(ii) $\sum_{x \in X} s(x) = \sum_{x \in X} g_L(\{x\}) = 1$ (24)

(iii) $^\forall A \subset X, n - |A| + (|A| - 1)L > 0$

$$\Rightarrow g_L(A) = \max_{x \in A}\left[s(x)\right] + \frac{(|A|-1)L\sum_{x \in A} s(x)}{\left[n - |A| + (|A|-1)L\right]}\left[1 - \max_{x \in A}\left[s(x)\right]\right] \tag{25}$$

[Property]

(i) L-measure has infinitely many solutions with closed form.

(ii) When L=0, the L-measure is just a P-measure with closed form.

(iii) g_L is an increasing function of L.

5. Fuzzy supports [6]

For given singleton measures s of a fuzzy measure μ on a finite set X, if $\sum_{x \in X} s(x) = 1$, then s is called a fuzzy support measure of μ, or a fuzzy support of μ, or a support of μ.

5.1. γ- support [6]

Let μ be a fuzzy measure on a finite set, $X = \{x_1, x_2, ..., x_n\}$, y_i be global evaluation or response of subject i and $f_i(x_j)$ be the evaluation of subject i for singleton x_j, satisfying

$$0 < f_i(x_j) < 1,\ i = 1,2,...,N,\ j = 1,2,...,n$$

If
$$\gamma(x_j) = \frac{1 + r\left(f(x_j)\right)}{\sum_{k=1}^{n}\left[1 + r\left(f(x_k)\right)\right]},\ j = 1,2,...,n \tag{26}$$

where
$$r\left(f(x_j)\right) = \frac{S_{y,x_j}}{S_y S_{x_j}} \tag{27}$$

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - \frac{1}{N}\sum_{i=1}^{N} y_i\right)^2 \tag{28}$$

$$S_{x_j}^2 = \frac{1}{N}\sum_{i=1}^{n}\left[f_i(x_j) - \frac{1}{N}\sum_{i=1}^{N} f_i(x_j)\right]^2 \quad (29)$$

$$S_{y,x_j} = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - \frac{1}{N}\sum_{i=1}^{N} y_i\right)\left[f_i(x_j) - \frac{1}{N}\sum_{i=1}^{N} f_i(x_j)\right] \quad (30)$$

satisfying $0 \le \gamma(x_j) \le 1$ and $\sum_{j=1}^{n}\gamma(x_j) = 1$ (31)

Then the function $\gamma: X \to [0,1]$ satisfying: $\mu(\{x\}) = \gamma(x)$, $\forall x \in X$ is a fuzzy support of μ, called γ-support of μ.

6. Choquet integral regression models

6.1. Choquet integral [12]

Let μ be a fuzzy measure on a finite set X. The Choquet integral of $f_i : X \to R_+$ with respect to μ for individual i is denoted by

$$\int_C f_i d\mu = \sum_{j=1}^{n}\left[f_i(x_{(j)}) - f_i(x_{(j-1)})\right]\mu(A_{(j)}^i) \ , i = 1,2,...,N \quad (32)$$

Where $f_i(x_{(0)}) = 0$, $f_i(x_{(j)})$ indicates that the indices have been permuted so that

$$0 \le f_i(x_{(1)}) \le f_i(x_{(2)}) \le ... \le f_i(x_{(n)}) \quad (33)$$

$$A_{(j)} = \{x_{(j)}, x_{(j+1)}, ..., x_{(n)}\} \quad (34)$$

6.2. Choquet integral regression algorithms [1], [2], [3], [4], [5], [6]

Let $y_1, y_2, ..., y_N$ be global evaluations of N objects and $f_1(x_j), f_2(x_j), ..., f_N(x_j)$, $j = 1,2,...,n$, be their evaluations of x_j, where $f_i : X \to R_+$, $i = 1,2,...,N$.

Let μ be a fuzzy measure, $\alpha, \beta \in R$,

$$y_i = \alpha + \beta\int_C f_i dg_\mu + e_i \ , \ e_i \sim N(0,\sigma^2) \ , \ i = 1,2,...,N \quad (35)$$

$$(\hat{\alpha},\hat{\beta}) = \arg\min_{\alpha,\beta}\left[\sum_{i=1}^{N}\left(y_i - \alpha - \beta\int_C f_i dg_\mu\right)^2\right] \quad (36)$$

then $\hat{y}_i = \hat{\alpha} + \hat{\beta}\int f_i dg_\mu$, $i = 1,2,...,N$ is called the Choquet integral regression equation of μ based on γ-support, where

$$\hat{\beta} = S_{yf} / S_{ff} \quad (37)$$

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{\beta}\frac{1}{N}\sum_{i=1}^{N}\int f_i dg_\mu$$

$$S_{hy} = \frac{\sum_{i=1}^{N}\left[y_i - \frac{1}{N}\sum_{i=1}^{N} y_i\right]\left[\int f_i dg_{\mu^*} - \frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]}{N-1}$$

$$S_{hh} = \frac{\sum_{i=1}^{N}\left[\int f_i dg_{\mu^*} - \frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]^2}{N-1} \quad (38)$$

7. Choquet integral logistic regression algorithm

Let $y_1, y_2, ..., y_N$ be global evaluations of N objects, $f_1(x_j), f_2(x_j), ..., f_N(x_j)$, $j = 1,2,...,n$, be their evaluations of x_j, where $f_i : X \to R_+$, $i = 1,2,...,N$, μ be a fuzzy measure based on γ- support.

$\tilde{y}_i = \hat{\alpha} + \hat{\beta}\int f_i dg_\mu$, $i = 1,2,...,N$, be the Choquet integral regression equation of μ based on γ- support. Furthermore, let $(\tilde{y}_i, y_i)$, $i = 1,2,...,N$ be a sample data, satisfying

$$\tilde{y}_i \in R, \ y_i \in \{0,1\}, Y_i^{\perp\perp} \sim B(1, p_i), \ i = 1,2,...,N \quad (39)$$

The Choquet integral logistic regression model is denoted as follows

$$P_i = P(Y_i = 1 \mid \tilde{y}_i) = \frac{1}{1 + \exp\left[-(\delta + \phi\tilde{y}_i)\right]} \quad (40)$$

where δ, ϕ are parameters of two regression coefficients, and the Choquet integral logistic regression algorithm is given as below

$$\hat{P}_i = \hat{P}(Y_i = 1 \mid \tilde{y}_i) = \frac{1}{1 + \exp\left[-(\hat{\delta} + \hat{\phi}\tilde{y}_i)\right]} \quad (41)$$

$$\begin{bmatrix}\hat{\delta}\\ \hat{\phi}\end{bmatrix}_{k+1} = \begin{bmatrix}\hat{\delta}\\ \hat{\phi}\end{bmatrix}_k - \begin{bmatrix}\frac{\partial^2 l}{\partial\delta^2} & \frac{\partial^2 l}{\partial\delta\partial\phi}\\ \frac{\partial^2 l}{\partial\phi\partial\delta} & \frac{\partial^2 l}{\partial\phi^2}\end{bmatrix}_k^{-1}\begin{bmatrix}\frac{\partial l}{\partial\delta}\\ \frac{\partial l}{\partial\phi}\end{bmatrix}_k \quad (42)$$

Increment k; until $\left\|\begin{bmatrix}\hat{\delta}\\ \hat{\phi}\end{bmatrix}_{k+1} - \begin{bmatrix}\hat{\delta}\\ \hat{\phi}\end{bmatrix}_k\right\| < \varepsilon$ (43)

where $$\frac{\partial l}{\partial \delta}=\sum_{i=1}^{N}\left[\frac{1}{1+\exp\left[-\left(\delta+\phi\tilde{y}_i\right)\right]}-\left(1-y_i\right)\right] \tag{44}$$

$$\frac{\partial l}{\partial \phi}=\sum_{i=1}^{N}\left[\frac{1}{1+\exp\left[-\left(\delta+\phi\tilde{y}_i\right)\right]}-\left(1-y_i\right)\right]\tilde{y}_i \tag{45}$$

$$\frac{\partial^2 l}{\partial \delta^2}=-\sum_{i=1}^{N}\frac{\exp\left(\delta+\phi\tilde{y}_i\right)}{\left[1+\exp\left(\delta+\phi\tilde{y}_i\right)\right]^2} \tag{46}$$

$$\frac{\partial^2 l}{\partial \phi^2}=-\sum_{i=1}^{N}\frac{\tilde{y}_i^2\exp\left(\delta+\phi\tilde{y}_i\right)}{\left[1+\exp\left(\delta+\phi\tilde{y}_i\right)\right]^2} \tag{47}$$

$$\frac{\partial^2 l}{\partial \delta\partial \phi}=\frac{\partial^2 l}{\partial \phi\partial \delta}=-\sum_{i=1}^{N}\frac{\tilde{y}_i\exp\left(\delta+\phi\tilde{y}_i\right)}{\left[1+\exp\left(\delta+\phi\tilde{y}_i\right)\right]^2} \tag{48}$$

8. Experiment and result

A female breast cancer data set was downloaded from website,ftp://ftp.cs.wisc.edu/math-prog/cpo-dataset/machine-learn/cancer/WDBC/

The sample included 569 females; there are two classes of tumors, 357 benign tumors and 212 malignant tumors, and 30 characteristics of tumors.

The above real data is applied to evaluate the performances of the multiple logistic regression algorithm, The SVM algorithm, and three Choquet integral logistic regression algorithms with γ-support based on P-measure, λ-measure and L-measure, respectively, by using Leave-one-out Cross-Validation method to compute the accuracies of the response category variable.

Since for any fuzzy measures, once a fuzzy support of the fuzzy measure is selected, all of the event measures of this fuzzy measure can be found, and then, the Choquet integral based on this fuzzy measure and the Choquet integral regression equation based on this fuzzy measure can also be found.

For three fuzzy measures, P-measure, λ-measure and L-measure, suppose the same fuzzy support, γ-support, is first selected. We can obtain the γ-support of the 30 futures of the Breast Cancer Data by using the equations (25)~(30) as Table 1

The performances of the Choquet integral logistic regression algorithms with γ-support based on P-measure λ-measure and L-measure, respectively, a multiple logistic regression algorithm and a SVM algorithm are compared by using Leave-one-out Cross-Validation accuracy. The experimental results of five classification regression algorithms are listed in Table 2. We can find that the Choquet integral logistic regression algorithm with γ-support based on L-measure outperforms other classification algorithms, and the multiple logistic regression algorithm is better than the SVM algorithm.

Table 1 γ-support of the thirty futures

No	γ -support	No	γ -support	No	γ -support
1	0.0181	11	0.0260	21	0.0154
2	0.0393	12	0.0624	22	0.0375
3	0.0171	13	0.0269	23	0.0146
4	0.0194	14	0.0320	24	0.0178
5	0.0396	15	0.0661	25	0.0367
6	0.0237	16	0.0411	26	0.0243
7	0.0180	17	0.0469	27	0.0179
8	0.0141	18	0.0345	28	0.0125
9	0.0421	19	0.0630	29	0.0386
10	0.0616	20	0.0526	30	0.0402

Table 2 Leave-one-out CV accuracy of six Classification algorithms

Classification algorithm	No. of mistrial	Accuracy
Choquet integral logistic regression with P-measure	60	0.8946
SVM	52	0.9086
Logistic Regression	47	0.9174
Choquet integral logistic regression with λ-measure	45	0.9209
Choquet integral logistic regression with L-measure	36	0.9367

9. Conclusions and future works

In classification problem, two well-known classification algorithms, multiple logistic regression algorithm and SVM algorithm are popular used. However, when the multicollinearity between independent variables occurs in above two algorithms, the performance of these two methods will always be not good. An enhanced classification algorithm combining the Choquet integral with respect to the λ-measure based on γ-support is proposed by our previous work. In this research, we took the place of the more sensitive fuzzy measure, L-measure with the λ-measure in the above enhanced classification algorithm, and a further enhanced algorithm, called Choquet integral logistic regression algorithm based on L-measure and γ-support was obtained. A real data experiment by using Leave-one-out Cross-Validation accuracy is conducted for evaluating the performances of the SVM, logistic regression and the Choquet integral logistic regression algorithm with γ-support based on P-measure, λ-measure and L-measure, respectively. And experimental result shows that our new algorithm, Choquet integral logistic regression algorithm based on L-measure and γ-support, has the best performance. The performance of multiple logistic regression algorithm is better than that of SVM algorithm.

In the future we will apply the proposed Choquet integral regression model with the better measure based on the best fuzzy support, γ-support, to develop multiple classifier system.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 96-2413--H-468-001).

References

[1] Hsiang-Chuan Liu, Chin-Chun Chen, Der-Bang Wu, and Yu-Du Jheng, “A new weighting method for detecting outliers in IPA based on Choquet integral”, Proceeding of IEEM2007 Conference, Singapore, pp. 2015-2019, Dec. 2007.

[2] Hsiang-Chuan Liu, “The Choquet integral regression model based on r-complete measure”, Journal of educational research and development, Chinese, Vol. 2, No. 4, pp 87-107, 2006.

[3] Hsiang-Chuan Liu, Wen-Chih Lin, and Wei-Sheng Weng, “A Choquet Integral Regression Model Based on a New Fuzzy Measure”, The 12th International conference on Fuzzy Theohery & Technology, Salt Lake City, Utah, US, July 2007.

[4] Hsiang-Chuan Liu, Wen-Chih Lin, Kei-Yi Chang, and Wei-Sheng Weng, “A Nonlinear Regression Model Based on Choquet Integral with ε -Measure”, WSEAS2007 International Conferences, Venice, Italy, November 2007.

[5] Hsiang-Chuan Liu, Yu-Du Jheng, Wen-Chih Lin, Guey-Shya Chen, “A novel fuzzy measure and its Choquet Integral Regression Model”, Proceedings of ICMLC2007 Conference, Hong Kong, Vol 7, No. 3, pp. 1394-1398, August 2007.

[6] Hsiang-Chuan Liu, Yu-Chieh Tu, Chin-Chun Chen. Wei-Sheng Weng. “The Choquet integral with respect to λ-measure and γ-support”, Proceeding of ICMLC2008 Conference, Kunming, China, July 2008 (accepted).

[7] B. E. Boser, I. Guyon, and V. Vapnik, “A training algorithm for optimal margin classifiers”, In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM Press, pp. 144-152, 1992.

[8] C. Cortes, and V. Vapnik, “Support-vector network”, Machine Learning Vol. 20, pp. 273-297, 1995.

[9] Z. Wang, and G. J. Klir, Fuzzy Measure Theory, Plenum Press, New York, 1992.

[10] L. A. Zadeh, “Fuzzy Sets and Systems”, Vol. 1, pp. 3, 1978.

[11] M. Sugeno, “Theory of fuzzy integrals and its applications”, unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[12] G. Choquet, “Theory of capacities”, Annales de l'Institut Fourier, Vol. 5, pp. 131-295, 1953.

Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, 30-31 Aug. 2008

FUZZY POSSIBILITY C-MEAN CLUSTERING ALGORITHMS BASED ON COMPLETE MAHALANOBIS DISTANCES

HSIANG-CHUAN LIU[1], JENG-MING YIH [2], DER-BANG WU[3], SHIN-WU LIU[4]

[1]Department of Bioinformatics, Asia University, Taiwan
[2]Gradguate institute of educational measurement, Taichung University, Taiwan
[3]Department of Mathematics Education, Taichung University, Taiwan
[4]National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA
E-MAIL: lhc@asia.edu.tw, yih@mail.ntcu.edu.tw, wudb@hotmail.com, fpan0366@yahoo.com.tw

Abstract:

Two well known fuzzy partition clustering algorithms, FCM and FPCM are based on Euclidean distance function, which can only be used to detect spherical structural clusters. GK clustering algorithm and GG clustering algorithm, were developed to detect non-spherical structural clusters, but both of them fail to consider the relationships between cluster centers in the objective function, needing additional prior information.. In our previous studies, we developed two improved algorithms, FCM-M and FPCM-M, based on unsupervised Mahalanobis distance without any additional prior information. And FPCM-M is better than FCM-M, since the former has the more information about the typicalities than the later. In this paper, an improved new unsupervised algorithm, "fuzzy possibility c-mean based on complete Mahalanobis distance without any prior information (FPCM-CM)", is proposed. In our new algorithm, not only the local covariance matrix of each cluster but also the overall covariance matrix was considered. It can get more information and higher accuracy by considering the additional overall covariance matrix than FPCM-M. A real data set was applied to prove that the performance of the FPCM-CM algorithm is better than those of the traditional FCM and FPCM algorithm and our previous FCM-M.

Keywords:

FCM; CM; FCM-M; PCM-M; FPCM-CM

1. Introduction

The clustering analysis plays an important role in data analysis and interpretation. It groups the data into classes or clusters so that the data objects within a cluster have high similarity in comparison to one another, but are very dissimilar to those data objects in other clusters.

Fuzzy partition clustering is a branch in cluster analysis, it is widely used in pattern recognition field. The well known fuzzy Possibility partition clustering algorithms, PCM [4], and FPCM [6] are proposed to improve the problems of outlier and noise in FCM [1], but the above three algorithms were based on Euclidean distance function, which can only be used to detect spherical structural clusters.

Extending Euclidean distance to Mahalanobis distance, Gustafson-Kessel (GK) clustering algorithm [2] and Gath-Geva (GG) clustering algorithm [3], are developed to detect non-spherical structural clusters, but both of them fail to consider the relationships between cluster centers in the objective function, needing additional prior information. In our previous studies, we developed two improved algorithms, FCM-M and FPCM-M , based on unsupervised Mahalanobis distance without any additional prior information, and FPCM-M is better than FCM-M, since the former has the more information about the typicalities than the later.

In this paper, an improved new unsupervised algorithm, "fuzzy possibility c-mean based on complete Mahalanobis distance without any prior information (FPCM-CM)", is proposed. In our new algorithm, not only the local covariance matrix of each cluster but also the overall covariance matrix were considered. It can get more information and higher accuracy by considering the additional overall covariance matrix than FPCM-M.

A real data set was applied to prove that the performance of the FPCM-CM algorithm is better than those of the traditional FCM and FPCM algorithm and our previous FCM-M.

This paper is organized as followings: Fuzzy c-mean algorithm is introduced in section 2, Fuzzy possibility c-mean algorithm is introduced in section 3, FCM-M algorithm is introduced in section 4. FPCM-M algorithm is described in section 5. FPCM-CM algorithm is described in section 6, Experiment and result are described in section 7 and final section is for conclusions and future works.

978-1-4244-2239-5/08/$25.00 ©2008 IEEE

2. Fuzzy c-Mean Algorithm [1, 5]

The objective function used in FCM is given by Equation (1)

$$J_{FCM}^{m}(U,A,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}d_{ij}^{2}=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left\|\underline{x}_{j}-\underline{a}_{i}\right\|^{2} \qquad (1)$$

$\mu_{ij}\in[0,1]$ is the membership degree of data object $\underline{x}_j$ in cluster C_i, and it satisfies the following constraint given by Equation (2-2)

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,...,n \qquad (2)$$

C is the number of clusters, m is the fuzzifier, m>1,which controls the fuzziness of the method. They are both parameters and need to be specified before running the algorithm. $d_{ij}^2=\left\|\underline{x}_j-\underline{a}_i\right\|^2$ is the square Euclidean distance between data object $\underline{x}_j$ to center $\underline{a}_i$.

Minimizing objective function (1) with constraint (2) , the updating function for $\underline{a}_i$ and μ_{ij} is obtained as (3) and (4),

$$\underline{a}_i=\frac{\sum_{j=1}^{n}\mu_{ij}{}^{m}\underline{x}_j}{\sum_{j=1}^{n}\mu_{ij}{}^{m}} \quad i=1,2,\ldots,c \qquad (3)$$

$$\mu_{ij}=\left[\sum_{l=1}^{c}\left[\frac{\left(\underline{x}_j-\underline{a}_i\right)'\left(\underline{x}_j-\underline{a}_i\right)}{\left(\underline{x}_j-\underline{a}_l\right)'\left(\underline{x}_j-\underline{a}_l\right)}\right]^{\frac{1}{m-1}}\right]^{-1} \qquad (4)$$

3. Fuzzy Possibility C-Mean Algorithm [6]

The improved fuzzy partition clustering algorithms "Fuzzy Possibility C-Mean (FPCM)" is given by Equation (5)

$$J_{FPCM}^{m}(U,T,A,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\left\|\underline{x}_j-\underline{a}_i\right\|^2 \qquad (5)$$

constraints: membership

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,\ldots,n, \qquad (6)$$

typicality $\sum_{j=1}^{n}t_{ij}=1,\ \forall i=1,2,\ldots,c$ (7)

Minimizing objective function (5) with constraint (6) and (7) , the updating function for $\underline{a}_i$, μ_{ij} and t_{ij} is obtained as (8) , (9) and (10)

$$\underline{a}_i=\frac{\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)\underline{x}_j}{\sum_{j=1}^{n}\left(\mu_{ij}^{m}+t_{ij}^{\delta}\right)},\quad i=1,2,\ldots,c \qquad (8)$$

$$\mu_{ij}=\left[\sum_{l=1}^{c}\left[\frac{\left(\underline{x}_j-\underline{a}_i\right)'\left(\underline{x}_j-\underline{a}_i\right)}{\left(\underline{x}_j-\underline{a}_l\right)'\left(\underline{x}_j-\underline{a}_l\right)}\right]^{\frac{1}{m-1}}\right]^{-1}, \qquad (9)$$

$$i=1,2,\ldots,c,\ j=1,2,\ldots,n$$

$$t_{ij}=\left[\sum_{l=1}^{n}\left[\frac{\left(\underline{x}_j-\underline{a}_i\right)'\left(\underline{x}_j-\underline{a}_i\right)}{\left(\underline{x}_l-\underline{a}_i\right)'\left(\underline{x}_l-\underline{a}_i\right)}\right]^{\frac{1}{\delta-1}}\right]^{-1} \qquad (10)$$

$$i=1,2,\ldots,c.\ j=1,2,\ldots,n$$

4. FCM-M Algorithm [7]

For improving the above problems of GK algorithm, based on unsupervised Mahalanobis distance without any additional prior information, we added the class covariance matrix and a regulating factor of covariance matrix, $-\ln\left|+\Sigma_i^{-1}\right|$, to each class in objective function (1). The improved new algorithm, "Fuzzy C-Mean based on Mahalanobis distance (FCM-M)", is obtained, and the objective function of FCM-M is given as (11) and constraints (12);

$$J_{FCM-M}^{m}(U,A,\Sigma,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|\right] \qquad (11)$$

Constrain $\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,\ldots,n$ (12)

Minimizing objective function (11) with constraint (12) , the updating function for $\underline{a}_i$, μ_{ij} and Σ_i is obtained as (13), (14) and (15)

$$\underline{a}_i = \left[\sum_{j=1}^{n} \mu_{ij}^m \Sigma_i^{-1} \right]^{-1} \sum_{j=1}^{n} \mu_{ij}^m \Sigma_i^{-1} \underline{x}_j \qquad (13)$$

$$i = 1, 2, \ldots, c$$

$$\mu_{ij} = \left[\sum_{s=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right|}{\left(\underline{x}_j - \underline{a}_s\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_s\right) - \ln\left|\Sigma_i^{-1}\right|} \right]^{\frac{1}{m-1}} \right] \qquad (14)$$

$$\Sigma_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m \left(\underline{x}_j - \underline{a}_i\right)\left(\underline{x}_j - \underline{a}_i\right)'}{\sum_{j=1}^{n} \mu_{ij}^m} \qquad (15)$$

5. FPCM-M Algorithm [8]

For improving the FPCM algorithm, we added the class covariance matrix and a regulating factor of covariance matrix, $-\ln\left|+\Sigma_i^{-1}\right|$, to each class in objective function (5). The improved new algorithm, "Fuzzy Possibility C-Mean based on Mahalanobis distance (FPCM-M)", is obtained, and the objective function of FCM-M is given as (16) and constraints (17);

$$J_{FPCM-M}^m(U,T,A,\Sigma,X) = \sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^m + t_{ij}^\delta\right)\left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1}\left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right|\right] \qquad (16)$$

constraints $\sum_{i=1}^{c} \mu_{ij} = 1$, $\forall j = 1, 2, \ldots, n$ (17)

$$\sum_{j=1}^{n} t_{ij} = 1, \ \forall i = 1, 2, \ldots, c \qquad (18)$$

Minimizing objective function (16) with constraint (17) , (18) the updating function for $\underline{a}_i$, μ_{ij} , t_{ij} and Σ_i is obtained as (19), (20), (21) and (22)

$$\underline{a}_i = \left[\sum_{j=1}^{n} \mu_{ij}^m \Sigma_i^{-1} \right]^{-1} \sum_{j=1}^{n} \mu_{ij}^m \Sigma_i^{-1} \underline{x}_j \qquad (19)$$

$$i = 1, 2, \ldots, c$$

$$\mu_{ij} = \left[\sum_{s=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right|}{\left(\underline{x}_j - \underline{a}_s\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_s\right) - \ln\left|\Sigma_i^{-1}\right|} \right]^{\frac{1}{m-1}} \right] \qquad (20)$$

$$t_{ij} = \left[\sum_{s=1}^{n} \left[\frac{\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| -}{\left(\underline{x}_s - \underline{a}_i\right)' \Sigma_i^{-1} \left(\underline{x}_s - \underline{a}_i\right) - \ln\left|\Sigma_s^{-1}\right|} \right]^{\frac{1}{\delta-1}} \right]^{-1} \qquad (21)$$

$$\Sigma_i = \frac{\sum_{j=1}^{n} \left(\mu_{ij}^m + t_{ij}^\delta\right)\left(\underline{x}_j - \underline{a}_i\right)\left(\underline{x}_j - \underline{a}_i\right)'}{\sum_{j=1}^{n} \left(\mu_{ij}^m + t_{ij}^\delta\right)} \qquad (22)$$

6. FPCM-CM Algorithm

In this paper, for improving the FPCM-M algorithm, , we added a overall scatter matrix, $-(\underline{a}_i - \underline{a}_t)' \Sigma_t^{-1} (\underline{a}_i - \underline{a}_t)$, in objective function (16). The improved new algorithm, "Fuzzy Possibility C-Mean based on Complete Mahalanobis distance (FPCM-CM)", is obtained, and the objective function of FCM-M is given as (23) and constraints (24);

$$J_{FPCM-CM}^m(U,T,A,\Sigma,X) = \sum_{i=1}^{c}\sum_{j=1}^{n}\left(\mu_{ij}^m + t_{ij}^\delta\right)\left[\left(\underline{x}_j - \underline{a}_i\right)' \Sigma_i^{-1}\left(\underline{x}_j - \underline{a}_i\right) - \ln\left|\Sigma_i^{-1}\right| - \left(\underline{a}_i - \underline{a}_t\right)' \Sigma_t^{-1}\left(\underline{a}_i - \underline{a}_t\right)\right] \qquad (23)$$

constraints：menbership

$$\sum_{i=1}^{c} \mu_{ij} = 1, \ \forall j = 1, 2, \ldots, n,$$

typicality $\sum_{j=1}^{n} t_{ij} = 1$, $\forall i = 1, 2, \ldots, c$ (24)

where

$$\underline{a}_t = \frac{1}{n}\sum_{j=1}^{n} \underline{x}_j, \ \Sigma_t = \frac{1}{n}\sum_{j=1}^{n}\left(\underline{x}_j - \underline{a}_t\right)\left(\underline{x}_j - \underline{a}_t\right)' \qquad (25)$$

Using the Lagrange multiplier method, to minimize the objective function (23) with constraints (24) respect to parameters $\underline{a}_i$, μ_{ij}, t_{ij}, Σ_i , we can obtain the updating function as (26), (27), (28),and(29),

$$\underline{a}_i = F^{-1}\left[\sum_{j=1}^{n}\left(\mu_{ij}^m + t_{ij}^\delta\right)\left[\Sigma_t^{-1}\underline{a}_t - \Sigma_i^{-1}\underline{x}_j\right]\right] \qquad (26)$$

where $F = \left[\sum_{j=1}^{n} \left[\mu_{ij}^{m} + t_{ij}^{\delta} \right] \left(\Sigma_i^{-1} - \Sigma_t^{-1} \right) \right]$

$$\mu_{ij} = \left[\sum_{s=1}^{c} \left[\frac{ \left(\underline{x}_j - \underline{a}_i \right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i \right) - \ln \left| \Sigma_i^{-1} \right| - \left(\underline{a}_i - \underline{a}_t \right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t \right) }{ \left(\underline{x}_j - \underline{a}_s \right)' \Sigma_s^{-1} \left(\underline{x}_j - \underline{a}_s \right) - \ln \left| \Sigma_s^{-1} \right| - \left(\underline{a}_s - \underline{a}_t \right)' \Sigma_t^{-1} \left(\underline{a}_s - \underline{a}_t \right) } \right]^{\frac{1}{m-1}} \right]^{-1} \tag{27}$$

$$t_{ij} = \left[\sum_{s=1}^{n} \left[\frac{ \left(\underline{x}_j - \underline{a}_i \right)' \Sigma_i^{-1} \left(\underline{x}_j - \underline{a}_i \right) - \ln \left| \Sigma_i^{-1} \right| - \left(\underline{a}_i - \underline{a}_t \right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t \right) }{ \left(\underline{x}_s - \underline{a}_i \right)' \Sigma_i^{-1} \left(\underline{x}_s - \underline{a}_i \right) - \ln \left| \Sigma_s^{-1} \right| - \left(\underline{a}_i - \underline{a}_t \right)' \Sigma_t^{-1} \left(\underline{a}_i - \underline{a}_t \right) } \right]^{\frac{1}{\delta-1}} \right]^{-1} \tag{28}$$

where

$$\underline{a}_t = \frac{1}{n} \sum_{j=1}^{n} \underline{x}_j, \quad \Sigma_t = \frac{1}{n} \sum_{j=1}^{n} \left(\underline{x}_j - \underline{a}_t \right) \left(\underline{x}_j - \underline{a}_t \right)'$$

$$\Sigma_i = \frac{ \sum_{j=1}^{n} \left[\mu_{ij}^{m} + t_{ij}^{\delta} \right] \left(\underline{x}_j - \underline{a}_i \right) \left(\underline{x}_j - \underline{a}_i \right)' }{ \sum_{j=1}^{n} \left[\mu_{ij}^{m} + t_{ij}^{\delta} \right] } \tag{29}$$

$i = 1, 2, \ldots, c$

The new fuzzy clustering algorithm (FPCM-CM) can be summarized in the following steps:

Step 1: Determining the number of cluster; c, let m=2, $\delta = 3$, Given converging error $\varepsilon > 0$ (such as $\varepsilon = 0.001$) choose the result membership matrix of FPCM-CM algorithm as the initial one and the normalized result typicality matrix of FPCM-CM algorithm as the initial one respectively;

let $\underline{a}_i^{(0)}, i = 1, 2, \ldots, c$ be the result centers of k-mean algorithm, and $d_{ij} = \left\| \underline{x}_j - \underline{a}_i^{(0)} \right\|$ be distances between data object $\underline{x}_j$ to center $\underline{a}_i^{(0)}$.

$$\mu_{ij}^{(0)} = \frac{ \left(d_M - d_{ij} \right) }{ \sum_{s=1}^{c} \left(d_M - d_{sj} \right) }, \tag{30}$$

$i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, n$

$$t_{ij}^{(0)} = \frac{ \left(d_M - d_{ij} \right) }{ \sum_{s=1}^{n} \left(d_M - d_{is} \right) }, \tag{31}$$

$i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, n$

$$\underline{a}_i^{(0)} = \left[F^{(0)} \right]^{-1} \left[\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(0)} \right]^{m} + \left[t_{ij}^{(0)} \right]^{\delta} \right) \left[\underline{x}_j - \underline{a}_t \right] \right] \tag{32}$$

$$F^{(0)} = \left[\sum_{j=1}^{n} \left[\left[\mu_{ij}^{(0)} \right]^{m} + \left[t_{ij}^{(0)} \right]^{\delta} \right] \left(I - \Sigma_t^{-1} \right) \right]$$

$$\Sigma_i^{(0)} = \frac{ \sum_{j=1}^{n} \left[\left[\mu_{ij}^{(0)} \right]^{m} + \left[t_{ij}^{(0)} \right]^{\delta} \right] \left(\underline{x}_j - \underline{a}_i^{(0)} \right) \left(\underline{x}_j - \underline{a}_i^{(0)} \right)' }{ \sum_{j=1}^{n} \left[\left[\mu_{ij}^{(0)} \right]^{m} + \left[t_{ij}^{(0)} \right]^{\delta} \right] } \tag{33}$$

Step 2: Find

$$\underline{a}_i^{(k)} = \frac{ \left[\left(\left[\mu_{ij}^{(k-1)} \right]^{m} + \left[t_{ij}^{(k-1)} \right]^{\delta} \right) \sum_{j=1}^{n} \left[\Sigma_t^{-1} \underline{a}_t - \left[\Sigma_i^{(k-1)} \right]^{-1} \underline{x}_j \right] \right] }{ \left[\sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k-1)} \right]^{m} + \left[t_{ij}^{(k-1)} \right]^{\delta} \right) \left[\Sigma_t^{-1} - \left[\Sigma_i^{(k-1)} \right]^{-1} \right] \right] },$$

$i = 1, 2, \ldots, c$

(34)

$$\Sigma_i^{(k)} = \frac{ \sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k-1)} \right]^{m} + \left[t_{ij}^{(k-1)} \right]^{\delta} \right) \left(\underline{x}_j - \left[\underline{a}_i \right]^{(k)} \right) \left(\underline{x}_j - \left[\underline{a}_i \right]^{(k)} \right)' }{ \sum_{j=1}^{n} \left(\left[\mu_{ij}^{(k-1)} \right]^{m} + \left[t_{ij}^{(k-1)} \right]^{\delta} \right) } \tag{35}$$

$i = 1, 2, \ldots, c$

$$\mu_{ij}^{(k)} = \left[\sum_{s=1}^{c} \left[\frac{ \left(\underline{x}_j - \left[\underline{a}_i \right]^{(k)} \right)' \left[\Sigma_i^{(k)} \right]^{-1} \left(\underline{x}_j - \left[\underline{a}_i \right]^{(k)} \right) - \ln \left| \left[\Sigma_i^{(k)} \right]^{-1} \right| - \left(\left[\underline{a}_i \right]^{(k)} - \underline{a}_t \right)' \Sigma_t^{-1} \left(\left[\underline{a}_i \right]^{(k)} - \underline{a}_t \right) }{ \left(\underline{x}_j - \left[\underline{a}_s \right]^{(k)} \right)' \left[\Sigma_s^{(k)} \right]^{-1} \left(\underline{x}_j - \left[\underline{a}_s \right]^{(k)} \right) - \ln \left| \left[\Sigma_s^{(k)} \right]^{-1} \right| - \left(\left[\underline{a}_s \right]^{(k)} - \underline{a}_t \right)' \Sigma_t^{-1} \left(\left[\underline{a}_s \right]^{(k)} - \underline{a}_t \right) } \right]^{\frac{1}{m-1}} \right]^{-1} \tag{36}$$

$i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, n$

$$t_{ij}^{(k)} = \left[\sum_{s=1}^{n} \left[\frac{ \left(\underline{x}_j - \left[\underline{a}_i \right]^{(k)} \right)' \left[\Sigma_i^{(k)} \right]^{-1} \left(\underline{x}_j - \left[\underline{a}_i \right]^{(k)} \right) - \ln \left| \left[\Sigma_i^{(k)} \right]^{-1} \right| - \left(\left[\underline{a}_i \right]^{(k)} - \underline{a}_t \right)' \Sigma_t^{-1} \left(\left[\underline{a}_i \right]^{(k)} - \underline{a}_t \right) }{ \left(\underline{x}_s - \left[\underline{a}_i \right]^{(k)} \right)' \left[\Sigma_i^{(k)} \right]^{-1} \left(\underline{x}_s - \left[\underline{a}_i \right]^{(k)} \right) - \ln \left| \left[\Sigma_i^{(k)} \right]^{-1} \right| - \left(\left[\underline{a}_i \right]^{(k)} - \underline{a}_t \right)' \Sigma_t^{-1} \left(\left[\underline{a}_i \right]^{(k)} - \underline{a}_t \right) } \right]^{\frac{1}{\delta-1}} \right]^{-1}$$

$i = 1, 2, \ldots, c$. $j = 1, 2, \ldots, n$ (37)

Step 3: Increment k; until $\max_{1\le i\le c}\left\|\underline{a}_i^{(k)}-\underline{a}_i^{(k-1)}\right\|<\varepsilon$

7. Experiment and Results

A real data set of 968 students from elementary schools was selected. These data included the 10 mathematics questions.

At first, the main factors of 968 data were calculated by using factor analysis. Next, according to the main factors, the samples were assigned to 4 clusters based on the clustering analysis using the k-mean clustering of SPSS for Windows 10.0. The results were shown in Table 1.

Table 1 The characteristics of 4 clusters

Cluster	samples size	Grade	average distance of the points from center of cluster
1	220	2	2.082132
2	435	4	1.433158
3	275	3	2.032674
4	56	1	2.356698

From Cluster 1, 15 samples randomly were selected, 15 from cluster 2, 15 from cluster 3, and 5 from cluster 4.

The combination the method of choosing the initial membership with distinct computing distance was shown in Table 2.

Table 2 Sample size of each group

Group	Number of Samples
1	15
2	15
3	15
4	5

The classification accuracies of testing samples were shown in Table 3.

From the data of Table 3, we found that the FPCM-CM algorithm could obtain the best results, and our previous algorithms, FPCM-M and FPCM-M are better than two well known algorithms, FPCM and FCM.

Table 3 Classification accuracies of testing samples.

Algorithms	Accuracies (%)
FCM	32
FPCM	30
FCM-M	56
FPCM-M	58
FPCM-CM	62

8. Conclusions

Two well known fuzzy partition clustering algorithms, FCM and FPCM are based on Euclidean distance function, which can only be used to detect spherical structural clusters. GK clustering algorithm and GG clustering algorithm, were developed to detect non-spherical structural clusters, but fail to consider the relationships between cluster centers in the objective function, needing additional prior information.. In our previous studies, we developed two improved algorithms, FCM-M and FPCM-M , based on unsupervised Mahalanobis distance without any additional prior information. And FPCM-M is better than FCM-M, since the former has the more information about the typicalities than the later. In this paper, an improved new unsupervised algorithm, “fuzzy possibility c-mean based on complete Mahalanobis distance without any prior information (FPCM-CM)”, is proposed. In our new algorithm, not only the local covariance matrix of each clusters but also the overall covariance matrix were considered. It can get more information and higher accuracy by considering the additional overall covariance matrix than FPCM-M. A real data set was applied to prove that the performance of the FPCM-CM algorithm is better than those of the traditional FCM and FPCM algorithm and our previous FCM-M, and our previous algorithms, FPCM-M and FPCM-M are better than two well known algorithms, FPCM and FCM.

In future, we will consider improve the initial value problem by using the swarm algorithm.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 96-2413--H-468-001).

References

[1] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, NY. 1981.

[2] D. E. Gustafson and W. C. Kessel, “Fuzzy clustering with a fuzzy covariance matrix,” Proc. IEEE Conf. Decision Contr., San Diego, CA, 1979..

[3] I. Gath, and A. B. Geva, “Unsupervised optimal fuzzy clustering,” IEEE Trans. Pattern Anal. Machine Intell., pp. 773-781, Jul. 1989.

[4] N. R. Pal, K. Pal and J. C. Bezdek, “A possibilistic approach to clustering,” IEEE Transactions on Fuzzy Systems, Val. 1, No 2, pp. 98-110, May 1993.
[5] J. C. Dunn, “A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters,” J. Cybern, vol. 3, no. 3 pp. 32-57, 1973.
[6] N. R. Pal, K. Pal and J.C. Bezdek, “A mixed c-mean clustering model”, Proceedings of the Sixth IEEE International conference on Fuzzy Systems, Vol. 1, pp. 11-21, Jul. 1997.
[7] Hsiang-Chuan Liu, Jeng-Ming Yih, Shin-Wu Liu (2007) Fuzzy c-mean algorithm based on Mahalanobis distances and better initial values, INFORMATION SCIENCES 2007, Proceedings of the 10th joint conferences, Salt Lake City, Utah, USA. 18-24 July, 2007, Page(s): 1398-1404.
[8] Hsiang-Chuan Liu, Jeng-Ming Yih, Tian-Wei Sheu, Shin-Wu Liu (2007).. A new fuzzy possibility c-mean algorithm based on unsuppervise Mahalanobis distances , Proceedings of International conference on Machine Learning and Cybernetics 2007, Vol. 7. No. 7 pp. 3939-3944, ISBN 1-4244-0973X.

Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, 30-31 Aug. 2008

AN IMPROVED SVM ALGORITHM BASED ON NORMALIZATION AND LIU–TRANSFORMATION

HSIANG-CHUAN LIU[1], YA-CHING CHIU[1], CHIEN-HSIUNG LIAO[1], TUNG-SHENG LIU[2]

[1]Department of Bioinformatics, Asia University, Taiwan
[2]Forensic Science Center of Military Police Command, Taiwan
E-MAIL:lhc@asia.edu.tw, ntctc123@yahoo.com.tw, ace.liao@gmail.com, tommy0366@yahoo.com.tw

Abstract:

The support vector machine (SVM) classifier is a popular and appealing classifier .It could be improved by taking some transformation about the original data before classification even sometimes its performance is not good,. In our previous paper, two transformations, NWFE-Transformation and Liu-Transformation are considered. The results showed that the SVM with our Liu-Transformation algorithm has the best performance.

In this paper, we considered the further improved SVM algorithm based on not only the Liu- transformation but also the well known normalization, For evaluating the performances of the SVM without any transformation and normalization, the SVM with NWFE-Transformation and Liu-Transformation, respectively, the SVM with one of above two transformations and the well known normalization, a real data experiment by using 5-fold and Leave-one-out Cross-Validation accuracy is conducted. Experimental result shows that the SVM with the proposed Liu-Transformation algorithm and the well known normalization algorithm has the best performance.

Keywords:

SVM; NWFE-Transformation; Liu-Transformation

1. Introduction

The support vector machine (SVM) classifier is a popular and appealing classifier [1], [2], [3], [4]. Due to sometimes its performance is not good, it can be improved by taking some transformation about the original data before classification, two transformations can be considered, one is NWFE-Transformation proposed by B. C. Kuo & D. A. Landgrebe in 2001 [5], [6], the other is Liu-Transformation proposed by our previous work in 2008 [7], [8]. The results of our previous paper [8] showed that the SVM with our Liu-Transformation algorithm has the best performance. In this paper, we considered the further improved SVM algorithm based on not only the Liu-transformation but also the well known normalization, For evaluating the performances of the SVM without any transformation and normalization, the SVM with NWFE-Transformation and Liu-Transformation, respectively, the SVM with one of above two transformations and the well known normalization, a real data experiment by using 5-fold and Leave-one-out Cross-Validation accuracy is conducted. Experimental result shows that the SVM with the proposed Liu-Transformation algorithm and the well known normalization algorithm has the best performance.

For evaluating the performances of the SVM without any transformation and normalization, the SVM with NWFE-Transformation and Liu-Transformation, respectively, the SVM with one of above two transformations and the well known normalization, a real data experiment by using 5-fold and Leave-one-out Cross-Validation accuracy is conducted. Experimental result shows that the SVM with the proposed Liu-Transformation algorithm and the well known normalization algorithm has the best performance

This paper is organized as followings: support vector machine classifier is introduced in section 2, NWFE-Transformation is introduced in section 3, Liu-Transformation is introduced in section 4. Normalization algorithm is described in section 5. Experiment and result are described in section 6 and final section is for conclusions and future works.

2. Support vector machine (SVM) [1], [2], [3], [4]

Given the training set of instance-labeled pairs $(\underline{x}_i, y_i), i = 1, 2, ..., N$, where

$$\underline{x}_i \in R^n, y_i \in \{1, -1\}, i = 1, 2, ..., N \quad (1)$$

The support vector machine (SVM) algorithm (Boser, Guyon, and Vapnik 1992, Cortes and Vapnik 1995) requires

978-1-4244-2239-5/08/$25.00 ©2008 IEEE

$$\min_{\underline{w},b,\xi} \frac{1}{2}\underline{w}'\underline{w}+c\sum_{i=1}^{N}\xi_i$$
$$\text{subject to } y_i\left(\underline{w}'\phi\left(\underline{x}_i\right)+b\right)\geq 1-\xi_i,$$
$$\xi_i \geq 0, \tag{2}$$
$$where \quad b,c\in R, \underline{w},\phi\left(\underline{x}_i\right)\in R^m$$
$$\phi: R^n \to R^m$$

For any testing point $\underline{x}_i \in R^n, y_i \in \{1,-1\}$, we can make an assignment according to the following formula:

$$f(\underline{x}_i)=\underline{w}'\varphi(\underline{x}_i)+b-(1-\xi_i)$$
$$y_i=\begin{cases}+1, & \text{if } f(\underline{x}_i)\geq 0\\ -1, & \text{if } f(\underline{x}_i)<0\end{cases} \tag{3}$$

3. NWFE-Transformation [5], [6]

The main ideas of nonparametric weighted feature extraction transformation (NWFE-Transformation)(Kuo, B. C. and Landgrebe, 2002, 2004) are putting different weights on every sample to compute the "local means" and defining new nonparametric weighted between-class and within-class scatter matrices to get more features.

The nonparametric weighted between-class scatter matrix, S_b^{NW} and the nonparametric weighted within-class scatter matrix, S_w^{NW}, of NWFE-Transformation are defined as

$$S_b^{NW}=\sum_{i=1}^{c}p_i\sum_{i\neq j=1}^{c}\sum_{k=1}^{n_i}\frac{\lambda_k^{(i,j)}}{n_i}\left[B_{k,j}^{(i)}\right]\left[B_{k,j}^{(i)}\right]^T$$
$$\left[B_{k,j}^{(i)}\right]=\left(\underline{x}_k^{(i)}-M_j\left(\underline{x}_k^{(i)}\right)\right) \tag{4}$$

$$S_w^{NW}=\sum_{i=1}^{c}p_i\sum_{k=1}^{n_i}\frac{\lambda_k^{(i,i)}}{n_i}\left[B_{k,i}^{(i)}\right]\left[B_k^{(i)}\right]^T$$
$$\left[B_{k,i}^{(i)}\right]=\left(\underline{x}_k^{(i)}-M_i\left(\underline{x}_k^{(i)}\right)\right) \tag{5}$$

where
$$M_j\left(\underline{x}_k^{(i)}\right)=\sum_{l=1}^{n_j}w_{k,l}^{(i,j)}\underline{x}_l^{(j)} \tag{6}$$

$$\lambda_k^{(i,j)}=\frac{d\left(\underline{x}_k^{(i)},M_j\left(\underline{x}_k^{(i)}\right)\right)^{-1}}{\sum_{l=1}^{n_i}d\left(\underline{x}_l^{(i)},M_j\left(\underline{x}_l^{(i)}\right)\right)^{-1}} \tag{7}$$

$$w_{k,l}^{(i,j)}=\frac{d\left(\underline{x}_k^{(i)},\underline{x}_l^{(j)}\right)^{-1}}{\sum_{l=1}^{n_i}d\left(\underline{x}_k^{(i)},\underline{x}_l^{(j)}\right)^{-1}} \tag{8}$$

C is the number of classes, p_i is the prior probability of class i, n_i is the training sample size of class i; $\underline{x}_k^{(i)}$ is the sample vector k with dimension d in class i; $M_j\left(\underline{x}_k^{(i)}\right)$ is the nonparametric weighted local mean of $\underline{x}_k^{(i)}$ in class j; $d\left(\underline{x},\underline{y}\right)$ is the Euclidean distance from x to y.

The goal of NWFE-transformation is to find a linear transformation $A\in R^{d\times p}, p\leq d$, which maximizes the between-class scatter and minimizes the within-class scatter. The columns of A are the optimal features by optimizing the following criterion

$$A=\arg\max_A tr\left[\left(A^T S_w^{NW}A\right)^{-1}A^T S_b^{NW}A\right] \tag{9}$$

This maximizing is equivalent to find the eigen-pairs $\left(\lambda_i,\underline{v}_i\right), i=1,2,...,d, \lambda_1\geq\lambda_2\geq...\geq\lambda_d$ for the generalized eigenvalue problem

$$S_b^{NW}\underline{v}=\lambda S_w^{NW}\underline{v} \tag{10}$$

4. Liu-Transformation [7]

The main ideas of Liu-transformation proposed by our previous work (Hsiang-Chuan Liu, 2008) [7] are putting different weights on every sample to compute the "weighted means" by referring the distances of the points from the 'outmost points' and defining new nonparametric weighted between-class and within-class scatter matrices to get more features.

Let $X_{p\times n}$ be the data set with n sample points and c classes, n_i be size of class i, satisfying $n=\sum_{i=1}^{c}n_i$; p_i be proportion of class i, satisfying $p_i=\frac{n_i}{n}$; $\underline{x}_k^{(i)}\in R^d$ be sample k in class i; $\underline{m}_i=\frac{1}{n_i}\sum_{k=1}^{n_i}\underline{x}_k^{(i)}$ be the original mean of class i; $\underline{x}_*^{(j)}$ be the outmost point of class j, satisfying

$$\underline{x}_*^{(j)}=\arg\max_{k=1,2,...,n_j}\sum_{i=1}^{c}d^2\left(\underline{x}_k^{(j)},\underline{m}_i\right) \tag{11}$$

$\underline{m}_j^*$ be weighted mean of class j by referring the distances of the sample points from the outmost point of class j satisfying

$$\underline{m}_j^* = \frac{\sum_{k=1}^{n_j} u_k^{(j)} \underline{x}_k^{(j)}}{\sum_{k=1}^{n_j} u_k^{(j)}} \tag{12}$$

where $u_k^{(j)} = \max_{l=1,2,\ldots n_j} dist^2\left(\underline{x}_l^{(j)}, \underline{x}_*^{(j)}\right) - dist^2\left(\underline{x}_k^{(j)}, \underline{x}_*^{(j)}\right)$ (13)

The nonparametric weighted between-class scatter matrix, S_b^L and the nonparametric weighted within-class scatter matrix, S_w^L, of Liu-Transformation are defined as

$$S_b^L = \sum_{i=1}^{c} p_i \sum_{i \neq j=1}^{c} \sum_{k=1}^{n_i} \frac{\lambda_k^{(i,j)}}{n_i} \left(\underline{x}_k^{(i)} - \underline{m}_j^*\right)\left(\underline{x}_k^{(i)} - \underline{m}_j^*\right)^T \tag{14}$$

where

$$\lambda_k^{(i,j)} = \frac{\max_{l=1,2,\ldots,n_i} dist^2\left(\underline{x}_l^{(i)}, \underline{m}_j^*\right) - dist^2\left(\underline{x}_k^{(i)}, \underline{m}_j^*\right)}{\sum_{l=1}^{n_i}\left[\max_{l=1,2,\ldots,n_i} dist^2\left(\underline{x}_l^{(i)}, \underline{m}_j^*\right) - dist^2\left(\underline{x}_k^{(i)}, \underline{m}_j^*\right)\right]} \tag{15}$$

$$S_w^{NW} = \sum_{i=1}^{c} p_i \sum_{k=1}^{n_i} \frac{\mu_k^{(i)}}{n_i} \left(\underline{x}_k^{(i)} - \underline{m}_i^*\right)\left(\underline{x}_k^{(i)} - \underline{m}_i^*\right)^T \tag{16}$$

$$\mu_k^{(i)} = \max_{l=1,2,\ldots n_i} dist^2\left(\underline{x}_l^{(i)}, \underline{m}_i^*\right) - dist^2\left(\underline{x}_k^{(i)}, \underline{m}_i^*\right) \tag{17}$$

The goal of Liu-transformation is to find a linear transformation $A \in R^{d \times p}, p \le d$, which maximizes the between-class scatter and minimizes the within-class scatter. The columns of A are the optimal features by optimizing the following criterion

$$A = \arg\max_A tr\left[\left(A^T S_w^L A\right)^{-1} A^T S_b^L A\right] \tag{18}$$

This maximizing is equivalent to find the eigen-pairs $(\lambda_i, \underline{v}_i), i = 1,2,\ldots,d, \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_d$ for the generalized eigenvalue problem

$$S_b^L \underline{v} = \lambda S_w^L \underline{v} \tag{19}$$

5. Normalization algorithm

Given the training set of instance-labeled pairs $(\underline{x}_i, y_i), i = 1,2,\ldots,N$. Let $\underline{x}_i = (x_{i,1}, x_{i,2}, x_{i,3}, \ldots, x_{i,n}) \in R^n$, then the normalization of $\underline{x}_i$ is $\underline{z}_i = (z_{i,1}, z_{i,2}, z_{i,3}, \ldots, z_{i,n}) \in R^n$ satisfying

$$z_{i,j} = \frac{x_{ij} - \overline{x}_j}{s_j}, i = 1,2,\ldots,N, J = 1,2,\ldots,n \tag{20}$$

where $\overline{x}_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij}, s_j^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_{ij} - \overline{x}_j\right)^2$ (21)

6. Experiment and result

A wine data set was downloaded from website, ftp://ftp.ics.uci.edu/pub/machine-learning-databases. The sample included 178 instances, 3 classes of wine, and 13 features for each instance.

The above real data is applied to evaluate the performances of the Support Vector Machine (SVM) algorithm without any transformation, the SVM algorithm with NWFE-Transformation, the SVM algorithm with Liu-Transformation, the SVM algorithm with normalization, the SVM algorithm with normalization and NWFE-Transformation, and the SVM algorithm with normalization and Liu-Transformation by using 5-fold and Leave-one-out Cross-Validation method to compute the accuracies of the response category variable.

Table 1 Accuracy of six Classification algorithms

Classification algorithm	5-fold CV accuracy	Leave-one-out CV accuracy
SVM	45.763	46.633
SVM_NWFE	93.023	96.305
SVM_N	97.740	98.740
SVM_ Liu	99.080	98.773
SVM_N_NWFE	100	100
SVM_N_Liu	100	100

The experimental results of six classification algorithms are listed in Table 1. For both 5-fold CV and Leave-one-out CV accuracy, we can find the same situations as following:

(i) The SVM algorithm with normalization and Liu-Transformation and the SVM algorithm with normalization and NWFE-Transformation had the same performance, better than others.

(ii) The SVM algorithm with just one of transformation or normalization is better than the SVM algorithm without any transformation and normalization.

(iii) The performance of the SVM algorithm without any transformation and normalization is not always good.

7. Conclusions and future works

The support vector machine (SVM) classifier is a popular and appealing classifier. Because sometimes its performance is not good, it could be improved by taking some transformation about the original data before classification. Two transformations can be considered, one is NWFE-Transformation proposed by B. C. Kuo & D. A. Landgrebe in 2001 [5], [6], the other is Liu-Transformation proposed by our previous work in 2008 [7], [8]. The results of our previous paper [8] showed that the SVM with our Liu-Transformation algorithm has the best performance. In this paper, we considered the further improved SVM algorithm based on not only the Liu- transformation but also the well known normalization.

For evaluating the performances of the SVM without any transformation and normalization, the SVM with NWFE-Transformation and Liu-Transformation, respectively, the SVM with one of above two transformations and the well known normalization, a real thyroid data included 178 instances, 3 classes of wine, and 13 features for each instance is conducted.

The above real data is applied to evaluate the performances of the Support Vector Machine (SVM) algorithm without any transformation, the SVM algorithm with NWFE-Transformation, the SVM algorithm with Liu-Transformation, the SVM algorithm with normalization, the SVM algorithm with normalization and NWFE-Transformation, and the SVM algorithm with normalization and Liu-Transformation by using 5-fold and Leave-one-out Cross-Validation method to compute the accuracies of the response category variable.

The experimental results of six classification algorithms are listed in Table 1. Both 5-fold CV and Leave-one-out CV accuracy, we can find the same situations as following;

(i) The SVM algorithm with Liu-Transformation is better than the SVM algorithm with NWFE-Transformation had and the SVM algorithm without any transformation.

(ii) The SVM algorithm with normalization and Liu-transformation and the SVM algorithm with normalization and NWFE-Transformation are same better than others.

In future, we will apply our Liu-Transformation with normalization to improve the performances of other classifiers.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 96-2413--H-468-001).

References

[1] Boser, B.E., Guyon, I.M., and Vapnik, V. "A training algorithm for optimal margin classifiers", In Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, 1992. ACM.

[2] C. Cortes, and V. Vapnik, "Support-vector network", Machine Learning, Vol. 20, pp. 273-297, 1995.

[3] Vapnik, V., The Nature of Statistical Learning Theory. New York, NY. Springer-Verlay, 1995.

[4] Chang, C.-C., and Lin, C.-J. LIBSVM; a library for support vector machine Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2004.

[5] B. C. Kuo, and D. A. Landgrebe, "Improved Statistics Estimation and Feature Extraction for Hyperspectral Data Classification", Technical Report, Purdue University, West Lafayette. IN., TR-ECE 01-6, December, 2001.

[6] B. C. Kuo, and D. A. Landgrebe, "Nonparametric Weighted Feature Extraction for Classification", IEEE Trans. on Geosience and Remove Sensing, Vol. 42, No. 5, 1096-1105, May, 2004.

[7] Hsiang-Chuan Liu, "A novel nonparametric weighted feature extraction transformation algorithm based on the outmost points", Journal of Taichung University (JNTCU), Vol. 22, No. 1, pp. 1-7, June 2008.

[8] Hsiang-Chuan Liu, Chien-Hsiung Liao, Tung-Sheng Liu, and Ya-Ching Chiu, "An improved SVM algorithm based Liu–Transformation", 2008 International Conference on Management and Technology, Yunlin, Taiwan, June 2008.

Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, 30-31 Aug. 2008

A NOVEL CLASSIFIER FOR INFLUENZA A VIRUSES BASED ON SVM AND LOGISTIC REGRESSION

HSIANG-CHUAN LIU[1], SHIN-WU LIU[2], PEI-CHUN CHANG[1], WEN-CHUN HUANG[3], CHIEN-HSIUNG LIAO[1]

[1]Department of Bioinformatics, Asia University, Taiwan
[2]National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA
[3]Graduate Institute of Educational Measurement and Statistics, Taichung University, Taiwan
E-MAIL: lhc@asia.edu.tw, liushin@mail.nih.gov, pcchang@asia.edu.tw, huang.hwc@gmail.com, ace.liao@gmail.com

Abstract:

In search of good classifier of hosts of influenza A viruses is an important issue to prevent pandemic flu. The hemagglutinin protein in the virus genome is the major molecule that determining the range of hosts. In this paper, a novel classification algorithm of hemagglutinin proteins integrating SVM and logistic regression based on 4 kinds of Hurst exponents for each protein sequence is proposed. This method not used before is the first one integrating the physicochemical properties, fractal property, SVM and logistic regression classifier. For evaluating the performance of this new algorithm, a real data experiment by using 5-fold Cross-Validation accuracy is conducted. Experimental result shows that this new classification algorithm is useful and batter than SVM and logistic regression, respectively.

Keywords:

Influenza A viruses; Hurst exponent; SVM; Logistic regression; SVM-Logistic regression

1. Introduction

Influenza A viruses are negative-strand RNA viruses that infect a wide variety of animals in the nature. The infection of human may cause significant mortality and morbidity worldwide [1]. The hemagglutinin (HA) protein in the virus genome is the major molecule that determining the range of hosts. The natural reservoir of influenza virus such as avian flu may emerges in strains infectious to human by mutation of HA protein and brings pandemic flu, therefore, in search of good classification algorithm of HA proteins is an important issue to prevent pandemic flu. In this paper, a novel classification algorithm of HA proteins combining Hurst exponents, SVM and logistic regression is proposed [2], [3], [4], [5]. This method not used before is the first one integrating the physicochemical properties, fractal property, support vector machine (SVM) and logistic regression classifier.

The protein residues were coded according to its physicochemical quantities of acidity, Van der waal volume, surface area and hydrophobicity in the situation of single amino acid [6], [7]

First step, the HA sequence data of serotype H5 of influenza A viruses with two classes used in this research were downloaded from public databases: Influenza Sequence Database (http://www.flu.lanl.gov). The sample included 90 HA protein sequences of human infections and 90 HA protein sequences of bird infections.

Second step, to replace each residue of amino acid in the sequences of the HA proteins with 4 physicochemical quantities.

Third step, computing the Hurst exponents of each non-symbolic sequences of the HA proteins, we can obtained four features of Hurst exponents in each sequences of the HA protein [2], [6], [7].

Last step, two well known and appealing classifiers, Support Vector Machine (SVM) and Logistic regression (LR), and our new hybrid classifier combining SVM and LR were used to discriminate the correct class of the 180 HA proteins with four features of Hurst exponents.

For evaluating the performance of above three classifiers, the above HA proteins data experiment by using 5-fold Cross-Validation accuracy is conducted.

This paper is organized as followings: four physicochemical quantities of 20 amino acids are introduced in section 2, Hurst exponent is introduced in section 3, support vector machine classifier is introduced in section 4, logistic regression is introduced in section 5, the new hybrid classifier combining SVM and logistic regression is introduced in section 6, experiment and result are described in section 7 and final section is for conclusions and future works.

2. Four physicochemical properties of amino acids

There are four physicochemical quantities of acidity, Van der waal volume, surface area and hydrophobicity in the situation of single amino acid showed as Table 1 [2], [3]

Table 1. 20 amino acids and its 4 physicochemical quantities

Amino acid	Acidity	Van der waal Volume	Surface area	HydroPh-obicity
A	7.0	67	115	0.616
C	3.4	86	135	0.680
D	3.9	67	150	0.028
E	4.1	109	190	0.043
F	7.0	135	210	1.000
G	7.0	48	75	0.501
H	6.0	118	195	0.165
I	7.0	124	175	0.943
K	10.5	135	200	0.283
L	7.0	124	170	0.943
M	7.0	124	185	0.738
N	7.0	148	160	0.236
P	7.0	90	145	0.711
Q	7.0	114	180	0.251
R	12.5	167	225	0.000
S	7.0	73	115	0.359
T	7.0	93	140	0.450
V	7.0	105	155	0.825
W	7.0	163	255	0.878
Y	10.5	141	230	0.880

3. Hurst exponent

The Hurst exponent occurs in several areas of applied mathematics, including fractals and chaos theory, long term memory processes and spectral analysis [8]. Hurst exponent estimation has been applied in areas ranging from biophysics to computer networking. Estimation of the Hurst exponent was originally developed in hydrology. However, the modern techniques for estimating the Hurst exponent comes from fractal mathematics.

Estimating the Hurst exponent for a data set provides a measure of whether the data is a pure random walk or has underlying trends.

The Hurst exponent (H) is a statistical measure used to classify time series. H=0.5 indicates a random series while H>0.5 indicates a trend reinforcing series. The larger the H value is, the stronger the trend. Experiments with backpropagation Neural Networks show that series with large Hurst exponent can be predicted more accurately than those with H value close to 0.50. Thus the Hurst exponent provides a measure for predictability.

Three methods were used most often for the estimation of the Hurst exponent: the R/S method, the roughness–length (R–L) method and a variogram. The R/S method (Hurst et al., 1965) [9] is commonly perceived as the most suitable for the time series analysis, because it presents the relationship between irregular (singular) rescaled ranges, signal value and their local statistical properties relative to the scale factor.

In this study R/S method is used. R/S method [10] is based on empirical observations by Hurst and estimates H are based on the R/S statistic. It indicates (asymptotically) second-order self-similarity. H is roughly estimated through the slope of the linear line in a log-log plot, depicting the R/S statistics over the number of points of the aggregated series. That is, given a time sequence of observations, w_t define the Series

$$W(t,\tau)=\sum_{u=1}^{t}(w_u-\bar{w}_\tau),\ 1\le t\le\tau \tag{5}$$

where

$$\bar{w}_\tau=\frac{1}{\tau}\sum_{t=1}^{\tau}w_t \tag{6}$$

Define

$$R(\tau)=\max_{t=1}^{\tau}W(t,\tau)-\min_{t=1}^{\tau}W(t,\tau) \tag{7}$$

and

$$S(\tau)=\sqrt{\left(\frac{1}{\tau}\sum_{t=1}^{\tau}(w_t-\bar{w}_\tau)^2\right)} \tag{8}$$

In plotting $\log\frac{R(\tau)}{S(\tau)}$ against $\log\tau$, we expect to get a line whose slope determines the Hurst exponent.

4. Support vector machine (SVM) [11~14]

Given the training set of instance-labeled pairs $(\underline{x}_i, y_i), i=1,2,...,N$, where

$$\underline{x}_i\in R^n, y_i\in\{1,-1\}, i=1,2,...,N \tag{9}$$

The support vector machine (SVM) algorithm (Boser, Guyon, and Vapnik 1992 [11], Cortes and Vapnik 1995 [12]) requires

$$\min_{\underline{w},b,\xi} \frac{1}{2}\underline{w}'\underline{w}+c\sum_{i=1}^{N}\xi_i$$
$$\text{subject to } y_i\left(\underline{w}'\phi\left(\underline{x}_i\right)+b\right)\geq 1-\xi_i,$$
$$\xi_i \geq 0, \tag{10}$$
$$where \quad b,c\in R, \underline{w},\phi\left(\underline{x}_i\right)\in R^m$$
$$\phi: R^n \rightarrow R^m$$

For any testing point $\underline{x}_i \in R^n$, $y_i \in \{1,-1\}$, we can make an assignment according to the following formula.

$$d(\underline{x}_i)=\left[\underline{w}'\varphi\left(\underline{x}_i\right)+b-\left(1-\xi_i\right)\right]$$
$$y_i=\begin{cases}+1, & \text{if } d(\underline{x}_i)\geq 0\\ -1, & \text{if } d(\underline{x}_i)<0\end{cases} \tag{11}$$

5. Multiple Logistic regression classifier

5.1. Multiple logistic regression model [4], [5]

Let $\left(x_{i1},x_{i2}....,x_{in}y_i\right), i=1,2,...,N$ be a sample data, satisfying $\underline{x}_i=\left(x_{i1},x_{i2},...,x_{in}\right)\in R^n$, $y_i\in\{0,1\}$,

$$Y_i^{\perp\perp}\sim B\left(1,p_i\right), i=1,2,...,N \tag{12}$$

The multiple logistic regression model is denoted as follows

$$P_i=P\left(Y_i=1\mid x_i\right)=\frac{1}{1+\exp\left[-\left(\alpha+\underline{\beta}'\underline{x}\right)\right]}, i=1,2,...,N \tag{13}$$

where $\underline{\beta}'=\left(\alpha,\beta_1,\beta_2,...,\beta_n\right)$ are parameters vector of regression coefficients.

5.2. Multiple logistic regression classifier [5]

We can obtain the likelihood function and log likelihood function as following equations (14) and (15)

$$L\left(p_1,p_2,...,p_N\right)=\prod_{i=1,2,...,N} p_i^{y_i}\left(1-p_i\right)^{1-y_i} \tag{14}$$

$$l=\log L\left(p_1,p_2,...,p_N\right)=\sum_{i=1}^{N}\left[y_i\log p_i+\left(1-y_i\right)\left(1-\log p_i\right)\right] \tag{15}$$

And we can get

$$l=l\left(\alpha,\underline{\beta}\right)=\sum_{i=1}^{N}\left[y_i\log p_i+\left(1-y_i\right)\left(1-\log p_i\right)\right]$$
$$=-\sum_{i=1}^{N}\left[\log\left(1+exp\left[-\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]\right)+\left(1-y_i\right)\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right] \tag{16}$$

Where $\alpha\in R$, $\underline{\beta}'=\left(\beta_1,\beta_2,...,\beta_n\right)\in R^n$

Using Newton-Raphson's iterative algorithm, we can get the estimated regression coefficients of the multiple logistic regression model and the estimated multiple logistic regression equation as follows:

$$\hat{P}_i=\hat{P}\left(Y_i=1\mid x_i\right)=\frac{1}{1+\exp\left[-\left(\hat{\alpha}+\underline{\hat{\beta}}'\underline{x}\right)\right]} \tag{17}$$

$$\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_n\end{bmatrix}_{k+1}=\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_n\end{bmatrix}_{k}-\begin{bmatrix}\frac{\partial^2 l}{\partial\alpha^2} & \frac{\partial^2 l}{\partial\alpha\partial\beta_1} & \cdots & \frac{\partial^2 l}{\partial\alpha\partial\beta_n}\\ \frac{\partial^2 l}{\partial\beta_1\partial\alpha} & \frac{\partial^2 l}{\partial\beta_1^2} & \cdots & \frac{\partial^2 l}{\partial\alpha\partial\beta_n}\\ \vdots & \ddots & & \vdots\\ \frac{\partial^2 l}{\partial\beta_n\partial\alpha} & \frac{\partial^2 l}{\partial\beta_n\partial\beta_1} & \cdots & \frac{\partial^2 l}{\partial\beta_n^2}\end{bmatrix}_k^{-1}\begin{bmatrix}\frac{\partial l}{\partial\alpha}\\ \frac{\partial l}{\partial\beta_1}\\ \vdots\\ \frac{\partial l}{\partial\beta_n}\end{bmatrix}_k \tag{18}$$

where

$$\frac{\partial l}{\partial\alpha}=\sum_{i=1}^{N}\left[\frac{1}{1+\exp\left[-\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]}-\left(1-y_i\right)\right] \tag{19}$$

$$\frac{\partial l}{\partial\beta_j}=\sum_{i=1}^{N}\left[\frac{1}{1+\exp\left[-\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]}-\left(1-y_i\right)\right]x_{ij}, j=1,2,...,n \tag{20}$$

$$\frac{\partial^2 l}{\partial\alpha^2}=-\sum_{i=1}^{N}\frac{\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2} \tag{21}$$

$$\frac{\partial^2 l}{\partial\beta_j^2}=-\sum_{i=1}^{N}\frac{x_{ij}^2\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2}, j=1,2,...,n \tag{22}$$

$$\frac{\partial^2 l}{\partial\alpha\partial\beta_j}=\frac{\partial^2 l}{\partial\beta_j\partial\alpha}=-\sum_{i=1}^{N}\frac{x_{ij}\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2}, j=1,2,...,n \tag{23}$$

$$\frac{\partial^2 l}{\partial\beta_j\partial\beta_k}=\frac{\partial^2 l}{\partial\beta_k\partial\beta_j}=-\sum_{i=1}^{N}\frac{x_{ij}x_{ik}\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)}{\left[1+\exp\left(\alpha+\underline{\beta}'\underline{x}_i\right)\right]^2}, j,k=1,2,...,n$$

Increment k; until $\left\| \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}_{k+1} - \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}_k \right\| < \varepsilon$ (24)

Then $y_i = \begin{cases} 1 & if\ \hat{P}_i = \hat{P}(Y_i = 1 \mid x_i) \geq 0.5 \\ 0 & if\ \hat{P}_i = \hat{P}(Y_i = 1 \mid x_i) < 0.5 \end{cases}$ (25)

6. SVM-Logistic regression classifier

In this paper, an improved hybrid classifier combining SVM and logistic regression is proposed here.

First step, using the SVM classifier, we can find the signed distance, $d(\underline{x}_i)$, between the point $\underline{x}_i = (x_{i1}, x_{i2}, ..., x_{in})$ and the hyperplane in SUM.

Second step, to consider the sample data $(d(\underline{x}_i), y_i), i = 1,2,...,N$, using the simple logistic regression to classify y_i.

6.1. Mathematical formulas

Let $(x_{i1}, x_{i2}...., x_{in} y_i), i = 1,2,..., N$ be a sample data, satisfying

$$\underline{x}_i = (x_{i1}, x_{i2}, ..., x_{in}) \in R^n,\ y_i \in \{0,1\} \quad (26)$$

Using the above support vector machine (SVM) algorithm, from equation (11), for any point $\underline{x}_i \in R^n$, we can obtain the signed distance as below

$$d(\underline{x}_i) = \left[\underline{w}' \varphi(\underline{x}_i) + b - (1 - \xi_i) \right] \quad (27)$$

6.2. Simple logistic regression classifier of the working sample data

Let the working sample data $(d(\underline{x}_i), y_i), i = 1,2,..., N$ satisfying $d(\underline{x}_i) \in R,\ y_i \in \{1,0\}$

$$Y_i^{\perp\perp} \sim B(1, p_i), i = 1,2,..., N \quad (28)$$

The simple logistic regression model is denoted as follows

$$P_i = P(Y_i = 1 \mid d(\underline{x}_i)) = \frac{1}{1 + \exp\left[-(\alpha + \beta d(\underline{x}_i)) \right]}, i = 1,2,..., N \quad (29)$$

Similarly as multiple logistic regression classifier, we can get log likelihood function, the estimated regression coefficients of the simple logistic regression model and the estimated simple logistic regression equation as follows:

$$l = l(\alpha, \beta) = \sum_{i=1}^{N} \left[y_i \log p_i + (1 - y_i)(1 - \log p_i) \right]$$
$$= -\sum_{i=1}^{N} \left[\log\left(1 + exp\left[-(\alpha + \beta d(\underline{x}_i)) \right]\right) + (1 - y_i)(\alpha + \beta d(\underline{x}_i)) \right] \quad (30)$$

$$\hat{P}_i = \hat{P}(Y_i = 1 \mid d(\underline{x}_i)) = \frac{1}{1 + \exp\left[-(\hat{\alpha} + \hat{\beta} d(\underline{x}_i)) \right]} \quad (31)$$

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}_{k+1} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}_k - \begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} \\ \frac{\partial^2 l}{\partial \beta \partial \alpha} & \frac{\partial^2 l}{\partial \beta^2} \end{bmatrix}_k^{-1} \begin{bmatrix} \frac{\partial l}{\partial \alpha} \\ \frac{\partial l}{\partial \beta} \end{bmatrix}_k \quad (32)$$

where

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{N} \left[\frac{1}{1 + \exp\left[-(\alpha + \beta d(\underline{x}_i)) \right]} - (1 - y_i) \right] \quad (33)$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{N} \left[\frac{1}{1 + \exp\left[-(\alpha + \beta d(\underline{x}_i)) \right]} - (1 - y_i) \right] d(\underline{x}_i) \quad (34)$$

$$\frac{\partial^2 l}{\partial \alpha^2} = -\sum_{i=1}^{N} \frac{\exp(\alpha + \beta d(\underline{x}_i))}{\left[1 + \exp(\alpha + \beta d(\underline{x}_i)) \right]^2} \quad (35)$$

$$\frac{\partial^2 l}{\partial \beta^2} = -\sum_{i=1}^{N} \frac{\left[d(\underline{x}_i) \right]^2 \exp(\alpha + \beta d(\underline{x}_i))}{\left[1 + \exp(\alpha + \beta d(\underline{x}_i)) \right]^2} \quad (36)$$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta} = \frac{\partial^2 l}{\partial \beta \partial \alpha} = -\sum_{i=1}^{N} \frac{d(\underline{x}_i) \exp(\alpha + \beta d(\underline{x}_i))}{\left[1 + \exp(\alpha + \beta d(\underline{x}_i)) \right]^2} \quad (37)$$

Increment k; until $\left\| \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix}_{k+1} - \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}_k \right\| < \varepsilon$ (38)

Then $y_i = \begin{cases} 1 & if\ \hat{P}_i = \hat{P}(Y_i = 1 \mid d(\underline{x}_i)) \geq 0.5 \\ 0 & if\ \hat{P}_i = \hat{P}(Y_i = 1 \mid d(\underline{x}_i)) < 0.5 \end{cases}$ (39)

7. Experiment and result

The sequence data of serotype H5 of Influenza A viruses with two classes used in this research were obtained from public databases: Influenza Sequence Database (http://www.flu.lanl.gov). The sample included 90 HA protein sequences of human infections and 90 HA protein sequences of bird infections.

The protein residues were coded according to its physicochemical quantities of acidity, Van der waal volume, surface area and hydrophobicity in the situation of single amino acid as Table 1.

Computing the Hurst exponents of each non-symbolic sequences of the HA proteins, we can obtain four features represented as Hurst exponents respectively in each sequences of the HA protein.

The above real data with four features in terms of Hurst exponents is applied to evaluate the performances of the Support Vector Machine (SVM) algorithm, logistic regression and the proposed classifier combining SVM and logistic regression classifier by using 5-fold Cross-Validation method to compute the accuracies of the response category variable.

The experimental results for Accuracies of above three classifiers are listed in Table 2. We can find that our new classification algorithm is useful and batter than SVM and logistic regression, respectively.

Table 2 Accuracies of three classifiers

Classifier	5-fold CV accuracy
SVM	0.8056
LR	0.8833
SVM-LR	0.9000

8. Conclusions and future works

In search of good classifier of influenza viruses is an important issue to prevent pandemic flu. In this paper, a novel classification algorithm of HA proteins integrating SVM and logistic regression based on 4 kinds of Hurst exponents for each protein sequence is proposed. This method not used before is the first one integrating the physicochemical properties, fractal property, SVM and logistic regression classifier. For evaluating the performance of this new algorithm, a real data experiment by using 5-fold Cross-Validation accuracy is conducted. Experimental result shows that this new classification algorithm is useful and batter than SVM and logistic regression, respectively.

Our proposed new classifier can be used to classify not only the data of Influenza A viruses but also the data of other biological sequences.

In future, we will consider look for some further improving classification algorithms by using Hurst exponent and other hybrid Classifiers.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 96-2413--H-468-001).

References

[1] P. Pale, “ Influenza: old and new threats”, Nat. Med , Vol.10, pp. 82–87, 2004.

[2] H. E. Hurst, “Long term storage capacity of reservoirs”, Transactions of the American Society of Civil Engineers 116, pp. 770-799, 1951.

[3] C. Cortes, and V., Vapnik, “Support-vector network”, Machine Learning, Vol. 20, pp. 273-297, 1995.

[4] D. R. Cox, and E. J. Snell, The analysis of binary data (2^{nd} ed.) London, Chapman & Hall, 1989.

[5] Hsiang-Chuan Liu, Yu-Du Jheng, Guey-Shya Chen, Bai-Cheng Jeng, “A new classification algorithm combining Choquet integral and logistic regression”, 2008 International Conference on Machine Learning and Cybernetics, 12-15 July 2008 Kunming, China (accepted).

[6] R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers., and Y. Kawaoka, “Evolution and ecology of influenza A viruses”, Microbiol. Rev., Vol. 56, pp. 152-179, 1992.

[7] N. J. Cox and K. Subbarao, “Global epidemiology of influenza: Past and present”, Annu. Rev. Med., Vol. 5*1,* pp. 407-421, 2000.

[8] T. Di Matteo, T. Aste and M. M. Dacorogna, "Longterm memories of developed and emerging markets: using the scaling analysis to characterize their stage of development", Journal of Banking & Finance 29/4, pp. 827-851, 2005.

[9] H. E. Hurst, R. Black, Y. M. Sinaika, “Long term storage capacity of reservoirs” , An experimental study Constable, London, 1965.

[10] Roger Kalden, Sami Ibrahim, “Searching for Self-Similarity in GPRS”, PAM, pp. 83-92, 2004.

[11] B.E. Boser, I.M. Guyon, and V. Vapnik, “A training algorithm for optimal margin classifiers”, In Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, 1992. ACM.

[12] C. Cortes, and V. Vapnik, “Support-vector network”, Machine Learning, Vol. 20, pp. 273-297, 1995.

[13] V. Vapnik, The Nature of Statistical Learning Theory. New York, NY. Springer-Verlay, 1995.

[14] C.-C. Chang, and C.-C. Lin, LIBSVM; a library for support vector machine Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2004

The Choquet Integral with Respect to R-Measure Based on γ-Support

Hsiang-Chuan Liu[1], Yu-Chieh Tu[2], Wen-Chun Huang[2], Chin-Chun Chen[2,3]
[1]*Department of Bioinformatics, Asia University, Taiwan*
[2]*Graduate Institute of Educational Measurement and Statistics, Taichung University, Taiwan*
[3]*General Education, Min-Hwei College of Health Care Management, Taiwan*
lhc@asia.edu.tw, tuyuchieh@yahoo.com.tw, huang.hwc@gmail.com, cc_chen0336@yahoo.com.tw

Abstract

When the multicollinearity within independent variables occurs in the multiple regression models, its performance will always be poor. Replacing the above models with the ridge regression model is the traditional improved method. In our previous work, we found that, the Choquet integral regression model with λ-measure based on the new support, γ-support, proposed by us has the best performance than before. In this study, for finding the further improved model, we replaced two well known fuzzy measures, P-measure and λ-measure with our new fuzzy measure, R-measure in Choquet integral regression model with the new support, γ-support. For comparing the Choquet integral regression model with P-measure, λ-measure and R-measure based on two different fuzzy supports, V-support and γ-support, respectively, the traditional multiple regression model and the ridge regression model, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted. Experimental result shows that the Choquet integral regression model with R-measure based on γ-support has the best performance.

1. Introduction

When interactions among independent variables exist in forecasting problems, the performance of the multiple linear regression models is poor. The traditional improved methods exploited the ridge regression models [1]. Recently, some Choquet integral regression models based on different fuzzy measures were used by our previous works to further improve this situation [2], [3], [4], [5].

In our previous works [6], we found that if the Choquet integral regression model based on the same fuzzy measure is derived from different fuzzy support, then it may have different performances, in other words, the better performance of a Choquet integral regression model is not only derived from a better fuzzy measure but also first derived from a better fuzzy support. Hence, before we find the better fuzzy measure of a Choquet integral regression model, we need first to find a better fuzzy support of the same fuzzy measure of that Choquet integral regression model. And we found that the Choquet integral regression model with λ-measure based on the new support, γ-support, proposed by us has the best performance than before.

In this study, the Choquet integral regression model with two well known fuzzy measures, P-measure and λ-measure and our new fuzzy measure, R-measure based on the V-support and γ-support, respectively, were considered. For comparing the performances of the above different Choquet integral regression models with the multiple regression model and the ridge regression model, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted.

This paper is organized as followings: The multiple linear regression and ridge regression are introduced in section 2, two well known fuzzy measure, P-measure and λ-measure are introduced in section 3, R-measures are introduced in section 4, two kind fuzzy supports: V-support and γ-support are described in section 5. The Choquet integral regression model based on fuzzy measures are described in section 6. Experiment and result are described in section 7, and final section is for conclusions and future works.

2. The multiple linear regression, ridge regression [1]

Let $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{\varepsilon} \sim \mathrm{N}\left(\underline{0}, \sigma^2 I_n\right)$ be a multiple linear model, $\underline{\hat{\beta}} = \left(X'X\right)^{-1} X'Y$ be the estimated regression coefficient vector, and $\underline{\hat{\beta}}_k = \left(X'X + kI_n\right)^{-1} X'Y$ be the estimated ridge regression coefficient vector, Kenard and Baldwin [1] suggested

$$\hat{k} = \frac{n\hat{\sigma}^2}{\underline{\hat{\beta}}'\underline{\hat{\beta}}}. \tag{1}$$

3. Fuzzy measures

978-0-7695-3305-6/08 $25.00 © 2008 IEEE
DOI 10.1109/FSKD.2008.545

The well known fuzzy measures, P-measure proposed by Zadah in 1978, and the λ-measure proposed by Sugeno in 1974, are concise introduced as follows.

3.1. Fuzzy measures [7], [8], [9]

A fuzzy measure μ on a finite set X is a set function $\mu : 2^X \to [0,1]$ satisfying the following axioms:

(i) $\mu(\phi) = 0$, $\mu(X) = 1$ (boundary conditions) (2)

(ii) $A \subseteq B \Rightarrow \mu(A) \le \mu(B)$ (monotonicity) (3)

3.2. Singleton measures [4], [5]

A singleton measure of a fuzzy measure μ on a finite set X is a function $s : X \to [0,1]$ satisfying:

$$s(x) = \mu(\{x\}), x \in X \tag{4}$$

$s(x)$ is called the density of singleton x .

3.3. P-measure [10]

For given singleton measures s, a P-measure, g_P , is a fuzzy measure on a finite set X, satisfying:

$$^{\forall} A \in 2^X \Rightarrow g_P(A) = \max_{x \in A} s(x) = \max_{x \in A} g_P(\{x\}) \tag{5}$$

3.4. λ-measure [8], [9]

For given singleton measures s, a λ-measure, g_λ , is a fuzzy measure on a finite set X, satisfying:

(i) $A, B \in 2^X$, $A \cap B = \phi$, $A \cup B \ne X$

$$\Rightarrow g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B) \tag{6}$$

(ii) $\prod_{i=1}^{n}\left[1 + \lambda s(x_i)\right] = \lambda + 1 > 0, s(x_i) = g_\lambda(\{x_i\})$ (7)

Note that once the singleton measure is known, we can obtain the values of λ uniquely by using the previous polynomial equation. In other words, λ-measure has a unique solution without closed form.

4. R-measure [4]

For given singleton measure s, a R-measure, g_R , is a fuzzy measure on a finite set X, $|X| = n$, satisfying:

(i) $R \in [0, \infty)$ (8)

(ii) $\sum_{x \in X} s(x) = \sum_{x \in X} g_R(\{x\}) = 1$ (9)

(iii) $^{\forall} A \subset X, n - |A| + (|A| - 1)R > 0$

$$\Rightarrow g_R(A) = \max_{x \in A}\left[s(x)\right] + \frac{(|A| - 1)R \sum_{x \in A} s(x)}{\left[n - |A| + (|A| - 1)R\right]}\left[1 - \max_{x \in X}\left[s(x)\right]\right] \tag{10}$$

[Property]

(i) R-measure has infinitely many solutions with closed form.

(ii) When R=0, the R-measure is just a P-measure with closed form.

(iii) g_R is an increasing function of R.

5. Fuzzy supports

For given singleton measures s of a fuzzy measure μ on a finite set X, if $\sum_{x \in X} s(x) = 1$, then s is called a fuzzy support measure of μ, or a fuzzy support of μ, or a support of μ. Two kinds of fuzzy supports are introduced as below.

5.1. V-support [6]

Let μ be a fuzzy measure on a finite set, $X = \{x_1, x_2, ..., x_n\}$ be the set of n courses, $f_1(x_j), f_2(x_j), ..., f_N(x_j),\ j = 1,2,...,n$ be the evaluating scores of subject i for singleton x_j , satisfying:

$$0 < f_i(x_j) < 1, i = 1,2,...,N,\ j = 1,2,...,n \tag{11}$$

If $V(x_j) = \dfrac{V_{ar}(f(x_j))}{\sum_{k=1}^{n} V_{ar}(f(x_k))},\ j = 1,2,...,n$ (12)

where $V_{ar}(f(x_j)) = \dfrac{1}{N}\sum_{i=1}^{N}\left[f_i(x_j) - \dfrac{1}{N}\sum_{i=1}^{N} f_i(x_j)\right]^2$ (13)

satisfying $0 \le V(x_j) \le 1$ and $\sum_{j=1}^{n} V(x_j) = 1$ (14)

then the function $V : X \to [0,1]$ satisfying $\mu(\{x\}) = V(x)$, $\forall x \in X$ is a fuzzy support of μ, called V-support of μ.

5.2. γ- support [6]

Let μ be a fuzzy measure on a finite set $X = \{x_1, x_2, ..., x_n\}$, y_i be global response of subject i and $f_i(x_j)$ be the evaluation of subject i for singleton x_j , satisfying:

$$0 < f_i\left(x_j\right) < 1, i = 1,2,...,N,\ j = 1,2,...,n \quad (15)$$

If $$\gamma\left(x_j\right) = \frac{1 + r\left(f\left(x_j\right)\right)}{\sum_{k=1}^{n}\left[1 + r\left(f\left(x_k\right)\right)\right]},\ j = 1,2,...,n, \quad (16)$$

where $$r\left(f\left(x_j\right)\right) = \frac{S_{y,x_j}}{S_y S_{x_j}} \quad (17)$$

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - \frac{1}{N}\sum_{i=1}^{N} y_i\right)^2 \quad (18)$$

$$S_{x_j}^2 = \frac{1}{N}\sum_{i=1}^{n}\left[f_i\left(x_j\right) - \frac{1}{N}\sum_{i=1}^{N} f_i\left(x_j\right)\right]^2 \quad (19)$$

$$S_{y,x_j} = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - \frac{1}{N}\sum_{i=1}^{N} y_i\right)\left[f_i\left(x_j\right) - \frac{1}{N}\sum_{i=1}^{N} f_i\left(x_j\right)\right] \quad (20)$$

Satisfying $0 \le \gamma\left(x_j\right) \le 1$ and $\sum_{j=1}^{n}\gamma\left(x_j\right) = 1$ (21)

then the function $\gamma : X \to [0,1]$ satisfying $\mu\left(\{x\}\right) = \gamma(x)$, $\forall x \in X$ is a fuzzy support of μ, called γ-support of μ.

6. Choquet integral regression models

6.1. Choquet integral [4], [9], [10]

Let μ be a fuzzy measure on a finite set X. The Choquet integral of $f_i : X \to R_+$ with respect to μ for individual i is denoted by

$$\int_C f_i d\mu = \sum_{j=1}^{n}\left[f_i\left(x_{(j)}\right) - f_i\left(x_{(j-1)}\right)\right]\mu\left(A_{(j)}^i\right),\ i = 1,2,...,N \quad (22)$$

where $f_i\left(x_{(0)}\right) = 0$, $f_i\left(x_{(j)}\right)$ indicates that the indices have been permuted so that

$$0 \le f_i\left(x_{(1)}\right) \le f_i\left(x_{(2)}\right) \le ... \le f_i\left(x_{(n)}\right) \quad (23)$$

$$A_{(j)} = \left\{x_{(j)}, x_{(j+1)}, ..., x_{(n)}\right\} \quad (24)$$

6.2. Choquet integral regression models [2], [3], [4], [5], [6]

Let $y_1, y_2, ..., y_N$ be global evaluations of N objects and $f_1\left(x_j\right), f_2\left(x_j\right), ..., f_N\left(x_j\right),\ j = 1,2,...,n$, be their evaluations of x_j, where $f_i : X \to R_+$, $i = 1,2,...,N$.

Let μ be a fuzzy measure, $\alpha, \beta \in R$,

$$y_i = \alpha + \beta\int_C f_i dg_\mu + e_i,\quad e_i \sim N\left(0, \sigma^2\right),\quad i = 1,2,...,N \quad (25)$$

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg\min_{\alpha,\beta}\left[\sum_{i=1}^{N}\left(y_i - \alpha - \beta\int_C f_i dg_\mu\right)^2\right] \quad (26)$$

then $\hat{y}_i = \hat{\alpha} + \hat{\beta}\int f_i dg_\mu,\ i = 1,2,...,N$ is called the Choquet integral regression equation of μ, where

$$\hat{\beta} = S_{yf} / S_{ff}$$

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{\beta}\frac{1}{N}\sum_{i=1}^{N}\int f_i dg_\mu \quad (27)$$

$$S_{hy} = \frac{\sum_{i=1}^{N}\left[y_i - \frac{1}{N}\sum_{i=1}^{N} y_i\right]\left[\int f_i dg_{\mu^*} - \frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]}{N-1}$$

$$S_{hh} = \frac{\sum_{i=1}^{N}\left[\int f_i dg_{\mu^*} - \frac{1}{N}\sum_{k=1}^{N}\int f_k dg_{\mu^*}\right]^2}{N-1} \quad (28)$$

7. Experiment and result

A real data set with 59 samples from a junior high school in Taiwan including the independent variables, examination scores of four courses, and the dependent variable, the score of the Basic Competence Test of junior high school listed in Table 2 is applied to evaluate the performances of three Choquet integral regression models with P-measure, λ-measure, and R-measure based on V-support, and γ-support respectively, a ridge regression model, and a multiple linear regression model by using 5-fold cross validation method to compute the mean square error (MSE) of the dependent variable. The formulas of MSE is

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (29)$$

For any fuzzy measure, μ-measures, once the fuzzy support of the μ-measure is given, all the event measures of μ can be found, and then, the Choquet integral based on μ and the Choquet integral regression equation based on μ can also be found.

The singleton measures, V-support and γ-support of the P-measure, λ-measure, and R-measure can be obtained by using the formulas (12) and (16), respectively.

The experimental results of eight forecasting models are listed in Table I. We can find that the Choquet integral regression model with R-measure outperforms other forecasting regression models.

Table 1 MSE of regression models

Regression model			5-fold CV MSE
	measure	support	
Choquet Integral Regression model	P	V	70.4011
		γ	68.9878
	λ	V	61.0440
		γ	57.5449
	R	V	60.5317
		γ	56.2746
Ridge regression			63.1253
Multiple linear regression			69.7094

8. Conclusions and future works

When the sub-tests of a composite test are with interaction, the performance of the traditional additive scale method is poor. Non-additive fuzzy measures and fuzzy integral can be applied to improve this situation. In this study, a real data set from a junior high school including the independent variables, test scores of four courses with interaction, and the dependent variable, junior high school graduates' scores of the Basic Competence Test (BCT) are applied to evaluate the performances of the Choquet integral regression model with three well known fuzzy measures, P-measure, λ-measure, and R-measure based on two different supports, V-support, and γ-support respectively, the traditional multiple linear regression model, and the ridge regression model. Experimental result shows that the following situations:

Choquet integral regression model with R-measure based on γ-support has the best performance.

(ii) Based on the same fuzzy support, not only the γ-support but also the V-support, the Choquet integral regression model with R- measure is better than which with fuzzy measure, λ-measure and P-measure.

(iii) The Choquet integral regression model with the same measure, P-measure, λ-measure, and R-measure, respectively, the performance of which is derived from the γ-support is better than which from the V-support.

(iv) The Choquet integral regression model with λ-measure, and R-measure based on V-support and γ-support, respectively, are all better than the ridge regression and the multiple regression model.

(v) The Choquet integral regression model with P-measure is not a good model.

In future we will apply the proposed Choquet integral regression model with the better measure based on the best fuzzy support, γ-support, to develop multiple classifier system.

9. References

[1] A. E. Hoerl, R. W. Kenard, and K. F. Baldwin, "Ridge regression: Some simulation", *Communications in Statistics*, Vol. 4, No. 2, pp. 105-123, 1975.

[2] Hsiang-Chuan Liu, Chin-Chun Chen, Der-Bang Wu, and Yu-Du Jheng, "A new weighting method for detecting outliers in IPA based on Choquet integral", *IEEE International conference on Industrial Engineering and Engineering Management 2007*, December 2-5, 2007, Singapore.

[3] Hsiang-Chuan Liu, "The Choquet integral regression model based on r-complete measure", *Journal of educational research and development*, Vol. 2, No. 4, pp 87-107, 2006 (in Chinese).

[4] Hsiang-Chuan Liu, Wen-Chih Lin, and Wei-Sheng Weng, "A Choquet Integral Regression Model Based on a New Fuzzy Measure", *The 12th International conference on Fuzzy Theohery & Technology*, July 19-24, 2007, Salt Lake City, Utah, U.

[5] Hsiang-Chuan Liu, Wen-Chih Lin, Kei-Yi Chang, and Wei-Sheng Weng, "A Nonlinear Regression Model Based on Choquet Integral with ε -Measure", *2007 WSEAS International Conferences*, Venice, Italy, November 21-24, 2007.

[6] Hsiang-Chuan Liu, Yu-Chieh Tu, Chin-Chun Chen, and Wei-Sheng Weng (2008), "The Choquet integral with respect to λ-measure based on γ-support", *2008 International Conferences on Machine Learning and Cybernetics*, Kuming, China, July 12-15, 2008 (Accepted).

[7] G. Choquet, "Theory of capacities", *Annales de l'Institut Fourier*, Vol. 5, pp. 131-295, 1953.

[8] M. Sugeno, "Theory of fuzzy integrals and its applications", unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[9] Z. Wang, and G. J. Klir, *Fuzzy Measure Theory*, Plenum Press, New York, 1992.

[10] L. A. Zadeh, “Fuzzy Sets and Systems”, Vol. 1, pp. 3, 1978.

Table 2 The data set with four courses and science scores of the BCT

No.	C1	C2	C3	C4	BCT	No.	C1	C2	C3	C4	BCT
1	77	75	79	83	31	31	74	70	80	75	35
2	71	72	78	75	26	32	56	61	75	68	22
3	78	86	86	86	33	33	62	68	72	74	29
4	58	64	68	66	32	34	86	80	82	81	35
5	48	59	65	68	16	35	63	78	88	83	31
6	68	74	77	80	28	36	56	66	76	71	21
7	62	72	84	78	47	37	77	74	80	76	42
8	51	53	65	59	9	38	73	78	84	81	24
9	62	64	76	70	36	39	63	60	68	69	17
10	63	70	81	75	41	40	53	68	80	74	31
11	66	68	75	74	25	41	74	86	87	88	44
12	66	72	80	76	23	42	78	83	81	85	50
13	75	75	85	80	39	43	47	58	66	62	15
14	74	63	69	75	12	44	51	60	63	64	18
15	68	78	85	75	27	45	60	65	75	70	23
16	71	74	80	77	26	46	68	68	80	74	26
17	49	60	69	64	13	47	52	60	70	65	20
18	73	78	84	81	39	48	57	65	75	70	24
19	68	70	74	76	40	49	70	66	70	74	13
20	54	56	62	68	7	50	53	68	74	80	30
21	53	68	74	71	11	51	68	68	78	76	35
22	56	63	69	75	21	52	57	60	68	64	23
23	70	80	78	70	31	53	61	62	70	70	25
24	51	74	82	75	49	54	59	70	80	76	37
25	61	66	72	78	33	55	59	62	70	78	29
26	67	70	80	75	35	56	52	64	76	70	27
27	59	75	80	82	27	57	68	70	80	75	33
28	53	56	70	63	22	58	71	76	74	78	38
29	56	56	65	61	6	59	72	66	78	72	19
30	52	57	67	62	15						

Fuzzy c-Mean Algorithm Based on Complete Mahalanobis Distances and Separable Criterion

Hsiang-Chuan Liu
Asia University, Taiwan
lhc@asia.edu.tw

Der-Bang Wu
Taichung University, Taiwan
wudb@hotmail.com

Jeng-Ming Yih
Taichung University, Taiwan
yih@mail.ntcu.edu.tw

Shin-Wu Liu
Rutgers University, U.S.A.
fpan0366@yahoo.com.tw

Abstract

The well known fuzzy partition clustering algorithms are most based on Euclidean distance function, which can only be used to detect spherical structural clusters. GK clustering algorithm and GG clustering algorithm, were developed to detect non-spherical structural clusters, but both of them fail to consider the relationships between cluster centers in the objective function, needing additional prior information. In our previous studies, we developed two improved algorithms, FCM-M and FCM-CM based on unsupervised Mahalanobis distance without any additional prior information. And FCM-CM is better than FCM-M, since the former has the more information about the overall covariance matrix than the later. In this paper, an improved new unsupervised algorithm, "fuzzy c-mean based on complete Mahalanobis distance and separable criterion without any prior information (FCM-CMS)", is proposed. In our new algorithm, not only the local and overall covariance matrices of all clusters but also an additional separable criterion were considered. It can get more information and higher accuracy by considering the additional separable criterion than FCM-CMx. A real data set was applied to prove that the performance of the FCM-CMS algorithm is better than those of the traditional FCM algorithm and our previous FCM-M.

1. Introduction

In the 1930s, as an Indian statistician, Mahalanobis developed the distance, so called "Mahalanobis distance" which is a distance by using the inverse of the covariance matrix as the metric. Mahalanobis distance is a distance in the geometrical sense because the covariance matrices as well as its inverse are positive definite matrices [1].

As we known, the clustering plays an important role in data analysis and interpretation. It groups the data into classes or clusters so that the data objects within a cluster have high similarity in comparison to one another, but are very dissimilar to those data objects in other clusters.

Fuzzy partition clustering is a branch in cluster analyses. It is widely used in pattern recognition field. The well known one, such as, C. Bezdek's "Fuzzy C-Mean (FCM)" [2], is all based on Euclidean distance function. The fuzzy partition clustering algorithm can only be used to detect the data classes with same super spherical shapes.

Extending Euclidean distance to Mahalanobis distance, the well known fuzzy partition clustering algorithms, Gustafson-Kessel (GK) clustering algorithm [4] and Gath-Geva (GG) clustering algorithm [3] were developed to detect non- spherical structural clusters, but these two algorithms fail to consider the relationships between cluster centers in the objective function, GK algorithm must have prior information of shape volume in each data class, otherwise, it can only be considered to detect the data classes with same volume. GG algorithm must have prior probabilities of the clusters.

In our previous works [7], [8], we added a regulating factor of covariance matrix to each class in objective function, and deleted the constraint of the determinants of covariance matrices in GK Algorithm, we developed two new unsupervised algorithms, FCM-M and FCM-CM, *And FCM-CM is better than FCM-M, since the former has the more information about the overall covariance matrix than the later.*

In this paper, an improved new unsupervised algorithm, "fuzzy c-mean based on complete Mahalanobis distance and separable criterion without any prior information (FCM-CMS)", is proposed. It can get more information and higher accuracy by considering the additional separable criterion than FCM-CM.

A real data set was applied to prove that the performance of the FCM-CMS algorithm is better than those of the traditional FCM algorithm and our previous FCM-CM and FCM-M'

This paper is organized as followings: The FCM algorithm is introduced in section 2, FCM-M is introduced in section 3, FCM-CM is introduced in section 4, FCM-CMS is described in section 5. Experiment and result are described in section 6, and final section is for conclusions and future works.

978-0-7695-3305-6/08 $25.00 © 2008 IEEE
DOI 10.1109/FSKD.2008.34

2. Fuzzy c-Mean Algorithm

Fuzzy c-Mean Algorithm (FCM) is the most popular objective function based fuzzy clustering algorithm, it is first developed by Dunn [6] and improved by Bezdek [3]. The objective function used in FCM is given by Equation (1)

$$J_{FCM}^{m}(U,A,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}d_{ij}^{2}=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left\|\underline{x}_j-\underline{a}_i\right\|^2 \tag{1}$$

$\mu_{ij}\in[0,1]$ is the membership degree of data object $\underline{x}_j$ in cluster C_i and it satisfies the following constraint given by Equation (2)

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall j=1,2,\ldots,n \tag{2}$$

C is the number of clusters, m is the fuzzifier, m>1, which controls the fuzziness of the method. They are both parameters and need to be specified before running the algorithm. $d_{ij}^2=\left\|\underline{x}_j-\underline{a}_i\right\|^2$ is the square of the Euclidean distance between data object $\underline{x}_j$ to center $\underline{a}_i$. Minimizing objective function Eq. (1) with constraint Eq. (2) is a non-trivial constraint nonlinear optimization problem with continuous parameters $\underline{a}_i$ and discrete parameters μ_{ij}. So there is no obvious analytical solution. Therefore an alternating optimization scheme, alternatively optimizing one set of parameters while the other set of parameters are considered as fixed, is used here. Then the updating function for $\underline{a}_i$ and μ_{ij} is obtained as Eq. (3) ~ (4)

$$\underline{a}_i=(\sum_{j=1}^{n}\mu_{ij}^{m}\underline{x}_j)(\sum_{j=1}^{n}\mu_{ij}^{m})^{-1},\quad i=1,2,\ldots,c \tag{3}$$

$$\mu_{ij}=\left[\sum_{l=1}^{c}\left[\left[(\underline{x}_j-\underline{a}_i)'(\underline{x}_j-\underline{a}_i)\right]\left[(\underline{x}_j-\underline{a}_l)'(\underline{x}_j-\underline{a}_l)\right]^{-1}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{4}$$

.

3. FCM-M Algorithm

In our previous study [7], for improving the above two problems, we added a regulating factor of covariance matrix, $-\ln\left|+\Sigma_i^{-1}\right|$, to each class in objective function, and deleted the constraint of the determinant of covariance matrices, $|M_i|=\rho_i$, in GK Algorithm as the objective function (5). We can obtain the objective function of Fuzzy c-Mean based on adaptive Mahalanobis distance (FCM-M) as following:

$$J_{FCM-M}^{m}(U,A,\Sigma,X)=$$
$$\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left[(\underline{x}_j-\underline{a}_i)'\Sigma_i^{-1}(\underline{x}_j-\underline{a}_i)-\ln\left|\Sigma_i^{-1}\right|\right] \tag{5}$$

To minimize the objective function (5) with constraint (2) respect to parameters $\underline{a}_i$, α_j, μ_{ij}, Σ_i., we can obtained the updating equations as bellows

$$\underline{a}_i^{(1)}=(\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right])^{-1}(\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right]\underline{x}_j),i=1,2,\ldots,c \tag{6}$$

$$\Sigma_i=\frac{\sum_{j=1}^{n}\mu_{ij}^{m}(\underline{x}_j-\underline{a}_i)(\underline{x}_j-\underline{a}_i)'}{\sum_{j=1}^{n}\mu_{ij}^{m}} \tag{7}$$

$$\mu_{ij}=\left[\sum_{s=1}^{c}\left[\frac{(\underline{x}_j-\underline{a}_i)'\Sigma_i^{-1}(\underline{x}_j-\underline{a}_i)-\ln\left\|\Sigma_i^{-1}\right\|}{(\underline{x}_j-\underline{a}_s)'\Sigma_s^{-1}(\underline{x}_j-\underline{a}_s)-\ln\left\|\Sigma_s^{-1}\right\|}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{8}$$

4. FCM-CM Algorithm

In our previous study [8, 9], for improving our proposed **FCM-M,** we added a regulating factor about the overall covariance matrix in objective function (5), and we can get the following new objective function

$$J_{FCM-CMS}^{m}(U,A,\Sigma,X)$$
$$=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left[(\underline{x}_j-\underline{a}_i)'\Sigma_i^{-1}(\underline{x}_j-\underline{a}_i)-\ln\left|\Sigma_i^{-1}\right|-(\underline{a}_i-\underline{a}_t)'\Sigma_t^{-1}(\underline{a}_i-\underline{a}_t)\right] \tag{9}$$

we can obtained the updating equations as bellows

$$\underline{a}_i^{(1)}=(\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right])^{-1}(\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right]\underline{x}_j),i=1,2,\ldots,c \tag{10}$$

$$\Sigma_i=\frac{\sum_{j=1}^{n}\mu_{ij}^{m}(\underline{x}_j-\underline{a}_i)(\underline{x}_j-\underline{a}_i)'}{\sum_{j=1}^{n}\mu_{ij}^{m}} \tag{11}$$

$$\mu_{ij}=\left[\sum_{s=1}^{c}\left[\frac{(\underline{x}_j-\underline{a}_i)'\Sigma_i^{-1}(\underline{x}_j-\underline{a}_i)-\ln\left\|\Sigma_i^{-1}\right\|}{(\underline{x}_j-\underline{a}_s)'\Sigma_s^{-1}(\underline{x}_j-\underline{a}_s)-\ln\left\|\Sigma_s^{-1}\right\|}\right]^{\frac{1}{m-1}}\right]^{-1} \tag{12}$$

5. FCM-CMS Algorithm

Now, for improving the algorithm FCM-CM, we added a separable factor in objective function (9), and we can get the following new objective function of the new algorithm FCM-CMS,

$$J^{m}_{FCM-CMS}(U,A,\Sigma,X)$$
$$=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left[\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)\right]$$
$$-\frac{1}{c(c-1)}\sum_{i=1}^{c}\sum_{l=1}^{c}v_{il}^{m}\left(\underline{a}_i-\underline{a}_l\right)'\left(\underline{a}_i-\underline{a}_l\right) \quad (13)$$

Constraints: membership,

$$\sum_{i=1}^{c}\mu_{ij}=1,\ \forall\ j=1,2,\ldots,n, \quad (14)$$

where

$$v_{il}^{(k)}=\frac{w_{il}^{k-1}-\min\limits_{1\le r,s\le c}w_{rs}^{k-1}}{\max\limits_{1\le r,s\le c}w_{rs}^{k-1}-\min\limits_{1\le r,s\le c}w_{rs}^{k-1}},$$

and

$$w_{rs}^{k-1}=\left\|\underline{a}-\underline{a}_r^{(k-1)}\right\|^2+\left\|\underline{a}-\underline{a}_s^{(k-1)}\right\|^2$$

Using the Lagrange multiplier method, to minimize the objective function (13) with constraint (14) respect to parameters $\underline{a}_i$, μ_{ij}, Σ_i, we can obtain the updating functions for $\underline{a}_i$, μ_{ij}, and Σ_i are obtained as (15), (16).

$$\underline{a}_i=F^{-1}\left[\sum_{j=1}^{n}\mu_{ij}^{m}\left(\Sigma_i^{-1}\underline{x}_j-\Sigma_t^{-1}\underline{a}_t\right)-\frac{1}{c(c-1)}\sum_{l=1}^{c}v_{il}^{m}\underline{a}_l\right] \quad (15)$$

where $F=\left[\sum_{j=1}^{n}\mu_{ij}^{m}\left(\Sigma_i^{-1}-\Sigma_t^{-1}\right)-\frac{1}{c(c-1)}\sum_{l=1}^{c}v_{il}^{m}I\right]$

$$\mu_{ij}=\left[\sum_{s=1}^{c}\left[\frac{\left(\underline{x}_j-\underline{a}_i\right)'\Sigma_i^{-1}\left(\underline{x}_j-\underline{a}_i\right)-\ln\left|\Sigma_i^{-1}\right|-\left(\underline{a}_i-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_i-\underline{a}_t\right)}{\left(\underline{x}_j-\underline{a}_s\right)'\Sigma_s^{-1}\left(\underline{x}_j-\underline{a}_s\right)-\ln\left|\Sigma_s^{-1}\right|-\left(\underline{a}_s-\underline{a}_t\right)'\Sigma_t^{-1}\left(\underline{a}_s-\underline{a}_t\right)}\right]^{\frac{1}{m-1}}\right]^{-1} \quad (16)$$

where

$$\underline{a}_t=\frac{1}{n}\sum_{j=1}^{n}\underline{x}_j,\ \Sigma_t=\frac{1}{n}\sum_{j=1}^{n}\left(\underline{x}_j-\underline{a}_t\right)\left(\underline{x}_j-\underline{a}_t\right)'$$

$$\Sigma_i=\frac{\sum_{j=1}^{n}\mu_{ij}^{m}\left(\underline{x}_j-\underline{a}_i\right)\left(\underline{x}_j-\underline{a}_i\right)'}{\sum_{j=1}^{n}\mu_{ij}^{m}}$$

$$\Sigma_i=\sum_{s=1}^{p}\lambda_{si}\Gamma_{si}\Gamma'_{si},\ i=1,2,\ldots,c$$

$$\Rightarrow\ {}_{+}\Sigma_i^{-1}=\sum_{s=1}^{p}\left(\lambda_{si}^{-1}\right)^{+}\Gamma_{si}\Gamma'_{si},$$

$$\left(\lambda_{si}^{-1}\right)^{+}=\begin{cases}\lambda_{si}^{-1} & if\ \lambda_{si}>0\\ 0 & if\ \lambda_{si}\le 0\end{cases}$$

$$\left|{}_{+}\Sigma_i^{-1}\right|=\prod_{1\le s\le p,\lambda_{si}>0}\lambda_{si}^{-1}$$

The new fuzzy clustering algorithm (FCM-CMS) can be summarized in the following steps:

Step 1: Determining the number of cluster; c and m-value (let m=2), given converging error, $\varepsilon>0$ (such as $\varepsilon=0.001$).

Method 1: choose the result membership matrix of FCM algorithm as the initial one.

Method 2: let $\underline{a}_i^{(0)}, i=1,2,\ldots,c$ be the result centers of k-mean algorithm, and $d_{ij}=\left\|\underline{x}_j-\underline{a}_i^{(0)}\right\|$ be distances between data object x_j to center $\underline{a}_i^{(0)}$.

$$d_M=\max_{1\le i\le c,1\le j\le n}d_{ij}, \quad (17)$$

$$\mu_{ij}^{(0)}=\frac{\left(d_M-d_{ij}\right)}{\sum_{s=1}^{c}\left(d_M-d_{sj}\right)},\ i=1,2,\ldots,c,\ j=1,2,\ldots,n \quad (18)$$

$$\underline{a}_i^{(0)}=\left(\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right]\underline{x}_j\right)\left(\sum_{j=1}^{n}\left[\mu_{ij}^{(0)}\right]\right)^{-1},\quad i=1,2,\ldots,c$$

$$\Rightarrow\ \Sigma_i^{(0)}=\frac{\sum_{j=1}^{n}\left(\mu_{ij}^{(0)}\right)^{m}\left(\underline{x}_j-\underline{a}_i^{(0)}\right)\left(\underline{x}_j-\underline{a}_i^{(0)}\right)'}{\sum_{j=1}^{n}\left(\mu_{ij}^{(0)}\right)^{m}} \quad (19)$$

$$U^{(0)}=\begin{bmatrix}\mu_{11}^{(0)}\ \mu_{12}^{(0)}\ldots\mu_{1n}^{(0)}\\ \mu_{21}^{(0)}\ \mu_{22}^{(0)}\ldots\mu_{2n}^{(0)}\\ \ldots\quad\ldots\quad\ldots\\ \mu_{c1}^{(0)}\ \mu_{c2}^{(0)}\ldots\mu_{cn}^{(0)}\end{bmatrix} \quad (20)$$

$$A^{(0)}=\left[\underline{a}_1^{(0)}\ \underline{a}_2^{(0)}\ldots\underline{a}_c^{(0)}\right] \quad (21)$$

$$\Sigma^{(0)}=\left[\Sigma_1^{(0)},\Sigma_2^{(0)},\ldots,\Sigma_c^{(0)}\right] \quad (22)$$

$$w_{rs}^{(0)}=\left\|\underline{a}-\underline{a}_r^{(0)}\right\|^2+\left\|\underline{a}-\underline{a}_s^{(0)}\right\|^2$$

Step 2: Find

$$v_{il}^{(k)} = \frac{w_{il}^{k-1} - \min\limits_{1 \le r,s \le c} w_{rs}^{k-1}}{\max\limits_{1 \le r,s \le c} w_{rs}^{k-1} - \min\limits_{1 \le r,s \le c} w_{rs}^{k-1}},$$

where

$$w_{rs}^{k-1} = \left\| \underline{a} - \underline{a}_r^{(k-1)} \right\|^2 + \left\| \underline{a} - \underline{a}_s^{(k-1)} \right\|^2$$

$$\mu_{ij}^{(k)} = \left[\sum_{s=1}^{c} \left[\frac{\left(\underline{x}_j - \underline{a}_i^{(k)}\right)' \left[\Sigma_i^{(k)}\right]^{-1} \left(\underline{x}_j - \underline{a}_i^{(k)}\right) - \ln\left|\Sigma_i^{(k)}\right|^{-1} - \left(\underline{a}_i^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_i^{(k)} - \underline{a}_t\right)}{\left(\underline{x}_j - \underline{a}_s^{(k)}\right)' \left[\Sigma_s^{(k)}\right]^{-1} \left(\underline{x}_j - \underline{a}_s^{(k)}\right) - \ln\left|\Sigma_s^{(k)}\right|^{-1} - \left(\underline{a}_s^{(k)} - \underline{a}_t\right)' \Sigma_t^{-1} \left(\underline{a}_s^{(k)} - \underline{a}_t\right)} \right]^{\frac{1}{m-1}} \right]^{-1} \tag{23}$$

$$a_i^{(k)} = \left[F^{(k)}\right]^{-1} \left[\sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)}\right]^m \left[\left[\Sigma_i^{(k-1)}\right]^{-1} x_j - \Sigma_t^{-1} \underline{a}_t \right] - \frac{1}{c(c-1)} \sum_{l=1}^{c} \left[v_{il}^{(k)}\right]^m \underline{a}_l^{(k-1)} \right]$$

$$,i = 1,2,\ldots,c \tag{24}$$

$$F^{(k)} = \sum_{j=1}^{n} \left[\mu_{ij}^{(k-1)}\right]^m \left[\left[\Sigma_i^{(k-1)}\right]^{-1} - \Sigma_t^{-1} \right] - \frac{1}{c(c-1)} \sum_{l=1}^{c} \left[v_{il}^{(k)}\right]^m I \tag{25}$$

$$\Rightarrow \Sigma_i^{(k)} = \frac{\sum\limits_{j=1}^{n} \mu_{ij}^m \left(\underline{x}_j - \underline{a}_i^{(k)}\right)\left(\underline{x}_j - \underline{a}_i^{(k)}\right)'}{\sum\limits_{j=1}^{n} \mu_{ij}^m} \tag{26}$$

where

$$\begin{aligned} \Sigma_i^{(k)} &= \sum_{s=1}^{p} \lambda_{si}^{(k)} \Gamma_{si}^{(k)} \left(\Gamma_{si}^{(k)}\right)', \\ \left[\lambda_{si}^{-1}\right]^{(k)} &= \begin{cases} \left[\lambda_{si}^{(k)}\right]^{-1} & if \ \lambda_{si}^{(k)} > 0 \\ 0 & if \ \lambda_{si}^{(k)} = 0 \end{cases} \\ \left[\Sigma_i^{(k)}\right]^{-1} &= \sum_{s=1}^{p} \left[\lambda_{si}^{(-1)}\right]^{(k)} \Gamma_{si}^{(k)} \left(\Gamma_{si}^{(k)}\right)' \\ \left|\Sigma_i^{(k)}\right|^{-1} &= \prod_{1 \le s \le p, \lambda_{si}^{(k)} > 0} \left[\lambda_{si}^{(k)}\right]^{-1} \end{aligned} \tag{27}$$

Step 3: Increment k until

$$\max_{1 \le i \le c} \left\| \underline{a}_i^{(k)} - \underline{a}_i^{(k-1)} \right\| < \varepsilon. \tag{28}$$

6. Numerical Example

A real data set of students with sample size 146 from elementary schools was selected. The main factors of the data were calculated by using factor analysis. According to the main factors, the samples were assigned to 4 clusters based on the clustering analysis. The results were shown in Table 1.

Table 1. The characteristics of 4 clusters

Cluster	samples size	Concepts	average distance of the points from center of cluster
1	36	Partition	-.14984
2	89	Unit	.21161
3	16	Fraction	-.30416
4	5	Unknown unit	-.74490

Each 15 sample points were randomly drawn from Cluster 1, cluster 2, and cluster 3, respectively, and 5 from cluster 4.

The classification accuracies of testing samples were shown in Table 2.

Table 2. Classification accuracies of testing samples.

Algorithms	Accuracies (%)
FCM	36
FCM-M	38
FCM-CM	30
FCM-CMS	44

From the data of Table 2, we found that using the Fuzzy Clustering Algorithm of FCM-CMS could obtain the best results, even better than that of our previous research [8].

7. Conclusions and future works

The well known fuzzy partition clustering algorithms are most based on Euclidean distance function, which can only be used to detect spherical structural clusters. GK clustering algorithm and GG clustering algorithm, were developed to detect non-spherical structural clusters, but both of them needed additional prior information .in their objection functions. In our previous studies, we proposed two improved algorithms, FCM-M and FCM-CM based on unsupervised Mahalanobis distance without any additional prior information. And FCM-CM is better than FCM-M. In this paper, we proposed a further improved new unsupervised algorithm, "fuzzy c-mean based on complete Mahalanobis distance and an additional separable criterion without any prior information (FCM-CMS)". This new algorithm, not only the local and overall covariance matrices of all clusters but also an additional separable criterion were considered. It can get more information and higher accuracy by considering the additional separable criterion than FCM-CMx. A real data set was applied to prove that the performance of the FCM-CMS algorithm is better than those of the traditional FCM algorithm and our previous FCM-M and FCM-CM

In future, we will consider improve the initial value problem by using the swarm algorithm.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 96-2413-H-468-001).

References

[1] S. D. Gupta, "The evolution of the D2-statistics of Mahalanobis", *Indian J. Pure Appl. Math.*,Vol. 26 no.6, PP. 485-50, 1995.

[2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms,* Kluwer Academic Publishers: Norwell, MA, 1981.

[3] I. Gath, and A. B. Geva, "Unsupervised optimal fuzzy clustering,"*, IEEE Trans. Pattern Anal. Machine Intell. 11,* 773, 1989.

[4] D. E. Gustafson and W. C. Kessel, "Fuzzy Clustering with a Fuzzy Covariance Matrix," *Proc. IEEE Conf. Decision Contr*. San Diego, CA, 761, 1979.

[5] Dunn J C. "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters". *Journal of Cybernetics*, 3(1): 32-57 1973

[6] Fisher, R.A., "The use of multiple measurements in taxonomic problems". *Annals of Eugenics 7*, 179-188, 1936.

[7] H. C. Liu, J. M. Yih, and S. W. Liu, "Fuzzy C-mean Algorithm Based on Mahalanobis Distances and Better Initial Values," *12th International Conference on Fuzzy Theory & Technology*, JCIS, PP. 1398-1404. Salt Lake City, Utah, U.S.A., Jul., 2007.

[8] H. C. Liu, J. M. Yih, D. B. Wu, and S. W. Liu, "Fuzzy C-mean Algorithm Based on Complete Mahalanobis Distances," *International Conference in the field of Machine Learning and Cybernetics*, Kunming, China., Jul. 13-16, 2008. (Accepted).

[9] H. C. Liu, "Fuzzy Partition Clustering Algorithms Based on Alternative Mahalanobis Distances," *Journal of Educational Measurement and Statistics*, Taichung, Taiwan, Taichung University, June., 2008. (Accepted).

A Novel Classification Algorithm of Thermostable Proteins by Using Hurst Exponent and SVM Classifier

Hsiang-Chuan Liu[1], Horng-jinh Chang[2], Kuei-Jen Lee[1], Jiunn-I Shieh[3]
Wen-Chun Huang[4], Shin-Ming Huang[1]
[1]*Department of Bioinformatics, Asia University, Taiwan*
[2]*Department o f Psychology, Asia University, Taiwan*
[3]*Department o f Information Science and Applications, Asia University, Taiwan*
[4]*Graduate Institute of Educational Measurement and Statistics, Taichung University, Taiwan*
lhc@asia.edu.tw, chj@asia.edu.tw, kjlee@asia.edu.tw, jishieh@yahoo.com.tw
huang.hwc@gmail.com, majastia@gmail.com

Abstract

In search of good classification algorithm of thermostable proteins is an important issue. In this paper, a novel classification algorithm of thermostable proteins by using Hurst exponent and SVM classifier is proposed. This method not used before is the first one integrating the physics-chemistry properties, fractal property and support vector machine (SVM) classifier. For evaluating the performance of this new algorithm, a real data experiment by using 5-fold and Leave-one-out Cross-Validation accuracy is conducted. Experimental result shows that this new classification algorithm is feasible and useful.

1. Introduction

In search of good classification algorithm of thermostable proteins is an important issue. In this paper, a novel classification algorithm of thermostable proteins by using Hurst exponent and SVM classifier is proposed. This method not used before is the first one integrating the physics-chemistry properties, fractal property and support vector machine (SVM) classifier.

First step, a thermostable proteins data set with two classes was downloaded from the Protein Data Bank (PDB), http://www.rcsb.org.

Second step, replacing four feature scores with each residue of amino acid in sequence of the thermostable proteins by using the four feature scaling estimators, we can obtained four non-symbolic sequences of the thermostable proteins.

Third step, computing the Hurst exponents of each non-symbolic sequences of the thermostable proteins, we can obtained four features of Hurst exponents in each sequences of the thermostable protein.

Last step, the well known and appealing classifier, Support Vector Machine (SVM), is used to discriminate the correct class of the 40 thermostable proteins with four features of Hurst exponents For evaluating the performance of this new algorithm, the above thermostable proteins data experiment by using 5-fold and Leave-one-out Cross-Validation accuracy is conducted.

This paper is organized as followings: four feature scoring estimators are introduced in section 2. Hurst exponent is introduced in section 3, support vector machine classifier is introduced in section 4, experiment and result are described in section 7 and final section is for conclusions and future works.

2. Four feature scaling estimators

2.1. Solvent accessible surface area (ASA)

Residues classified as buried or exposed are conventionally described by a geometric parameter calculated using the solvent-accessible surface area (ASA), which is generated by rolling a spherical probe with a radius of 1.4 Å over the surface of a protein. The ASA of a protein was obtained using POPS [1], [2] on the web side (mathbio.nimr.mrc.ac.uk/~ffranca/ POPS/), selecting output residue areas (POPS_R).

978-0-7695-3304-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICNC.2008.852

Both the polar (hydrophilic) and apolar (hydrophobic) surface areas can be obtained from the output residue areas, which were then changed to the percentage of apolar area for each residue in a protein.

2.2. Exposed/ Buried

The solvent accessibility percentages of the residues were obtained using the ASAView [3] data base (www.netasa.org/asaview/). Residues were classified to be buried in a protein core as the values between 0-50%, and those were considered to be exposed to solvent when the percentage exceeded 50%.

2.3. Electrostatic interactions

The number of ion pairs (electrostatic interactions) was calculated according to the following criterion [4]: two oppositely charged residues were considered an ion pair if the distance between the oppositely charged atoms of these residues was less than 6 Å. Asp, Glu, Arg, Lys and His residues were used to calculate the ion pairs.

2.4. Contact energies

A 20×20 matrix of effective contact energies, the interaction energies between all amino acids pairs, was developed by Miyazawa and Jernigan [5], [6], which was also called MJ matrix. The MJ effective energy (eij), which is the element of MJ matrix, was derived from all the possible interaction energies, including hydrophobic and solvation energies. Furthermore, the hydrophobic interaction is the dominant contribution to the MJ effective energy. The eij can be presented as the following equation

$$e_{ij} = e'_{ij} + \frac{e_{ii} + e_{jj}}{2} \tag{1}$$

The e'ij is the mixing term, which is the free energy change upon the mixing of residues of type i and residues of type j when the contacts in self-pairs i-i and j-j are separated to form i-j pairs. The eii or ejj is the free energy change after the desolvation of residue i or of residue j to form the self-pairs i-i or j-j. The values of eii or ejj should have high correlation with the hydrophobicity of residue type i or residue type j [5], [6].

The average contact energy of each type of amino acid, ei, was used in this work, and it is defined as: [5], [6].

$$e_i = \frac{\sum_{j=1}^{20} e_{ij} N_{ij}}{N_{ir}} \tag{2}$$

where

$$N_{ij} = \sum_p n_{ij;p} \tag{3}$$

and

$$n_{ir} = \sum_{j \neq 0} n_{ij} \tag{4}$$

The supscript p denotes the total number of contacts in all proteins, nij is the total number of contacts between i and j types of amino acid residues, and nir is the total number contacts made by residue type i.

3. Hurst exponent

The Hurst exponent occurs in several areas of applied mathematics, including fractals and chaos theory, long memory processes and spectral analysis [7], [8]. Hurst exponent estimation has been applied in areas ranging from biophysics to computer networking. Estimation of the Hurst exponent was originally developed in hydrology. However, the modern techniques for estimating the Hurst exponent comes from fractal mathematics.

Estimating the Hurst exponent for a data set provides a measure of whether the data is a pure random walk or has underlying trends. Another way to state this is that a random process with an underlying trend has some degree of autocorrelation. Furthermore, when the autocorrelation has a very long (or mathematically infinite) decay this kind of Gaussian process is sometimes referred to as a long memory process.

The Hurst exponent (H) is a statistical measure used to classify time series. H=0.5 indicates a random series while H>0.5 indicates a trend reinforcing series. The larger the H value is, the stronger the trend. Experiments with backpropagation Neural Networks show that series with large Hurst exponent can be predicted more accurately than those with H value close to 0.50. Thus the Hurst exponent provides a measure for predictability.

Three methods were used most often for the estimation of the Hurst exponent: the R/S method, the roughness–length (R–L) method and a variogram. The R/S method (Hurst et al., 1965) [9] is commonly perceived as the most suitable for the time series analysis on the stock market or an optimal volume of water reservoirs, because it presents the relationship between irregular (singular) rescaled ranges, signal value and their local statistical properties relative to the scale factor. In this study R/S method is used. R/S method [10] is based on empirical observations by Hurst and estimates H are based on the R/S statistic. It indicates (asymptotically) second-order self-similarity. H is roughly estimated through the slope of the linear line in a log-log plot, depicting the R/S statistics over the number of points of the aggregated series. That is, given a time sequence of observations, w_t define the series

$$W(t,\tau)=\sum_{u=1}^{t}\left(w_u-\overline{w}_\tau\right), 1\le t\le\tau \quad (5)$$

where

$$\overline{w}_\tau=\frac{1}{\tau}\sum_{t=1}^{\tau}w_t \quad (6)$$

Define

$$R(\tau)=\max_{t=1}^{\tau}W(t,\tau)-\min_{t=1}^{\tau}W(t,\tau) \quad (7)$$

and

$$S(\tau)=\sqrt{\left(\frac{1}{\tau}\sum_{t=1}^{\tau}\left(w_t-\overline{w}_\tau\right)^2\right)} \quad (8)$$

In plotting $\log\frac{R(\tau)}{S(\tau)}$ against $\log\tau$, we expect to get a line whose slope determines the Hurst exponent.

4. Support vector machine (SVM) [11], [12], [13], [14]

Given the training set of instance-labeled pairs $\left(\underline{x}_i, y_i\right), i=1,2,...,N$, where

$$\underline{x}_i\in R^n, y_i\in\{1,-1\}, i=1,2,...,N \quad (9)$$

The support vector machine (SVM) algorithm (Boser, Guyon, and Vapnik 1992, Cortes and Vapnik 1995) requires

$$\begin{aligned}&\min_{\underline{w},b,\xi}\frac{1}{2}\underline{w}'\underline{w}+c\sum_{i=1}^{N}\xi_i\\ &\text{subject to } y_i\left(\underline{w}'\phi\left(\underline{x}_i\right)+b\right)\ge 1-\xi_i,\\ &\qquad \xi_i\ge 0, \qquad (10)\\ &\textit{where}\quad b,c\in R, \underline{w},\phi\left(\underline{x}_i\right)\in R^m\\ &\qquad \phi: R^n\to R^m\end{aligned}$$

For any testing point $\underline{x}_i\in R^n, y_i\in\{1,-1\}$, we can make an assignment according to the following formula.

$$\begin{aligned}f(\underline{x}_i)&=\operatorname{sign}\left[\underline{w}'\varphi\left(\underline{x}_i\right)+b-\left(1-\xi_i\right)\right]\\ &=\begin{cases}+, & \text{if } y_i=+1\\ -, & \text{if } y_i=-1\end{cases}\end{aligned} \quad (11)$$

5. Experiment and result

A thermostable proteins data set with two classes was downloaded from the Protein Data Bank (PDB), http://www.rcsb.org. The sample included 40 instances, 20 instances are higher thermostable proteins, and the other 20 instances are lower thermostable proteins.

Replacing four feature scores called solvent accessible surface area, exposed/ buried, electrostatic interactions, and contact energies, with each residue of amino acid in sequence of the thermostable proteins by using the four feature scaling estimators, we can obtained four non-symbolic sequences of the thermostable proteins.

Computing the Hurst exponents of each non-symbolic sequences of the thermostable proteins, we can obtained four features represented as Hurst exponents respectively in each sequences of the thermostable protein. The transformed data is listed in Table 2

The above real data with four features in terms of Hurst exponents is applied to evaluate the performances of the Support Vector Machine (SVM) algorithm by using 5-fold and Leave-one-out Cross-Validation method to compute the accuracies of the response category variable.

The experimental results for Accuracies of SVM classifier are listed in Table 1. We can find that both 5-fold CV and Leave-one-out CV accuracy had the similar result, the SVM classifier based on Hurst exponents is a feasible and useful algorithm.

Table 1 Accuracies of SVM classifier

Classification algorithm	5-fold CV accuracy	Leave-one-out CV accuracy
SVM_HE	71.4286	62.5000

6. Conclusions and future works

In search of good classification algorithm of thermostable proteins is an important issue. In this paper, a novel classification algorithm of thermostable proteins combining four feature scaling estimators, Hurst exponent and SVM classifier is proposed. For evaluating the performance of this new algorithm, a thermostable proteins data set by using 5-fold and Leave-one-out Cross-Validation accuracy is conducted. Experimental result shows that this new classification algorithm is feasible and useful.

In future, we will consider look for some improving classification algorithm of thermostable proteins by using Hurst exponent and other Classifiers.

7. Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 96-2413--H-468-001).

8. References

[1] Franca Fealernali and Luigi Cavallo, “Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome”, Nucleic Acids Res 2002 Vol. 30, No 13, PP. 2950-2960, Oxford University press.

[2] Luigi Cavallo, Jens Kleinjung and Franca Fraternali, “POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level”, Nucleic Acids Reasearch, 2003, Vol. 31. No. 13. pp. 3364-3366. DOI:10.1093/nar/gkg601.

[3] S. Ahmad, M. Gromiha, H. Fawareh, “A. Sarai, ASAView: Database and tool for solvent accessibility representation in proteins”, BMC bioinformatics 5, 51-55, 2004.

[4] A. Szilagyi, P. Zavodszky, “Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey”, Structure Fold. Des. 8, 493-504, 2000.

[5] Miyazawa, S., Jernigan, R. L., “Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation”. Macromolecules, 18, 534-552, 1985.

[6] Miyazawa, S., Jernigan, R. L., “Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading”. J. Mol. Biol. 256, 623-644, 1996.

[7] T. Di Matteo, T. Aste and M. M. Dacorogna, "Long term memories of developed and emerging markets: using the scaling analysis to characterize their stage of development", Journal of Banking & Finance 29/4, 827-851, 2005.

[8] Hurst, H. E. : “Long term storage capacity of reservoirs”, Transactions of the American Society of Civil Engineers 116: 770-799, 1951

[9] Hurst, H. E., Black, R., Sinaika, Y. M., “Long term storage capacity of reservoirs” An experimental study Constable, London, 1965.

[10] Roger Kalden, Sami Ibrahim, “Searching for Self-Similarity in GPRS” PAM 83-92, 2004.

[11] Boser, B.E., Guyon, I.M. and Vapnik, V. “A training algorithm for optimal margin classifiers”. In Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, 1992. ACM.

[12] C. Cortes, and V. Vapnik, “Support-vector network”, Machine Learning Vol. 20, pp. 273-297, 1995.

[13] Vapnik, V., The Nature of Statistical Learning Theory. New York, NY. Springer-Verlay, 1995.

[14] Chang, C.-C.,and Lin, C.-J. LIBSVM; a library for support vector machine Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2004.

Table 2 Hurst exponents of four feature scaling of Thermostable Proteins

Code of Proteins	Class	ASA	Electrostatic interactions	Contact energy	Exposed/ Buried
010	0	0.3832	0.4125	0.3335	0.7335
011	1	0.4691	0.6572	0.3411	0.5636
020	0	0.4129	0.3985	0.3315	0.4772
021	1	0.5119	0.537	0.4524	0.5345
030	0	0.5079	0.4489	0.2766	0.5512
031	1	0.4224	0.4183	0.4881	0.5818
040	0	0.4463	0.3964	0.3807	0.6805

041	0	0.4936	0.5039	0.5010	0.6290
050	0	0.3577	0.4718	0.4509	0.6078
051	1	0.4751	0.5812	0.4299	0.5699
060	0	0.2847	0.5155	0.3618	0.6286
061	1	0.4314	0.6042	0.3244	0.4178
070	0	0.5432	0.5496	0.4466	0.6699
071	1	0.2550	1	0.3501	0.5807
080	0	0.3573	0.5062	0.1892	0.6273
081	1	0.3461	0.5167	0.2403	0.4284
090	0	0.5153	0.4902	0.3485	0.6008
091	1	0.3989	0.4223	0.4090	0.5215
100	0	0.4641	0.4432	0.3261	0.5817
101	1	0.4460	0.5810	0.2082	0.5506
110	0	0.3832	0.4125	0.3335	0.7335
111	1	0.4691	0.6572	0.3411	0.5636
120	0	0.4129	0.3985	0.3315	0.4772
121	1	0.5119	0.5371	0.4524	0.5345
130	0	0.5079	0.4489	0.2766	0.5512
131	1	0.4224	0.4183	0.4881	0.5818
140	0	0.4463	0.3964	0.3807	0.6805
141	1	0.4936	0.5039	0.5010	0.6290
150	0	0.3577	0.4718	0.4509	0.6078
151	1	0.4751	0.5812	0.4299	0.5699
160	0	0.2847	0.5155	0.3618	0.6286
161	1	0.4314	0.6042	0.3244	0.4178
170	0	0.5432	0.5496	0.4466	0.6699
171	1	0.2550	0.5078	0.3501	0.5807
180	0	0.3573	0.5062	0.1892	0.6273
181	1	0.3461	0.5167	0.2403	0.4284
190	0	0.5153	0.4902	0.3485	0.6008
191	1	0.3989	0.4223	0.4090	0.5215
200	0	0.4641	0.4432	0.3261	0.5817
201	1	0.4460	0.5810	0.2082	0.5506

Physiochemical Constraints in Influenza A Hemagglutinin

Pei-Chun Chang
Department of Bioinformatics
Asia University, Taichun, Taiwan, R.O.C.
pcchang@asia.edu.tw

Kuei-Jen Lee
Department of Health and Nutrition Biotechnology
Asia University, Taichun, Taiwan, R.O.C.
kjlee@asia.edu.tw

Jiunn-I Shieh
Department of Information Science and Applications
Asia University, Taichun, Taiwan, R.O.C.
jishieh@asia.edu.tw

Chung-Hung Li
Department of Bioinformatics
Asia University, Taichun, Taiwan, R.O.C.
leonchlee@hotmail.com

Jing-Doo Wang
Department of Computer Science and Information Engineering
Asia University, Taichun, Taiwan, R.O.C.
jdwang@asia.edu.tw

Hsiang-Chuan Liu
Department of Bioinformatics
Asia University, Taichun, Taiwan, R.O.C.
lhc@asia.edu.tw

Abstract

Influenza A viruses are negative-strand RNA viruses. The gene of hemagglutinin (HA) protein in the virus genome is the major molecule that determines the range of hosts. Mutation of HA gene may bring infection cross species. In this paper, we studied physicochemical constraints during the variations of HA gene. Fuzzy measure and Choquet integral were used to estimate the combining effect of different physicochemical properties for single residue in HA protein that related to infective events. With this method, an HA sequence was quantified residue by residue and produced a series of values. Finally, the Hurst exponent was adopted to infer the constraints in the series. We found that the physicochemical constraints in HA sequences mainly falling into two classes of interdependence strength during gene variation, that was distinct from the diversity of clusters in the phylogenetic analysis.

1. Introduction

Influenza A viruses are negative-strand RNA viruses that infect a wide variety of animals in the nature. The infection of human may cause significant mortality and morbidity worldwide [1]. The gene of hemagglutinin (HA) protein in the virus genome is the major molecule that determining the range of hosts. The natural reservoir of influenza virus such as avian flu may emerge in strains infectious to human by mutation of HA gene [2,3]. Owing to that, it is important to understand the variation nature of HA gene. In the past, the researches in this field mainly have been focused on the phylogenetic reconstructions [4,5]. As shown in the explosive information on HA sequences, the reconstruction of a phylogenetic tree can provide abundant evolution information, and help in understanding the drifts of influenza hosts [6]. However, the feature and tendency about physicochemical properties of gene variations for specific host are never been discussed.

Fuzzy measure theory considers a number of special classes of measurements, each of which is characterized by a special property. In the fuzzy measure theory, the conditions are precise, but the information about an element alone is insufficient to determine which special classes of measure should be used. The fuzzy measure estimates the possible interactions among the special classes of measurements [7]. Choquet integral is a tightly related concept with fuzzy measure. It assesses the integrated effect for some issue based on the concept of fuzzy measure [7,8]. The Hurst exponent (H) is a statistical measure used to classify time series [9]. For example, H=0.5 indicates a random series while H>0.5 indicates a constrained reinforcing series. The larger the H value is, the stronger the constraint. In this paper, we studied the physicochemical constraints of HA protein of Influenza A viruses regarding to serotypes H1, H3, and H5. We concerned three types of physicochemical property for each residue that have acidity, Van der waal volume, and hydrophobicity [10]. Pearson's correlation coefficient was used to quantify the dependence of physicochemical properties on infection hosts, human or avian. For each residue, there were three values of Pearson's correlation coefficient corresponding to three

978-0-7695-3304-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICNC.2008.712

types of physicochemical properties. Based on the coefficients, Sugeno λ-measure [11] was adopted to calculate the fuzzy measure. Subsequently, the Choquet integral was applied to assess the integrated effect of physicochemical properties on infection hosts for each residue. A protein sequence implies a series of integral values. Finally, we used Hurst exponent to analyze the value series for exploring the integrated physicochemical constraints in the protein sequence.

2. Methods

2.1 Sequence data collection

The sequence data of Influenza A viruses used in this research were obtained from public databases: Influenza Sequence Database (http://www.flu.lanl.gov). All HA nucleotide sequences of human and birds in this databases were downloaded on October 16, 2006. The HA sequences were extracted, of which less than 900 nucleotides were considered as partial sequences and were excluded from this study. Identically coded sequences are considered as duplicates and only the earliest isolated strain among the duplicates was used as a representative sequence in the group. In total, we had 831 H1 sequences, 3018 H3 sequences and 1376 H5 sequences for our analysis. All sequences were isolated between 1963 and 2006 from locations around the globe. The exact isolation time (calendar year), host type and location can be found in the strain names.

2.2. Residue coding

The sequence alignment processes were implemented in ClustalX 3.14 [12] regarding to H1, H3, and H5. After alignment, the sequence length regarding to H1, H3, and H5 were 565, 567, and 583 amino acids respectively. The protein residues were coded according to its values of acidity, Van der waal volume, and hydrophobicity in the situation of single amino acid [10, 13], as shown in table 1. For every physicochemical property, we had a matrix size of 831x565 for H1 group, 3018x567 for H3 group, and 1376x583 for H5 group.

Table 1. The residue codes regarding to acidity, Van der waal volume, and hydrophobicity.

Amino acid	Acidity	Van der waal volume	Hydrophobicity
Alanine	7.0	67.	0.616
Cysteine	8.4	86.	0.68
Aspartic acid	3.9	67.	0.028
Glutamic acid	4.1	109.	0.043
Phenylalanine	7.0	135.	1.
Glycine	7.0	48.	0.501
Histidine	6.0	118.	0.165
Isoleucine	7.0	124.	0.943
Lysine	10.5	135.	0.283
Leucine	7.0	124.	0.943
Methionine	7.0	124.	0.738
Asparagine	7.0	148.	0.236
Proline	7.0	90.	0.711
Glutamine	7.0	114.	0.251
Arginine	12.5	167.	0.
Serine	7.0	73.	0.359
Threonine	7.0	93.	0.45
Valine	7.0	105.	0.825
Tryptophan	10.5	163.	0.878
Tyrosine	7.0	141.	0.88

2.3. Inference of physicochemical constraints

Choquet integral is defined to integrate functions with respect to the fuzzy measure [7]. It is very useful in assessment of the effect that results from the nonlinear interactions. The definitions of fuzzy measure and Choquet integral are as follows:

Definition 1. Let N be a finite set of criteria. A discrete fuzzy measure on N is a set function $v: 2^N \to [0,1]$ which satisfies the following axioms:

(i) $v(\phi) = 0$, $v(N) = 1$ (boundary conditions)

(ii) $A \subseteq B$ implies $v(A) \le v(B)$ (monotonicity) for $A, B \in 2^N$.

For each subset of criteria $S \subseteq N$, $v(S)$ can be interpreted as the weight of the coalition S.

The Sugeno λ-measure is a special case of fuzzy measures. It has the following definition.

Definition 2. Let $N = \{X_1, X_2, \cdots, X_n\}$ be a finite set and $\lambda \in (-1, \infty)$. A Sugeno λ-measure is a function ν from 2^N to [0, 1] with properties:

(i) ν(N) = 1.

(ii) if $A, B \subseteq 2^N$ with $A \cap B = \phi$ then $v(A \cup B) = v(A) + v(B) + \lambda v(A) \bullet v(B)$.

As a convention, the value of ν at a singleton set $\{X_i\}$ is called a density and is denoted by $v\{X_i\}$. In addition, we have that λ satisfies the property

$$\lambda + 1 = \prod_{i=1}^{n} \left(1 + \lambda v(X_i)\right) \qquad (1)$$

Tahani and Keller [14] as well as Wang and Klir [15] have showed that that once the densities are known, it is possible to use the previous polynomial to obtain the values of λ uniquely.

Definition 3. Let v be a fuzzy measure on N. The discrete Choquet integral of function x: $N \to R$ with respect to v is defined by

$C_v(x) = \sum_{i=1}^{n} x_{(i)} \left[v(A_{(i)}) - v(A_{(i+1)}) \right]$, where $(\cdot)$ indicates a permutation on N such that $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$. Also $A_{(i)} = \{x_{(i)}, \cdots, x_{(n)}\}$, and $A_{(n+1)} = \phi$. For instance, if $x_1 \le x_3 \le x_2$, then rank x_1, x_2, x_3 from the smallest one to the largest one. The result is $x_{(1)} = x_1$, $x_{(2)} = x_3$, $x_{(3)} = x_2$. Finally,

$$C_v(x_1, x_2, x_3) = x_1 * [v(\{x_1, x_2, x_3\})] + (x_3 - x_1) * [v(\{x_2, x_3\})] + (x_2 - x_3) * [v(\{x_2\})] \quad (2)$$

The discrete Choquet integral takes into account the interaction by means of the fuzzy measure v. If the criteria are independent, the fuzzy measure is additive. Then, the discrete Choquet integral coincides with the weighted arithmetic mean method. That is, $C_v(x)$ = $\sum_{i=1}^{n} v(i) x_i$, $x \in R^n$. In this study, the correlation-based method proposed by Hsiang-Chuan Liu in 2006 [16,17] to construct the fuzzy measures in the discrete Choquet integral was used.

The Hurst exponent occurs in several areas of applied mathematics, including fractals and chaos theories, long memory processes and spectral analysis. Hurst exponent estimation has been applied in areas ranging from biophysics to computer networking. Estimation of the Hurst exponent was originally developed in hydrology. However, the modern techniques for estimating the Hurst exponent come from fractal mathematics.

Estimating the Hurst exponent for a data set provides a measure of whether the data is a pure random walk or has underlying trends. Another way to state this is that a random process with an underlying trend has some degree of autocorrelation. Furthermore, when the autocorrelation has a very long (or mathematically infinite) decay this kind of Gaussian process is sometimes referred to as a long memory process.

The Hurst exponent (H) is a statistical measure used to classify time series. H=0.5 indicates a random series while H>0.5 indicates a trend reinforcing series. The larger the H value is, the stronger the trend. In this paper we investigate the use of the Hurst exponent to classify series of financial data representing different periods of time. Experiments with back propagation Neural Networks show that series with large Hurst exponent can be predicted more accurately than those with H value close to 0.50. Thus the Hurst exponent provides a measure for predictability.

Three methods were used most often for the estimation of the Hurst exponent: the R/S method, the roughness–length (R–L) method and variogram. The R/S method [18] is commonly perceived as the most suitable for the time series analysis on the stock market or an optimal volume of water reservoirs, because it presents the relationship between irregular (singular) rescaled ranges, signal value and their local statistical properties relative to the scale factor. In this study R/S method is used. R/S method [19] is based on empirical observations by Hurst and estimates H are based on the R/S statistic. It indicates (asymptotically) second-order self-similarity. H is roughly estimated through the slope of the linear line in a log-log plot, depicting the R/S statistics over the number of points of the aggregated series. That is, given a time sequence of observations w_t, define the series $W(t,\tau) = \sum_{u=1}^{t} (w_u - \overline{w}_\tau)$, where $\overline{w}_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} w_t$.

Define $R(\tau) = \max_{t=1}^{\tau} W(t,\tau) - \min_{t=1}^{\tau} W(t,\tau)$ and $S(\tau) = \sqrt{\left(\frac{1}{\tau} \sum_{t=1}^{\tau} (w_t - \overline{w}_\tau)^2 \right)}$. In plotting $\log \frac{R(\tau)}{S(\tau)}$ against $\log \tau$, we expect to get a line whose slope determines the Hurst exponent.

There is a 7-step to make hurst exponent analyze:

Step 1. With quantizing three properties each amino acid of each protein sequence, we have three time series for each protein sequence.

Step 2. For each property, normalize the data for each position which the same position of aligned protein sequences for affecting human and birds. That is, label elements in the sample by *l* and treat each position in aligned protein sequence as a random variable. Assume the size of the sample is k. For the element l, let *i*-th position of aligned protein sequences for property *m* be a random variable $X_i^{l,m}$ where $1 \leqq l \leqq k$, $1 \leqq m \leqq 3$, and *n* is the length of aligned protein sequences. If $\max_l \{X_i^{l,m}\} - \min_l \{X_i^{l,m}\} \neq 0$, then $Z_i^{l,m} = \frac{X_i^{l,m} - \min_l \{X_i^{l,m}\}}{\max_l \{X_i^{l,m}\} - \min_l \{X_i^{l,m}\}}$. Otherwise, set $Z_i^{l,m} = 0$.

Step 3. Let Y^l be a random variable which is 1 if affecting the human and 0 otherwise for the element *l*. Let

$X_i^m = (X_i^{1,m}, X_i^{2,m}, \cdots, X_i^{k,m})'$ and $Y = (Y^1, Y^2, \cdots, Y^k)'$. For each *m*, compute corr(X_i^m , Y) where "corr" is the Pearson correlation coefficient. For each *m*, define the weight w_i^m to be $\frac{1+corr(X_i^m, Y)}{2}$ for each *i*. That is, $v(\{X_i^m\}) = w_i^m$ for $1 \leqq m \leqq 3$ and $1 \leqq i \leqq n$.

Step 4. For using Sugeno λ-measure, solve (1) for λ. Then, for each *i* compute $v(\{X_i^1, X_i^2\})$, $v(\{X_i^1, X_i^3\})$, $v(\{X_i^2, X_i^3\})$ by Sugeno λ-measure. Note that $v(\{X_i^1, X_i^2, X_i^3\}) = 1$.

Step 5. Combined the three properties to be one, compute the Choquet integral for each position by equation (2). Then we get one time series for each aligned protein sequence.

Step 6. Calculate Hurst exponent for each aligned protein sequence.

Step 7. Analyze the results.

The above steps were calculated using Matlab package, except for Hurst exponent was obtained from the website: http://www.mathworks.com/matlabcentral/.

3. Results

We calculated the Hurst exponent regarding to H1, H3, and H5 to infer the physicochemical interdependency among the residues in the HA protein. The serotype H1 are shown in Fig.1, there are 2 clusters in the frequency distributions of Hurst exponents for human strains and avian strains. The Hurst exponent is nearby 1 for one cluster, and nearby 0.5 for another cluster. That mean some variations are constrained strongly, and some variations are random-like. The tendency of H3 is shown in Fig.2 and similar to H1, but the Hurst exponents in the two clusters are closer and away from 1 and 0.5. The results about H5 are shown in Fig.3, the distribution pattern is different from H1 and H3 for avian strains. There are three clusters in the frequency distribution.

The phylogenetic analysis is based on the mutation frequency between residues regarding homologous proteins. The evolution of quantitative property during the process of residue changes is ambiguous. In this study, we proposed a method based on the quantitative properties of residues regarding to infection issue of Influenza A viruses to estimate the constrain strength in the HA proteins. The distribution of constrain strength are distinct from the diversity of clusters in the phylogenetic analysis.

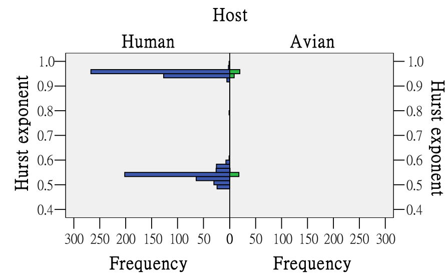

Figure 1. The frequency distribution of H1 Hurst exponents for human strains and avian strains.

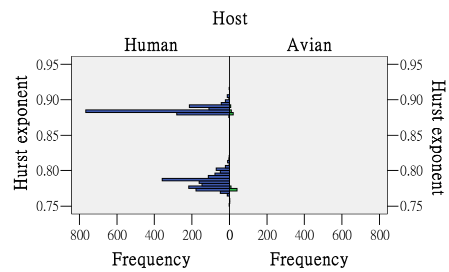

Figure 2. The frequency distribution of H3 Hurst exponents for human strains and avian strains.

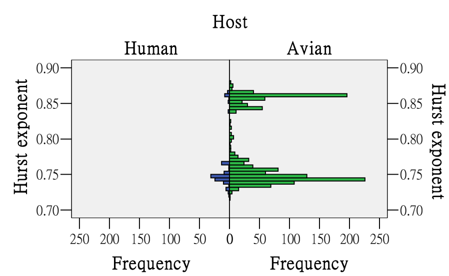

Figure 3. The frequency distribution of H5 Hurst exponents for human strains and avian strains.

4. Discussion

The gene of HA protein in the virus genome is the major molecule that determining the range of hosts. Basically, the infection process is physicochemical interaction between receptor of host and HA protein. For the sake of successful infection, the gene variations must follow certain rules under physicochemical base. Higher value of Hurst exponent implies more constraints or intra-structure in the sequence properties. As to that, the gene variations are apt to destroy the intra-structure with high value of Hurst exponent. The variation tolerance is different for the same serotype of HA corresponding to the different clusters of Hurst exponents.

The constraints in HA sequences mainly fall into two classes of Hurst strength during gene variations. That imply the variation tolerance of HA gene is diverse in the same serotype of HA.

5. References

[1] P. Palese, "Influenza: old and new threats", *Nat. Med.,* Vol. 10, 2004, pp. s82–s87.

[2] R.G. Webster, W.J. Bean, O.T. Gorman, and T.M. Chambers, Y. Kawaoka, "Evolution and ecology of influenza A viruses", *Microbiol. Rev.,* Vol. 56, 1992, pp. 152–179.

[3] N.J. Cox and K. Subbarao, "Global epidemiology of influenza: Past and present", *Annu. Rev. Med.,* Vol. 51, 2000, pp. 407–421.

[4] W.M. Fitch, R.M. Bush, C.A. Bender, and N.J. Cox, "Long term trends in the evolution of H(3) HA1 human influenza type A", *Proc. Nat. Acad. Sci.,* Vol. 94, 1997, pp. 7712–7718.

[5] R.M. Bush, W.M. Fitch, C.A. Bender, and N.J. Cox, "Positive selection on the H3 hemagglutiningene of human influenza virus A",. *Mol. Biol. Evol.,* Vol. 16, 1999, pp. 1457–1465.

[6] R.M. Bush, C.A. Bender, K. Subbarao, N.J. Cox, and W.M. Fitch, "Predicting the evolution of human influenza A", *Science,* Vol. 286, 1999, pp. 1921–1925.

[7] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measure and the Choquet integral as an integral with respect to a fuzzy measure", *Fuzzy Sets and Systems,* Vol 29, 1989, pp. 201–227.

[8] T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar, *Aggregation operators: New trends and applications*, Physica-Verlag, Springer, 2002.

[9] T. Di Matteo, T. Aste, and M.M. Dacorogna,."Long term memories of developed and emerging markets: using the scaling analysis to characterize their stage of development", *Journal of Banking & Finance,* Vol 29, 2005, pp. 827-851.

[10] D. Whitford, *Proteins: structure and function.* John Wiley & Sons Ltd., 2005.

[11] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measure and the Choquet integral as an integral with respect to a fuzzy measure", *Fuzzy Sets and Systems,* Vol 29, 1989, pp. 201–227.

[12] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins, "The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools", *Nucl. Acid Res.* Vol 24, 1997, pp. 4876–4882.

[13] S.D. Black and D.R. Mould, "Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Post- or Cotranslational Modifications", *Anal. Biochem.,* Vol 193, 1991, pp. 72–82.

[14] H. Tahani and J. Keller, "Information Fusion in Computer Vision Using the Fuzzy Integral", *IEEE Transactions on Systems, Man and Cybernetic,* Vol 20, 1990, pp. 733-741.

[15] Z. Wang and G.J. Klir, *Fuzzy Measure Theory*, Plenum Press, New York, 1991.

[16] Hsiang-Chuan Liu, Chin-Chun Chen, Der-Bang Wu, and Yu-Du Jheng, "A new weighting method for detecting outliers in IPA based on Choquet integral", *IEEE International conference on Industrial Engineering and Engineering Management 2007,* Singapore, Dec. 2~5 2007.

[17] Hsiang-Chuan Liu, "The Choquet integral regression model based on r-complete measure",. *Journal of educational research and development,* Vol 2, 2006, pp. 87-107.

[18] H.E. Hurst, R. Black, Y.M. Sinaika, *Long-Term Storage in Reservoirs: An experimental Study,* Constable, London, 1965.

[19] R. Kalden and S. Ibrahim, "Searching for Self-Similarity in GPRS", *PAM,* 2004, pp. 83-92.

Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009

CHOQUET INTEGRAL REGRESSION MODEL BASED ON HIGH-ORDER L-MEASURE

HSIANG-CHUAN LIU[1], WEI-SUNG CHEN[2], YU-CHIEH TU[3], YEN-KUEI YU[3]

[1]Department of Bioinformatics, Asia University, Wufeng, Taiwan
[2]Department of Computer Science and Information Engineering, Asia University, Wufeng, Taiwan
[3]Graduate Institute of Educational Measurement and Statistics, Taichung University, Taichung, Taiwan
E-MAIL: lhc@asia.edu.tw, james@asia.edu.tw, tuyuchieh@yahoo.com.tw, haikuei@gmail.com

Abstract:

The well known fuzzy measures, λ-measure and P-measure, have only one formulaic solution, the former is not a closed form, and the later is not sensitive. An improved multivalent fuzzy measure with infinitely many solutions of closed form, called L-measure, is proposed by our previous work. In this paper, expend the L-measure for being more choice, and get an improved fuzzy measures, called "hth-order L-measure", denoted as L^h-measure , and a new Choquet integral regression model based on this L^h-measure is also proposed. For evaluating the proposed regression models with different fuzzy measures, a real data experiment by using a 5-fold cross-validation mean square error (MSE) is conducted. The performances of Choquet integral regression models with fuzzy measure based on λ-measure, P-measure, L-measure and L^h-measure, respectively, a ridge regression model, and a multiple linear regression model are compared. Experimental result shows that the Choquet integral regression models with L^h-measure based on γ-support outperforms others forecasting models.

Keywords:

λ-measure; P-measure; L-measure; L^h-measure; Choquet integral regression model

1. Introduction

When interactions among independent variables exist in forecasting problems, the performance of the multiple linear regression models is poor. The traditional improved methods exploited the ridge regression models [1]. In this paper, we suggest use the Choquet integral regression models based on some single or compounded fuzzy measures to improve this situation. The well known fuzzy measures, λ-measure and P-measure, have only one formulaic solution of fuzzy measure, the former is not a closed form, and the later is not sensitive. Recently, some Choquet integral regression models based on different fuzzy measures were used by our previous works to further improve this situation [2, 3, 4, 5, 6].

In our previous works [7, 8, 9], we found that the Choquet integral regression model with L-measure based on γ-support has the best performances. In this paper, we proposed a new fuzzy measure, Hth-order L-measure, denoted as L^h-measure, which has infinitely many solutions of fuzzy measure with closed form and apply it to form a Choquet integral regression model.

This paper is organized as followings: Two well known fuzzy measure, λ-measure, P-measure and our L-measure are introduced in section 2; our new measure, L^h-measure, is introduced in section 3; the fuzzy support, γ-support are described in section 4; the Choquet integral regression model based on fuzzy measures are described in section 5; experiment and result are described in section 6; and final section is for conclusions and future works

2. Fuzzy Measures

The well known fuzzy measures, the λ-measure proposed by Sugeno in 1974, and P-measure proposed by Zadah in 1978, are concise introduced as follows.

2.1. Fuzzy Measures [10, 11, 12]

A fuzzy measure μ on a finite set X is a set function $\mu : 2^X \to [0,1]$ satisfying the following axioms:

1) $\mu(\phi) = 0$, $\mu(X) = 1$ *(boundary conditions)* (1)

2) $A \subseteq B \Rightarrow \mu(A) \le \mu(B)$ *(monotonicity)* (2)

2.2. Singleton Measures [4, 5, 6]

A singleton measure of a fuzzy measure μ on a finite set X is a function $s : X \to [0,1]$ satisfying:

$$s(x)=\mu(\{x\}),\ x\in X \tag{3}$$

$s(x)$ is called the density of singleton x.

2.3. λ-measure [11, 12]

For given singleton measures s, a λ-measure, g_λ, is a fuzzy measure on a finite set X, satisfying:

1) $A,B\in 2^X$, $A\cap B=\phi,\ A\cup B\neq X$

$$\Rightarrow g_\lambda(A\cup B)=g_\lambda(A)+g_\lambda(B)+\lambda g_\lambda(A)g_\lambda(B) \tag{4}$$

2) $$\prod_{i=1}^{n}\left[1+\lambda s(x_i)\right]=\lambda+1>0,\ s(x_i)=g_\lambda(\{x_i\}) \tag{5}$$

Note that once the singleton measure is known, we can obtain the values of λ uniquely by using the previous polynomial equation. In other words, λ-measure has a unique solution without closed form.

2.4. P-measure [13]

For given singleton measures s, a P-measure, g_P, is a fuzzy measure on a finite set X, satisfying:

$$^{\forall}A\in 2^X\Rightarrow g_P(A)=\max_{x\in A}s(x)=\max_{x\in A}g_P(\{x\}) \tag{6}$$

2.5. L-measure [6, 9]

For given singleton measure s, a L-measure, g_L, is a fuzzy measure on a finite set X, $|X|=n$, satisfying:

1) $$L\in[0,\infty) \tag{7}$$

2) $$\sum_{x\in X}s(x)=\sum_{x\in X}g_L(\{x\}) \tag{8}$$

3) $^{\forall}A\subset X, n-|A|+(|A|-1)L>0$

$$g_L(A)=\max_{x\in A}\left[s(x)\right]+\frac{(|A|-1)L\sum_{x\in A}s(x)}{\left[n-|A|+(|A|-1)L\right]\sum_{x\in X}s(x)}\left[1-\max_{x\in A}\left[s(x)\right]\right] \tag{9}$$

[Note]

i) L-measure has infinitely many solutions with closed form.

ii) When L=0, the L-measure is just a P-measure with closed form.

3. Hth-Order L-measure, L^h-measure

3.1. Definition of Hth-Order L-measure, L^h-measure

For given singleton measure s, a Hth-order L-measure, g_{L^P}, is a fuzzy measure on a finite set X, $|X|=n$, satisfying:

1) $$L\in[0,\infty),\ P\in\mathbb{N} \tag{10}$$

2) $$\sum_{x\in X}s(x)=\sum_{x\in X}g_{L^P}(\{x\}) \tag{11}$$

3) $^{\forall}A\subset X, n-|A|+(|A|-1)L>0$

$$g_{L^P}(A)=\max_{x\in A}\left[s(x)\right]+\frac{(|A|-1)L\left\{\sum_{x\in A}\left[s(x)\right]^{1/h}\right\}^h\left[1-\max_{x\in A}\left[s(x)\right]\right]}{\left[n-|A|+(|A|-1)L\right]\left(\left\{\sum_{x\in A}\left[s(x)\right]^{1/h}\right\}^h+\sum_{x\in(X-A)}s(x)\right)} \tag{12}$$

3.2. Important Properties of hth-Order L-measure, L^h-measure

1) Property 1

For given singleton measure s, $^{\forall}L\in[0,\infty)$, Hth-order L-measure is a fuzzy measure.

[Proof]

When L=0, L^h-measure is just the P-measure.

When L>0;

i) $\forall A\subset X$,Since

$$0\le\frac{(|A|-1)L\left\{\sum_{x\in A}\left[s(x)\right]^{1/h}\right\}^h}{\left[n-|A|+(|A|-1)L\right]\left(\left\{\sum_{x\in A}\left[s(x)\right]^{1/h}\right\}^h+\sum_{x\in(X-A)}s(x)\right)}\le 1 \tag{13}$$

We get $0\le g_{L^P}(A)\le 1$, The boundary condition is satisfied.

ii) $\forall A\subset B\subset X, |A|\le|B|\le|X|=n$

$$\Rightarrow \max_{x\in A}\left[s(x)\right]\le\max_{x\in B}\left[s(x)\right] \tag{14}$$

$$0\le\left\{\sum_{x\in A}\left[s(x)\right]^{1/h}\right\}^h\le\left\{\sum_{x\in B}\left[s(x)\right]^{1/h}\right\}^h \tag{15}$$

$$0\le\sum_{x\in(X-B)}s(x)\le\sum_{x\in(X-A)}s(x) \tag{16}$$

$$\Rightarrow 0 \le \frac{\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h}{\left(\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-A)} s(x)\right)}$$

$$\le \frac{\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h}{\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-B)} s(x)\right)} \le 1 \tag{17}$$

$$(|A|-1) \le (|B|-1),\ n-|B| \le n-|A| \tag{18}$$

$$\Rightarrow (|A|-1)L[n-|B|] \le (|B|-1)L[n-|A|] \tag{19}$$

$$\Rightarrow 0 \le \frac{(|A|-1)L}{[n-|A|+(|A|-1)L]} \le \frac{(|B|-1)L}{[n-|B|+(|B|-1)L]} \le 1 \tag{20}$$

$$\Rightarrow 0 \le \frac{(|A|-1)L\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h}{[n-|A|+(|A|-1)L]\left(\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-A)} s(x)\right)}$$

$$\le \frac{(|B|-1)L\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h}{[n-|B|+(|B|-1)L]\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-B)} s(x)\right)} \le 1 \tag{21}$$

If $\max_{x\in A}[s(x)] = \max_{x\in B}[s(x)] \Rightarrow g_{L^p}(A) \le g_{L^p}(B)$ (22)

If $\max_{x\in A}[s(x)] < \max_{x\in B}[s(x)]$ (23)

Let $\max_{x\in B}[s(x)] = \max_{x\in A}[s(x)] + c,\ 0 < c \le 1$ (24)

$$g_{L^h}(B) - g_{L^h}(A)$$

$$= c + \frac{(|B|-1)L\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h\left[1-\left(\max_{x\in A}[s(x)]+c\right)\right]}{[n-|B|+(|B|-1)L]\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-B)} s(x)\right)}$$

$$- \frac{(|A|-1)L\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h\left[1-\max_{x\in A}[s(x)]\right]}{[n-|A|+(|A|-1)L]\left(\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-A)} s(x)\right)}$$

$$= c\left[1 - \frac{(|B|-1)L\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h}{[n-|B|+(|B|-1)L]\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-B)} s(x)\right)}\right]$$

$$+L\left[1-\max_{x\in A}[s(x)]\right]\left[\frac{(|B|-1)\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h}{[n-|B|+(|B|-1)L]\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-B)} s(x)\right)} - \frac{(|A|-1)\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h}{[n-|A|+(|A|-1)L]\left(\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-A)} s(x)\right)}\right] \tag{25}$$

Since $(|B|-1)[n-|A|+(|A|-1)L] - (|A|-1)[n-|B|+(|B|-1)L]$

$$= (|B|-|A|)(n-1) \ge 0 \tag{26}$$

and
$$\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h\left(\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-A)} s(x)\right)$$

$$-\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-B)} s(x)\right)$$

$$= \left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h \sum_{x\in(X-A)} s(x) - \left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h \sum_{x\in(X-B)} s(x) \ge 0 \tag{27}$$

$$\Rightarrow (|B|-1)[n-|A|+(|A|-1)L]\times$$

$$\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h\left(\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h + \sum_{x\in(X-A)} s(x)\right)$$

$$\geq (|A|-1)\left[n-|B|+(|B|-1)L\right]\times$$
$$\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h+\sum_{x\in(X-B)}s(x)\right) \tag{28}$$

$$\Rightarrow\left[\frac{(|B|-1)\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h}{\left[n-|B|+(|B|-1)L\right]\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h+\sum_{x\in(X-B)}s(x)\right)}\right.$$
$$\left.\frac{(|A|-1)\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h}{\left[n-|A|+(|A|-1)L\right]\left(\left\{\sum_{x\in A}[s(x)]^{1/h}\right\}^h+\sum_{x\in(X-A)}s(x)\right)}\right]\geq 0 \tag{29}$$

Since

$$\left[1-\frac{(|B|-1)L\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h}{\left[n-|B|+(|B|-1)L\right]\left(\left\{\sum_{x\in B}[s(x)]^{1/h}\right\}^h+\sum_{x\in(X-B)}s(x)\right)}\right]\geq 0 \tag{30}$$

and $L\left[1-\max_{x\in A}[s(x)]\right]\geq 0$ (31)

We get $g_{L^P}(B)-g_{L^P}(A)\geq 0$ (32)

The monotonicity condition is satisfied.

2) Property 2

When h=1, Hth-order L-measure is L-measure; that is, L-measure is the special case of Hth-order L-measure.

4. γ–support [7, 8, 9]

For given singleton measure s of a fuzzy measure μ on a finite set X, if $\sum_{x\in X}s(x)=1$, then s is called a fuzzy support measure of μ, or a fuzzy support of μ, or a support of μ. Two kinds of fuzzy supports are introduced as below.

Let μ be a fuzzy measure on a finite set $X=\{x_1,x_2,...,x_n\}$, y_i be global response of subject i and $f_i(x_j)$ be the evaluation of subject i for singleton x_j, satisfying:

$$0<f_i(x_j)<1,\ i=1,2,...,N,\ j=1,2,...,n \tag{33}$$

If

$$\gamma(x_j)=\frac{1+r(f(x_j))}{\sum_{k=1}^{n}\left[1+r(f(x_k))\right]},\ j=1,2,...,n \tag{34}$$

where $r(f(x_j))=\frac{S_{y,x_j}}{S_y S_{x_j}}$ (35)

$$S_y^2=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N}y_i\right)^2 \tag{36}$$

$$S_{x_j}^2=\frac{1}{N}\sum_{i=1}^{n}\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N}f_i(x_j)\right]^2 \tag{37}$$

$$S_{y,x_j}=\frac{1}{N}\sum_{i=1}^{n}\left(y_i-\frac{1}{N}\sum_{i=1}^{N}y_i\right)\left[f_i(x_j)-\frac{1}{N}\sum_{i=1}^{N}f_i(x_j)\right] \tag{38}$$

satisfying $0\leq\gamma(x_j)\leq 1$ and $\sum_{j=1}^{n}\gamma(x_j)=1$ (39)

then the function $\gamma:X\to[0,1]$ satisfying $\mu(\{x\})=\gamma(x)$, $\forall x\in X$ is a fuzzy support of μ, called γ-support of μ.

6. Choquet Integral Regression Models

6.1. Choquet Integral [4, 12, 13]

Let μ be a fuzzy measure on a finite set X. The Choquet integral of $f_i:X\to R_+$ with respect to μ for individual i is denoted by

$$\int_C f_i d\mu=\sum_{j=1}^{n}\left[f_i(x_{(j)})-f_i(x_{(j-1)})\right]\mu(A_{(j)}^i)\ ,i=1,2,...,N \tag{40}$$

where $f_i(x_{(0)})=0$, $f_i(x_{(j)})$ indicates that the indices have been permuted so that

$$0\leq f_i(x_{(1)})\leq f_i(x_{(2)})\leq...\leq f_i(x_{(n)}) \tag{41}$$

$$A_{(j)}=\{x_{(j)},x_{(j+1)},...,x_{(n)}\} \tag{42}$$

6.2. Choquet Integral Regression Models [4, 5, 6, 7, 8, 9]

Let $y_1, y_2, ..., y_N$ be global evaluations of N objects and $f_1(x_j), f_2(x_j), ..., f_N(x_j), j = 1, 2, ..., n$, be their evaluations of x_j, where $f_i : X \to R_+$, $i = 1, 2, ..., N$.

Let μ be a fuzzy measure, $\alpha, \beta \in R$,

$$y_i = \alpha + \beta \int_C f_i dg_\mu + e_i \ , \ e_i \sim N(0, \sigma^2) \ , \ i = 1, 2, ..., N \quad (43)$$

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \left[\sum_{i=1}^{N} \left(y_i - \alpha - \beta \int_C f_i dg_\mu \right)^2 \right] \quad (44)$$

then $\hat{y}_i = \hat{\alpha} + \hat{\beta} \int f_i dg_\mu, \ i = 1, 2, ..., N$ is called the Choquet integral regression equation of μ, where

$$\hat{\beta} = S_{yf} / S_{ff} \quad (45)$$

$$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^{N} y_i - \hat{\beta} \frac{1}{N} \sum_{i=1}^{N} \int f_i dg_\mu \quad (46)$$

$$S_{yf} = \frac{\sum_{i=1}^{N} \left[y_i - \frac{1}{N} \sum_{i=1}^{N} y_i \right] \left[\int f_i dg_{\mu^*} - \frac{1}{N} \sum_{k=1}^{N} \int f_k dg_{\mu^*} \right]}{N-1} \quad (47)$$

$$S_{ff} = \frac{\sum_{i=1}^{N} \left[\int f_i dg_{\mu^*} - \frac{1}{N} \sum_{k=1}^{N} \int f_k dg_{\mu^*} \right]^2}{N-1} \quad (48)$$

7. Experiment and Result

A real data set with 59 samples from a junior high school in Taiwan including the independent variables, examination scores of four courses, and the dependent variable, the score of the Basic Competence Test of junior high school is applied to evaluate the performances of four Choquet integral regression models with P-measure, λ-measure, L-measure and L^h-measure based on γ-support respectively, a ridge regression model, and a multiple linear regression model by using 5-fold cross validation method to compute the mean square error (MSE) of the dependent variable. The formulas of MSE is

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (49)$$

For any fuzzy measure, μ-measures, once the fuzzy support of the μ-measure is given, all event measures of μ can be found, and then, the Choquet integral based on μ and the Choquet integral regression equation based on μ can also be found.

The singleton measures, γ-support of the P-measure, λ-measure, L-measure and L^h-measure can be obtained by using the formulas (30).

The experimental results of six forecasting models are listed in Table 1 and Table 2. We can find that the Choquet integral regression model with L^h-measure based on γ-support outperforms other forecasting regression models.

TABLE 1.MSE OF CHOQUET INTEGRAL REGRESSON MODELS WITH PTH-ORDER L-MEASURE BASE ON γ-SUPPORT

MSE of Choquet Integral Regression Models with Hth-order L-measure based on γ-support			
h	*5-fold CV MSE*	*h*	*5-fold CV MSE*
1	56.2711	10	53.5390
2	54.7839	11	53.5371
3	54.1228	12	53.5361
4	53.8145	13	53.5357
5	53.6690	14	53.5354
6	53.5999	15	53.5353
7	53.5668	16	53.5352
8	53.5507	17	53.5352
9	53.5429	18	53.5352

TABLE 2. MSE OF REGRESSON MODEL

Regression model		**5-fold CV MSE**
	measure	
Choquet Integral Regression model	***P***	68.9878
	λ	57.5449
	L	56.2711
	Min L^h	53.5352
Ridge regression		63.1263
Multiple linear regression		69.7094

8. Conclusions and Future Works

In this paper, a novel fuzzy measure, second-order L-measure, Choquet integral regression models with fuzzy measure are proposed. An educational data experiment is conducted for comparing the performances of a ridge regression model, a multiple linear regression model, and the proposed Choquet integral regression model with P-measure, λ-measure, L-measure and second-order L-measure based on γ-support. Experimental result shows that the Choquet integral regression models with the proposed second-order L-measure based on γ-support outperforms other forecasting models.

In future, we will apply the proposed Choquet integral regression model with fuzzy measure based on γ-support to develop multiple classifier system.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 97-2410-H-468-014)

References

[1] A. E. Hoerl, R. W. Kenard, and K. F. Baldwin, "Ridge regression: Some simulation," Communications in Statistics, vol. 4, No. 2, pp. 105-123, 1975.

[2] H.-C. Liu, C.-C. Chen, D.-B. Wu, and Y.-D. Jheng, "A new weighting method for detecting outliers in IPA based on Choquet integral," IEEE International conference on Industrial Engineering and Engineering Management 2007, Singapore, December 2007.

[3] H.-C. Liu, "The Choquet integral regression model based on r-complete measure," Journal of educational research and development, vol. 2, No. 4, pp. 87-107, 2006 (in Chinese).

[4] H.-C. Liu, W.-C. Lin, and W.-S. Weng, "A Choquet Integral Regression Model Based on a New Fuzzy Measure," The 12th International Conference on Fuzzy Theohery & Technology, Salt Lake City, Utah, U, July 2007.

[5] H.-C. Liu, W.-C. Lin, K.-Y. Chang, and W.-S. Weng, "A Nonlinear Regression Model Based on Choquet Integral with ε-Measure," 2007 WSEAS International Conferences, Venice, Italy, November 2007.

[6] H.-C. Liu, Y.-D. Jheng, W.-C. Lin, and G.-S. Chen, "A novel fuzzy measure and its Choquet Integral Regression Model," Proceedings of International conference on Machine Learning and Cybernetics 2007, Hong Kong, China, August 2007, vol. 7, No.3, pp. 1394-1398, ISBN 1-4244-0973X.

[7] H.-C. Liu, Y.-C. Tu, C.-C. Chen, and W.-S. Weng, "The Choquet integral with respect to λ-measure based on γ-support," 2008 International Conferences on Machine Learning and Cybernetics, Kuming, China, July 2008.

[8] H.-C. Liu, Y.-C. Tu, W.-C. Huang, and C.-C. Chen, "The Choquet integral with respect to R-measure based on γ-support,"ICNC'08-FSKD'08, Jinan, China, October 18-20, 2008.

[9] H.-C. Liu, Y.-C. Tu, W.-C. Lin, and C.-C. Chen, "Choquet integral regression model based on L-measure and γ-support," 2008 International Conferences on Wavelet Analysis and Pattern Recognition, Hong Kong, China,August 2008.

[10] G. Choquet, "Theory of capacities," Annales de l'Institut Fourier, vol. 5, pp. 131-295, 1953.

[11] M. Sugeno, "Theory of fuzzy integrals and its applications," unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[12] Z. Wang, and G. J. Klir, Fuzzy Measure Theory, Plenum Press, New York, 1992.

[13] L. A. Zadeh, "Fuzzy Sets and Systems,"vol. 1, pp. 3, 1978.

A COMPARISON ON CHOQUET INTEGRAL WITH RESPECT TO DIFFERENT INFORMATION-BASED FUZZY MEASURES

HORNG-JINH CHANG[1], HSIANG-CHUAN LIU[2], SHANG-WEN TSENG[3], FENGMING M. CHANG[4]

[1]Department of Business Administration, Asia University, Wufeng, Taiwan
[2]Department of Bioinformatics, Asia University, Wufeng, Taiwan
[3]Department of Computer Science and Information Engineering, Asia University, Wufeng, Taiwan
[4]Department of Information Science and Applications
E-MAIL: lhc@asia.edu.tw, chj@asia.edu.tw, hugo2460@ms51.hinet.net, spyable@gmail.com

Abstract:

In this paper, for grouped data, three kinds of the Choquet integral regression models with fuzzy measures based on joint entropy, complexity and multiple mutual information is considered. The above three fuzzy measures are called, E-measure, C-measure and M-measure, respectively. For evaluating the Choquet integral regression models with these three information-based fuzzy measures, a real grouped data experiment by using a 5-fold cross validation accuracy is conducted. The performances of the Choquet integral regression models based on these three fuzzy measures, respectively, and the traditional multiple linear regression model are compared. Experimental result shows that the Choquet integral regression model based on our proposed M-measure has the best performance and it outperforms the Choquet integral regression model based on our previous proposed C-measure.

Keywords:

E-measure; C-measure; M-measure; Choquet integral; Choquet integral regression model

1. Introduction

When interactions among independent variables exist in forecasting problems, the well known multiple linear regression method is unable to overcome the undesirable phenomenon. In contrast, the Choquet integral takes into account the interactions among criteria [1]. In addition, there is a key issue unsolved in the application of fuzzy integral with the determination of density values to decide the fuzzy measures in the fusion process [2], [3]. In this paper, for grouped data, three kinds of fuzzy measures based on information theory are considered, the first one is the joint entropy-based fuzzy measure, called E-measure [4], [5], [6], the second one is the complexity-based fuzzy measure proposed by or previous study, called C-measure [7], and the third one is our proposed multiple mutual information-based fuzzy measure, called M-measure.

For evaluating the Choquet integral regression models with these three information-based fuzzy measures, a real grouped data experiment by using a 5-fold cross validation accuracy is conducted. The performances of the Choquet integral regression models with these three information-based fuzzy measures, respectively, and the traditional multiple linear regression model are compared.

This paper is organized as followings: three kinds of information-based fuzzy measures are introduced in section 2, Choquet integral is described in section 3. Choquet integral regression models with respect to different fuzzy measures are described in section 4, Experiment and result are described in section 5 and final section is for conclusions and future works.

2. Information-based fuzzy measures

2.1. Fuzzy measure [2], [3]

The Choquet integral takes into account the interaction by means of fuzzy measure, to compute the Choquet integral we need to compute a fuzzy measure first. The formal definition of fuzzy measure is given as below

[Definition 1] A fuzzy measure μ on a finite set X is a set function $\mu : 2^X \to [0,1]$ satisfying the following axioms:

1) $\mu(\phi) = 0$, $\mu(X) = 1$ (boundary conditions) (1)

2) $A \subseteq B \Rightarrow \mu(A) \le \mu(B)$ (monotonicity) (2)

In this paper, entropy-based method and two of our proposed methods, complexity-based method and multiple mutual information- based method, are discussed.

978-1-4244-3703-0/09/$25.00 ©2009 IEEE

2.2. Joint entropy-based fuzzy measure, E-measure

2.2.1. Joint entropy [4]

[Definition 2] Let $Y, X_1, X_2, ..., X_m$ be (m+1) random variables, the joint entropy of Y and $X_1, X_2, ..., X_m$, denoted as $H(Y, X_1, X_2, ..., X_m)$, is defined as follows

$$H(Y, X_1, X_2, ..., X_m) = -E\left[\log f_{Y, X_1, ..., X_m}(y, x_1, ..., x_m)\right] \quad (3)$$

[Property 1]

$$0 \le H(Y, X_1) \le H(Y, X_1, X_2) \le ... \le H(Y, X_1, X_2, ..., X_m) \quad (4)$$

2.2.2. Joint entropy-based measure, E-measure [5], [6]

[Definition 3] The joint entropy-based measure, E-measure, on a finite set $X = \{X_1, X_2, ..., X_n\}$ is a set function satisfying the following conditions

1) $y_i, i = 1, 2, .., N$ are scores of dependent variable Y for examinee i, $X^i = \{X_1^i, X_2^i, ..., X_n^i\}$ is the set of n independent variables for any individual $x_j^i : X_j \to R^+$, $i = 1, 2, ..., N$, $j = 1, 2, ..., n$ are scores of independent variable j for individual i.

2) $\{x_{(1)}^i, x_{(2)}^i, ..., x_{(n)}^i\}$ is a permutation of $\{x_1^i, x_2^i, ..., x_n^i\}$ for examinee i, satisfying

$$x_{(1)}^i \le x_{(2)}^i \le ... \le x_{(n)}^i \quad (5)$$

3) $A_{(j)}^i = \{X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\}, j = 1, 2, ..., n \quad (6)$

4)
$$E\left(A_{(j)}^i\right) = \frac{H\left(Y, A_{(j)}^i\right)}{H(Y; X)} = \frac{H\left(Y, X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\right)}{H(Y; X_1, X_2, ..., X_n)} \quad (7)$$

where $A_{(j)}^i \subset X$

$$H\left(Y, A_{(j)}^i\right) = H\left(Y, X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\right) = -E\left[\log f_{\left(Y, X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\right)}\left(y, x_{(j)}^i, x_{(j+1)}^i, ..., x_{(n)}^i\right)\right]$$

$$H(Y, \phi) = 0 \quad (8)$$

[Property 2]

$$1 \ge E\left(Y, A_{(1)}^i\right) \ge E\left(Y, A_{(2)}^i\right) \ge ... \ge E\left(Y, A_{(n)}^i\right) \ge 0 \quad (9)$$

2.3. Complexity-based fuzzy measure, C-measure

2.3.1. Complexity

[Definition 4] Let $Y, X_1, X_2, ..., X_m$ be (m+1) random variables, the complexity of Y and $X_1, X_2, ..., X_m$, denoted as $N(Y, X_1, X_2, ..., X_m)$, is defined as the number of different patterns of the outcomes of $(Y, X_1, X_2, ..., X_m)$

[Property 3]

$$0 = N(Y, \phi) \le N(Y, X_1) \le N(Y, X_1, X_2) \le ... \le N(Y, X_1, X_2, ..., X_m) \quad (10)$$

2.3.2. Complexity-based fuzzy measure, C-measure [7]

[Definition 5] The complexity-based fuzzy measure, C-measure, on a finite set $X = \{X_1, X_2, ..., X_n\}$ is a set function $C : 2^X \to [0, 1]$ satisfying the following conditions

1) $y_i, i = 1, 2, .., N$ are scores of dependent variable Y for examinee i, $X^i = \{X_1^i, X_2^i, ..., X_n^i\}$ is the set of n independent variables for any individual $x_j^i : X_j \to R^+$, $i = 1, 2, ..., N$, $j = 1, 2, ..., n$ are scores of independent variable j for individual i.

2) $\{x_{(1)}^i, x_{(2)}^i, ..., x_{(n)}^i\}$ is a permutation of $\{x_1^i, x_2^i, ..., x_n^i\}$ for examinee i, satisfying

$$x_{(1)}^i \le x_{(2)}^i \le ... \le x_{(n)}^i \quad (11)$$

3) $A_{(j)}^i = \{X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\}, j = 1, 2, ..., n \quad (12)$

4)
$$C\left(A_{(j)}^i\right) = \frac{N\left(Y, A_{(j)}^i\right)}{N(Y; X)} = \frac{N\left(Y, X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\right)}{N(Y; X_1, X_2, ..., X_n)} \quad (13)$$

[Property 4]

$$1 = C\left(A_{(1)}^i\right) \ge C\left(A_{(2)}^i\right) \ge ... \ge C\left(A_{(n)}^i\right) \ge 0 \quad (14)$$

2.4. Multiple mutual information-based fuzzy measure, M-measure [8]

2.4.1. Multiple mutual information

[Definition 6] Let $Y, X_1, X_2, ..., X_m$ be (m+1) random variables, the multiple mutual information of Y and $X_1, X_2, ..., X_m$, denoted as $I(Y, X_1, X_2, ..., X_m)$, is defined as follows

$$I(Y, X_1, X_2, ..., X_m) = \sum_{y} \sum_{x_1, x_2, ..., x_m} f_{Y, X_1, ..., X_m}(y, x_1, ..., x_m) \log \frac{f_{Y, X_1, ..., X_m}(y, x_1, ..., x_m)}{f_Y(y) f_{X_1, ..., X_m}(x_1, ..., x_m)} \quad (15)$$

[Property 5]

$$0 \le I(Y, X_1) \le I(Y, X_1, X_2) \le ... \le I(Y, X_1, X_2, ..., X_m) \quad (16)$$

2.4.2. M-measure

[Definition 7] The M-measure on a finite set $X = \{X_1, X_2, ..., X_n\}$ is a set function $M : 2^X \to [0,1]$ satisfying the following conditions

1) $y_i, i = 1, 2, .., N$ are scores of dependent variable Y for examinee i, $X^i = \{X_1^i, X_2^i, ..., X_n^i\}$ is the set of n independent variables for any individual $x_j^i : X_j \to R^+$, $i = 1, 2, ..., N$, $j = 1, 2, ..., n$ are scores of independent variable j for individual i.

2) $\{x_{(1)}^i, x_{(2)}^i, ..., x_{(n)}^i\}$ is a permutation of $\{x_1^i, x_2^i, ..., x_n^i\}$ for examinee i, satisfying

$$x_{(1)}^i \le x_{(2)}^i \le ... \le x_{(n)}^i \quad (17)$$

3) $$A_{(j)}^i = \{X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\}, j = 1, 2, ..., n \quad (18)$$

4) $$M\left(A_{(j)}^i\right) = \frac{I\left(Y, A_{(j)}^i\right)}{I(Y; X)} = \frac{I\left(Y, X_{(j)}^i, X_{(j+1)}^i, ..., X_{(n)}^i\right)}{I(Y; X_1, X_2, ..., X_n)} \quad (19)$$

where $I(Y, \phi) = 0$, $A_{(j)}^i \subset X$

[Property 6]

$$1 = M\left(A_{(1)}^i\right) \ge M\left(A_{(2)}^i\right) \ge ... \ge M\left(A_{(n)}^i\right) \ge 0 \quad (20)$$

3. Choquet integral [3], [4]

[Definition 8] Let μ be a fuzzy measure on a finite set X. The Choquet integral of $f_i : X \to R_+$ with respect to μ for individual i is denoted by

$$\int_C f_i d\mu = \sum_{j=1}^{n} \left[f_i\left(x_{(j)}\right) - f_i\left(x_{(j-1)}\right) \right] \mu\left(A_{(j)}^i\right), i = 1, 2, ..., N \quad (21)$$

where $f_i\left(x_{(0)}\right) = 0$, $f_i\left(x_{(j)}\right)$ indicates that the indices have been permuted so that

$$0 \le f_i\left(x_{(1)}\right) \le f_i\left(x_{(2)}\right) \le ... \le f_i\left(x_{(n)}\right),$$

$$A_{(j)} = \left\{x_{(j)}, x_{(j+1)}, ..., x_{(n)}\right\} \quad (22)$$

4. Choquet integral regression models [8]

[Definition 9] Let $y_1, y_2, ..., y_N$ be global evaluations of N objects (or by N individuals), and $f_1(x_j), f_2(x_j), ..., f_N(x_j), j = 1, 2, ..., n$, be their evaluations of x_j, where $f_i : X \to R_+, i = 1, 2, ..., N$. Let μ be a fuzzy measure, $\alpha, \beta \in R$

$$y_i = \alpha + \beta \int_c f_i dg_\mu + e_i, \quad e_i \sim N\left(0, \sigma^2\right), i = 1, 2, ..., N \quad (23)$$

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg\min_{\alpha, \beta} \left[\sum_{i=1}^{N} \left(y_i - \alpha - \beta \int_c f_i dg_\mu \right)^2 \right] \quad (24)$$

then $\hat{y}_i = \hat{\alpha} + \hat{\beta} \int f_i dg_\mu$, $i = 1, 2, ..., N$ is called the optimal Choquet integral regression equation of μ, where

$$\hat{\beta} = S_{yf} / S_{ff} \quad (25)$$

$$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^{N} y_i - \hat{\beta} \frac{1}{N} \sum_{i=1}^{N} \int f_i dg_\mu \quad (26)$$

$$S_{yf} = \frac{\sum_{i=1}^{N} \left[y_i - \frac{1}{N} \sum_{i=1}^{N} y_i \right] \left[\int f_i dg_{\mu^*} - \frac{1}{N} \sum_{k=1}^{N} \int f_k dg_{\mu^*} \right]}{N-1} \quad (27)$$

$$S_{ff} = \frac{\sum_{i=1}^{N} \left[\int f_i dg_{\mu^*} - \frac{1}{N} \sum_{k=1}^{N} \int f_k dg_{\mu^*} \right]^2}{N-1} \quad (28)$$

5. Experiment and result

A real raw data set comes from a class with 59 students in a junior high school in Taiwan, and each student took 3 courses (namely physics and chemistry, biology, and geoscience) for natural science. The credit hours for these three courses are 16, 4, and 4, respectively. The maximum score for each course is 100 points. Later, all students took a Basic Competence Test of natural science for all junior high school students. The maximum and minimum scores of the Basic Competence Test are 60 and 1. To simplify the notations, the scores of physics and chemistry, biology, and geoscience are denoted by C1, C2, and C3, while the scores of natural science in the Basic Competence Test is denoted by BCT. The detailed information is depicted in TA BLE 2.

For computing each joint entropy and mutual information, we need to consider the distributions and the joint probability about the data, and we need to decide the number of level to be used to classify the raw data into the level of the score for each criterion. In our study, the sample size n is 59, according to the Sturge's formula [9]

$$m = 1 + 3.3\log_{10}(n) = 1 + 3.3\log_{10}(59) \tag{29}$$

We can obtain the possible candidates m=6 or 7, in this study, set m=3, 4, 5, 6, 7, 8.

Then, to transform the scores of the raw data of the courses into the level of the scores for each item when m=3, 4, 5, 6, 7, 8, for example, the transformed data for m=6 is listed in TABLE 3.

And then, we can compute the above three kinds of fuzzy measures, E-measure; C-measure; M-measure, and their Choquet integrals, furthermore, we can obtain all of the estimated overall performance values for m=3, 4, 5, 6, 7, 8.

Next, transform the results into the level of the scores for each m=3, 4, 5, 6, 7, 8. For example, the transformed data of four forecasting regression models for m=6 is listed in TABLE 4.

Finally, by using 5-fold cross validation method to compute the accuracy of BCT for four methods, the results are listed in TABLE 1.

From TABLE 1, we know that all of the three Choquet integral methods are better than the traditional regression model, Our proposed the Choquet integral methods based on M-measure is better than other two methods, so our proposed Choquet integral methods based on M-measure has the best performance.

TABLE 1 The accuracy of each method for m=3, 4, 5, 6, 7, 8

m	Regression	Choquet integral with E-measure	Choquet integral with C-measure	Choquet integral with M-measure
3	0.5254	0.6102	0.5763	0.6102
4	0.4576	0.4915	0.4237	0.4915
5	0.3051	0.3729	0.3390	0.3729
6	0.3051	0.3559	0.3898	0.3729
7	0.2542	0.3220	0.3390	0.3390
8	0.2542	0.2712	0.2881	0.2881

6. Conclusions and future works

When the sub-tests of a composite test are with interaction, the performance of the traditional additive scale method is poor. Non-additive fuzzy measures and fuzzy integral can be applied to improve this situation. In this study, a real data set from a junior high school including the independent variables, test scores of three courses with interaction, and the dependent variable, junior high school graduates' scores of the Basic Competence Test (BCT) are applied to evaluate the performances of the Choquet integral regression model with three fuzzy measures, E-measure, C-measure, M-measure, and traditional multiple linear regression model. Experimental result shows that Choquet integral regression model with M-measure has the best performance, the rest in order are Choquet integral regression model with C–measure, Choquet integral regression model with E-measure and the multiple linear regression model.

The Choquet integral regression model with M-measure can be used to not only the interval variables but also the nominal variables. In future we will apply the proposed Choquet integral regression model based on the new measure to develop multiple classifier system.

Acknowledgements

This paper is partially supported by the National Science Council grant (NSC 97-2410-H-468-014).

References

[1] G. Choquet, "Theory of capacities", Annales de l'Institut Fourier, Vol. 5, pp. 131-295, 1954.

[2] Z. Wang, and G. J. Klir, Fuzzy measure theory, Plenum Press, New York, 1992.

[3] M. Sugeno, "Theory of fuzzy integrals and its applications", unpublished doctoral dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[4] T. Calvo, A. Kolesarova, M. Komornikova and R. Mesiar, "Agregation operators: properties, classes and construction methods", Fuzziness and Soft Computing, Vol. 97, pp. 1-103, 2001.

[5] I. Kojadinovic, "Estimation of the weights of interaction criteria from the set of profile by means of information-theoretic functionals", European Journal of Operational Research, Vol. 155, pp. 271–350, June 2004.

[6] K. Yu, Z. Wen, X. Xu, & M Ester, "Feature Weighting and Instance Selection for Collaborative Filtering", In Proceedings of 2nd International Workshop on Management of Informaon on the Web-Web Data and Text Mining (MIV'O1) 2001.

[7] J-I. Shieh, H-H. Wu, and H-C. Liu, “Applying A Compexity-Based Choquet Integral to Evaluate Students’ Performance”, Expert Systems with Applications, Vol. 36, pp. 5100-5106, April 2009.

[8] H-C. Liu, H-J. Chang, W-C. Lin, K.-Y. Chang . “A Choquet Integral Regression Model with a New Fuzzy Based on Multiple Mutual-infomation”, 2008 International Conference on Machine Learnig and Cybernetics, Kuming, China, July, 2008.

[9] D. W. Scott, Multivariate density estimation: theory, practice, and visualization. New York: John Wiley, 1992.

TABLE 2 Raw data of the scores of 59 students

Student	C1	C2	C3	BCT	Student	C1	C2	C3	BCT	Student	C1	C2	C3	BCT
1	77	75	79	31	21	53	68	74	11	41	74	86	87	44
2	71	72	78	26	22	56	63	69	21	42	78	83	81	50
3	78	86	86	33	23	70	80	78	31	43	47	58	66	15
4	58	64	68	32	24	51	74	82	49	44	51	60	63	18
5	48	59	65	16	25	61	66	72	33	45	60	65	75	23
6	68	74	77	28	26	67	70	80	35	46	68	68	80	26
7	62	72	84	47	27	59	75	80	27	47	52	60	70	20
8	51	53	65	9	28	53	56	70	22	48	57	65	75	24
9	62	64	76	36	29	56	56	65	6	49	70	66	70	13
10	63	70	81	41	30	52	57	67	15	50	53	68	74	30
11	66	68	75	25	31	74	70	80	35	51	68	68	78	35
12	66	72	80	23	32	56	61	75	22	52	57	60	68	23
13	75	75	85	39	33	62	68	72	29	53	61	62	70	25
14	74	63	69	12	34	86	80	82	35	54	59	70	80	37
15	68	78	85	27	35	63	78	88	31	55	59	62	70	29
16	71	74	80	26	36	56	66	76	21	56	52	64	76	27
17	49	60	69	13	37	77	74	80	42	57	68	70	80	33
18	73	78	84	39	38	73	78	84	24	58	71	76	74	38
19	68	70	74	40	39	63	60	68	17	59	72	66	78	19
20	54	56	62	7	40	53	68	80	31					

C1 : physics and chemistry
C2 : biology
C3 : geoscience
BCT : Basic Competence Test

TABLE 3 Transformed data of the scores of 59 students for m=6

Student	C1	C2	C3	BCT	Student	C1	C2	C3	BCT	Student	C1	C2	C3	BCT
1	5	4	4	4	21	1	3	3	1	41	5	6	6	6
2	4	4	4	3	22	2	2	2	3	42	5	6	5	6
3	5	6	6	4	23	4	5	4	4	43	1	1	1	2
4	2	2	2	4	24	1	4	5	6	44	1	2	1	2
5	1	2	1	2	25	3	3	3	4	45	2	3	3	3
6	4	4	4	3	26	4	4	5	4	46	4	3	5	3
7	3	4	6	6	27	2	4	5	3	47	1	2	2	2
8	1	1	1	1	28	1	1	2	3	48	2	3	3	3
9	3	2	4	5	29	2	1	1	1	49	4	3	2	1
10	3	4	5	5	30	1	1	2	2	50	1	3	3	4
11	3	3	3	3	31	5	4	5	4	51	4	3	4	4
12	3	4	5	3	32	2	2	3	3	52	2	2	2	3
13	5	4	6	5	33	3	3	3	4	53	3	2	2	3
14	5	2	2	1	34	6	5	5	4	54	2	4	5	5
15	4	5	6	3	35	3	5	6	4	55	2	2	2	4

16	4	4	5	3	36	2	3	4	3	56	1	2	4	3
17	1	2	2	1	37	5	4	5	5	57	4	4	5	4
18	4	5	6	5	38	4	5	6	3	58	4	5	3	5
19	4	4	3	5	39	3	2	2	2	59	4	3	4	2
20	2	1	1	1	40	1	3	5	4					

C1 : physics and chemistry
C2 : biology
C3 : geoscience
BCT : Basic Competence Test

TABLE 4 Transformed data of the estimated scores by each method for m=6

Student	R	E	C	M	BCT	Student	R	E	C	M	BCT
1	4	4	4	4	4	31	4	4	5	5	4
2	4	4	4	4	3	32	3	3	3	3	3
3	5	5	5	5	4	33	3	3	3	3	4
4	3	2	3	3	4	34	5	5	5	5	4
5	2	2	2	2	2	35	5	4	4	4	4
6	4	4	4	4	3	36	4	3	3	3	3
7	4	4	4	4	6	37	4	4	4	4	5
8	2	2	2	2	1	38	5	5	5	5	3
9	3	3	3	3	5	39	3	3	3	3	2
10	4	4	4	4	5	40	4	3	3	3	4
11	3	3	3	3	3	41	5	5	5	5	6
12	4	4	4	4	3	42	5	5	5	5	6
13	4	5	5	5	5	43	2	2	2	2	2
14	3	3	3	3	1	44	2	2	2	2	2
15	5	5	4	5	3	45	3	3	3	3	3
16	4	4	4	4	3	46	4	4	4	4	3
17	3	2	2	2	1	47	3	2	2	2	2
18	5	5	5	5	5	48	3	3	3	3	3
19	4	4	4	4	5	49	3	3	3	3	1
20	2	2	2	2	1	50	3	3	3	3	4
21	3	3	3	3	1	51	3	4	4	4	4
22	3	2	3	3	3	52	3	2	3	3	3
23	4	4	4	4	4	53	2	3	3	3	3
24	3	4	3	3	6	54	4	4	4	4	5
25	3	3	3	3	4	55	2	2	3	2	4
26	4	4	4	4	4	56	3	3	3	3	3
27	4	4	3	4	3	57	4	4	4	4	4
28	2	2	2	2	3	58	4	4	4	4	5
29	2	2	2	2	1	59	4	4	4	4	2
30	2	2	2	2	2						

R : estimated transformed -scores of BCT by using the regression model
E : estimated transformed -scores of BCT by using the Choquet integral with E-measure
C : estimated transformed -scores of BCT by using the Choquet integral with C-measure
M : estimated transformed -scores of BCT by using the Choquet integral with M-measure
BCT : classified scores of the Basic Competence Test

A Novel Approach for Evaluating Class Structure Ambiguity

Jing-Doo Wang[1], Hsiang-Chuan Liu[2], Yao-Chug Shi[1]

[1]Department of Computer Science and Information Engineering, Asia University, Taichung 41354, Taiwan
[2]Department of Bioinformatics, Asia University, Taichung 41354, Taiwan
E-MIAL: wangjingdoo@gmail.com, lhc@asia.edu.tw, aoc5202000@yahoo.com.tw

Abstract

It is attractive and worthy to estimate the ambiguity of one existing class structure such that one could give suggestions to domain experts when and how to reorganize the original class structure. In this paper *Class Structure Ambiguity* (CSA) was proposed to estimate the quality of one class structure. To inspect whether the CSA did tell the quality of class structure or not, the Pearsons correlation between classification accuracies achieved by a linear SVM classifier and the values of CSA were evaluated according to two types of datasets, one generated randomly and another selected from the LIBSVM. The experimental results showed that the CSA did reveal the degree of the ambiguities among classes. To our knowledge, we were the first to discuss the problem of class structure ambiguity.

Keywords:

classification, class structure, class ambiguity.

1 Introduction

Classification was well known as a supervised problem[2, 12, 17, 20]. The class structures usually were determined in the beginning by some domain experts manually or automatically [5, 6, 7, 13, 14, 15, 19, 21, 22] and it was difficult to challenge these experts to verify whether that structures were well designed or not. Therefore, it is attractive and worthy to estimate the ambiguity of one existing class structure, especially when that structure had been existed for a long time, or when the characteristics of its contents become diverse as more and more instances put into it. That is, it is important to know how to evaluate the ambiguity of class structure such that one could give suggestions to those domain experts when and how to reorganize the original class structure.

The classifiers achieving the highest classification accuracy for some well known datasets available in the public domain were regularly proposed and published in the papers. However, focusing on searching such classifier from a lot of classifiers was somewhat another kind of overfitting problem that found one specific classifier fitting to that datasets, even with k-fold cross-validation[2]. On the other hand, the problem of class ambiguity becomes significant for the class structures designed only with a few number of classes in the beginning such that the coming instances might be put to a unsuitable classes for storage by the distributors or automatic classification software at that moment. Although there were studies for discussing the quality of clustering approaches [8, 23], it was seldom to discuss the class ambiguous problem that how to evaluate the degree of class ambiguity between two classes, and how to determine the quality of one existing class structure. In our previous study [18], the estimation unit using the standard derivation of the distances from every instances in one class to the centroid of that class were coarse and it was only for estimating the ambiguity between two classes, but not for evaluating the ambiguities for the whole class structure.

In this paper an approach was proposed to estimate the quality of one class structure according to the value of *Class Structure Ambiguity* (CSA). Generally speaking, the less ambiguity of one class structure was, the higher accuracy one classifier could achieve. In other words, the value of CSA might suppose to be high if the value of classification accuracy was low. To inspect whether the CSA did reveal the ambiguity of structure class or not, it was expected that the degree of class ambiguity increased as the value of classification accuracy decreased. To show that expectation in this paper, the Pearsons correlation [1, 3] between classification accuracies and the values of CSA was computed according to two types of datasets, one generated randomly and another selected from the LIBSVM [4]. The accuracies achieved by linear SVM classifier derived from the LIBSVM [4]. Note that linear SVM was well-known and achieved high accuracy [8, 10, 11].

The experimental results showed that the values of Pearsons correlation as described above were above -0.9. This observation told that the correlation was a negative linear relationship between accuracy and CSA. That is, the higher classification accuracies were, the lower the values of CSA were. In other words, the CSA proposed in this paper did reveal the degree of class structure ambiguity. To our knowledge, we were the first to discuss the problem of class structure ambiguity.

The remainder of this paper is organized as follows. Section 2 gives the notations and the computation of CSA. Section 3 gives experimental results. Section 4 gives conclusions and discussions.

2 Methods

It was expectable that there was a negative linear relationship between the classification accuracy and the class structure ambiguity. In other words, the higher classification accuracy was, the less class structure ambiguity was. In this paper the ambiguity of one class structure was measured according to the summation of the class ambiguities of all pairs of any two classes within that class structure. To evaluate whether the measurement of class structure ambiguity works or not, the Pearsons correlation [1, 3] was computed between the classification accuracies achieved by linear SVM classifier from LIBSVM[4] and the values of class structure ambiguity according to different datasets. Note that linear SVM classifier was well-known for achieving high classification accuracy[2]. In the following the notations used in this paper was given in section 2.1, and then described the computation of class structure ambiguity in section 2.2.

2.1 Notations

Let $\{C_1, C_2, \ldots, C_c\}$ be an actual partition of a data set Y as

$$Y = \left\{ \begin{array}{l} y_{1,1}, y_{1,2}, \cdots, y_{1,n_1}, \\ y_{2,1}, y_{2,2}, \cdots, y_{2,n_2}, \\ \cdots, \\ y_{c,1}, y_{c,2}, \cdots, y_{c,n_c}. \end{array} \right\}. \quad (1)$$

where $y_{i,l} \in R^m$, $i = 1, 2, \ldots, c$; $l = 1, 2, \ldots, n_i$; $n = \sum_{i=1}^{c} n_i$; $\{y_{i,1}, y_{i,2}, \ldots, y_{i,n_i}\} \in C_i$; R is for real number; m is the dimension in the vector model; c is the number of classes. Let the centroid(mean) of C_i be as $\overline{y_i} = \frac{1}{n_i} \sum_{l=1}^{n_i} y_{i,l}$ and the standard deviation of C_i be S_i as following:

$$S_i = \sqrt{\frac{1}{n_i} \sum_{l=1}^{n_i} (y_{i,l} - \overline{y_i})^T (y_{i,l} - \overline{y_i})}. \quad (2)$$

Let $d_{i,m}(d_{i,M})$ be the minimum(maximum) of the distances from all instances in C_i to the centroid $\overline{y_i}$ as shown in equation 3(4)

$$d_{i,m} = arg \;\; min_{l=1,2,\ldots,n_i} \; d(y_{i,l}, \overline{y_i}). \quad (3)$$

$$d_{i,M} = arg \;\; max_{l=1,2,\ldots,n_i} \; d(y_{i,l}, \overline{y_i}). \quad (4)$$

Let $P_i(y_{j,l})$ present the ambiguous degree of the lth instance $y_{j,l}$ of class C_j relative to class C_i. The value of $P_i(y_{j,l})$ was defined as equation 5 where $D_i = max\{5S_i, d_{i,M}\}$. The D_i was used to exclude the instances $y_{j,l}$ in class C_j that were far away from the centroid $\overline{y_i}$ of class C_i. Note that $0 \leq P_i(y_{j,l}) \leq 1$.

$$P_i(y_{j,l}) = \left\{ \begin{array}{ccll} 0 & if & d(y_{j,l}, \overline{y_i}) & > D_i \\ \sqrt{\frac{D_i - d(y_{j,l}, \overline{y_i})}{D_i - d_{i,m}}} & if & d_{i,m} \leq d(y_{j,l}, \overline{y_i}) & \leq D_i \\ 1 & if & d(y_{j,l}, \overline{y_i}) & < d_{i,m} \end{array} \right. \quad (5)$$

The $U_i(y_{j,l})$, as equation 6, was used to filter out the instances in class C_j that were closer to the centroid of class C_i than that of class C_j.

$$U_i(y_{j,l}) = \left\{ \begin{array}{ll} 1 & d(y_{j,l}, \overline{y_i}) \leq d(y_{j,l}, \overline{y_j}) \\ 0 & d(y_{j,l}, \overline{y_i}) > d(y_{j,l}, \overline{y_j}) \end{array} \right. \quad (6)$$

2.2 Class Structure Ambiguity (CSA)

To provide a measure how ambiguous one class (C_j) relative to another class (C_i) is, the *Ambiguity Ratio* $AR(C_i/C_j)$ of class C_j relative to class C_i is defined as follows:

$$AR(C_i/C_j) = \frac{CA(C_i/C_j)}{CA(C_j/C_j)}. \quad (7)$$

where $CA(C_i/C_j) = \frac{\sum_{l=1}^{n_j} U_i(y_{j,l}) * P_i(y_{j,l})}{n_j}$. Intuitively, the $CA(C_i/C_j)$ provide the ambiguous degree of class C_j to class C_i. That is, the more instances in class C_j closer to the centroid of class C_i than that of class C_j are, the more ambiguous of class C_j to class C_i is. Note that the relation of CA is not symmetric[16].

The *Class Ambiguity* $CA(C_i)$, as equation 8, was to summing up all the ambiguities of the neighbors of class C_i relative to itself while with the weighting proportional to the ratio of the number of instances n_j over that of the other instances not in class C_i.

$$CA(C_i) = \sum_{i \neq j=1}^{c} \frac{n_j}{n - n_i} * AR(C_i/C_j). \quad (8)$$

To have an overall estimation of the ambiguities of class structure Ψ among classes, the *Class Structure Ambiguity* $CSA(\Psi)$ sums up the values of $CA(C_i), 1 \leq i \leq c$, with the weighting as the ratio of n_i over n. The definition of $CSA(\Psi)$ was given as following:

$$CSA(\Psi) = \sum_{i=1}^{c} \frac{n_i}{n} CA(C_i) \quad (9)$$

.

3 Experimental Results

It is hard to have an objective point to tell whether the quality of class structure is good or not because even domain experts might have different estimations of the quality of class structure. Hence, the assumption was made in this paper that the higher classification accuracy one excellent classifier could achieve, the lower the ambiguous degree of class structure was.

To estimate the effectiveness of the class structure ambiguity (CSA) proposed in this paper, the Pearsons correlation [1, 3] between the values of classification accuracy achieved by linear SVM classifier derived from the LIBSVM [4] and the values of CSA was computed. Pearson's Correlation Coefficient can take on the values from -1.0 to 1.0. Where -1.0 is a perfect negative (inverse) correlation, 0.0 is no correlation, and 1.0 is a perfect positive correlation [1]. Note that the SVM classifier was known as an excellent classifier with proper training parameters [8, 10, 11].

There were two types of resources for experiments in this paper. One type of resources consisted of random-generated datasets and another consisted of the datasets selected from the LIBSVM [4]. The details of experimental results were given in Section 3.1 and Section 3.2, respectively.

3.1 Resources From Randomly Generation

First of all, to verify the probability of the above assumption, there were datasets generated in terms of different degree of class ambiguity. Each of these datasets contained only two classes with randomly generated instances as 2-dimension vectors with normal distribution[3] in each dimension, and the centroids of that two classes were M distance apart in order to simulate the degree of the ambiguity via the M. That is, the less value of the M was, the higher degree of class ambiguity was.

There were n instances generated randomly for two classes in 2-dimension vector space, X-axis and Y-axis, as normal distribution where $S_1 = S_2 = 1$, and the centroids of C_1 and C_2 were M distance apart. Intuitively, the degree of class ambiguity between C_1 and C_2 increased when the value of M decreased. As shown in Fig.1 and Fig.2, there were instance distributions for $M = 3$ and $M = 5$ and the instances were marked as "$*$" for C_1 or "$+$" for C_1, respectively. It was observable that the degree of class ambiguity when $M = 3$ as in Fig.2 was higher than that when $M = 5$ as in Fig.1.

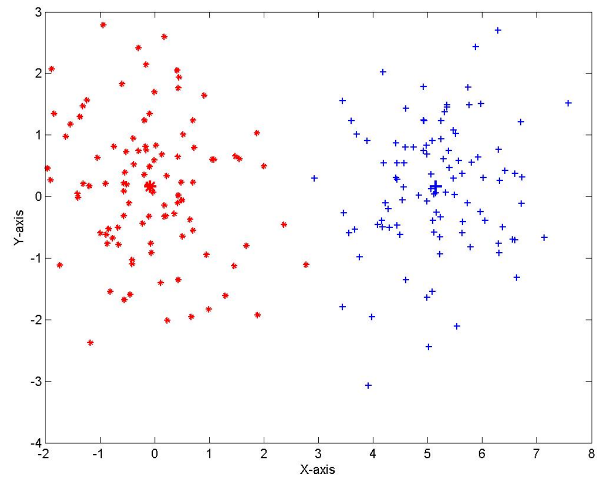

Figure 1. An example for the distribution of $C_1(*)$ and $C_2(+)$ when $M = 5$.

For a given constant M, there were 10 datasets generated in which each class contained 100 instances ($n = 100$). The value of accuracy and the value of CSA were based on the average of these 10 datasets. As shown in Table 1, the value of Pearson's Correlation Coefficient were about -0.99 according to the values of accuracy and that of CSA while the value of M ranged from 1 to 6 with an increasing step as 0.25. This result told that there was almost a perfect negative linear relationship between the values of accuracy and that of CSA when $c = 2$ and $m = 2$, as shown in Fig. 3. According to the previous assumption that the higher classification accuracy one excellent classifier could achieve, the lower the ambiguous degree of class structure was, the value of CSA proposed in this paper did coincide with the degree of the ambiguity of class structure. Note that the accuracy was achieved by inside-testing that used the same dataset for training and testing.

3.2 Resources from the LIBSVM

The statistics of the resources selected from the LIBSVM [4] were shown in Table 2. The resources were selected if both the training and testing set were available and the number of features under 10000 ($m < 10000$) due to the limitation of memory size of our PCs using MATLAB [9] for computing the values of CSA. The scatter diagram of accuracy and CSA was shown in Fig.4 and the value of Pearson's Correlation Coefficient was -0.47 with all resources. However, as shown in Fig.5, the value changed to

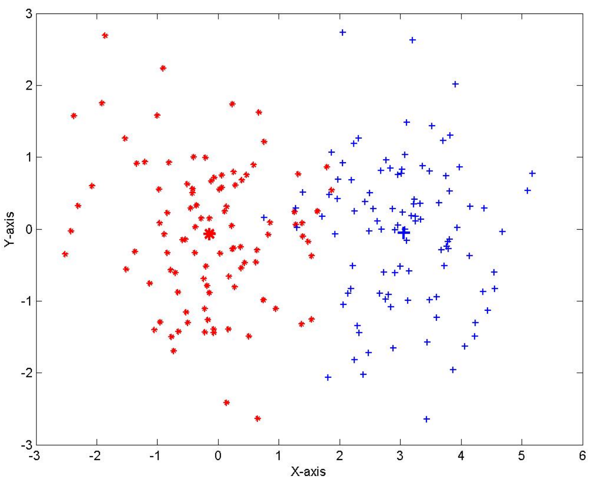

Figure 2. An example for the distribution of $C_1(*)$ and $C_2(+)$ when $M = 3$.

Table 1. The statistics of Accuracy and CSA with random-generated Resources.

	M	Accuracy(%)	CSA
1	1.00	70	0.30679
2	1.25	74.8	0.2663205
3	1.50	77.25	0.2347083
4	1.75	82.05	0.1869931
5	2.00	84.25	0.1622219
6	2.25	88.5	0.1192407
7	2.50	89.4	0.1099905
8	2.75	92.1	0.0796791
9	3.00	93.9	0.0627044
10	3.25	95.25	0.0508211
11	3.50	95.8	0.0425605
12	3.75	97.25	0.0281125
13	4.00	97.75	0.0216393
14	4.25	98.5	0.0146757
15	4.50	98.6	0.0126738
16	4.75	99.3	0.007593
17	5.00	99.5	0.0068342
18	5.25	99.6	0.0051605
19	5.50	99.9	0.0017158
20	5.75	99.85	0.0016424
21	6.00	99.9	0.0017002
	Correlation=	-0.9998177	

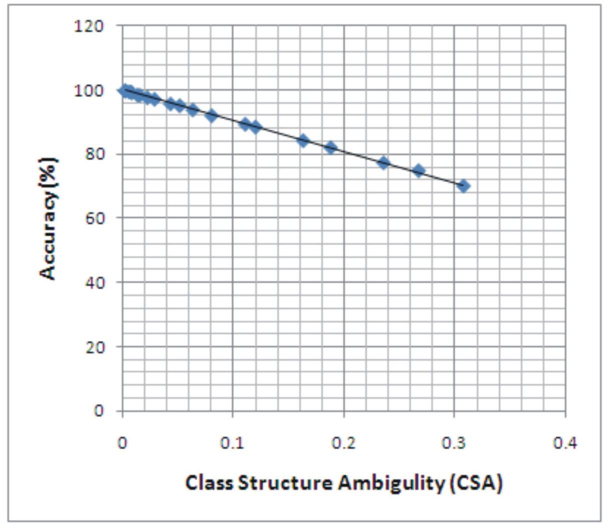

Figure 3. The scatter diagram of Accuracy and CSA with random-generated resources.

-0.78 if the values from two datasets, named as "vowel" and "letter", were excluded from the evaluation of Pearson's Correlation Coefficient. Generally speaking, the CSA could reveal the degree of class ambiguity according to the real datasets from the LIBSVM. Note that the range of the number of classes c was from 2 to 26 and that of the dimension m was from 4 to 780 with resources from the LIBSVM while it was fixed as $c = 2$ and $m = 2$ with resources randomly gererated in Section 3.1.

According to above experimental results, the effectiveness of Class Structure Ambiguity (CSA) with real-life datasets seemed not as significant as that with datasets generated randomly in Section 3.1. It was because the sizes of training and testing instances seemed too small to have a robust estimation, hence, the distribution of the instances for these datasets were quite sparse, especially when the dimension was high. Considering the dataset "vowel", for example, the number c of classes is 11 and the dimension m is 10 while the dataset "vowel" contained only 528 instances for training and 462 instance for testing. On the other hand, the dataset "letter", $c = 26$ and $m = 16$, contained $15,000$ instances for training and 5000 ones for testing while achieving the value of CSA as small as 0.078419 but the value of accuracy as low as 69.88%, which was supposed to be higher than that.

4. Conclusions and Discussions

In this paper Class Structure Ambiguity(CSA) was proposed and evaluated by inspecting the Pearsons correlation between classification accuracy achieved by linear SVM

Table 2. The statistics of the resources from the LIBSVM[4].

name	class (= c)	training size	testing size	feature (=m)	accuracy(%)	CSA
a1a	2	1,605	30956	123	83.8674	0.204466
a2a	2	2,265	30296	123	83.1386	0.214995
a3a	2	3,185	29376	123	84.232	0.201449
a4a	2	4,781	27780	123	84.4276	0.196471
a5a	2	6,414	26147	123	84.4724	0.200164
a6a	2	11,220	21341	123	84.8273	0.200804
a7a	2	16,100	16461	123	84.9766	0.19771
a8a	2	22,696	9865	123	85.4232	0.199223
a9a	2	32,561	16281	123	84.9456	0.199804
ijcnn1	2	49,990	91701	22	91.7973	0.248187
splice	2	1,000	2175	60	85.2414	0.205029
svmguide1	2	3,089	4000	4	94.4	0.08438
svmguide3	2	1,243	41	21	46.3415	0.368848
w1a	2	2,477	47272	300	97.7471	0.09953
w2a	2	3,470	46279	300	98.0618	0.113519
w3a	2	4,912	44837	300	98.2871	0.11469
w4a	2	7,366	42383	300	98.3838	0.110344
w5a	2	9,888	39861	300	98.3919	0.114258
w6a	2	17,188	32561	300	98.6395	0.124572
w7a	2	24,692	25057	300	98.679	0.128744
w8a	2	49,749	14951	300	98.669	0.136425
dna	3	2,000	1186	180	94.6037	0.047316
SensIT(acoustic)	3	78,823	19705	50	67.5565	0.224589
SensIT(combined)	3	78,823	19705	50	80.1827	0.168187
SensIT(seismic)	3	78,823	19705	100	69.9569	0.248261
shuttle	7	43,500	14500	9	92.2897	0.029769
mnist	10	60,000	10000	780	91.82	0.043286
usps	10	7,291	2007	256	91.5296	0.031424
vowel	11	528	462	10	40.4762	0.101382
letter	26	15,000	5000	16	69.88	0.078419

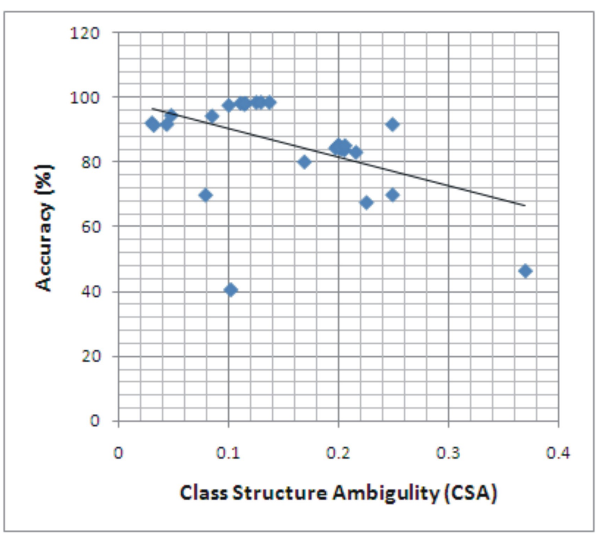

Figure 4. The scatter diagram of Accuracy and CSA for LIBSVM datasets.

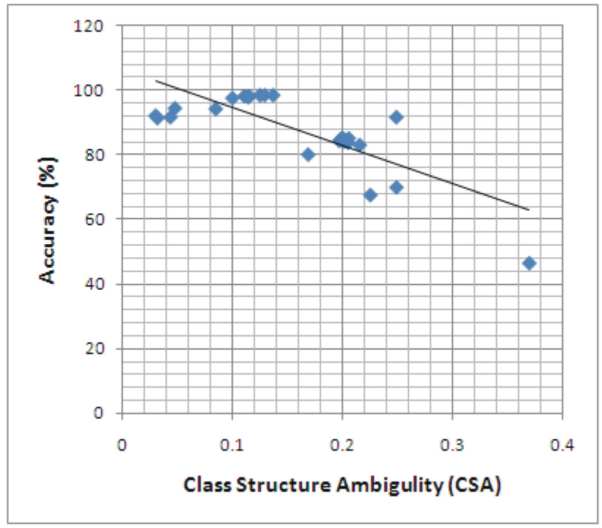

Figure 5. The scatter diagram of Accuracy and CSA for LIBSVM datasets excluding "vowel" and "letter".

and the values of CSA according to experiments with two types of datasets as randomly-generated and selected from LIBSVM[4]. The experimental results showed that the correlations described above were a negative linear relationship between accuracy and CSA with both of two types datasets. That is, the higher classification accuracies were, the lower the values of CSA were. The observations told that the evaluation of CSA proposed in this paper did reveal the ambiguities among classes. To our knowledge, we are the first to address the problem of class ambiguity for classification problems although there were studies to discuss the quality (purity) of clusters [23].

There are still many works for further study. First of all, it is too optimistic to assume that the distribution of the instances in one class in high dimension vector space is as normal distribution. Indeed, it needs a lot of cost to have instances with class-label from the real world. Therefore, the distribution of instances for one class in high dimension could be very sparse and should not suppose to be normal distribution. On the other hand, how to decide the suitable number of instances to have the statistic as normal distribution for a given m dimension is hard to predict. Secondly, the classifier used in this paper was linear SVM. Therefore, what the measures should be with different classifiers is unknown. Thirdly, the class structure ambiguity discussed in this paper was only based on the relationship between two classes. It might be more reasonable to take all the instances of the neighboring classes into consideration but not just two classes. Finally, it is desirable to estimate the degree of class structure ambiguity not only for flat class structure but for hierarchical one. It is our future works to tackle these problems as described above.

Acknowledgements

This work was partially supported by the research project of Asia University under grant $97 - I - 07$. Thanks for the reviewer's valuable suggestions.

References

[1] Statistics Tutorial. http://www.texasoft.com/winkpear.html.

[2] Ethem Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2004.

[3] M.G. Bulmer. *Principles of Statistics*. Dover Publications, Inc., 180 Varick Street, NewYork, N.Y. 10014, 2 edition, 1979.

[4] Chih-Chung Chang and Chih-Jen Lin. Libsvm : a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm., 2001.

[5] Shui-Lung Chuang and Lee-Feng Chien. Taxonomy generation for text segments: A practical web-based approach. *ACM Trans. Inf. Syst.*, 23(4):363–396, 2005.

[6] Bin Gao, Guang Feng, Tao Qin, and Qian-Sheng Cheng. Hierarchical taxonomy preparation for text categorization using consistent bipartite spectral graph copartitioning. *IEEE Trans. on Knowl. and Data Eng.*, 17(9):1263–1273, 2005.

[7] Stephen C. Gates, Wilfried Teiken, and Keh-Shin F. Cheng. Taxonomies by the numbers: building high-performance taxonomies. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 568–577. ACM, 2005.

[8] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques, 2nd ed.* Morgan Kaufmann, 2005.

[9] William J. Palm III. *A concise Introduction to MATLAB*. McGRAW Hill, 2008.

[10] Bing Liu. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer, January 2007.

[11] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[12] Tom M. Mitchell. *Machine Learning*. The McGraw-Hill Companies, Inc, 1997.

[13] Kunal Punera and Joydeep Ghosh. CLUMP: A scalable and robust framework for structure discovery. In *ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 757–760. IEEE Computer Society, 2005.

[14] Kunal Punera, Suju Rajan, and Joydeep Ghosh. Automatically learning document taxonomies for hierarchical classification. In *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 1010–1011. ACM, 2005.

[15] Kunal Punera, Suju Rajan, and Joydeep Ghosh. Automatic construction of n-ary tree based taxonomies. In *ICDMW '06: Proceedings of the Sixth IEEE International Conference on Data Mining - Workshops*, pages 75–79. IEEE Computer Society, 2006.

[16] Kenneth A. Ross and Charles R. Wright. *Discrete Mathematics*. Prentice Hall, Englewood Cliffs, NJ, 5th edition edition, 2003.

[17] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition edition, 2003.

[18] Jing-Doo Wang. Evaluating the ambiguities of class structure via euclidean distance. In *2008 International Computer Symposium (ICS2008)*, volume 2, pages 293–298, 2008.

[19] Ke Wang, Senquiang Zhou, and Shiang Chen Liew. Building hierarchical classifiers using class proximity. In *Proceedings of VLDB-99, 25th International Conference on Very Large Data Bases*, pages 363–374. Morgan Kaufmann Publishers, 1999.

[20] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*. Elsevier, 2005.

[21] Li Zhang, Tao Li, Shixia Liu, and Yue Pan. An integrated system for building enterprise taxonomies. *Inf. Retr.*, 10(4-5):365–391, 2007.

[22] Li Zhang, ShiXia Liu, Yue Pan, and LiPing Yang. Infoanalyzer: a computer-aided tool for building enterprise taxonomies. In *CIKM '04: Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 477–483. ACM, 2004.

[23] Y. Zhao and G. Karypis. Criterion functions for document clustering: Experiments and analysis, 2001.