Expert Systems with Applications 36 (2009) 12547–12553

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Customer churn prediction by hybrid neural networks

Chih-Fong Tsai [a,*], Yu-Hsin Lu [b]

[a] Department of Information Management, National Central University, Taiwan
[b] Department of Accounting and Information Technology, National Chung Cheng University, Taiwan

ARTICLE INFO

Keywords:
Hybrid data mining
Self-organizing maps
Back-propagation neural networks
Churn prediction

ABSTRACT

As churn management is a major task for companies to retain valuable customers, the ability to predict customer churn is necessary. In literature, neural networks have shown their applicability to churn prediction. On the other hand, hybrid data mining techniques by combining two or more techniques have been proved to provide better performances than many single techniques over a number of different domain problems. This paper considers two hybrid models by combining two different neural network techniques for churn prediction, which are back-propagation artificial neural networks (ANN) and self-organizing maps (SOM). The hybrid models are ANN combined with ANN (ANN + ANN) and SOM combined with ANN (SOM + ANN). In particular, the first technique of the two hybrid models performs the data reduction task by filtering out unrepresentative training data. Then, the outputs as representative data are used to create the prediction model based on the second technique. To evaluate the performance of these models, three different kinds of testing sets are considered. They are the general testing set and two fuzzy testing sets based on the filtered out data by the first technique of the two hybrid models, i.e. ANN and SOM, respectively. The experimental results show that the two hybrid models outperform the single neural network baseline model in terms of prediction accuracy and Types I and II errors over the three kinds of testing sets. In addition, the ANN + ANN hybrid model significantly performs better than the SOM + ANN hybrid model and the ANN baseline model.

1. Introduction

Enterprises in the competitive market mainly rely on the profits which come from customers. Therefore, customer relationship management (CRM) always concentrates on confirmed customers that are the most fertile source of data for decision making. This data reflects customers' actual individual behavior. This kind of behavioral data can be used to evaluate customers' potential value (Hung & Tsai, 2008), assess the risk that they will stop paying their bills, and anticipate their future needs (Berry & Linoff, 2003). Besides, because customer churning will likely to result in the loss of businesses, churn prediction has received increasing attention in the marketing and management literature over the past time. In addition, it shows that a small change in the retention rate can result in significant impact on businesses (Van den Poel & Larvie'rе, 2004).

In order to effectively manage customer churn for companies, it is important to build a more effective and accurate customer churn prediction model. In literature, statistical and data mining techniques have been used to create the prediction models.

The data mining task can be used to describe (i.e. discover interesting patterns or relationships in the data), and predict (i.e. predict or classify the behavior of the model based on available data) (Fayyad & Uthurusamy, 1996; Ngai, Xiu, & Chau, 2009). In other words, it is an interdisciplinary field with a general goal of predicting outcomes and employing sophisticated algorithms to discover mainly hidden patterns, associations, anomalies, and/or structure from extensive data stored in data warehouses or other information repositories and filter necessary information from large datasets (Han & Kamber, 2001).

In literature, hybrid data mining models by combining clustering and classification data mining techniques can improve the performance of the single clustering or classification techniques individually. In particular, they are composed of two learning stages, in which the first one is used for 'pre-processing' the data and the second one for the final prediction output (Lenard, Madey, & Alam, 1998).

However, few studies examining the performance of hybrid data mining techniques for customer churn prediction. Therefore, in this paper two different combination methods to create the hybrid models are examined in terms of customer churn prediction. The first one is based on combining clustering, i.e. self-organizing maps (SOM) and classification techniques, i.e. back-propagation artificial neural networks (ANN), which is SOM + ANN, and the

* Corresponding author. Tel.: +886 3 4267250; fax: +886 3 4254604.
E-mail address: cftsai@mgt.ncu.edu.tw (C.-F. Tsai).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.05.032

second one is to combine two classification techniques (ANN), which is ANN + ANN.

The remainder of this paper is organized as follows. In Section 2, we review the literature related to customer churn and data mining techniques used in this paper. Section 3 describes the research methodology, and Section 4 presents the experimental results. Finally, conclusion is provided in Section 5.

2. Literature review

2.1. Customer churn

In the today's competitive market, many companies are coming to a full realization of the importance of the customer-oriented business strategy for sustaining their competitive edge and maintaining a stable profit level. That is, companies mainly rely on the income which comes from customers. However, to create and retain customers is difficult and costly in term of marketing. As new account setup, credit searches, and advertising and promotional expenses can add up to several times the cost of efforts that might enable the firms to retain a customer (Keaveney, 1995), it is becoming an industry-wide belief that the best core marketing strategy for the future is to retain existing customers and avoid customer churn (Kim, Park, & Jeong, 2004; Kim & Yoon, 2004).

Burez and Van den Poel (2007) indicate that there are two types of targeted approaches to managing customer churn: reactive and proactive. When a company adopts a reactive approach, it waits until customers ask the company to cancel their service relationship. In this situation, the company will offer the customer an incentive to stay. On the other hand, when a company adopts a proactive approach, it tries to identify customers who are likely to churn before they do so. The company then provides special programs or incentives for these customers to keep the customers from churning. Targeted proactive programs have potential advantages of having lower incentive costs. However, these systems may be very wasteful if churn predictions are inaccurate, because companies are wasting incentive money on customers who will not churn. Therefore, it is important to build a customer-churn prediction model as accurately as possible (Burez & Van den Poel, 2007; Van den Poel & Larivie're, 2004).

2.2. Data mining techniques

In order to establish effective and accurate customer-churn prediction model, many data mining methods have been recently considered (e.g. Coussement & Van den Poel, 2008; Hung, Yen, & Wang, 2006). The two primary goals of data mining in practice tend to be description and prediction. Description focuses on finding human-interpretable patterns describing the data, and prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest (Fayyad, Piatetsky, & Smyth, 1996; Fayyad & Uthurusamy, 1996). The goals of description and prediction can be achieved using a variety of particular data mining methods include classification, clustering, regression, and so on (Berry & Linoff, 2003).

2.2.1. Artificial neural network

Classification is one of the commonly used data mining methods, and called as supervised learning techniques. It calculates the value of some variables, and classifies according to results. The algorithms of classification include decision trees, artificial neural networks, and so on (Han & Kamber, 2001; Tou & Gonzalez, 1974), in which artificial neural networks are the most widely considered technique in many business problems (Wong, Bodnovich, & Selvi, 1997).

Artificial neural networks (ANN) attempt to simulate biological neural systems which learn by changing the strength of the synaptic connection between neurons upon repeated stimulations by the same impulse (Li & Tan, 2006). Neural networks can be distinguished into single-layer perception and multilayer perception (MLP). The multilayer perception consists of multiple layers of simple, two taste, sigmoid processing nodes or neurons that interact by using weighted connections. In addition, the neural network contains one or more several intermediary layers between the input and output layers. Such intermediary layers are called hidden layers and nodes embedded in these layers are called hidden nodes shown in Fig. 1. Based on prior research results (e.g. Cybenco, 1998; Hung et al., 2006; Zhan, Patuwo, & Hu, 1998), multilayer perception is a relatively accurate neural network model.

2.2.2. Self-organizing maps

Clustering is an unsupervised learning method that partitions a set of patterns into groups (or clusters). Cluster analysis refers to the grouping of a set of data object into clusters. In particular, no predefined classes are assigned (Jain, Murty, & Flyn, 1999).

Kohonen (1987) proposed and demonstrated a new form of a neural network architecture called self-organizing map (SOM), which has proved extremely useful when the input data are high dimensionality and complexity. SOM is used to discover associations in a dataset and cluster data according to the similarity of data (i.e., similar expression patterns) where the model creators cannot predict the nature of the classification, or they consider that there may be more than one way to categorize the characteristics

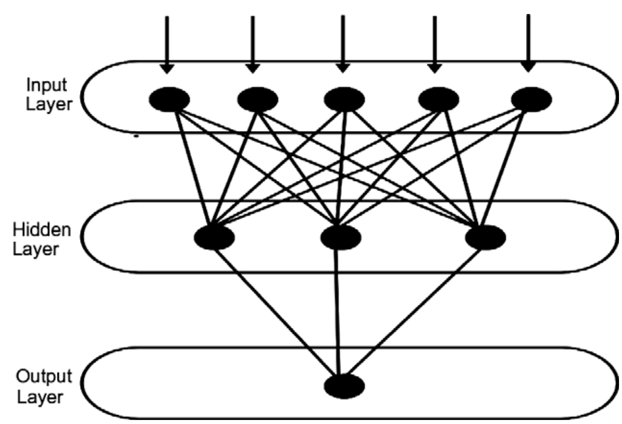

Fig. 1. Multilayer neural network.

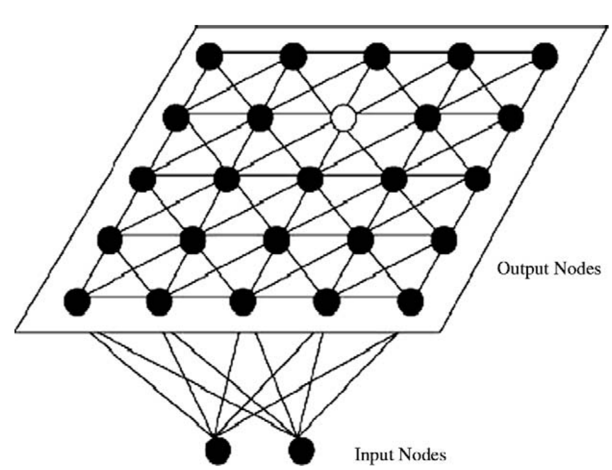

Fig. 2. A 4 × 4 Kohonen's self-organizing map.

of a dataset (Zhang, Edwards, & Harding, 2007). Fig. 2 shows an example of a 4×4 SOM.

2.2.3. Hybrid approaches

In general, hybrid models are based on combining two or more data mining techniques. For example, the clustering and classification techniques can be serially combined. That is, clustering can be used as a pre-processing stage to identify pattern classes for subsequent supervised prediction or classification (Jain et al., 1999). Therefore, the clustering result can be used for either pre-classification of unlabelled collections or identifying major populations of a given dataset.

Then, the clustering result becomes the training set to train and create a prediction model. After the model is created, it is able to classify or predict new (unknown) instances. In other words, the first component of the hybrid models can simply perform the task of outlier detection and the second for prediction.

Therefore, there are two possible combination methods for hybrid models (Lenard et al., 1998) as shown in Figs. 3 and 4, respectively.

As the churn prediction task is based on supervised classification techniques, the second component or prediction model of the above mentioned two hybrid approaches are all based on classification techniques. Therefore, two hybrid models considered in this paper are the classification + classification and cluster + classification hybrid models (c.f. Section 3.2).

2.3. Related work

Building an effective customer churn prediction model has become an important topic for business and academics in recent years. In order to understand how related work constructs their prediction models, this papers reviews some of current related studies shown in Table 1.

Coussement and Van den Poel (2008) applies support vector machines in a newspaper subscription context in order to construct a churn model. The predictive performance of the support vector machine model is benchmarked to logistic regression and random forests.

Burez and Van den Poel (2007) built a prediction model for European pay-TV company by using Markov chains and a random forest model benchmarked to a basic logistic model. Other studies (e.g. Kim & Yoon, 2004) only use a binomial logit regression to construct the prediction model.

According to prior literature about customer churn, much related work focus on using only one data mining method such as classification or clustering to mine the customer retention data. Few studies (Hung et al., 2006) apply more than one technology which are cluster analysis and classification.

In this paper, we consider to develop hybrid data mining models by combining clustering and classification data mining techniques which may outperform one single methods. In particular, in the data mining process, the first clustering or classification technique is used during the pre-processing stage for data reduction, which aims at filtering out unrepresentative training data.

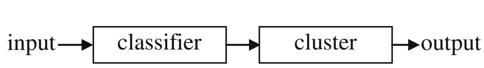

Fig. 3. A classifier combined with a cluster.

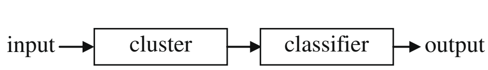

Fig. 4. A cluster combined with a classifier.

Besides, we also contemplate using cross-validation which can improve the performance and reflect the 'true' performance when validating the created models for unseen data that the validation set stays untouched (Coussement & Van den Poel, 2008).

As ANN and SOM are the two mostly used neural network techniques, there are two approaches to develop the hybrid prediction model for customer churn, which are SOM + ANN and ANN + ANN. The baseline model for comparisons is based on the single ANN model.

3. Research methodology

3.1. Dataset

For the purpose of this paper, we consider a CRM dataset [1] provided by American telecom companies, which focuses on the task of customer churn prediction. Specifically, the dataset contains 51,306 subscribers, including 34,761 churners and 16,545 non-churners, from July 2001 to January 2002. In addition, the subscribers have to be mature customers who were with the telecom company for at least six months. Churn was then calculated based on whether the subscriber left the company during the period 31–60 days after the subscriber was originally sampled.

Besides, this paper considers five-fold cross-validation. It is based on dividing five equal parts of a dataset, in which 80% of the dataset performs model training, and the other for model testing. Therefore, every subset will be trained and tested five times, and the average prediction performance can be obtained consequently.

3.2. Model development

3.2.1. The baseline

At first, we use the original dataset to train a MLP neural network as the baseline ANN model for comparisons, which is similar to related work, such as Buckinx and Van den Poel (2005) and Hung et al. (2006).

In addition, four different learning epochs (50, 100, 200, and 300) and five different hidden layer nodes (8, 12, 16, 24, and 32) are used in order to obtain the best ANN baseline model. Table 2 shows the settings of the learning epochs and numbers of hidden layer nodes. As a result, there are twenty different ANN models developed for comparisons.

3.2.2. ANN + ANN

The first hybrid model is based on cascading two ANN models, in which the first one performs the data reduction task and the second one for churn prediction. That is, the original training set is used to 'test' the fist created ANN model, which is based on the 'best' baseline model identified above. As there is no 100% accuracy, there are a number of correctly and incorrectly predicted data from the training set by the first ANN model. Consequently, the incorrectly predicted data can be regarded as outliers since the ANN model cannot predict them accurately. Then, the correctly predicted data by the first ANN model are used to train the second ANN model as the prediction model for later prediction. Fig. 5 shows the process of combing two ANN models.

3.2.3. SOM + ANN

For the second hybrid model, a self-organizing map (SOM), which is a clustering technique, is used for the data reduction task. Then, the clustering result is used to train the second model based on ANN.

[1] http://www.fuqua.duke.edu/centers/ccrm/datasets/download.html#data.

Table 1
Related literature about customer churn.

Author	Data set	Prediction method	Preprocessing	Cross-validation
Coussement and Van den Poel (2008)	Subscriber database	Support vector machines random forests logistic regression	X	V
Burez and Van den Poel (2007)	Pay-TV company	Logistic regression Markov chains random forests	X	X
Hung et al. (2006)	Wireless telecom. company	Classification (decision tree, neural network) clustering (K-means)	V	V
Buckinx and Van den Poel (2005)	Retailing dataset	Neural networks, logistic regression	X	X
Van den Poel and Larivie're (2004)	European financial services company	Hazard model survival analysis	X	X
Kim and Yoon (2004)	Five mobile carriers in Korea	Logistic regression	X	X
Chiang et al. (2003)	Network banking	Association rules	V	X
Wei and Chiu (2002)	Taiwan wireless telecommunications company	Classification (decision tree)	X	V

To develop the SOM, the map size is set by 2 * 2, 3 * 3, 4 * 4, and 5 * 5 respectively in order to obtain the best clustering result, i.e. the highest rate of prediction accuracy. That is, two clusters of SOM which contain the highest proportion of the churner and non-churner groups respectively are selected as the clustering result. For the example of 2 * 2 SOM, there are four clusters generated, in which one of the four clusters contains the highest proportion of the churner group and another for the non-churner group. Fig. 6 shows the process of combining SOM and ANN.

3.3. Evaluation methods

To evaluate the above developed models, prediction accuracy and the Type I and II errors are considered. They can be measured by a confusion matrix shown in Table 3. The rate of prediction accuracy is defined as $\frac{a+d}{a+b+c+d}$.

The Type I error is the error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature. In this paper, it means that the event was occurred when the model classified the event occurred group into the non-event occurred group. On the other hand, the Type II error is defined as the error of rejecting a null hypothesis when it is the true state of nature. It means that the event was occurred when the model classified the non-event occurred group into the event occurred group.

In addition to using the general testing set to assess the prediction performances of these models, we also consider the fuzzy testing data (FTD) to examine the rate of predication accuracy and the Type I and II errors. The fuzzy testing data is based on the filtered out data by the first model of the two hybrid models. Therefore, there are two different kinds of fuzzy testing data. The first one comes from the incorrectly predicted data of the ANN model by

Table 2
Parameters settings.

Hidden nodes	Learning epochs			
	50	100	200	300
8	X1	X6	X11	X16
12	X2	X7	X12	X17
16	X3	X8	X13	X18
24	X4	X9	X14	X19
32	X5	X10	X15	X20

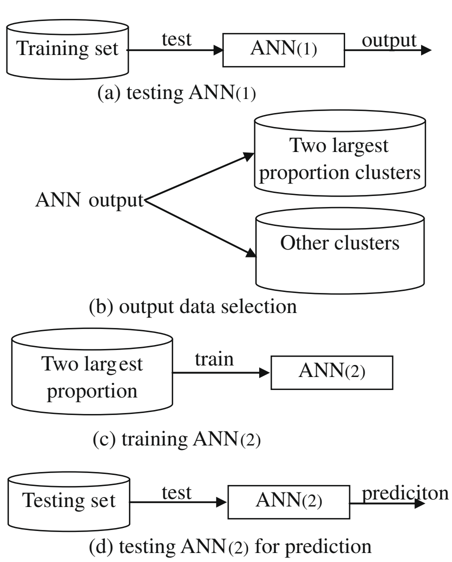

Fig. 5. The process of combining two ANN models.

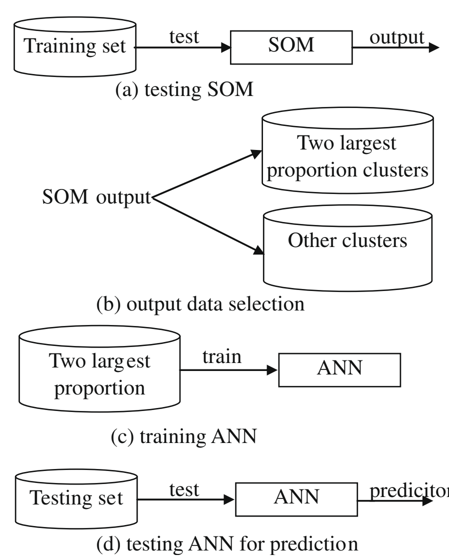

Fig. 6. The process of combining SOM and ANN.

Table 3
Confusion matrix.

		Actual	
		Non-churners	Churners
Predict	Non-churners	a	b (II)
	Churners	c (I)	d

Table 4
Prediction performances of the MLP baseline models.

Learning epochs	50					100				
Hidden nodes	8	12	16	24	32	8	12	**16**	24	32
CRM										
Subset 1	92.56	92.58	92.55	91.06	**92.80**	92.60	92.60	92.55	92.41	92.68
Subset 2	90.15	90.17	90.28	90.02	90.30	91.03	90.59	90.24	90.41	90.22
Subset 3	**91.81**	91.38	91.40	90.34	91.37	91.45	91.45	91.26	91.59	91.72
Subset 4	87.43	87.20	87.34	86.99	87.38	86.45	87.11	**88.08**	87.41	86.43
Subset 5	86.96	**86.99**	86.37	86.42	86.38	86.48	86.48	85.18	86.60	86.24
Avg	89.78	89.66	89.59	88.97	89.65	89.60	89.65	89.46	89.68	89.46
Learning epochs	**200**					**300**				
Subset 1	92.53	92.55	92.00	92.62	92.49	92.43	92.55	92.10	92.30	92.55
Subset 2	92.19	92.56	92.37	92.35	92.16	92.20	92.43	92.31	92.34	**92.75**
Subset 3	91.22	91.73	91.70	91.63	90.94	91.45	91.71	91.65	90.90	91.33
Subset 4	86.24	87.47	86.89	87.71	87.81	87.28	86.76	87.70	88.07	87.33
Subset 5	85.45	86.71	86.61	86.66	86.90	86.01	86.83	86.71	86.91	85.99
Avg	89.53	90.20	89.91	90.19	90.06	89.87	90.06	90.09	90.10	89.99

ANN + ANN, namely FTD (ANN). The second one comes from the clustered data of SOM, namely FTD (SOM). These fuzzy testing data can be thought of as ‘real world data’, which are much more challenging and more difficult to be predicted for the models if compared with the general testing set. This is because these data are difficulty to be recognized by either ANN or SOM.

4. Experimental results

4.1. The baseline

Table 4 shows the prediction performance of the baseline ANN models based on five-fold cross-validation and twenty different parameter settings. On average, the baseline ANN models provide higher than 88% accuracy. In particular, the best parameter setting of the ANN model from each of the five subsets is chosen for later comparisons summarized in Table 5. In addition, they are used to develop the hybrid models.

4.2. ANN + ANN

For the hybrid model based on combining two ANN models, the first ANN model performs the data reduction task. Therefore, we consider the ANN model by the subset 1 since it performs the best over the others. As a result, correctly predicted data by the ANN model are collected. There are 37,454 subscribers, including 26,105 churners and 11,349 non-churners.

Table 5
The best parameter setting of the MLP baseline model.

Testing set	Learning epochs	Hidden nodes	Accuracy
Subset 1	50	32	92.80
Subset 2	300	32	92.75
Subset 3	50	8	91.81
Subset 4	100	16	88.08
Subset 5	50	12	86.99

Table 6
Prediction performance of ANN + ANN hybrid models.

Testing set	Learning epochs	Hidden nodes	Accuracy
Subset 1	50	32	94.32
Subset 2	300	32	93.70
Subset 3	50	8	93.68
Subset 4	100	16	93.71
Subset 5	50	12	90.18

Consequently, the collected data are used to train the second ANN model by using the best parameter settings from each of the five subsets. Table 6 shows the prediction accuracy of the ANN + ANN hybrid model. The result indicates that the ANN + ANN hybrid models perform better than the baseline ANN models over the five subsets. In other words, combining two ANN models outperforms single ANN models.

4.3. SOM + ANN

To construct the second hybrid model by combining SOM with ANN, 2 * 2, 3 * 3, 4 * 4, and 5 * 5 SOMs are used to cluster the data at first. We found that 3 * 3 SOM performs the best which can provide the highest rate of accuracy for two clusters, i.e. churner and non-churner groups. For these two clusters, there are 18,668 subscribers, including 8,438 churners and 10,230 non-churners.

Table 7
Prediction performance of SOM + ANN hybrid models.

Testing set	Learning epochs	Hidden nodes	The test accuracy
Subset 1	50	32	93.06
Subset 2	300	32	92.32
Subset 3	50	8	92.48
Subset 4	100	16	91.52
Subset 5	50	12	88.86

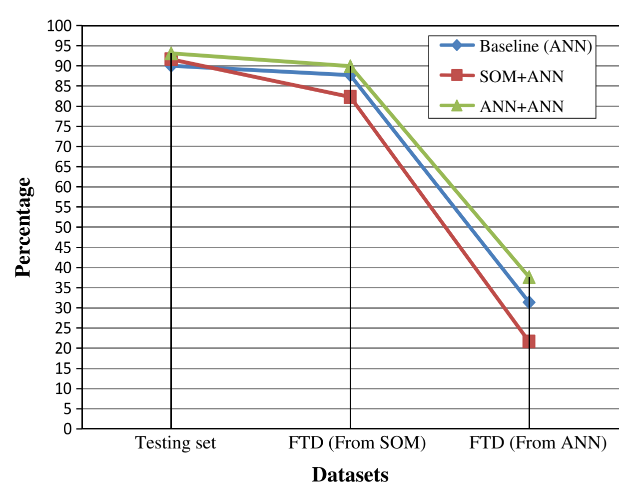

Fig. 7. Prediction accuracy of the testing and FTD sets.

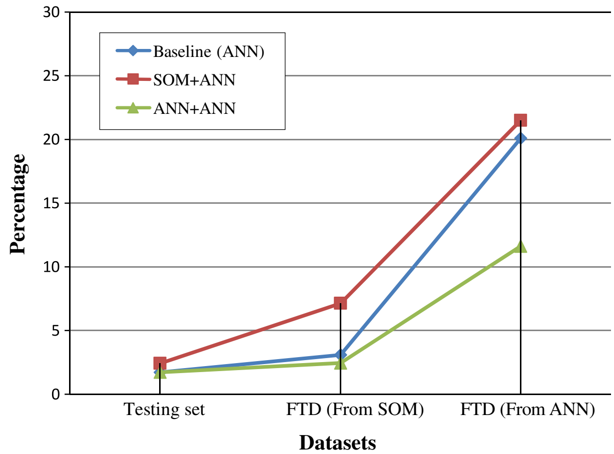

Fig. 8. The Type I error by the testing and FTD sets.

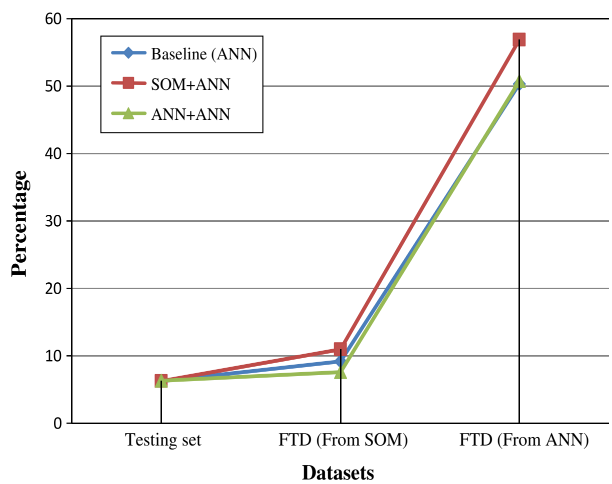

Fig. 9. The Type II error by the testing and FTD sets.

Then, these collected data are used to train the best ANN models from each of the five subsets. Table 7 shows the prediction performance of the SOM + ANN hybrid model. The result indicates that the accuracy rates of the SOM + ANN hybrid models are higher than the ANN baseline models. Therefore, hybrid models by combining clustering and classification neural network techniques outperform single ANN models.

4.4. Model evaluation by fuzzy testing data (FTD)

In this paper, we also consider using the fuzzy testing data (FTD) based on the outliers identified by ANN and SOM respectively to examine the rate of predication accuracy and Type I and II errors of the baseline ANN and hybrid models.

4.4.1. Prediction accuracy

Fig. 7 shows the rate of prediction accuracy of the baseline and two hybrid models over the general testing set and two different FTD sets. As we can see, using FTD sets decrease the prediction performances. This is because these fuzzy testing data are based on the outliers. In addition, the FTD set identified by ANN is much more difficulty to be predicted. Particularly, the ANN + ANN hybrid model performs better and decrease more slightly than the other two models. It is interesting that the SOM + ANN hybrid model does not perform better than the baseline ANN model over the two FTD sets.

4.4.2. Type I and II errors

In addition to prediction accuracy, the Type I and II errors of these models over the testing and FTD sets are also examined shown in Figs. 8 and 9, respectively. The results indicate that using FTD sets increase the error rates of these models. In addition, we can see that the ANN + ANN hybrid model performs the best which provides the lowest error rates. Similar to prediction accuracy, the SOM + ANN hybrid model does not perform better than the baseline ANN model.

For the managerial and marketing purposes, the prediction model should not only provide as higher accuracy as possible, but also lower the Type I error rate since it can retain more valuable customers. Therefore, the result in Fig. 8 suggests that the ANN + ANN hybrid model can provide better prediction accuracy and lower Type I errors, which is a better model for churn prediction.

4.4.3. Paired t-test

Finally, this paper applies the paired *t* test to examine if there are any significant differences between these models in terms of prediction accuracy and the Type I and II errors. As shown in Table 8, the ANN + ANN hybrid model has a significant level of difference between the SOM + ANN hybrid model and single ANN model ($p < 0.05$) by prediction accuracy and Type I and II errors. Therefore, we can conclude that the ANN + ANN hybrid model significantly performs better than the SOM + ANN hybrid model and ANN baseline model.

5. Conclusion

Churn prediction and management is very important for enterprises in the competitive market to predict possible churners and take proactive actions to retain valuable customers and profit. Therefore, to build an effective customer churn prediction model, which provides a certain level of accuracy, has become a research problem for both academics and practitioners in recent years.

In this paper, we consider two different hybrid data mining techniques by neural networks to examine their performances

Table 8
The *t*-test result.

Type of model	ANN	*t*-value	ANN + ANN	*t*-value	SOM + ANN	*t*-value
Panel A: Prediction accuracy						
ANN			0.013***	−4.254	0.045**	−2.876
ANN + ANN					0.001***	−8.056
Panel B: Type I error						
ANN			0.033**	−5.350	0.209	−1.830
ANN + ANN					0.028**	−5.871
Panel C: Type II error						
ANN			0.002***	−21.757	0.095*	−3.009
ANN + ANN					0.001***	−31.237

** represents a high level of significant difference.
*** represents a very high level of significant difference.

for telecom churn prediction. In particular, back-propagation artificial neural networks (ANN) and self-organizing maps (SOM) are considered. Consequently, ANN + ANN and SOM + ANN hybrid models are developed, in which the first component of the hybrid models aims at filtering out unrepresentative data or outliers. Then, the representative data as the outputs are used to create the prediction model. To assess the prediction performance of these models, this paper considers not only the general testing set, but also two kinds of fuzzy testing sets based on the outliers identified by ANN and SOM, respectively, i.e. the first technique of the hybrid models.

The experimental results indicate that the hybrid models outperform the single neural network baseline model in term of prediction accuracy and the Type I and II errors. In particular, the ANN + ANN hybrid model performs the best. However, when the two fuzzy testing sets are used, the SOM + ANN hybrid model does not perform better than the baseline ANN model. Therefore, we can conclude that the hybrid model by combining two ANN techniques can perform better than the baseline model and the hybrid model by combining SOM and ANN. In addition, the ANN + ANN hybrid model performs more stably than the other two models.

For future work, several issues can be considered. First, as the pre-processing stage in data mining is a very important step for the final prediction performances, the dimensionality reduction or feature selection step can be involved in addition to data reduction. Second, besides neural networks, other prediction techniques can be applied, such as support vector machines, genetic algorithms, etc. Finally, other domain datasets about churn prediction can be used for further comparisons.

References

Berry, M. J. A., & Linoff, G. S. (2003). *Data mining techniques: For marketing, sales, and customer support*. John Wiley & Sons.

Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research, 164*(1), 252–268.

Burez, J., & Van den Poel, D. (2007). Crm at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications, 32*, 277–288.

Chiang, D. A., Wang, Y. F., Lee, S. L., & Lin, C. J. (2003). Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications, 25*, 293–302.

Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications, 34*, 313–327.

Cybenco, H. (1998). Approximation by super-positions of sigmoidal function. *Mathematical of Control and Signal Systems, 2*, 303–314.

Fayyad, U., & Uthurusamy, R. (1996). Data mining and knowledge discovery in databases. *Communications of the ACM, 39*, 24–27.

Fayyad, U., Piatetsky, S. G., & Smyth, P. (1996). *From data mining to knowledge discovery: An overview in advances in knowledge discovery and data mining*. AAAI/MIT Press.

Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

Hung, C., & Tsai, C.-F. (2008). Segmentation based on hierarchical self-organizing map for markets of multimedia on demand. *Expert Systems with Applications, 34*(1), 780–787.

Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications, 31*, 515–524.

Jain, A., Murty, M., & Flyn, P. (1999). Data clustering: A review. *ACM Computing Surveys, 31*, 264–323.

Keaveney, S. M. (1995). Customer switching behavior in service industries: An exploratory study. *Journal of Marketing, 59*, 71–82.

Kim, H. S., & Yoon, C. H. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy, 28*, 751–765.

Kim, M., Park, M., & Jeong, D. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications Policy, 28*, 145–159.

Kohonen, T. (1987). Adaptive, associative, and self-organizing functions in neural computing. *Applied Optics, 26*(23), 4910–4918.

Lenard, M. J., Madey, G. R., & Alam, P. (1998). The design and validation of a hybrid information system for the auditor's going concern decision. *Journal of Management Information Systems, 14*(4), 219–237.

Li, C. T., & Tan, Y. H. (2006). Adaptive control of system with hysteresis using neural networks. *Journal of Systems Engineering and Electronics, 17*, 163–167.

Ngai, E. W. T, Xiu, L., & Chau, D. C. K. 2009. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, *36*(2), 2592–2602.

Tou, J. T., & Gonzalez, R. C. (1974). *Pattern recognition principles*. Addison-Wesley.

Van den Poel, D., & Larivie're, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research, 157*, 196–217.

Wei, C. P., & Chiu, I. T. (2002). Turning telecommunications call details to churn prediction: A data mining approach. *Expert Systems with Applications, 23*, 103–112.

Wong, B. K., Bodnovich, T. A., & Selvi, Y. (1997). Neural network applications in business: a review and analysis of the literature (1988–1995). *Decision Support Systems, 19*, 301–320.

Zhan, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural network: The state of the art. *International Journal of Forecasting, 14*, 35–62.

Zhang, X., Edwards, J., & Harding, J. (2007). Personalised online sales using web usage data mining. *Computers in Industry, 58*, 772–782.