# Ensembling Classifier Based on Context and AHP and Its Application in Customer Churning Management of Phone Company

Yibing Chen[1,2], Yuhong Guo[3], Lingling Zhang[1,2,*] and Yong Shi[2,4]
[1]School of Management, Graduate University of Chinese Academy of Sciences, Beijing, China.
[2]Research Centre on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, China.
[3]Department of Information Science and Technology, University of International Relations, Beijing, China.
[4]College of Information Science and Technology, University of Nebraska at Omaha, Omaha, USA

chenchen2304@163.com, yhguo@uir.cn, zhangll@gucas.ac.cn, yshi@gucas.ac.cn
*Corresponding author: Phn +86-13661125028,E-mail:zll933@163.com, zhangll@gucas.ac.cn

**Keywords**: ensembling classifier; AHP (analytic hierarchy process); context knowledge; customer churning

**Abstract**. Classification is an important task in data mining. How to ensemble individual classifiers to improve classification performance has great significance. Previous ensembling classifier models are coarse and lack scientific quantitative direction on weights computation, leading low accuracy and bad explanation. Further, previous researches care only about the single index of overall classification accuracy, ignoring other vital indexes in certain application context. This paper proposes a new ensembling classifier based on context and AHP (Analytic hierarchy process). It uses AHP based on context knowledge to scientifically compute the weights of excellent base-classifiers and the weights of indexes in the certain application context. Thus, an ensembling classifier orienting multiple indexes is built. Experiments on an American phone company's customer churning data show that the ensembling classifier based on context and AHP can improve classification accuracy, satisfy preference selection and balance requirements of multiple indexes in certain application context.

## Introduction

Classification is an important task in data mining. It is the progress to build a classifier based on existing massive data, and classify them into different categories according to their attributes. Thus the model can be used to analyze existing data and predict new record belongs to which category. Classification is widely used in various fields including disease discovery, credit card fraud detection and network intrusion detection, etc.

We may consider those common classifiers such as decision tree, Bayesian classifier, neural network, SVM (Support Vector Machine) and regression model as specific map functions. They reflect the instances in the data warehouse to different class labels based on their own methods. Among the classifiers, no one always performs better than the others, and some of them are even weak in some certain tasks. Therefore, constructing an ensemble of classifiers can improve classification performance. It is held that, accuracy of ensembling classifier can be improved if the individual classifiers are precise and diverse, and outperforms any of the individual classifiers (Hansen&Salamon, 1990).Framework of ensembling classifier is depicted as Fig 1. Many researchers focus on how to ensemble individual classifiers' predictions to give a final decision of class label.

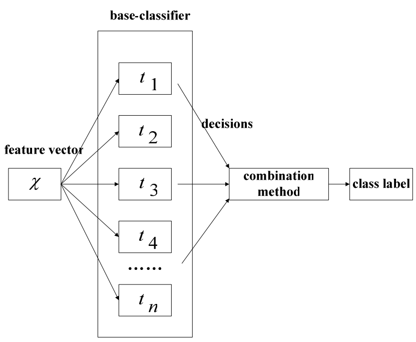


Fig.1 Framework of ensembling classifier

Previous ensembling classifier models are coarse and lack scientific quantitative direction on weights computation, leading low accuracy and bad explanation. Further, they care only about the single index of overall classification accuracy, ignoring other vital indexes in certain application context. This paper proposes a new ensembling classifier based on context and AHP (Analytic hierarchy process), in attempt to integrate context knowledge with data mining process to improve classification performance. It firstly determines important indexes in certain application context, and then uses AHP based on the context to scientifically compute the weights of indexes in the application context and the weights of excellent base-classifiers. Thus, an ensembling classifier orienting multi indexes can be built. Finally, we take weighted sum of excellent individual classifier as the final decision.

The rest of the paper is organized as follows. Section 2 reviews related work and points out the importance to integrate context knowledge with data mining process. Section 3 builds ensembling classifier based on context and AHP and describes the main procedures, followed by a case study used to verify our model in Section 4. Section 5 compares experiment results of our model and other ensembling classifiers, and Section 6 concludes the paper.

**Related work review**

For the past many years, researchers put forward various methods to combine the decisions of multiple individual classifiers, which fall into two fields.

The first field focused on the training and selection of base-classifiers, i.e. how to train and select excellent base-classifiers with high accuracy to give an overall decision. Some classic methods, such as bagging, boosting, and stacking, belong to the field (Witten&Frank, 2005). Cherkaucer (1996) trained 32 neural network model based on 8 different subsets of attributes and 4 different neural network structures. He took the majority voting of the 32 models as the final decision. In the field of spam filtering, Neumayer (2006) built a few classifiers based on feature clustering and machine learning, and select the best fitting classifier to decide the class labels of instances. Linemann et al (2008) optimize NMR spectra based on different configurations of data processing techniques, and select different subsets of attributes. Thus some different SVM models are built according to the selected attributes. Finally, majority voting is used to combine the decisions of multiple individual classifiers. Zhou et al (2002) believed that “many could be better than all”, i.e. ensemble of some excellent base-classifiers could outperform ensemble of all base-classifiers, and therefore selection of available individual classifiers should be paid attention to.

The other field centers on combination of base-classifier decisions, i.e. how to combine all the base-classifier decisions to give an overall decision of instance class label. Benchmarks in the field are AVG, MAJ and WAVG. Adaboost (Freund&Schapire, 1995,1996,1997,1998) firstly assigns random weights to the base-classifiers, and increases or decreases their weights according to their performances in the following testing process. The ensembling classifier’s decision is the weighted sum of all the base-classifiers. An interesting method based on information market to combine the decisions of multiple individual classifiers was proposed by Perols et al (2009). In their method, individual classifiers are considered as participants with some wealth in an information market where they place bets on different object classes. In training, each classifier estimates his prediction of class label of a certain object, and places their bets on them. Each single round finishing, those classifiers whose predications are correct would be paid, and thus their wealth increases. After training, the reciprocals of the market odds represent the ensembling classifier probability estimations of each class being the true object class.

Based on Zhou et al (2002)’s opinions, this paper insists that training and selection of base-classifiers and the combination of base-classifier decisions should be considered together in order to highlight the advantages of more efficient base-classifiers. Thus, we attempts to firstly select excellent base-classifiers rather than all base-classifiers to construct the ensemble, then we would use our method to combine excellent base-classifiers.

Further, previous researches depending on advanced algorithms in machine learning, care more about overall accuracy of the ensembling model, ignoring other vital indexes in certain application. They may take context knowledge into consideration in data preprocessing, but overlook them in modeling process which, however, is very important.

Universal and objective, context is closely connected with things and activities. It was researched in social sciences fields previously (Bunt, 1994; Srinivas, 1997; Murphy, 1996; Heidegger, 1962; Penco, 1999), and drew researchers’ attention in natural science later (Berthouzoz, 1999; Compton, 1988; Turney, 1996; Wobcke, 1999). As researches and applications in knowledge management (KM) go further, importance of context has also been recognized. Many scholars considered context as the key component in full comprehension of knowledge (Dieng et al, 1999; Brezillion&Pomerol, 1999; Goldkuhl&Braf, 2001). Despres and Chauval (2000) held that, as context is the perception of surroundings, without it, all kinds of knowledge would become meaningless. Thus, knowledge and knowledge management is only significant in certain context. Kakabadse (2003) pointed out that contextual requirement of knowledge is the issue to be discussed in future research of knowledge management.

As a vital part of knowledge management, knowledge discovery (or data mining) is also context-driven process. But to the best of our knowledge, few researches incorporate context knowledge to data mining task, and majority of the limited papers focus on its function in data preprocessing or feature selection rather than in modeling process (Sinha&Zhao, 2008; Alonso et al, 2002; Pan et al, 2007). It is easy to find that models with high accuracy are not that fit for the certain data task. In real application, user cares not only overall accuracy but also some other important indexes such as specificity (the ability to detect negative instances) and sensitivity (the ability to detect positive instances). A simple instance is that a model with high accuracy but low sensitivity, lacking the ability to detect the “fraud” objects, is surely not ideal.

**Ensembling classifier based on context and AHP**

This paper attempts to adopt human-computer combination, and incorporates context knowledge to classification process to improve model’s performance. We believe that data mining task is context-driven, and context knowledge would make up the shortcoming of machine learning. The fundamental framework is shown as Fig. 2.

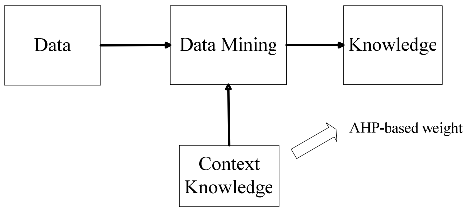


Fig. 2 fundamental framework of ensembling classifier based on context and AHP

We combine the existing two research areas of ensembling classifier. After first-step training and testing, more excellent base-classifiers are selected rather than all trained base-classifier to construct the ensemble. Then we adopt traditional decision model in computing excellent base-classifiers’ weights. Experts will determine which indexes should be considered in the certain context, and the indexes constitute the rule layer of decision model. The experts according to their experience and knowledge also determine the importance of the indexes in the context, and the comparison matrix of this layer is then built. The selected excellent base-classifiers form the scheme layer, and their performances on testing set are used to compute their weights to each index considered. Based on AHP method, we could easily work out the ultimate weights of all excellent base-classifiers, and their weighted sum would give the final decision.

The modeling process is shown as Fig. 3, and will be explained step by step afterwards.

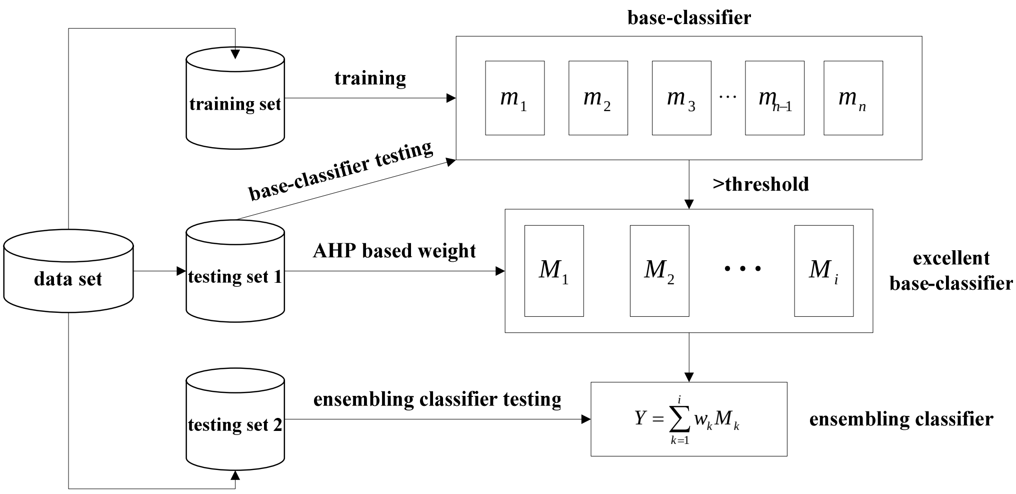


Fig. 3 modeling process of ensembling classifier based on context and AHP

Details of the modeling process are as follows:

1. Divide the original data set into 3 parts: (1) training set, for training multiple base-classifiers; (2) testing set 1, for testing the performance of all the base-classifiers and selecting excellent base-classifiers and computing their weights; (3) testing set 2, for testing the performance of our ensembling classifier.

2. Train multiple models based on training set. Individual classifiers are different as a result of different algorithms and model structures (there are choices of structure of neural network, decision tree, etc.) In this paper, we regard all the different models as individual base-classifiers and any subset of them can constitute the ensemble. Suppose after training, we will get $n$ base-classifiers, denoted as $m_1, m_2 \cdots m_n$.

3. Test all the base-classifiers’ performances on testing set 1. It should be ensured that all the base-classifiers should be tested by the same testing set. Then we will accumulate their performances on several indexes including overall accuracy, specificity, sensitivity, etc.

4. Select $i$ excellent base-classifiers who surpass the set threshold, and denote them $M_1, M_2 \cdots M_i$. They will construct the ensemble and their weighted sum is the final ensembling classifier’s decision,

i.e. $Y = \sum_{k=1}^{i} w_k M_k$ , $w_k$ is the ultimate weight of excellent base-classifier $M_k$ . If $Y \geq 0.5$ , we classify the object into class “1” according to the standard of Naïve Bayesian classifier, because $Y \geq 0.5$ indicates that probability the object belonging to class “1” is greater than it belonging to class “0”, and vise versa.

5. Compute each of the selected excellent base-classifier’s voting weight in final decision. We adopt the traditional decision model and set the goal as “selecting the best classifier”, displayed in Fig. 4. The important indexes chosen by experts form the rule layer and the excellent base-classifiers form the scheme layer.

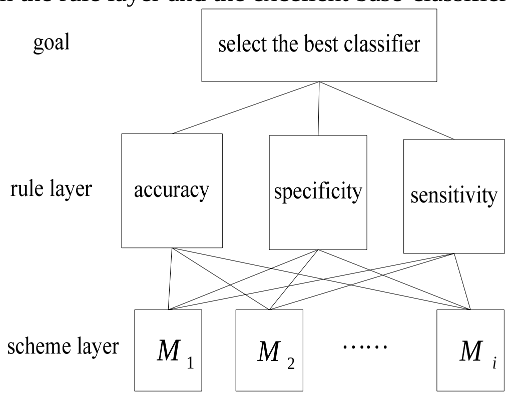


Fig. 4 decision model of selecting the best classifier

Different application contexts pay attention to different indexes and also the importance of each index varies from context to context. It is domain experts’ task to determine which indexes should be taken into account and how important they are. The very essence of AHP method is to transfer subject judgments into objective figures. Now that we get experts’ opinions about the indexes’ importance, then we can build the mutual matrix according to AHP method.

When we calculate the excellent base-classifiers’ weights to a certain index, their objective performances on it are used to construct the mutual matrix. The figures of matrix are the ratios of mutual performances on a certain index. For instance, ratio of two base-classifier’s overall accuracy is the element in the mutual matrix when computing base-classifiers’ weights to the index of overall accuracy.

Now that we have obtained the weight vector of index and several weight vectors of base-classifiers to indexes, we can easily work out the final weight vector of base-classifier to the goal based on AHP method. And thus equation $Y = \sum_{k=1}^{i} w_k M_k$ is the final ensembling classifier orienting multiple important indexes.

6. Test the ensembling classifier’s performance on testing set 2 and compare the result with other methods.

**Model application in customer churning management**

In this section, we will apply our method mentioned above to customer churning data from an American phone company to build a classification model and predict whether the customers will churn based on the model later.

Part of the original records of the American phone company customers are displayed as Table 1.

Table 1 original records of phone company customers

| state | account length | area code | phone number | international plan | voice mail_plan | number vmail_messages | total day minutes | total day_calls | total day_charge | total eve_minutes |
|---|---|---|---|---|---|---|---|---|---|---|
| KS | 128 | 415 | 382-4657 | 0 | 1 | 25 | 265.1 | 110 | 45.07 | 197.4 |
| OH | 84 | 408 | 375-9999 | 1 | 0 | 0 | 299.4 | 71 | 50.9 | 61.9 |
| OK | 75 | 415 | 330-6626 | 1 | 0 | 0 | 166.7 | 113 | 28.34 | 148.3 |
| AL | 118 | 510 | 391-8027 | 1 | 0 | 0 | 223.4 | 98 | 37.98 | 220.6 |
| LA | 117 | 408 | 335-4719 | 0 | 0 | 0 | 184.5 | 97 | 31.37 | 351.6 |
| WV | 141 | 415 | 330-8173 | 1 | 1 | 37 | 258.6 | 84 | 43.96 | 222 |
| RI | 74 | 415 | 344-9403 | 0 | 0 | 0 | 187.7 | 127 | 31.91 | 163.4 |
| ID | 85 | 408 | 350-8884 | 0 | 1 | 27 | 196.4 | 139 | 33.39 | 280.9 |
| VT | 93 | 510 | 386-2923 | 0 | 0 | 0 | 190.7 | 114 | 32.42 | 218.2 |
| VA | 76 | 510 | 356-2992 | 0 | 1 | 33 | 189.7 | 66 | 32.25 | 212.8 |
| TX | 73 | 415 | 373-2782 | 0 | 0 | 0 | 224.4 | 90 | 38.15 | 159.5 |
| FL | 147 | 415 | 396-5800 | 0 | 0 | 0 | 155.1 | 117 | 26.37 | 239.7 |
| AZ | 130 | 415 | 358-1958 | 0 | 0 | 0 | 183 | 112 | 31.11 | 72.9 |
| NE | 174 | 415 | 331-3698 | 0 | 0 | 0 | 124.3 | 76 | 21.13 | 277.1 |
| WY | 57 | 408 | 357-3817 | 0 | 1 | 39 | 213 | 115 | 36.21 | 191.1 |
| MO | 20 | 415 | 353-2630 | 0 | 0 | 0 | 190 | 109 | 32.3 | 258.2 |
| HI | 49 | 510 | 410-7789 | 0 | 0 | 0 | 119.3 | 117 | 20.28 | 215.1 |
| IL | 142 | 415 | 416-8428 | 0 | 0 | 0 | 84.8 | 95 | 14.42 | 136.7 |
| NH | 75 | 510 | 370-3359 | 0 | 0 | 0 | 226.1 | 105 | 38.44 | 201.5 |
| OK | 57 | 408 | 395-2854 | 0 | 1 | 25 | 176.8 | 94 | 30.06 | 195 |
| GA | 72 | 415 | 362-1407 | 0 | 1 | 37 | 220 | 80 | 37.4 | 217.3 |
| AK | 36 | 408 | 341-9764 | 0 | 1 | 30 | 146.3 | 128 | 24.87 | 162.5 |
| MA | 78 | 415 | 353-3305 | 0 | 0 | 0 | 130.8 | 64 | 22.24 | 223.7 |
| AK | 136 | 415 | 402-1381 | 1 | 1 | 33 | 203.9 | 106 | 34.66 | 187.6 |

The data set contains 3171 customers’ information with 21 attribute variables. The first 20 variables describe the customer’s present state of phone business, including international plan, voice plan, total minutes of international call, etc. And the last variable indicates whether the customer would churn. All the variables and their descriptions are listed in Table 2.

Table 2 attributes descriptions and explanations

| Attribute | Type | Description | Explanation |
|---|---|---|---|
| State | text | state of location | |
| Account_length | continuous | period of present business | |
| Area_code | discrete | area code | |
| Phone_number | text | phone number | |
| Voice_mail_plan | bool | voice mail plan | "1" yes, "0" no |
| Total_day_minutes | continuous | minutes of day calls | |
| Total_day_calls | continuous | calls of day | integer |
| Total_day_charge | continuous | charge of day calls | |
| Total_eve_minutes | continuous | minutes of evening calls | |
| Total_eve_calls | continuous | calls of evening | integer |
| Total_eve_charge | continuous | charge of evening calls | |
| Total_night_minutes | continuous | minutes of night calls | |
| Total_night_calls | continuous | calls of night | integer |
| Total_night_charge | continuous | charge of night calls | |
| Total_intl_minutes | continuous | international calls minutes | |
| Total_intl_calls | continuous | international calls | integer |
| Total_intl_charge | continuous | international calls charge | |
| Number_customer_service_calls | continuous | calls of customer service | integer |
| Class | bool | churn | "1" yes, "0" no |

Among the attributes, "state", "phone_number" and "area_code" are variables describing basic information of customers, but have little impact on customer churning. Besides them and "class", the remaining 17 attributes will be the input cells in the modeling process. After eliminating 138 records with null values, we divide the left 3033 records into 3 parts: training set with 1033 records, testing set 1 with 1000 and testing set 2 with 1000.

In practical application, instances belonging to the positive class (the class user cares more) are usually the minority of the all, which however causes "unbalanced sample". This may influence the modeling process and leading to model's poor We randomly sample records from the negative class and repeated randomly sample records from the positive to ensure that the number of instances belonging to the 2 classes are almost the same.

Based on the training set, we build several base-classifiers using different algorithms, including 4 decision trees, 6 neural networks, 2 logistic regression models and 2 naïve Bayesian classifiers. In this case, we focus on customers that would churn and cause loss of the company, thus the "churn" class is the positive class.

Experts determined that we should take accuracy, specificity and sensitivity into consideration in this application context. And the 3 indexes can be calculated by the following equations (Eq 1, Eq 2 and Eq 3). In the equations, $pos$ is the total number of positive instances and $neg$ is the total number of negative instances; $t\_pos$ is the number of positive instances labeled with "positive" by model and $t\_neg$ is the number of negative instances labeled with "negative" by model.

$$accuracy = \frac{t\_pos + t\_neg}{pos + neg} \tag{1}$$

$$specificity = \frac{t\_neg}{neg} \tag{2}$$

$$sensitivity = \frac{t\_pos}{pos} \tag{3}$$

We test all the 12 base-classifiers using testing set 1 and get their performances on the listed 3 indexes. The thresholds set for accuracy, sensitivity and specificity are 75%, 75% and 70% respectively. It is easy for us to select the excellent base-classifiers whose accuracy, specificity and sensitivity all surpass the thresholds and list them with performances on the 3 indexes in Table 3.

Table 3 four excellent base-classifiers and their performances on testing set 1

| Base-classifier | Accuracy | Specificity | Sensitivity |
|---|---|---|---|
| $M_1$ | 79.1% | 79.6% | 76.4% |
| $M_2$ | 84.2% | 86.2% | 73.0% |
| $M_3$ | 87.3% | 90.1% | 70.9% |
| $M_4$ | 78.4% | 79.4% | 73.0% |

Among the 3 indexes, experts assign largest weight to overall accuracy. And the context knowledge of customer churning management is that it will cost enterprises some money to manage customer relationship. Furthermore, cost of developing a new customer can reach as much as 5 times of maintaining an old one (Sheng&Liu, 2005; Su, 2005). Thus, misclassifying customer that would churn into the “will not churn” class is much serious. Therefore, it is believed sensitivity should be paid more attention to than specificity. Based on 1-9 principle in AHP method, the mutual matrix of accuracy, specificity and sensitivity are as following:

$$\begin{bmatrix} 1 & 7 & 2 \\ 1/7 & 1 & 1/5 \\ 1/2 & 5 & 1 \end{bmatrix}$$

And upon the matrix, their weight vector is $\left(0.592,\ 0.075,\ 0.333\right)$.

For those excellent base-classifiers to each index, their performances are used to construct the mutual matrix and then calculate the weight vector, shown as Table 4.

Table 4 mutual matrix and weight vector of base-classifier on each index

| | Mutual matrix | Weight vector |
|---|---|---|
| Accuracy | $\begin{bmatrix} 1 & 79.1/84.2 & 79.1/87.3 & 79.1/78.4 \\ 84.2/79.1 & 1 & 84.2/87.3 & 84.2/78.4 \\ 87.3/79.1 & 87.3/84.2 & 1 & 87.3/78.4 \\ 78.4/79.1 & 78.4/84.2 & 78.4/87.3 & 1 \end{bmatrix}$ | $\left(0.2404, 0.2559, 0.2654, 0.2383\right)$ |
| Specificity | $\begin{bmatrix} 1 & 79.6/86.2 & 79.6/90.1 & 79.6/79.4 \\ 86.2/79.6 & 1 & 86.2/90.1 & 86.2/79.4 \\ 90.1/79.6 & 90.1/86.2 & 1 & 90.1/79.4 \\ 79.4/79.6 & 79.4/86.2 & 79.4/90.1 & 1 \end{bmatrix}$ | $\left(0.2374, 0.2571, 0.2687, 0.2368\right)$ |
| Sensitivity | $\begin{bmatrix} 1 & 76.4/73.0 & 76.4/70.9 & 76.4/73.0 \\ 73.0/76.4 & 1 & 73.0/70.9 & 73.0/73.0 \\ 70.9/76.4 & 70.9/73.0 & 1 & 70.9/73.0 \\ 73.0/76.4 & 73.0/73.0 & 23.0/70.9 & 1 \end{bmatrix}$ | $\left(0.2605, 0.2489, 0.2417, 0.2489\right)$ |

Based on AHP method, we can calculate that the weight vector of 4 excellent base-classifiers to the goal is $(0.2469, 0.2536, 0.2578, 0.2417)$.

Therefore, the final ensembling classifier we built is $Y = 0.2469M_1 + 0.2536M_2 + 0.2578M_3 + 0.2417M_4$.

**Results comparison and discussion**

We test the ensembling classifier using testing set 2. The data task is to predict whether the 1000 customers would churn according to their attribute variables. Meanwhile individual classifiers and MAJ (Majority Voting), AVG (Averaging) also do the same task.

Performance comparison between ensembling classifier $Y$ and AHP and individual classifiers is shown in Table 5, and comparison of ensembling classifier $Y$ and MAJ, AVG is displayed in Table 6.

Table 6 performance comparison between ensembling classifier $Y$ and individual excellent classifiers

| Model | Accuracy | Specificity | Sensitivity |
|---|---|---|---|
| $M_1$ | 84.0% | 84.4% | 74.4% |
| $M_2$ | 83.7% | 84.0% | 76.8% |
| $M_3$ | 88.9% | 89.4% | 79.1% |
| $M_4$ | 78.8% | 78.8% | 79.1% |
| $Y$ | 89.1% | 89.6% | 79.1% |

Table 7 performance comparison between ensembling classifier $Y$ and MAJ, AVG

| Model | Accuracy | Specificity | Sensitivity |
|---|---|---|---|
| Majority Voting | 88.5% | 89.2% | 74.4% |
| Averaging | 84.3% | 84.9% | 72.1% |
| $Y$ | 89.1% | 89.6% | 79.1% |

We can easily draw some conclusions from the above results comparison:

1. The ensembling classifiers usually outperform individual classifiers, even excellent ones. This supports discovery and conclusions of previous researches, and indicates the significance of researching ensembling classifier.

2. The ensembling classifier based on context and AHP can improve classification accuracy, satisfy preference selection and balance requirements of multiple indexes in certain application context. This also illustrates that data mining task should be context-driven, and attach importance to context knowledge in modeling process.

**Conclusion**

This paper proposes a novel ensembling classifier based on context and AHP. We integrate context knowledge with data mining process to improve classification performance. Context knowledge is used to determine which are important indexes in certain circumstance. We then build a universal ensembling classifier based on hierarchal decision model, and adopt AHP to compute weights of individual classifiers. Finally, we take weighted sum of excellent individual classifier as the final decision.

A case of American phone company customer churn management is used to verify our model. We use our novel method to predict whether 1000 customers would churn. Experiment results show that our model outperforms individual classifiers and common ensembling classifier. It satisfies preference of index in certain context, and ensures good performance at more vital indexes.

This paper is an attempt to ensemble individual classifier's predications to get an overall decision. We think that application context should be taken into full consideration in classification modeling. And result has proved that human-computer combination can improve the classification performance.

**Acknowledgement**

This work was partially supported by the President Fund of Graduate University of Chinese Academy of Sciences（GUCAS)（A）(Grant No.085102HN00), National Natural Science Foundation of China (Grant No. 71071151, 70921061), National Key Technology R&D Program of China (No.2009BAK63B08), Youth Project Foundation of University of International Relations (Grant No. UIR2009-22).

**References**

Alonso, F., et al. (2002), "Combining expert knowledge and data mining in a medical diagnosis domain", *Expert Systems with Applications*, 23(4), pp367~375.

Berthouzoz, C. (1999), "A Model of Context Adapted to Domain-Independent Machine Translation", In: *Modeling and Using Context*, Second International and Interdisciplinary Conference (CONTEXT'99), Trento, Italy.

Brezillion, P., Pomerol, J. (1999) "Contextual knowledge sharing and cooperation in intelligent assistant systems", *Le Travail Humain*, 62(3):pp223~246.

Bunt, H. (1999), "Context and Dialogue Control", *Think Quarterly*, 3(1): pp19~31.

Cherkauer, K.J. (1996), "Human Expert-Level Performance on a Scientific Image Analysis Task by a System Using Combined Artificial Neural Networks", Working Notes of the AAAI Workshop on Integrating Multiple Learned Models, pp15~21.

Compton, R.J. (1988), "Knowledge in context: A strategy for expert system maintenance", In: proceedings of 2nd Australian Joint Artificial Intelligence Conference, Adelaide, Australia.

Despres, C, Chauval D. (2000), *The Present and the Promise of Knowledge Management*, Butterworth Heinemann, Boston, MA, pp55~86.

Dieng, R., et al. (1999) "Methods and Tools for Corporate Knowledge Management", *International Journal of Human Computer Studies*, 51(3).

Freund, Y., Schapire, R.E. (1995), "A Decision-Theoretic Generalization of Online Learning and an Application to Boosting", In: proceedings of the 2nd European Conference on Computational Learning Theory, pp23~37.

Freund, Y., Schapire, R.E. (1996), "Experiments with a New Boosting Algorithm", In: proceedings of the 13th International Conference on Machine Learning, pp148~156.

Freund, Y., Schapire, R.E.(1997), "Using and Combining Predictors that Specialize". In proceedings of the 29th annual ACM Symposium on Theory of Computing, pp334~343.

Freund, Y., Schapire, R.E., Bartlett, P. and Lee, W.S. (1998), "Boosting the margin: A New Explanation for the Effectiveness of Voting Methods", *Annals of Statistics*, 26(5):pp1651~1686.

Goldkuhl, G., Braf, E. (2001) "Contextual knowledge analysis—understanding knowledge and its relations to action and communication", In: proceedings of 2nd European Conference on Knowledge Management, IEDC-Bled School of Management, Slovenia.

Hansen, L., Salamon, P. (1990), "Neural Network Ensembles", In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):pp993~1001.

Heidegger, M. (1962), *Being and Time*, Harper & Row, New York.

Kakabadse, N.K., Kakabadse, A., Kouzmin, A. (2003), "Reviewing the knowledge management literature: Towards a taxonomy", *Journal of Knowledge Management*, 7(4):pp75~91.

Lienemann, K., Plotz, T. and Fink G.A. (2008), "SVM Ensemble Classification of NMR Spectra Based on Different Configurations of Data Processing Techniques", In: proceedings of the 19th International Conference on Pattern Recognition, pp1~4.

Murphy, L.D. (1996), "Information Product Evaluation as Asynchronous Communication in Context: A Model for Organizational Research", In: proceedings of the 1st ACM international conference on Digital Libraries.

Neumayer, R. (2006), "Clustering Based Ensemble Classification for Spam Filtering", In: proceedings of the 7th Workshop on Data Analysis, pp11~22.

Pan, H., et al. (2007), "Incorporating domain knowledge into medical image clustering", *Applied Mathematics and Computation*, 185(2):pp844~856.

Penco, C. "Objective and cognitive context", In: *Modeling and Using Context*, Second International and Interdisciplinary Conference (CONTEXT'99), Trento, Italy.

Perols, J., Chari, K. and Agrawal, M. (2009), "Information Market-Based Decision Fusion", *Management Science*, 55(5):pp827-842.

Sheng, Z.H., Liu, B.X. (2005), "Decision tree method for analysis of customer churning crisis", *Journal of Management Sciences in China*, 8(2):pp20~25. (in Chinese with English abstract)

Sinha, A.P. and H. Zhao. (2008), "Incorporating domain knowledge into data mining classifiers: An application in indirect lending", *Decision Support Systems*, 46(1):pp287~299.

Srinivas, K. (1997), "How is context represented in explicit and implicit memory", In: proceedings of the 2nd European conference on cognitive science, workshop on context, ECCS'97, Manchester, UK.

Su, H.J.(2005), "Data mining and its applications in telecom customer churning", Ph.D. Thesis, Changsha: Hunan University.(in Chinese with English abstract)

Turney, P. (1996), "The Identification of Context-Sensitive Features: A Formal Definition of Context for Concept Learning", In: Proceedings of 13th International Conference on Machine Learning (ICML'96), Workshop on Learning in Context-Sensitive Domains, Bari, Italy.

Witten, I. H., Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufman, San Francisco.

Wobcke, W. (1999), "The Role of Context in the Analysis and Design of Agent Programs", In: *Modeling and Using Context*, Second International and Interdisciplinary Conference (CONTEXT'99), Trento, Italy.

Zhou, Z.H., Wu, J.X. and Tang, W. (2002), "Ensembling Neural Networks: Many Could Be Better than All", *Artificial Intelligence*, 137(1): pp239~263.