

Relationship on the Price Sensitivity and Actual Market Acceptance Degree of Metallic Materials

Muxue Jing ^{1,*} and Xilai Jin ²

¹School of Management, Wuhan University of Technology, Wuhan, China, 430074

²School of Engineering, University of Florida, Gainesville, FL, US

* Corresponding author: jmuxue@126.com

Abstract. In order to find out the relationship between the price sensitivity and actual market acceptance degree of metallic materials, the database ensemble learning model is proposed in this paper. Due to the variety and class imbalance of customers, a database marketing model based on supervised clustering and ensemble learning is used for the model. The results show that the database ensemble learning model can thus improve the calculation accuracy and time-efficiency substantially.

1 Introduction

A key problem of database marketing is to correctly locate target clients. In the example of Baesens etc. [1] the accurate target client locating can add 500,000 Euro extra earnings for every extra percentage. Knott [2] etc point out that for a retail service bank, 0.7% extra percentage correct target client locating can help to improve 20% revenues for each customer. Judging from the perspective of data mining, the target client locating can be beckoned as classification problem, namely to predict the percentage for customers buying or to buy product based on their consumption characteristics. However, the diversity and unbalanced category of consumer groups restrict traditional classification predication technique. To begin with, among consumer groups, the number of target consumers is far lower the number of non-target clients, in other word, a class imbalance problem. The traditional classification predication technique aims to minimize the risk and it will be hard to effectively deal with class imbalance [3-4]; what's more, consumer groups are extensive and diverse, so single classification predication model cannot accurately reflect various consumption model as well as characteristics and over-fitting problem for learning model may appear [5].

In order to solve this problem, we consider targeting customers in database marketing as a classification and prediction problem in data mining, i.e., to predict whether a customer would purchase a product or the probability of purchasing based on his/her characteristics. Due to the variety and class imbalance of customers, a database marketing model based on supervised clustering and ensemble learning is proposed. The empirical study indicates that the proposed approach is able to improve the performance of database marketing.

Even though database marketing research has achieved important progress, an existing problem is to effectively optimize the difference and individual performance among various learning models so as to improve the accuracy of database marketing. The main idea of the learning model generated in existing ensemble studying research is to repeatedly and randomly take samples in sample space or feature space which yet cannot guarantee the big difference among various learning models. At the same time, the training subset created by random sampling cannot accurately depict the diversity and difference among consumer groups and as a result the accuracy of establishing learning model based on training subset will decrease. In order to solve above problems, this paper comes up with ensemble learning model based on supervised clustering which firstly adopts K-Means to gather the non-target consumers and divide them into several consumer subgroups with big differences. Then it combines subgroups with few respondent consumers so as to have several training subsets to train artificial neural network sample learning model in several subgroups to carry out integration and overcome the data imbalance problem meanwhile improve the learning performance of every single learner.

2 Basic Framework

The ensemble learning based on supervised clustering can be divided into three stages of data pre-processing, supervised clustering and ensemble learning.

The first stage is data pre-processing. The dimension property of consumer is different (for example, if the property of income is over 1000, the property of age is dozens). In clustering analysis dimension discrepancy will cause a big influence for those properties with big differences while for those with small differences, the influence can be ignored. Therefore, before carrying out clustering analysis, we have to carry out data pre-processing so as to make the dimension property consistent. This paper adopts min-max normalized methods to conduct linear conversion on data and the formula shows as follows:

$$\bar{x}_{ij} = \frac{x_{ij} - \min_j}{\max_j - \min_j}, (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \tag{1}$$

In this formula, n means the table cardinality and m refers to data attribute number (not including class label), x_{ij} refers to the original value of i record's j property, \bar{x}_{ij} refers to the value of i record's j property after standardized implementation, \min_j is the minimum value of j property while \max_j is the maximum value of j property. After standardizing formula (1) we can make the value range of all properties become $[0,1]$ so as to avoid the influence caused by different dimension properties.

The second stage is supervised clustering. Firstly calculate the proportion of majority class and minority class, set K is equal to the proportion in the K -Means algorithm and then we carry out clustering analysis on the majority class and cluster the majority class into K aggregates of data. Based on this, we recombine each aggregate of data with minority class to form K sub set sample with relevant balance.

Suppose the training sample set of N consumer information is $G = \{y_i, \mathbf{x}_i\}_{i=1}^N$. The individual information property of each consumer (such as age, sex and family members) shows in the form of $\mathbf{x}_i \in R^p$, among which p refers to the property number of customers. Suppose G_{pos} and G_{neg} ($G_{pos} \cup G_{neg} = G$) respectively refers to the customer set who response and do not response to marketing activities and in general. In order to explore the consumer subgroup with different consumption model and characteristics and at the same time overcome the problem of class imbalance we set the K value in K -Means algorithm is $|G_{neg}|/|G_{pos}|$ and adopts this value to carry out clustering analysis on the majority class (which does not affect consumers). Based on this, we combine every aggregate of data and minority class (which can affect consumers) to form K subgroups which are relevantly balanced. The algorithm flow chart for supervised clustering is shown as follows:

The third stage is to establish basic learning model and carry out ensemble learning. BP neural network has good learning ability and fault tolerant ability in classification predication and this paper adopts it as the basic learning algorithm.

In supervised clustering stage, we train the BP neural network of every sample set $subG^k, (k = 1, 2, \dots, K)$ so as to have K basic learning model $Learner^1, Learner^2, \dots, Learner^K$. Different basic learning models have different accuracy for different data samples and how to choose the ensemble learning method for basic learning model is a key issue. Comparing from choosing the optimal model among all basic learning models, a better approach is to choose different learning modes for different data samples (commonly known as dynamic integration)^[27]. This paper takes weighted voting based on sample neighborhood learning accuracy as the integration method which belongs to dynamic integration.

Concerning the unknown data record \mathbf{x} , the first step is to find out the close sample NN_x^k of $subG^k$ based on KNN algorithm so as to get the predict output of $Learner^1, Learner^2, \dots, Learner^K$ and \mathbf{x} as well as its close sample. P_x^k is the predict percentage on \mathbf{x} carried out by basic learning model $Learner^k$ while ϕ_x^k is the accuracy of $Learner^k$ in NN_x^k . When integrate the predict output of K basic models we can finally get the integrated forecasting output FP_x and the formula shows as follows:

$$FP_x = \sum_{k=1}^K P_x^k \frac{\phi_x^k}{\sum_{k=1}^K \phi_x^k} \tag{2}$$

$(k = 1, 2, \dots, K)$

The output of ensemble learning in whole testing set is $FP = (FP_1, FP_2, \dots, FP_x)$.

Based on the above three stages, the overall flow chart for database marketing model based on supervised clustering and ensemble learning shows in figure 1.

3 Discussion and Conclusion

This paper adopts the data of predict contest of COIL (Computational Intelligence and Learning) in 2000 as the empirical research. This data includes the sample data of 9822 European families buying the car insurance. What's more, this paper divides this data set into training set and testing set, among which training set includes 5822 data which is used to establish the ensemble learning model put forward in this paper while another 4000 data is used to assess the effect of this model. Among the 5822 training data set, there are only 348 minor class sets, occupying 5.97% of the training set. Among the 4000 testing data, there are 238 minor class sets, occupying 5.95%. Therefore, we can conclude that the data set has obvious class imbalance problem which can be beckoned as the data source of this paper.

Single ANN approach has the slowest lift speed which demonstrates that it has poor ability to deal with the imbalance class problem in database marketing while ANN based on SMOTE, ANN based on FN and GA/ANN approaches can create better effect. The model put forward in this paper has obvious advantages from 5%-35% depth and is worse than GA/ANN from 40%-50% depth. In reality, there is a dazzling array of customers in database marketing and therefore we pay more attention to the performance in smaller depth. In a summary, the above result indicates that the model in this paper can have higher hit rate in smaller depths which can be effectively affect the database marketing.

The ensemble learning method based on supervised clustering can have higher hit rate with the depth from 5%-35% comparing to other four methods especially from the depth 5%-25%, the hit rate is 32.00%, 23.25%, 18.67%, 15.75 and 14.30% which have obvious prediction advantages over other approaches. To conclude, the ensemble learning method based on supervised clustering is able to have higher Hit Rate with smaller depth which can improve the database marketing efficiency.

At present, both from the enterprise management practice and from the academic research level, brand management focus on relationship establishment and maintain, the attention to the relationship fracture customers is far less than the first two stages. In fact, with the increased competition of customer resources and customer development and maintenance cost of ascension, to re-develop the relationship of fracture customers should be the new focus on brand relationship management. Under the condition of buyer's market, due to the selective enhancement of consumers, customer resources become more competitive. The tentacles of enterprise brand relationship management have to extend to relationship revival of fracture customers. The premise of effective brand relationship revival strategy is correctly grasp of what factors drive the customer-brand relationship revival behavior and how these factors work. However, the academic study on brand relationship revival has just begun. Driving factors of the brand relationship revival, especially the thinking of mechanism of action is still very insufficient. The biggest problem of the only scattered results is the lack of a clear theoretical basis and analysis framework of the system.

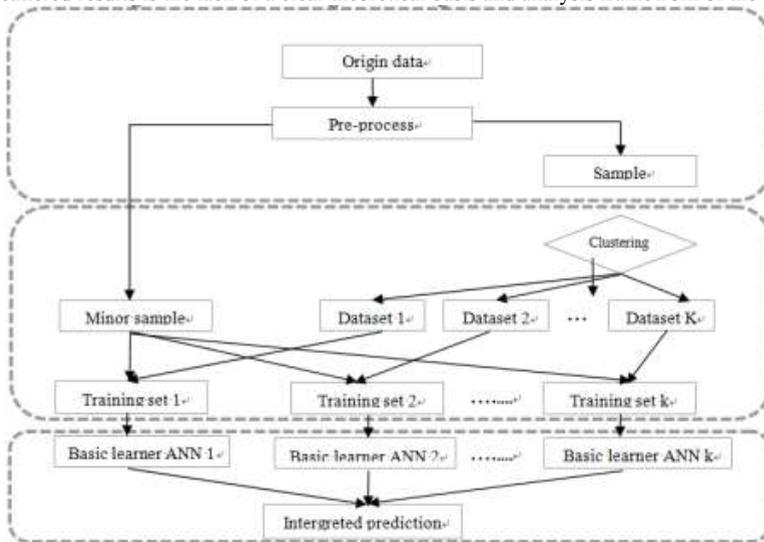


Fig. 1. Supervised clustering process.

This model believes that inside the data there are consumer aggregate of data with similar characteristics, so we adopt K-Means to carry out supervised clustering to conduct integration and classification with artificial neural network based on this. After carrying out experimental analysis and comparison among several learning methods, we can verify that the model in this paper is able to effectively deal with the imbalance class problem in database marketing and improve the accuracy. The following researches can be carried out from two aspects: firstly, extract the data characteristics to improve performance; secondly, carry out optimization as well as integration with other algorithms.

References

1. Gilles B, Cédric F, Benjamin G et al. Probabilistic Relational Verification for Cryptographic Implementations. ACM SIGPLAN Notices: A Monthly Publication of the Special Interest Group on Programming Languages, 49(2014): 193-205.
2. Wei Dai, Mostafa Bassiouni. An improved task assignment scheme for Hadoop running in the clouds. Journal of Cloud Computing, (2013), pp. 21-33. Babak A, Kristian K, Martin M et al. Exploiting symmetries for scaling loopy belief propagation and relational training. Machine learning, 92(2013):91-132.
3. Babak A, Kristian K, Sriraam N et al., Lifted Online Training of Relational Models with Stochastic Gradient Methods. Machine learning and knowledge discovery in databases, 5(2012):585-600.
4. Chingarande A, Mzumara M, Karambakuwa RT. Does MERSOSUR possess comparative advantage? Journal of Economics, 5(2014): 341-349.
5. Eric T. G. W, Jeffrey C. F. T, Varun G et al., Examining the Relational Benefits of Improved interfirm information processing capability in Buyer-Supplier Dyads. MIS quarterly, 37(2013):149-173.