第 24 卷第 2 期
2011 年 4 月
管理科学 ISSN 1672-0334
Journal of Management Science
Vol. 24 No. 2 20-29
April, 2011

A Cost-Sensitive Decision Tree Learning Model
——An Application to Customer-Value Based Segmentation

Zou Peng[1], Mo Jiahui[2], Kiang Melody[3], Ye Qiang[1]
1 School of Management, Harbin Institute of Technology, Harbin 150001, China
2 School of Management, The University of Texas at Dallas, TX Dallas 75080, USA
3 College of Business Administration, California State University, CA Long Beach 90840, USA

Abstract: The objective of this research is to extend the current decision tree learning model, to handle data sets with unequal misclassification costs. The research explores the issue of asymmetric misclassification costs through an application to customer-value based segmentation using empirical data collected from one of the largest credit card issuing banks in China. The data includes attributes from customer satisfaction survey and credit card transaction history is used to validate the proposed model. The results show that the proposed cost-sensitive decision tree for customer-value based segmentation is an effective method compared to the original decision tree learning model.

Keywords: cost-sensitive learning; asymmetric misclassification cost; decision tree; customer-value based segmentation

中图分类号:F713.5 **文献标识码**:A **文章编号**:1672-0334(2011)02-0020-10

1 Introduction

Customer segmentation refers to the process of classifying customers into subgroups that behave in similar ways or have similar preferences and needs[1]. Whereas, customervalue based segmentation is a process that groups customers based on the customer's current or potential value to the company[2]. A successful customer-value base system allows firms to effectively implement marketing strategies targeting at high value customer groups and take full advantage of the loyalty economics to achieve its marketing goals.

According to CFO Magazine, for most companies, one third of their customers are not profitable-and two-thirds of those never will be[3]. Another study in customer-value based marketing also showed that in many companies, when ranking customers by profitability, the bottom 20 percent of customers can deplete a firm's profit by 80 percent, while the top 20 percent customers can generate 150 percent of a firm's profit[4]. In today's ever increasing competitive business environment, it is imperative for a firm to target marketing efforts at the right group of customers. Although various classification techniques, both statistical and machine learning approaches, have been successfully applied to customer-value based segmentation, most of them fail to consider the effect of non-uniform cost per error and skewed class distribution on the overall performance of the model. Consequently the resulting model is biased towards the larger group with lower misclassification cost that is the low value customers that drain company profits. If implemented, the marketing strategies derived from the compromised system can perpetuate unprofitable relationships or, even worse, risking the loss of more profitable customers.

In this research, we propose a cost-sensitive decision tree learning model for customer-value based segmentation. The model is based on ID3, a classic machine learning approach, due to its simplicity and flexibility. It allows us to customize the splitting and pruning rules using the proposed cost-sensitive learning algorithm. We empirically test the

Received Date: October 8th, 2010 **Accepted Date**: January 13th, 2011

Funded Project: Supported by the National Natural Science Foundation of China(70802019)

Biography: Dr. Zou Peng, a Hubei Macheng native (1975 -), graduated from Harbin Institute of Technology and is a lecturer in the School of Management at Harbin Institute of Technology. His research interests include customer relationship management and business intelligence, etc. E-mail: zoupeng@hit.edu.cn

performance of our system using both customer survey data and transaction history collected from one of the largest credit card issuing banks in China. The original ID3 learning algorithm, based on information gain, is used as the benchmark for comparison. Results show the cost-sensitive learning method is able to identify high value customers more effectively and generate lower total misclassification cost.

The balance of this paper is organized as follows: Section 2 reviews previous studies related to customer-value based segmentation and decision tree learning. Section 3 presents the cost-sensitive learning tree and its application to customer-value based classification. Section 4 describes the data set and experimental procedures that applies this cost-sensitive decision tree to empirical data collected from one of the largest credit card issuing banks in China. Benchmark analysis is also performed to validate the effectiveness of our approach. Section 5 presents the managerial implications of the classification results. The paper concludes with a summary of our findings in Section 6.

2 Literature Review

Extant research in customer-value based segmentation has suggested different ways to group customers. Literature[5] divides customers into two groups, "churn permanently" and "churn temporarily", for calculating customer lifetime values. Literature[6] groups customers based on customer values at different stages of customer lifecycle. Literature[7] proposes using "present value", "potential value", and "loyalty" as three dimensions to segment customers.

Customer-value based segmentation problems usually involve a large amount of customer transaction data. Data mining techniques are designed for solving problems with large data sets, therefore are suitable for such tasks. A review of the literature shows various data mining techniques have been applied to customer-value based segmentation[8] such as the associate rule[9], decision tree[10-12], Chi-squared Automatic Interaction Detector (CHAID)[13], neural network[14], and Bayesian Network[15]. However, all of them assume equal misclassification cost and the algorithms are designed to minimize the total misclassification rate. The application of these simplified models to problems with significantly different misclassification costs will result in a deficient model. The 20:80 theory in market value analysis suggests often times the top 20 percent of high value customers account for 80 percent of the company's total profit[16]. Therefore it is more costly to misclassify a high value customer as a low value one than the other way around. In banking industry, particularly it is believed that the top 10% value customers create 90% of total profit.

The consequence of misclassify a low value customer as high value one is simply wastingmarketing resources and the impact usually is limited, whereas misclassify a high value customer as low value one may lead to higher top value customer churn rate, hence has adverse effect on a firm's image. Since the existing learning algorithms were designed to minimize the total misclassification rate (the sum of type I and type II errors)[17] and makes no distinction between the differences in the two types of error, the resulting model is unlikely to do well when measured by total misclassification cost.

Moreover, previous study in commercial market research suggests that data involve product ratings and customer satisfaction assessments tend to be markedly skewed. When the data are highly skewed, in our case the number of high value customer far less than that of low value ones, the problem aggravates. As the current learning algorithms try to balance type I and type II errors, the resulting model will bias towards the larger group, the low value customers. In other words, the learning algorithms tend to produce high predictive accuracy over the class with relatively large number of examples, but poor predictive accuracy over the class with relatively small number of examples[18]. Therefore it will increase the probability of misclassifying high value customer in order to reduce the probability of misclassifying the low value ones. Many examples of imbalanced class distribution can be found in business applications such as the percentage of fraud in credit card customers[19] and the percentage of bankrupt companies in bankruptcy prediction[20].

Decision tree learning is one of the most popular data mining techniques used in business applications. The objective of decision tree learning is to create a model that predicts the value of a target variable based on several input variables. Each interior node in the output tree corresponds to one of the input variables; while each leaf node corresponds to a value of the target variable given the values of the input variables represented by the path from the root to the leaf. During the learning process, the tree recursively partition itself by selecting the next best input variable to use for splitting the tree. The splitting process will divide the set of inputs into subsets that have the same value of the target variable. Different learning algorithms use different formulae to measure "best"[21]. These formulae are applied to each candidate subset, and the resulting values are combined to provide a measure of the quality of the split. The learning process is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of decision tree classifier does not require domain expert knowledge or subjective parameter settings, therefore is appropriate for exploratory knowledge discovery. The resulting tree diagram is intuitive and easy to assimilate by novice[22]. Moreover, decision tree provides users with a visual

Table 1 Type I and Type II Errors

表 1 类型 I 和类型 II 错误

Sample Size	Predicted Value(H)	Predicted Value(L)	Total
True Value (H)	*TP*	*FN* (Type II Error)	*P*
True Value (L)	*FP* (Type I Error)	*TN*	*N*

rendering of the market structure and can function as an effective decision support tool for prediction and classification.

The well-known decision tree learning algorithms based on information gain (i.e. ID3, C4.5, and C5.0), assume the misclassification costs of type I and type II errors are the same and the underlying group sizes are equal. Previous research found that implicitly ignoring asymmetric consequences in certain problem domains, such as medical diagnosis and bankruptcy prediction, could be costly[23]. To improve upon the current decision tree learning algorithms, we propose a cost-sensitive classification algorithm as the splitting rule for decision tree models.

An earlier study in cost-sensitive related problems focused on prediction with non-symmetric costs of error using statistical tools[24]. Post hoc threshold adjusting and weighting are two major approaches to cost-sensitive classifier learning[25]. For the post hoc threshold adjusting approach, Literature [2] proposes to first learn the decision tree from the training set, then compute the optimal decision based on minimizing total expected cost using the probability estimates from the learned tree. Literature [26] and [27] incorporate both misclassification cost and prediction accuracy in a genetic algorithm fitness function using simulated data sets. For the weighting approach, Literature [28] use minimizing total cost (both test and misclassification costs) as a decision tree's splitting criterion, but the focus is on minimizing test cost. Misclassification cost is cost incurred by misclassification error and test cost is cost incurred for obtaining attribute values. The method is designed especially for handling data sets with missing values.

In this research, Weka, an open-source data mining suite, contains various decision tree learning algorithms, is used to implement the proposed cost-sensitive learning. Comparative analysis of classification accuracy and total misclassification costs with ID3 are also conducted.

3 Research Methodologies

3.1 Problem Description

To simplify the problem, we define customer-value based segmentation as a binary classification task where high value customers are positive (H) examples and low value customers are negative (L). Table 1 shows the type I and type II errors for our customer-value based segmentation problem.

True Positive (TP) is the number of positive samples predicted as positive and True Negative (TN) is the number of negative samples predicted as negative. Both are correct predictions of the samples. False Positive (FP), also known as Type I error or α error, is the number of negative samples incorrectly predicted as positive. False Negative (FN), also known as type II error or β error, is the number of positive samples incorrectly predicted as negative. We also have $P = TP + FN$ and $N = FP + TN$. Table 2 depicts the misclassification costs associated with each type of prediction.

Table 2 Misclassification Costs

表 2 错误分类代价

Cost value	Predicted Value(H)	Predicted Value(L)
True Value(H)	0	C_{FN}
True Value(L)	C_{FP}	0

C_{FP} is the cost of incorrectly classifying negative (low value) samples as positive (high value) and C_{FN} is the cost of incorrectly classifying positive (high value) samples as negative (low value). In our case, C_{FN} should be a lot higher than C_{FP}.

3.2 Performance Evaluation

A popular statistical method for evaluating model prediction outcome is the *F-measure*, also known as *F-score*[29]. *F-measure* calculates the weighted average of precision and recall. Precision is the measure of exactness and is calculated as the number of true positive divided by the total number of positive predicted ($TP + FP$) (see equation (1)). Recall, also refer to as True Positive Rate (TPR) is the measure of completeness and is calculated as the number of true positive divided by the total number of positive samples (P) (see equation (2)).

$$Precision = \frac{TP}{FP + TP} \tag{1}$$

$$Recall(TPR) = \frac{TP}{P} \tag{2}$$

$$F\text{-}measure = \frac{2Recall \cdot Precision}{Recall + Precision} \tag{3}$$

$$\text{False Positive Rate }(FPR) = \frac{FP}{N} \tag{4}$$

$$Accuracy = \frac{TP + TN}{P + N} \tag{5}$$

F-measure as shown in equation (3), reaches its best value at 1 when $FN = FP = 0$, and worst at 0 when $TP = 0$. False Positive Rate (FPR) measures the percentage of low value customers that are misclassified as high value ones (see equation (4)). Accuracy (see equation (5)) calculates the total misclassification rate and is commonly used to measure the performance of a binary classifier.

3.3 Cost-Sensitive Learning Tree for Customer-value Based Classification

Existing decision tree algorithms, such as ID3, C4.5, and CART, use information gain as their attribute selection measure and the goal is to maximize accuracy as shown in equation (5). In this research, we propose a cost-sensitive learning algorithm that selects the splitting attribute by minimizing the misclassification cost, C, measured using the cost function suggested by Literature [30]:

$$C = C_{FP} \cdot FP + C_{FN} \cdot FN \tag{6}$$

$$AVG(C) = \frac{C}{P + N} \tag{7}$$

$AVG(C)$ calculates the per sample average cost and is used to evaluate model performance. Since the attributes collected for our study are mostly categorical, we design the cost-sensitive algorithm for handling categorical data as follows. In a binary classification with S samples each contains attributes $A_j (j = 1, \cdots, n)$. For each categorical attribute A_j, it has m distinct values $A_{ji} (i = 1, \cdots, m)$. The misclassification cost of a node splitting at A_j is calculated as:

$$C_j = \sum_{i=1}^{m} P(i) \cdot C(i) \tag{8}$$

Where $P(i)$ is the proportion of each branch i and $C(i)$ is the cost of branch i. For a non-leaf node i, the cost is the sum of all its branches. For a leaf node i, the cost of $C(i)$ is calculated as:

$$C(i) = P(\frac{P}{node_i}) \cdot C_p + P(\frac{N}{node_i}) \cdot C_N \tag{9}$$

where $P(\frac{P}{node_i}) = \frac{P}{S_i}$ is the proportion of positive samples at $node_i$ and S_i is the total number of samples at $node_i$. $P(\frac{N}{node_i}) = \frac{N}{S_i}$ is the proportion of negative samples at $node_i$. C_P is the cost when label a leaf node, $node_i$, as positive (high value customers) and C_N is the cost when lable a leaf node, $node_i$, as negative (low value customers). The decision tree tries to minimizing the total cost by labeling the leaf nodes according to $\min\{C_P, C_N\}$. In other words, if $C_P \geqslant C_N$, label the leaf node as negative, otherwise positive.

$$C_P = C_{FP} \cdot FP \tag{10}$$

$$C_N = C_{FN} \cdot FN \tag{11}$$

The original ID3 decision tree algorithm uses accuracy as the criterion to prune trees. We extend ID3 algorithm and implement the tree pruning feature based on our cost functions. The postpruning feature similar to the one in CART has also been employed using the above cost functions.

3.4 Measuring Customer Value

Previous research suggests that historical customer transaction data can be used as good estimates for customer values. Using the data we collect from the bank, we calculate each customer value as follows[31]:

$$M = M_v - M_c \tag{12}$$

Where M is the value of a customer, M_v is the revenue firm earned from the customer, and M_c is the expense generated by providing products or services to that customer. To simplify the problem, we assume that the revenue and cost generated by each customer are the same within the category (high/low values). We calculate the customer value for each category as follows:

$$M_{ji} = \begin{cases} M_{1V} - M_{1c} & i = j = 1 \\ N_{1V} - M_{0c} & i = 1, j = 0 \\ M_{0v} - M_{0c} & i = j = 0 \\ N_{0v} - M_{1c} & i = 0, j = 1 \end{cases} \tag{13}$$

Where M_{ji} is the customer value when predicting customer of class i as class j (True value (i); Predicted value (j)). M_{iv} is the positive revenue from a customer of class i when predicted correctly, while N_{iv} is the negative revenue from a customer of class i when predicted incorrectly. M_{ic} is the cost generated when providing products or services to class i customer. Table 3 shows the value matrix we use to calculate C_{FP} and C_{FN}.

Table 3 Value Matrix

表3 价值矩阵

	Predicted Value(1)	Predicted Value(0)
True Value(1)	$M_{1V} - M_{1c}$	$N_{1v} - M_{0c}$
True Value(0)	$N_{0v} - M_{1c}$	$M_{0v} - M_{0c}$

Where class 1 represents high value customers and class 0 represents low value customers. We also assume that by providing more products and services to low value customers will not result in increased revenue, whereas by providing less products and services to high value customers will result in loss of revenue or customer churn. The misclassification costs, C_{FP} and C_{FN}, are the opportunity costs, or the difference between the revenue the firm could have earned and the revenue earned given the wrong prediction.

$$C_{ji} = M_{ii} - M_{ji} = (M_{iV} - M_{ic}) - (N_{iv} - M_{jc})$$
$$i = 0,1 \quad j = 0,1 \quad \text{and} \quad i \neq j \tag{14}$$

Based on equation (14), we can represent the misclassification cost in Table 2 using the cost matrix as shown in Table 4.

Table 4 Cost Matrix for the Proposed Cost-sensitive Learning Method

表 4 代价敏感学习方法的代价矩阵

	Predicted Value(1)	Predicted Value(0)
True Value(1)	0	$(M_{1v} - M_{1c}) - (N_{1v} - M_{0c})$
True Value(0)	$(M_{0v} - M_{0c}) - (N_{0v} - M_{1c})$	0

The revenue and cost information for the cost matrix is derived from the bank's financial data, marketing budget, and customer transaction history and is used to implement the proposed cost-sensitive learning algorithm.

4 Empirical Studies and Performance Evaluation

4.1 Data

With the explosive growth of the credit card business, the competition among the credit card issuing banks to gain market shares in China has become more and more intense[32]. Therefore how to best target and retain credit card customers becomes a critical issue for both domestic and foreign credit card issuing banks operate in China.

Survey questionnaires regarding consumer loyalty and credit card transaction history are collected from one of the largest credit card issuing banks in China. The subjects contain customers reside in twelve largest cities in China. Surveys with a five-point Likert scale ranging from strongly agree (5) to strongly disagree (1) are used. These Questionnaires are sent to credit card holders of the bank through the bank's email system between May and October 2007. A total of 8 381 valid responses from 12 different cities are collected. In addition, more than 300 000 transaction records between January 2006 and May 2007 are recorded for calculating the customer values. We start with a total of 13 attributes as shown in Table 5.

We implement the input preprocessing function proposed by Literature[33], an attribute-oriented induction technique, to reduce the relative attribute set down to five attributes and they are: Age, Marital status, Income, Share wallet, and Loyalty. The descriptive statistics of the attributes are presented in Table 6.

Table 5 Thirteen Attributes Collected

表 5 初始属性集合

Source	Attributes
Account Application Form	Gender, Age, Marital status, Education, Income, Occupation, Years worked, Ownership of property, City
Transaction Records	Frequency of transaction, Amount of transaction
Survey Questionnaire	Wallet share, Customer loyalty

Table 6 Attributes Descriptive Statistics

表 6 描述性统计

Attribute	Category		Sample Proportion
Age	≤25	1	57.9%
	>25	2	42.1%
Income	0 ~ 30 000	1	31.4%
	30 001 ~ 50 000	2	30.0%
	50 001 ~ 80 000	3	19.4%
	80 001 ~ 120 000	4	10.7%
	120 001 ~ 200 000	5	4.7%
	>200 001	6	3.8%
Marital status	no	1	70.6%
	yes	2	29.4%
Loyalty	low	1	74.4%
	high	2	25.6%
Wallet share	low	1	48.8%
	high	2	51.2%
Customer value class	low	0	90.46%
	high	1	9.54%

In customer loyalty study, customer wallet share refers to the proportion of consumption volume a customer spend in a firm divided by the total consumption with all firms in the same category[34].

4.2 Calculating Customer Value and Misclassification Costs

One year of customer transaction data collected between 2006 and 2007 are used to calculate the revenues the bank earned from customers of the two classes (high value and low value), M_{0v} and M_{1v}. Bank's financial data from the same period is used to calculate the marketing costs of the two classes, M_{0c} and M_{1c}. We assume that the incorrect prediction of high value customer (false negative) will result in customer churn, therefore $N_{1v} = M_{0v}$, whereas false negative will not gain more revenue from low value customer due to their limited purchasing power, therefore $N_{0v} = 0$. We apply equation (14) and derived C_{FP} and C_{FN} in Table 7.

Table 7 Cost Matrix Calculated from Bank Data
表 7　基于银行数据计算的代价矩阵

	Predicted Value(1)	Predicted Value(0)
True Value(1)	0	107 008
True Value(0)	43 495	0

4.3 Cost-Sensitive Learning Algorithm

We use Weka 3.6.0, an open source data mining software program, to employ the proposed Cost-Sensitive Learning (CSL) algorithm. The data set is divided into training and test sets at 2:1 ratio. We use a 10-fold cross-validation to train and test the models. The data are randomly partitioned into ten mutually exclusive subsets, each of approximately equal size. Both the training and testing processes are performed ten times and the final classification rules are presented in Table 8. To be conservative, the results reported are the worst results from the ten runs. Table 9 shows the classification results from CSL decision tree. The type I error is 4.93% ($\frac{374}{7\ 581}$), the type II error is 42.00% ($\frac{336}{800}$), and the overall accuracy rate is 91.53% ($\frac{464+7\ 207}{800+7\ 581}$). Figure 1 shows the final decision tree derived from the proposed CSL.

In the next subsection, we compare the performance of CLS with the original ID3 algorithm.

4.4 Comparative Analyses

We compare CSL with ID3 to evaluate the performance of the models on data with unequal misclassification costs and the results are shown in Table 10. We expect the Accuracy of ID3 to be higher than CSL because ID3 algorithm is designed to minimize the total misclassification rate, or to maximize Accuracy, whereas CSL is designed to minimize the total misclassification cost . However , results show that

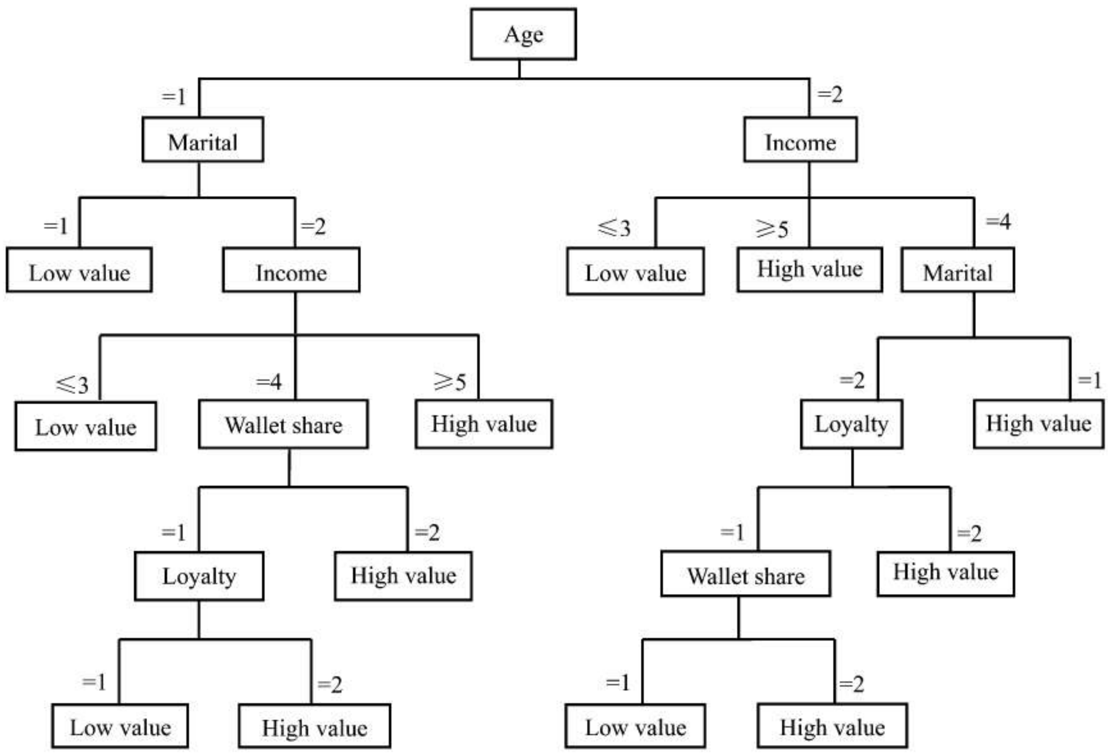

Figure 1 Decision Tree Generated from the Proposed Cost-sensitive Learning
图 1　代价敏感学习方法生成的决策树

Table 8 Classification Rules Derived from CSL

表8 分类规则

Class	Rules
Low value	Age < 25 and Marital Status = No
	Age < 25 and Marital Status = Yes and Income < 80 000
	Age < 25 and Marital = yes and 80 001 < Income < 120 000 and Wallet Share = Low and Loyalty = low
	Age > 25 and Income < 80 000
	Age > 25 and Marital Status = yes and 80 001 < Income < 120 000 and Wallet Share = low and Loyalty = Low
High value	Age < 25 and Marital Status = Yes and Income > 120 001
	Age < 25 and Marital Status = Yes and 80 001 < Income < 120 000 and Wallet Share = High
	Age < 25 and Marital Status = Yes and 80 001 < Income < 120 000 and Wallet Share = low and Loyalty = High
	Age > 25 and Income > 120 001
	Age > 25 and Marital Status = No and 80 001 < Income < 120 000
	Age > 25 and Marital Status = Yes and 80 001 < Income < 120 000 and Loyalty = High
	Age > 25 and Marital Status = Yes and 80 001 < Income < 120 000 and Loyalty = low and Wallet Share = High

Table 9 Prediction Outcome of CSL

表9 预测结果

Sample Size	Predicted Value(H)	Predicted Value(L)	Total
True Value(H)	464	336	800
True Value(L)	374	7 207	7 581

Table 10 Performance Comparision of the Two Approaches

表10 两种方法的性能对比

	$\frac{Recall}{TP}$ *Rate*	*Precision*	*F-Measure*	*FP Rate*	*Accuracy*	Average Cost
CSL	0.58	0.55	0.57	0.05	0.92	6 213.65
ID3	0.35	0.67	0.46	0.02	0.92	7 306.83

the Accuracy of ID3 (0.92) is the same as that of CSL (0.92), but the misclassification cost of CSL (6 213.65) is substantially lower than that of ID3 (7 306.83). *Recall* measures the true positive rate and in our case is the percentage of high value customers predicted correctly as high value. Therefore to increase *Recall* is more important than to reduce *FPR* and there are tradeoffs between the two. The results show CSL's *Recall* rate (0.58) is higher than ID3's (0.35) by 65.71%. *F-Measure* of CSL is also higher than that of ID3 by 23.91% ($\frac{0.57-0.46}{0.46}$). The average cost calculated using equation (7) shows ID3's average cost per customer is higher than CSL by 17.59% ($\frac{7\,306.83-6\,213.65}{6\,213.65}$). In summary, the results show that the proposed cost-sensitive learning algorithm is a more effective method in handling data with unequal misclassification costs than the original ID3 algorithm.

5 Managerial Implications

The rules extracted by CSL can be used to guide managers when devising marketing strategies. The rules shown in Table 8 suggest: ①Most single customers 25 years old or younger are low value customers because their income is limited. ②Customers with medium income (annual income between 80 000 and 120 000) can become high value customers if bank can devise marketing programs to enhance their loyalty or wallet share. ③Almost all customers with high income (annual income over 120 000) are high value customers. ④The customers who are low in both wallet share and loyalty have more low value than other customers, and the customers who are high in either wallet share or loyalty tend to be high value; therefore bank should focus on retaining this group of customers and design marketing programs to promote business to them. Moreover, from the collected data, we have made several interesting observations pertinent to Chinese credit card market: ①Most credit card holders 25 years of age or younger are either still in school or just started working, therefore their purchasing power is relatively weak. Banks should focus more on building long term relationship with these customers and less on promote spending. ②Becoming married is considered as one of the most important milestones in life for Chinese. Most people change their purchasing behavior after getting married. The most significant one is the purchase of a residential property. ③Demographic attributes such as income and marital status are the most important attributes to be considered when perform market segmentation. ④Low wallet share may mean potential promotion value so that firms may improve the customer's value by higher his or her wallet share or loyalty.

One of our contributions for database marketing is integrating the survey data with internal data. It is really important for firms to get this information, such as wallet share. For example, A and B are two customers who have the same transaction time and transaction amount in a company. If we evaluate their loyalty only by repeated purchasing, they are just the same. But in fact, all of A's spending power is used in his transaction with the company, while B only use a small part of his spending power in the company. In the light of "specificity" and "exclusiveness" of customer loyalty, it is obvious that A is more loyal than B, and that is the one-side effect of repeated purchasing behavior measurement. Secondly, from the strategy-making requirements of cross-sell and up-sell, it will mislead enterprises to assess the potential promotion value of customers. Customers who have a higher number of transactions and larger trading volume may transact little with other similar business, and these have little potential promotion. On the contrary, customers who have a fewer number of transactions and small trading volume may transact lot with other competitive firms, and these customers have great potential promoting value. And Attitude loyalty explains the real reason why customers repeatedly buy and how the customer loyalty forms. The extracted rules can be used at the application stage to decide what status or program to assign to the applicant. The segmentation result can help the manager to determine the amount of credit line to be assigned to a new customer, target a mail solicitation such as credit card promotions, and design loyalty program to improve the customer wallet share and persist longer customer relationship.

6 Conclusions and Future Research Directions

The rapid growth of credit card business in China represents both an opportunity and a threat to the banking industry in China. How to correctly identify and target valuable credit card customers becomes a challenging task. Previous research in market segmentation classifies customers based on assumption that all the misclassification costs are equal without considering different customer value.

In this research, we propose a cost-sensitive decision tree learning approach and apply it to credit card customer data from a major credit card issuing bank in China. Results show that the proposed cost-sensitive learning tree method outperforms the original ID3 learning algorithm when the data set contains unequal misclassification costs. Therefore the proposed model can serve as an effective technique for building a decision support system that handles problems with asymmetric misclassification costs, which is commonly seen in real-world applications.

The current model is designed to handle binary classification problems with categorical data. The model can be extended to handle problems with more than two classes. Future research will focus on extending the model to handle problems with numerical data.

References:

[1] Liautaud B, Hammond M. E-Business Intelligence turning information into knowledge into profit[M]. New York: McGraw Hill, 2000:38-39.

[2] Brodie J R, Whittome J R M, Brush G J. Investigating the service brand: A customer value perspective[J]. Journal of Business Research, 2009, 62(3): 345-355.

[3] Hyatt Josh. Playing favorites[EB/OL]. CFO Magazine. http://www.cfo.com/article.cfm/12835154/c_12838677. January 1, 2009.

[4] Selden L, Colvin G. Angel customers and demon customers: Discover which is which and turbo-charge your stock[M]. 1st ed. New York: Portfolio Hardcover, 2003:76-80.

[5] Dwyer F R. Customer lifetime valuation to support

marketing decision making [J]. Journal of Direct Marketing, 1997, 11(4): 6-13.

[6] Riechland F F, Sasser W E, Jr. Zero defections: Quality to services [J]. Harvard Business Review, 1990, 68(5): 105-111.

[7] Hwang H, Jung T, Suh E. An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry [J]. Expert Systems with Applications, 2004, 26 (2): 181-188.

[8] McCarty J A, Hastak M. Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression [J]. Journal of Business Research, 2007, 60(6): 656-662.

[9] Hsieh N C. An integrated data mining and behavioral scoring model for analyzing bank customers [J]. Expert Systems with Applications, 2004, 27 (4): 623-633.

[10] Ngai E W T, Li X, Chau D C K. Application of data mining techniques in customer relationship management: A literature review and classification [J]. Expert Systems with Applications, 2009, 36(2): 2592-2603.

[11] Jiang L X, Li C Q. An empirical study on attribute selection measures in decision tree learning [J]. Journal of Computational Information Systems, 2010, 6 (1): 105-112.

[12] Shi N Y, Chu X, XU J Y, Duan Y X. Research of the decision tree in data mining based on concept hierarchy [J]. Journal of Computational Information Systems, 2008, 4(2): 641-646.

[13] Jonkera J J, Piersmab N, Van den Poel D. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability [J]. Expert Systems with Applications, 2004, 27(2): 159-168.

[14] Pendharkar P C. A threshold varying bisection method for cost sensitive learning in neural networks [J]. Expert Systems with Applications, 2008, 34 (2): 1456-1464.

[15] Baesens B, Verstraeten G, Van den Poel D, Egmont-Petersen M, Van Kenhove P, Vanthienen J. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers [J]. European Journal of Operational Research, 2004, 156(2): 508-523.

[16] Bloomquist A. Small business, big profits [J]. Bank Marketing, 1998, 30(8): 18-20.

[17] Zhang Y, Chi Z X, Xie F D, Li N. A weights-based accuracy evaluation method for multi class multi-label classifier [J]. Journal of Computational Information Systems, 2008, 4(2): 589-594.

[18] Maloof M A. Learning when data sets are imbalanced and when costs are unequal and unknown [C] // ICML-2003 Workshop on Learning from Imbalanced Data Sets II. Washington D. C., 2003: 219-227.

[19] Syeda M, Zhang Y Q, Pan Y. Parallel granular neural networks for fast credit card fraud detection [C] // Proceedings of the 2002 IEEE International Conference on Fuzzy Systems. Honolulu HI USA, 2002: 572-577.

[20] Tam K Y, Kiang M Y. Managerial applications of neural networks: The case of bank failure predictions [J]. Management Science, 1992, 38(7): 926-947.

[21] Han J, Kamber M. Data mining concepts and techniques [M]. 2nd ed. Sinagpore: Elsever Pte Ltd, 2006: 268-270.

[22] Yue X, Li Q Z, Zhang C G, Xu H L. Approach to evaluate the imbalanced degree of the training data sets using community structures [J]. Journal of Computational Information Systems, 2007, 3 (5): 1929-1934.

[23] Granger C W J. Prediction with a generalized cost of error function [J]. Operational Research Quarterly, 1969, 20(2): 199-207.

[24] Zhao H M. Instance weighting versus threshold adjusting for cost-sensitive classification [J]. Knowledge & Information Systems, 2008, 15(3): 321-334.

[25] Elkan C. The Foundations of cost-sensitive learning [C] // Proceedings of the 17th International Joint Conference on Artificial Intelligence. Seattle Washingtom USA, 2001: 79-84.

[26] Turney P D. Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree algorithm [J]. Journal of Artificial Intelligence Research, 1995, 23(2): 369-409.

[27] Kwedlo W, Kretowski M, De Raedt L, Flach P. An evolutionary algorithm for cost-sensitive decision rule learning [C] // Proceeding EMCL '01 Proceedings of the 12th European Conference on Machine Learning. London: Springer-Verlag, 2001: 288-299.

[28] Ling C X, Yang Q, Wang J, Zhang S. Decision trees with minimal costs [C] // Proceedings of the 21st International Conference on Machine Learning. Bantt, 2004: 425-432.

[29] Burez J, Van den Poel D. Handling class imbalance in customer churn prediction [J]. Expert Systems with Applications, 2009, 36(3): 4626-4636.

[30] Joos P, Vanhoof K, Ooghe H. Credit classification: A comparison of logit models and decision [C] // European Conference on Machine Learning. Chemnitz, 1998: 59-72.

[31] Gupta S, Zeithaml V. Customer metrics and their impact on financial performance [J]. Marketing Science, 2006, 26(5): 718-739.

[32] Zhou J X, Arnold M J, Pereira A, Yu J. Chinese consumer decision-making styles: A comparison between the coastal and inland regions [J]. Journal of Business Research, 2010, 63(1): 45-51.

[33] Chen Yen-Liang, Shen Ching-Cheng. Mining generalized knowledge from ordered data through attribute-oriented induction techniques [J]. European Journal of Operational Research, 2005, 66(1): 221-245.

[34] Cooil B, Keiningham L T, Aksoy L, Hsu M. A longitudinal analysis of customer share of wallet [J]. Journal of Marketing, 2007, 71(1): 67-83.

[35] Bolton R N, Kannan P K, Bramlett M D. Implications of loyalty program membership and service experiences for customer retention and value [J]. Journal of the Academy of Marketing Science, 2000, 28(1): 95-108.

基于代价敏感决策树的客户价值细分

邹　鹏[1], 莫佳卉[2], 江亦华[3], 叶　强[1]

1 哈尔滨工业大学 管理学院, 哈尔滨 150001

2 德克萨斯大学达拉斯分校 管理学院, 达拉斯 75080

3 加利福尼亚州立大学 商学院, 长滩 91840

摘要: 由于错误分类代价差异和不同价值客户数量的不平衡分布, 基于总体准确率的数据挖掘方法不能体现由于客户价值不同对分类效果带来的影响。为了解决错误分类不平衡的数据分类问题, 利用代价敏感学习技术扩展现有决策树模型, 将这一方法应用在客户价值细分, 建立基于客户价值的错分代价矩阵, 以分类代价最小化作为决策树分支的标准, 建立分类的期望损失函数作为分类效果的评价标准, 采用中国某银行的信用卡客户数据进行实验。实验结果表明, 与传统决策树方法相比, 代价敏感决策树对客户价值细分问题有更好的分类效果, 可以更精确地控制代价敏感性和不同种分类错误的分布, 降低总体的错误分类代价, 使模型能更准确反映分类的代价, 有效识别客户价值

关键词: 代价敏感学习; 不对称错分代价; 决策树; 客户价值细分

收稿日期: 2010-10-08　　**修返日期:** 2011-01-13

基金项目: 国家自然科学基金(70802019)

作者简介: 邹鹏(1975-), 男, 湖北麻城人, 毕业于哈尔滨工业大学, 获博士学位, 现为哈尔滨工业大学管理学院讲师, 研究方向: 客户关系管理, 商务智能等。E-mail: zoupeng@hit.edu.cn

□