

Decision Tree Algorithm for Precision Marketing via Network Channel

Yulan Zheng*

College of Economic and Management, Fuzhou University, Fuzhou, Fujian 350108, China

With the development of e-commerce, more and more enterprises attach importance to precision marketing for network channels. This study adopted the decision tree algorithm in data mining to achieve precision marketing. Firstly, precision marketing and C4.5 decision tree algorithm were briefly introduced. Then e-commerce enterprise A was taken as an example. The data from January to June 2018 were collected. Four attributes including age, income, occupation and educational background were selected for calculation and decision tree was established to extract classification rules. The results showed that the consumers of the products of the company were high-income young and middle-aged people, middle-income young people, middle-income middle-aged and elderly people with a college degree or above, low-income middle-aged people with a college degree or above and low-income elderly people with a state-owned enterprise. After precision marketing to these customers, it was found that the monthly sales volume of the enterprise increased by 22.82% and the marketing cost decreased by 28.21%, which verified the effectiveness and application value of precision marketing and showed that the decision tree algorithm could provide enterprises with decision support in precision marketing.

Keywords: data mining, decision tree, precision marketing, C4.5

1. INTRODUCTION

The rapid development of Internet and e-commerce has brought challenges to traditional marketing methods. Only accurate marketing methods can help enterprises not be eliminated in the fierce market price competition [Luo and Li (2014)]. Marketing activities are inseparable from data [Elsalamony (2014)]. In a large amount of marketing data, there is a lot of valuable customer information. Through the statistics, collection and analysis of these data [Ozyirmidokuz, Uyar and Ozyirmidokuz (2015)], precision marketing can be realized, which can reduce marketing costs and improve marketing efficiency [Abbas (2015); Mitik, Korkmaz, Karagoz et al. (2017)]. In order to extract valuable information from massive data, data mining technology has been widely applied [Karakatsanis, Alkhader, Maccrory et al. (2017)], such as decision tree, association rule, neural network, etc. [Abessi, Yazdi, Abessi et al. (2014)]. You et al.

[You, Si, Zhang et al. (2015)] selected attributes through the RFM model. Then important attribute values were identified by CHAID decision tree and Pareto value, and customer groups were classified. Next targeted marketing strategies was put forward. Finally, it was proved by the case study that the scheme could provide enterprises with good precision marketing strategies. Wan et al. [Wang, Fang and Wang (2015)] combined EM clustering analysis and neural network algorithm. Firstly, the sales data are preprocessed by EM clustering, and then classified by neural network to predict the purchase behavior of customers, which effectively improved the accuracy of direct marketing of enterprises. Hongda [Hongda (2017)] applied the decision tree algorithm to build a classifier model by collecting historical data of power users and to mine customer characteristics and preferences. Through the research on the promotion of palm power APP, it is found that this way of precision marketing was effective. Zou and Wang [Zou and Wang (2008)] introduced the merchant gene bank model to improve the competitiveness of e-commerce enterprises. A customer preference model was proposed based on optimal

*Email: yulanzh23@163.com

neighbor and utility function. Then the genetic algorithm was improved and the two models were matched to achieve precision marketing. In the era of consumer supremacy, it is of great practical significance to grasp customer information through data mining technology and conduct targeted marketing for improving marketing effect and promoting enterprise development. Therefore, it is of great value to study the application of data mining technology in precision marketing. At present, the application of decision tree in precision marketing is seldom, but the algorithm of decision number has low computational complexity and good effect on processing large data. In this study, the decision tree algorithm in data mining technology was used. C4.5 algorithm was used for classifying the collected customer historical data, constructing decision tree and extracting classification rules, so as to obtain customer categories with high purchase probability and conduct precision marketing on them. Moreover an example analysis was carried to verify the reliability of the decision tree algorithm. This work provides some theoretical supports for the further application of decision tree algorithm in precision marketing and is beneficial to the better and faster development of precision marketing in the field of e-commerce.

2. DECISION TREE ALGORITHM

Data classification is a kind of data mining. Through data classification, precise marketing can be realized by classifying customer information data. The methods of data classification include genetic algorithm, decision tree and neural network. Decision tree with simple structure and high efficiency has a very good performance in data prediction [Wang, Li, Ran et al., 2018] and data classification [Berhane, Lane, Wu et al. (2018)]. In this study, C4.5 algorithm [Li, Jiang, Yao et al. (2017)] in the decision tree algorithm is adopted to achieve precision marketing.

Training sample set is expressed as S . The sample number of the i -th result in category $C_i (i = 1, 2, \dots, m)$ is expressed as s_i . The probability that any sample belongs to is expressed as ρ_i . Then it was found that:

$$\rho_i = \frac{S_i}{S}, \tag{1}$$

and the expected information for the result is:

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m \rho_i \log_2(\rho_i). \tag{2}$$

Suppose there are u different values of D , $D = \{d_1, d_2, \dots, d_u\}$. Attribute D is used to divide training sample set S , and $\{s_1, s_2, s_3, \dots, s_u\}$ is obtained. Suppose that S_j contains training sample set S and there is a sample with value d_j in attribute D . If D is the best splitting property, then these samples are branches growing out of the nodes in training sample set S . S_{ij} is set as the number of samples belonging to C_i in subset S_j , and the expected information of the subset divided according to attribute D is:

$$E(D) = \sum_{j=1}^u \frac{s_{1j} + s_{2j} + s_{3j} + \dots + s_{mj}}{S} I(s_{1j}, s_{2j}, s_{3j}, \dots, s_{mj})$$

$$I(s_{1j}, s_{2j}, s_{3j}, \dots, s_{mj}) = - \sum_{i=1}^m \rho_{ij} \log_2(\rho_{ij}), \tag{3}$$

where $\rho_{ij} = \frac{s_{ij}}{|S_j|}$ represents the probability that a sample in S_j belongs to C_j .

The information gain of attribute D is

$$Gain(D) = I(s_1, s_2, s_3, \dots, s_m) - E(D) \tag{4}$$

The split information of attribute D is

$$SplitInfo(D) = - \sum_{j=1}^m \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{S} \log_2 \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{S}. \tag{5}$$

Then, the information gain rate of the attribute D is

$$GainRatio(D) = \frac{Gain(D)}{SplitInfo(D)}. \tag{6}$$

The attribute with the largest information gain rate is selected. The data set is divided into different subsets, and the decision tree is established.

3. PRECISION MARKETING BASED ON DECISION TREE ALGORITHM

3.1 Data Acquisition

E-commerce enterprise A is taken as the research subject. The transaction pattern of e-commerce enterprise A is mainly on the line, which helps enterprise A to achieve communication with customers through the network platform. Its marketing mode is mainly in the form of Internet advertising, which attracts customers to buy by placing advertisements in search engines, WeChat, weibo and other online channels. In order to achieve precision marketing, this paper classifies customers through the decision tree to determine which customers may buy the products of the enterprise, and is worth putting in the marketing plan. The browsing and purchasing information of enterprise products from January to June 2018 was collected, and SQL Server was used to establish a database.

3.2 Data Processing

Since there may be a large number of vacant values and incomplete data in the collected data, it is necessary to clean up these data and screen the data to select the information which is valuable to precision marketing, in order to improve the data quality and improve the accuracy of data classification. In the perspective of consumer privacy, private information such as family status and marital status were not selected. In this study, attributes including age, income, occupation and education background were selected to establish the decision tree. Y was used to represent the final purchase, and N was used to represent no purchase. The classification is shown in Table 1.

Table 1 Client attribute classification.

Customer number	Age	Income	Occupation	Education background	Classification
1	34	5600	Private enterprise	College degree	Y
2	36	6700	State-owned enterprises	College degree	Y
3	21	1800	Private enterprise	High school diploma	N
4	34	5800	State-owned enterprises	College degree	N
5	56	4500	Individual	High school diploma	N
6	27	3600	Private enterprise	College degree	Y
.....
n	X_n	X_n	X_n	X_n	Y_n

Table 2 Training sample set.

Customer number	Age	Income	Occupation	Education background	Classification
1	Middle-aged	Medium income	Non-state-owned enterprises	Below university degree	N
2	Youth	High income	State-owned enterprises	University degree or above	Y
3	Youth	Low income	Non-state-owned enterprises	University degree or above	N
4	The elderly	Low income	State-owned enterprises	University degree or above	Y
5	Youth	High income	Non-state-owned enterprises	University degree or above	Y
6	Youth	Medium income	Non-state-owned enterprises	University degree or above	Y
.....
97	The elderly	High income	State-owned enterprises	Below university degree	Y
98	Middle-aged	Low income	Non-state-owned enterprises	University degree or above	Y
99	Youth	Low income	Non-state-owned enterprises	Below university degree	N
100	Youth	High income	Non-state-owned enterprises	University degree or above	Y

3.3 Establishment of Decision Tree

One hundred records were randomly selected from the database as the training sample set. Discrete processing was performed on the data, and the age was divided into the following groups: (1) under 30 years old: young people; (2) 30–50 years old: middle-aged; (3) over 50 years old: the elderly; Income is divided into the following groups: (1) under 3000: low income; (2) 3000–5000: medium income; (3) over 5000 yuan: high income; The occupations are divided into the following groups: (1) state-owned enterprises; (2) non-state-owned enterprises; Academic qualifications are divided into the following groups: (1) university degree or above; (2) university degree or below. The training sample set shown in Table 2.

The training sample set is sorted according to attributes, and the results are shown in Table 3.

Among them, 46 records were classified as “Y” and 54 records were classified as “N”, so the expected information of the result “Y” is $I = 0.978$.

The expected information of different attributes is calculated, and the results are as follows:

$$E(\text{age}) = 0.912$$

$$E(\text{income}) = 0.842$$

$$E(\text{occupation}) = 0.936$$

$$E(\text{education}) = 0.901$$

The information gain of different attributes is

$$\text{Gain}(\text{age}) = I - E(\text{age}) = 0.978 - 0.912 = 0.066$$

$$\text{Gain}(\text{income}) = I - E(\text{income}) = 0.978 - 0.842 = 0.136$$

$$\text{Gain}(\text{occupation}) = I - E(\text{occupation}) = 0.978 - 0.936 = 0.042$$

$$\text{Gain}(\text{education}) = I - E(\text{education}) = 0.978 - 0.901 = 0.077$$

Table 3 Training sample set after classification.

Attribute		Y	N
Age	Youth	22	15
	Middle-aged	24	17
	The elderly	0	22
Income	Low income	1	32
	Medium income	18	17
	High income	27	5
Occupation	State-owned enterprises	22	28
	Non-state-owned enterprises	24	26
Education background	University degree or above	28	10
	Below university degree	18	44

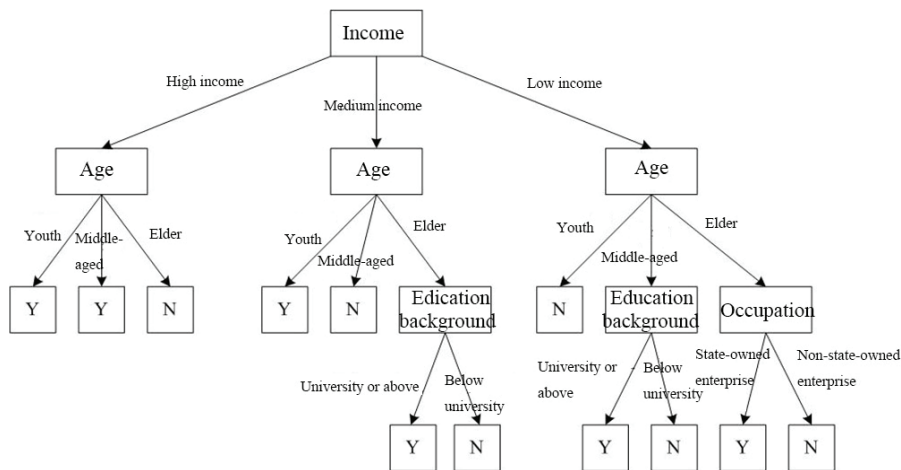


Figure 1 Precision marketing decision tree

The information gain rate is

$$GainRation(age) = 0.066$$

$$GainRation(income) = 0.091$$

$$GainRation(occupation) = 0.053$$

$$GainRation(education) = 0.056$$

It can be found that the attribute “income” has the largest information gain rate. Therefore, it is taken as the first level attribute of the decision tree, and then the classified samples are calculated again. The final decision tree is shown in the Figure 1.

According to the decision tree, the following classification rules can be extracted:

- (1) IF income = high income AND age = youth THEN category = purchase
- (2) IF income = high income AND age = middle-aged THEN category = purchase
- (3) IF income = high income AND age = the elderly THEN category = no buy
- (4) IF income = medium income AND age = youth THEN category = purchase
- (5) IF income = medium income AND age = middle-aged THEN category = no buy
- (6) IF income = medium income AND age = the elderly AND education = university degree or above THEN category = purchase
- (7) IF income = medium income AND age = the elderly AND degree = below university degree THEN category = no buy
- (8) IF income = low income AND age = youth THEN category = no buy
- (9) IF income = low income AND age = middle-aged AND education = university degree or above THEN category = purchase
- (10) IF income = low income AND age = middle-aged AND degree = below university degree THEN category = no buy
- (11) IF income = low income AND age = the elderly AND occupation = state-owned enterprise THEN category = purchase
- (12) IF income = low income AND age = the elderly AND occupation = non-state-owned THEN category = no buy

According to the above classification rules, it can be found that youth and middle-aged people with high income, youth with middle income, middle-aged and elderly people with university degree or above, middle-aged people with low-income and university degree or above and low-income elderly people in state-owned enterprises will buy the products of this enterprise,

Table 4 Changes of monthly sales volume and marketing cost before and after precision marketing.

Time	August, 2018	September, 2018
Monthly sales volume/n	6214	7632
Marketing cost/ten thousand yuan	2.34	1.68

i.e., precision marketing can be carried out for them to achieve good marketing effects.

3.4 Analysis of Precision Marketing Effect

In order to verify the effect of precision marketing, e-commerce enterprise A applied decision tree to conduct precision marketing to customers in September 2018. The monthly sales volume and marketing cost of August and September were compared. The results are shown in Table 4.

As shown in Table 4, through precision marketing, the monthly sales volume of e-commerce enterprise A in September 2018 is 7,632, which is 22.82% higher than that before precision marketing, and the marketing cost is 28.21% lower than that of last month, which suggests the good effect of precision marketing. After the implementation of precision marketing, the marketing cost of the enterprise is significantly reduced. This is because the enterprise only carries out the marketing plan for the designated customers, which largely avoids the invalid marketing plan and reduces the marketing cost. In addition to the reduction of marketing cost, the monthly sales volume of the enterprise has significantly increased, which indicates that the limited marketing plan has achieved good marketing effect and the precision marketing for customers with high purchase probability is practical and effective.

4. DISCUSSION

With the development of Internet technology, the ways of transaction and marketing have changed a lot. There are increasing transactions conducted through the Internet between enterprises, and between enterprises and consumers, with the rapid development of e-commerce. Enterprises also increasingly use the Internet as a tool to promote marketing [Andreopoulou, Tsekouropoulos, Koliouka et al. (2014)]. As the transactions go on, a lot of data is generated in the Internet. In the era of big data, how to use data for effective marketing is a problem widely valued by enterprises. With the transformation of marketing mode from traditional broadcast network to precision marketing, precision marketing oriented to network channels is also welcomed by more and more enterprises.

Precision marketing is a form of one-to-one marketing. It takes consumers as the center and pushes the influence scheme to the target audience accurately through the accurate positioning of the product target audience [Li (2014)]. Precision marketing oriented to network channels can show the enterprise's marketing plan to the audience through static web pages, dynamic web pages, floating Windows, AD links and other forms [Han (2015)], which not only saves the marketing cost to a large extent, but also can effectively improve the marketing effect and obtain greater returns.

Customer classification is the key to precision marketing. In the past, customers' information was often collected through questionnaires to understand their needs and classify them. However, this method is not only time-consuming and labor-intensive, but also does not have high accuracy of classification under the influence of various factors, so it cannot provide valuable decision support for enterprises. In the era of big data, especially transactions formed on the basis of the Internet have accumulated a large amount of data, which contains a lot of valuable information. If this part of information can be fully statistical, analyzed and used, you can get a lot of information beneficial to the enterprise. The development of technology makes the storage and analysis of massive data a reality. Data mining technology refers to the process of extracting potential and valuable information and knowledge from a large number of fuzzy and random data [Abessi, Yazdi, Abessi et al. (2014); Nawani, Gupta and Thakur (2014)]. It is widely used in finance [Moradi and Mokhatab Rafiei (2019)], industry [Borzooei, Teegavarapu, Abolfathi et al. (2019)], biomedicine [Huang, Chen, Liu et al. (2019)] and other fields, and can also provide good help for enterprise customers and market segmentation [Dutta, Bhattacharya and Guin (2014)], so as to achieve precision marketing.

In this study for precision marketing, the decision tree algorithm is selected to realize the accurate classification of customers. Decision tree algorithm is a method for classification and prediction. Its fast classification speed and high classification accuracy are widely favored by enterprises. Among them, C4.5 algorithm is one of the better decision tree algorithms. This paper first introduces the C4.5 algorithm, and then takes e-commerce enterprise A as an example to realize precision marketing through the decision tree algorithm. By classifying the customer information collected according to the four attributes of age, income, occupation and education background, and then calculating the information gain rate, it is found that in the four attributes, the information gain rate is in the order of income, age, education background and occupation from the largest to the smallest. Therefore, the decision tree is constructed with income as the first level attribute. By further calculating the final decision tree and extracting the classification rules, it can be found out that which customers are more likely to buy products and which customers are less likely to buy products. Then, the marketing plans are designed and pushed to customers with high purchase probability, which will further increase the purchase possibility of these customers and realize precision marketing. It was found from the analysis of precise marketing effect that the monthly sales volume of Enterprise A reached 6214 pieces, and the marketing cost reached 234,000 yuan before using precise marketing; after using precise marketing, the monthly sales volume of Enterprise A was 7632, increased by 22.82%, while the marketing cost was 168,000 yuan, decreased by 28.21%, indicating the application of precise marketing could effectively reduce the marketing cost of enterprises, and targeted marketing strategies also promoted the growth of enterprise

sales. The results proved the effectiveness of the precise marketing method.

Although this study has obtained some achievements, there are still some shortcomings, for example, the enterprise data obtained in the case analysis were not comprehensive, and the data storage and customer classification method need further research, which are the direction of future work.

5. CONCLUSION

In this study, C4.5 decision tree algorithm in data mining is applied to realize precision marketing for network channels. Taking e-commerce enterprise A as an example, A decision tree is constructed and classification rules are extracted through the calculation and analysis of collected customer information. Eventually, a customer category with a high probability of purchase is obtained. These customers are then marketed with precision. It is found that the sales volume of the enterprise significantly improves, and the marketing cost reduces, indicating that the decision tree constructed by using C4.5 algorithm is effective and reliable for precision marketing. The decision tree constructed by using C4.5 algorithm provides a basis for the marketing decision of enterprises and has great application value.

REFERENCES

1. Abbas, S. (2015): Deposit subscribe Prediction using Data Mining Techniques based Real Marketing Dataset. *International Journal of Computer Applications*, vol. 110, no. 3, pp. 1–7.
2. Abessi, M.; Yazdi, E. H.; Abessi, M.; et al. (2014): Marketing Data Mining Classifiers: The Criteria Selection Issues in Customer Segmentation. *International Journal of Computer Applications*, vol. 106, no. 10, pp. 5–10.
3. Abessi, M.; Yazdi, E. H.; Abessi, M.; et al. (2014): Marketing Data Mining Classifiers: The Criteria Selection Issues in Customer Segmentation. *International Journal of Computer Applications*, vol. 106, no. 10, pp. 5–10.
4. Andreopoulou, Z.; Tsekouropoulos, G.; Koliousska, C.; et al. (2014): Internet marketing for sustainable development and rural tourism. *International Journal of Business Information Systems*, vol. 16, no. 4, pp. 446–461.
5. Berhane, T.M.; Lane, C.R.; Wu, Q.S.; Autrey, B.; Anenkhonov, O.; Chepinoga, V.; Liu, H. X. (2018): Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. *Remote Sensing*, vol. 10, no. 4, pp. 580.
6. Borzooei, S.; Teegavarapu, R.; Abolfathi, S.; Amerlinck, Y.; Nopens, I.; Zanetti, M. C.: (2019): Data mining application in assessment of weather-based influent scenarios for a WWTP: getting the most out of plant historical data. *Water Air and Soil Pollution*, vol. 230, no. 1, pp. 5.
7. Dutta, S.; Bhattacharya, S.; Guin, K. K. (2014): Data Mining, in Market Segmentation: A Literature Review and Suggestions. *Advances in Intelligent Systems & Computing*, vol. 335, pp. 87–98.
8. Elsalamony, H. (2014): Bank Direct Marketing Analysis of Data Mining Techniques. *International Journal of Computer Applications*, vol. 85, no. 7, pp. 12–22.
9. Han, Z. (2015): Design and Implementation of Novel Precision Internet Marketing Patterns under the Big Data and Cloud Environment. *International Journal of Technology Management*, vol. 2015, no. 7, pp. 86–88.
10. Hongda, Z. (2017): Analysis of Electronic Channel Precision Marketing Efficiency Based on Decision Tree Algorithm. *Zhejiang Electric Power*, vol. 36, no. 12, pp. 22–26.
11. Huang, Q.; Chen, Y.; Liu, L.; Tao, D.; Li, X. (2019): On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification. *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1.
12. Karakatsanis, I.; Alkhader, W.; Maccrory, F.; et al. (2017): Data Mining Approach to Monitoring The Requirements of The Job Market: A Case Study. *Information Systems*, vol. 65, pp. 1–6.
13. Li, Y. (2014): Research on the Development Trend of Precision Marketing under the Background of China Mobile Internet. *Value Engineering*, vol. 7, no. 3, pp. 355–361.
14. Li, Y.; Jiang, Z. L.; Yao, L.; et al. (2017): Privacy -preserving C4.5 decision tree algorithm over horizontal and reflective dataset among multiple parties. *Cluster Computing*, vol. 2017, no. 2, pp. 1–13.
15. Luo, F. L.; Li, S. (2014): Precision Marketing and Modern Information Technology. *Applied Mechanics and Materials*, 687–6914, pp. 4.
16. Mitik, M.; Korkmaz, O.; Karagoz, P.; et al. (2017): Data Mining Approach for Direct Marketing of Banking Products with Profit/Cost Analysis. *Review of Socionetwork Strategies*, vol. 11, no. 1, pp. 1–15.
17. Moradi, S.; Mokhtab Rafiei, F. (2019): A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation*, vol. 5, no. 1, pp. 15.
18. Nawani, A.; Gupta, H.; Thakur, N. (2014): Prediction of Market Capital for Trading Firms through Data Mining Techniques. *International Journal of Computer Applications*, vol. 70, no. 18, pp. 7–12.
19. Ozyirmidokuz, E. K.; Uyar, K.; Ozyirmidokuz, M H. (2015): A Data Mining Based Approach to A Firm's Marketing Channel. *Procedia Economics & Finance*, vol. 27, pp. 7–84.
20. Wang, J.D.; Li, P.; Ran, R.; Che, Y.B.; Zhou, Y. (2018): A short-term photovoltaic power prediction model based on the gradient boost decision tree. *Applied Sciences*, vol. 8, no. 5, pp. 689.
21. Wang, W. C.; Fang, Y. J.; Wang, C.; et al. (2015): Application of EM Neural Network Mining Technology for Enterprise Direct Marketing. *Journal of Hunan Industry Polytechnic*, vol. 15, no. 5, pp. 15–17.
22. You, Z.; Si, Y. W.; Zhang, D.; et al. (2015): A decision-making framework for precision marketing. *Expert Systems with Applications An International Journal*, vol. 42, no. 7, pp. 3357–3367.
23. Zou, Q. Y.; Wang, X. F. (2008): Research and implementation of improved genetic algorithm based precision marketing in entity commerce. *Modern Electronics Technique*, vol. 41, no. 13, pp 177–181.