



Electric Power Customer Positioning Model Based on Decision Tree

Yingchun Guo

Department of Economics and Management, North China Electric Power University, Baoding 071003, China

College of Mathematics & Computer Science, Hebei University, Baoding 071002, China

E-mail: guoyc@hbu.edu.cn

Dongxiao Niu

Department of Economics and Management, North China Electric Power University, Baoding 071003, China

The research is supported by National Natural Science Foundation of China (No. 70671039). (Sponsoring information)

Abstract

The electric power consumer positioning analysis is a very important aspect of the electric power consumer relationship management, and the decision tree is a usual tool to implement classification analysis. Based on the segmentation of electric power consumers, this article mainly discusses how to construct the decision tree of electric power consumer positioning by ID3 algorithm, introduce this concrete operation process, and accordingly offers technical supports for the electric power marketing department to better hold consumer characters and implement individual services.

Keywords: Electric power consumer relationship management, Electric power consumer positioning, Decision tree, ID3 algorithm

1. Introduction

For electric power companies, consumer positioning analysis and developing potential consumers are very important. We divide electric power consumers into four classes, i.e. important consumer, risk consumer, developmental consumer and common consumer, and different classified consumers need different marketing strategies. For example, those important consumers with good credits will acquire more and better services, but for those risk consumers with bad credits, we should adopt certain risk prevention measures. Then when a new consumer occurs, we should judge the consumer's possible class according to his special behavior characters and adopt sales strategy with strong pertinences to obtain maximum benefits. But how can we predict the new consumer's possible class objectively, exactly and quickly? One scientific and quick method is to apply decision tree, mine consumer data in virtue of computer and make objective positioning analysis for the consumer.

In classification analysis, the decision tree model is one of the most popular models because it can expediently show the mining results by means of figure (tree structure) and suit decision-making for the management department of the enterprise. In this article, we will discuss how to apply the ID3 algorithm to construct the decision tree of electric power consumer positioning and introduce this concrete operation process.

2. The basic construction process of decision tree

The decision tree is a data structure similar to flow chart and every interior node denotes an attribution test. Every branch denotes a test output and every leaf node denotes class or class distribution.

The operation process of decision tree is seen in Figure 1.

- (1) The system collects consumer information from various nodes in the interior network, settles and integrates data information, and forms the consumer information resources with uniform structure.
- (2) Pretreat data resources and eliminate attributions independent of decisions, approximate numerical value attribution and deal with those attributions with vacant values, and form the training set of decision tree.
- (3) Train the training set, compute information increase of every attribution, select the attributions with maximum information increase as the present main attribution node, and construct a branch for every value of this attribution. Recursively implement above process to sample subsets included in this sub-node until the values of data records on the main attribution in the subset are same, or no attributions can be divided, and form the induction decision tree.
- (4) Extract classification rules from obtained decision tree. Establish a rule for every approach from root to leaf and form the rule set. Display the rule set to consumers, and store feasible rules selected by consumer into the rule database.
- (5) When new consumer logs in the company website, the system will analyze the new consumer's information material

by means of rules obtained from decision tree, predict which class the consumer belongs to, give the consumer a reasonable position, and accordingly offer assistants to establish individuation sales strategy for the company.

3. The ID3 algorithm to implement decision tree classification

In various decision tree classification algorithms, the most influential one is the ID3 algorithm put forward by Quinlan in 1979, which takes the descending speed of entropy as the standard to test attribution. This algorithm utilizes samples of known classes to give sequence to test attribution until all samples are classified exactly.

In the forming process of decision tree, we utilize one method of information theory, and this method can seek and offer maximum information increase at any time, i.e. it can offer the attribution with maximum entropy decrease. The entropy is denoted by character E .

To N samples, they are divided and respectively belong to class $C_i (i = 1, 2, \dots, c)$, the sample number in class C_i is N_i , every sample has K attributions, and every attribution has J_k values. The construction process of decision tree can be described as follows.

(1) Compute initial entropy.

$$E(I) = \sum_{i=1}^c -(N_i / N) / \log_2(N_i / N) = \sum_{i=1}^c -p_i \log_2 p_i$$

For the training set, the classes of all samples are known, so N samples compose the initial entropy of the system.

(2) Select one attribution as the root node of the decision tree.

First, for every attribution $A_k (k = 1, 2, \dots, K)$, divide original samples into the first class sample set according to J values a_{kj} of the attribution A_k . Though the branch of a_{kj} includes n_{kj} samples, but they may not belong to single class.

Second, for n_{kj} samples of every branch, the sample number belonging to class C_i is $n_{kj}(i)$, and the following formula can compute the entropy of this branch.

$$E(I, A_k, J) = \sum_{i=1}^c [-n_{kj}(i) / n_{kj}] \log_2 [n_{kj}(i) / n_{kj}]$$

$$E(I, A_k) = \sum_{j=1}^J (n_{kj}(i) / N) \sum_{i=1}^c [-n_{kj}(i) / n_{kj}] \log_2 [n_{kj}(i) / n_{kj}]$$

Third, compute the entropy decrease induced by testing attribution A_k , i.e. $\Delta E(k) = E(I) - E(I, A_k)$.

Fourth, select the attribution A_{k_0} which can produce maximum entropy decrease, i.e. A_{k_0} should fulfill the condition that $\Delta E(k_0) > \Delta E(k), (k = 1, 2, \dots, K, \text{ and } k \neq k_0)$.

Fifth, the attribution A_{k_0} is just the root of the decision tree.

(3) The attribution A_{k_0} will produce J_{k_0} leaf nodes, and based on that the sample set is divided into J_{k_0} subsets. To the sample subset of every leaf node, select one attribution $A_{k \neq k_0}$ as the down class utilizing above method in turn to make this leaf node obtain maximum entropy decrease.

(4) Continually construct the down class of the decision tree according to step (3) until all sample subsets have only one class, which indicates the entropy of the system is zero, and the construction process of the decision tree ends.

4. The decision tree model of electric power consumer positioning

4.1 Construction of decision tree

Before the decision tree is constructed, we must find out the main attribution of the decision tree. The main attribution of the decision tree is confirmed by the actual situation. The consumer positioning analysis through decision tree is mainly based on the consumer's class character, which attention is to find out consumer's class and differences through consumer's character index and action index and institute corresponding marketing strategies aiming at different consumers. So the main attribution of the consumer positioning decision tree

should be "consumer class", i.e. the consumer class through the consumer segmentation based on the consumer character. The consumer classes are divided into the important consumer (class I), the risk consumer (class II), the developmental consumer (class III) and the common consumer (class IV).

Based on the collection and analysis of electric power consumer data, we select three representative attributions, consumer power consumption level, consumer credit degree and consumer potential. For the selection of data, not all data can apply the requirements, and the data required by decision tree should have no noises and absences, so data

need to be collected and disposed. The collection and disposal can not only integrate data of different departments and different marketing stations, but also can approximate data, i.e. replace initial data on the low level to the concept on the high level, in order to implement data mining. Here, we approximate actual consumer power consumption level into high, middle and low stages, approximate consumer credit degree into high, middle and low classes, and approximate consumer potential into good, middle and bad sorts.

Through actual researches on the consumer data of Jinan Electric Power Company, we select the data from 20 consumers, dispose and approximate these data, and the result is seen in Table 1.

According to above construction process of decision tree, the actual computation process can be described as follows.

(1) Compute the initial entropy value of the sample class.

$$E(I) = -\frac{6}{20} \log_2 \left(\frac{6}{20} \right) - \frac{3}{20} \log_2 \left(\frac{3}{20} \right) - \frac{7}{20} \log_2 \left(\frac{7}{20} \right) - \frac{4}{20} \log_2 \left(\frac{4}{20} \right) = 1.9261.$$

(2) Compute the information entropy of every attribution.

If taking the power consumption level as the testing attribution, so we have

$$E(I, \text{level}) = \frac{5}{20} \left[-\frac{4}{5} \log_2 \left(\frac{4}{5} \right) - \frac{1}{5} \log_2 \left(\frac{1}{5} \right) \right] + \frac{6}{20} \left[-\frac{2}{6} \log_2 \left(\frac{2}{6} \right) \times 3 \right] \\ + \frac{9}{20} \left[-\frac{5}{9} \log_2 \left(\frac{5}{9} \right) - \frac{4}{9} \log_2 \left(\frac{4}{9} \right) \right] = 1.1020.$$

If taking the credit degree as the testing attribution, so we have

$$E(I, \text{credit}) = \frac{7}{20} \left[-\frac{2}{7} \log_2 \left(\frac{2}{7} \right) - \frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{4}{7} \log_2 \left(\frac{4}{7} \right) \right] \times 2 \\ + \frac{6}{20} \left[-\frac{3}{6} \log_2 \left(\frac{3}{6} \right) - \frac{1}{6} \log_2 \left(\frac{1}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right] = 1.4029.$$

If taking the potential as the testing attribution, so we have

$$E(I, \text{potential}) = \frac{6}{20} \left[-\frac{3}{6} \log_2 \left(\frac{3}{6} \right) - \frac{1}{6} \log_2 \left(\frac{1}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right] \\ + \frac{6}{20} \left[-\frac{1}{6} \log_2 \left(\frac{1}{6} \right) \times 2 - \frac{4}{6} \log_2 \left(\frac{4}{6} \right) \right] + \frac{8}{20} \left[-\frac{3}{8} \log_2 \left(\frac{3}{8} \right) \times 2 - \frac{1}{8} \log_2 \left(\frac{1}{8} \right) \times 2 \right] = 1.5377.$$

(3) Respectively compute information plus of various attribution.

$$\Delta E(\text{level}) = E(I) - E(I, \text{level}) = 1.9261 - 1.1020 = 0.8241$$

$$\Delta E(\text{credit}) = E(I) - E(I, \text{credit}) = 1.9261 - 1.4029 = 0.5232$$

$$\Delta E(\text{potential}) = E(I) - E(I, \text{potential}) = 1.9261 - 1.5377 = 0.3884$$

(4) Because the “power consumption level” has maximum information plus, so we take it as the first class attribution of the decision tree, and hereby divide this sample set into large, middle and small subsets, produce three leaf nodes, apply above method to every leaf node in turn, so the decision tree seen in Figure 2 is formed.

4.2 Rule extraction of decision tree

Establish a rule for every approach from root to leaf, form classification rule with the form of “If-Then”, and compose the rule set. Form a conjunction item for the part of “If” along every pair of “attribution-value” on the confirmed approach, the leaf nodes including class prediction forms the part of “Then”, and store rules into the rule base.

From Figure 2, we can confirm approximate deployment of character attribution of four sorts of consumer, and according to known information, we can approximately know these consumers’ positions in four sorts from these characters. We present the information of the tree diagram by means of the form of “If-Then”. For example,

(1) If the power consumption level = “high” and (the credit degree = “high” or the credit degree = “middle”), then the consumer belongs to class I, i.e. the important consumer.

(2) If the power consumption level = “high” and the credit degree = “low”, then the consumer belongs to class II, i.e. the risk consumer.

(3) If the power consumption level = “low” and the credit degree = “high” and (the potential = “good” or the potential = “middle”), then the consumer belongs to class III, i.e. the developmental consumer.

(4) If the power consumption level = “low” and the credit degree = “high” and the potential = “bad”, then the consumer belongs to class IV, i.e. the common consumer.

4.3 Application of decision tree

When new consumer logs in the company website, the system will record the new consumer’s information, analyze and position the new consumer’s information material by means of rules obtained from decision tree, put in the positioning result to the decision-making personnel for the company, and accordingly offer assistants to establish individuation sales strategy for the company. For example,

(1) If the new consumer belongs to class I, i.e. the important consumer, so this consumer is the key consumer locked by the electric marketing department, the key object that the electric marketing department adopts individuation service, because he will bring large economic benefits for the electric power company.

(2) If the new consumer belongs to class II, i.e. the risk consumer, so the consumer is not stable, and the electric marketing department should adopt measures to prevent risks according to actual situation, such as reducing power fees return cycle or requiring offering caution money and so on.

(3) If the new consumer belongs to class III, i.e. the developmental consumer, so this consumer is the key development object for the electric marketing department, and the electric company should follow the consumer’s development process, dig potential, and timely offer individual power consumption strategy suiting for the consumer, such as implementing favor policies, advising changing power consumption period and so on.

(4) If the new consumer belongs to class IV, i.e. the common consumer. Though the consumer can not bring high economic benefits, but because this sort of consumer occupies large proportion, so the electric marketing department also should offer good client services and try to fulfill consumer’s demands to enhance social benefits of the electric power company.

5. Conclusions

The electric power consumer relationship management is a sort of new management mechanism which aim is to strengthen the relation between the electric power company and consumers. The key in the research of consumer relationship management is how to utilize the data mining technology to deeply analyze consumer data, know consumer character, dig consumer potential, and offer appropriate products and services at the appropriate time. In this article, based on the collection and analysis to electric power consumer information, we put forward constructing the electric power consumer positioning decision tree by means of the ID3 algorithm, which can mine the positioning knowledge about consumers, fully realize consumers’ demands of difference and levity, make the marketing strategy made by the electric power marketing department more suit for the demands of the market and change with consumers’ demands, enhance service quality, and accordingly acquire higher economic and social benefits for the electric power company.

References

- Berson A, Stephen S & Kurt T, interpreted by Heqi. (2001). *Construct Data Mining Application Facing CRM*. Beijing: People's Post and Telecommunication Press.
- Diao, Baiqing and Zhou, Zunguo. (2004). *Consumer Relationship Management under the Environment of Electric Power Reform*. Beijing: China Electric Power Press. December, 2004.
- Jiawei Han & Micheline Kamber, interpreted by Fanming and Meng, Xiaofeng. (2001). *Data Mining: Concepts and Technology*. Beijing: China Machine Press. August, 2001.
- Kangjian, Liang, Yunrong. (2003). Application of Classification Data Mining in Financial Customer Relationship Management. *Journal of Beijing Institute of Technology (Natural Science Edition)*. 23 (2). p. 207-211.
- Michael J. S, Chandrasekar S & Gek W. T. (2001). Knowledge Management and Data Mining for Marketing. *Decision Support Systems*. 31(1). p. 127-137.
- S.C. Hui & G. Jha. (2000). Data Mining for Customer Service Support. *Information and Management*. 38 (1). p. 1-13.
- Wan, Zhihua and Xu, Zhongjian. (2004). Application of Data Mining in CRM. *Computer Engineering and Design*. 25 (12). p. 2324-2326.
- Zhou, Danchen, Yin, Guofu & Long, Hongneng. (2004). Research on Application of Data Classification Technology for Knowledge Discovery in CRM. *Computer Applications and Software*. 21 (2). p. 9-11.

Table 1. Data collection of consumers

Consumer number	Level	Credit	Potential	Class
1	high	high	middle	important consumer
2	high	high	bad	important consumer
3	high	middle	middle	important consumer
4	high	middle	bad	important consumer
5	high	low	bad	risk consumer
6	middle	high	good	important consumer
7	middle	high	middle	important consumer
8	middle	middle	good	developmental consumer
9	middle	middle	middle	developmental consumer
10	middle	low	good	risk consumer
11	middle	low	middle	risk consumer
12	low	high	good	developmental consumer
13	low	high	middle	developmental consumer
14	low	high	bad	common consumer
15	low	middle	good	developmental consumer
16	low	middle	middle	developmental consumer
17	low	middle	bad	common consumer
18	low	low	good	developmental consumer
19	low	low	middle	common consumer
20	low	low	bad	common consumer

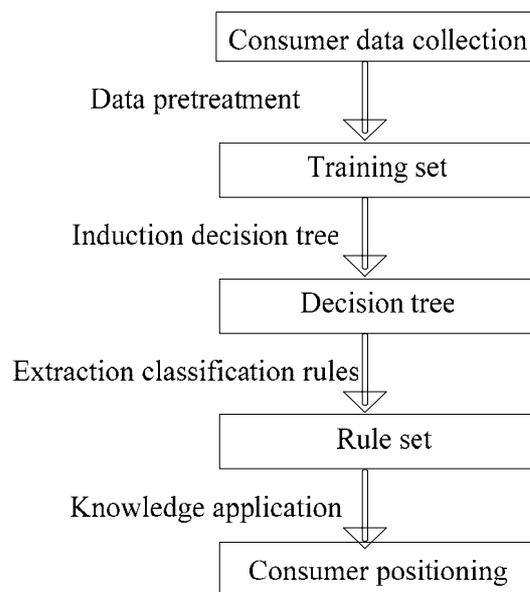


Figure 1. Forming Process of Decision Tree

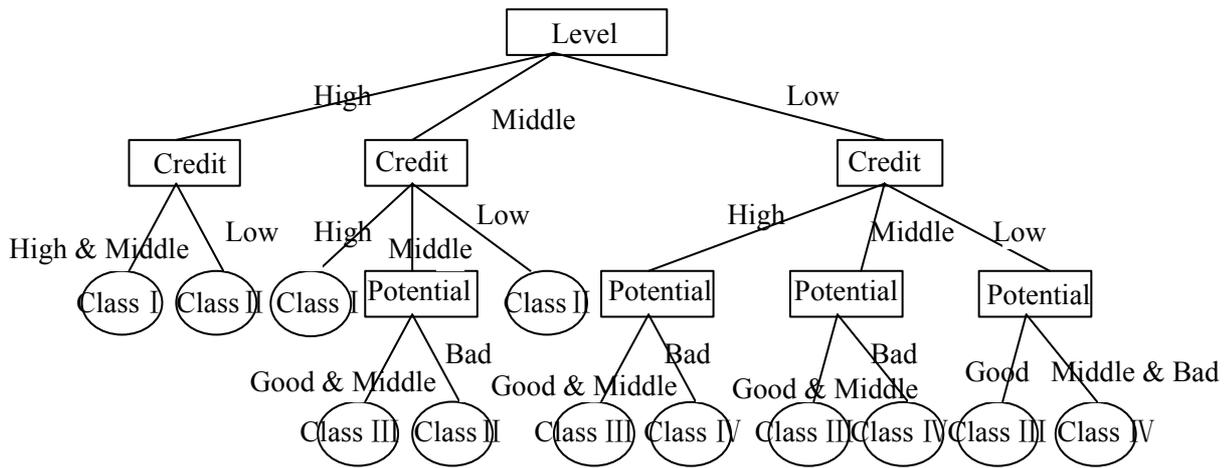


Figure 2. Consumer Positioning Decision Tree