A Study on the Application of Decision Tree Algorithm in Mobile Marketing

Danning He1,*

1Sichuan Vocational College of Culture and Communication, 611230,Chongzhou, Sichuan, People R China.

*Corresponding author e-mail: hedanningl@scvc.edu.cn

Abstract. The decision tree algorithm is an inductive learning algorithm based on examples. It infers a classification rule represented by a tree structure through the learning of the training set. Mobile telecom operators have accumulated a large amount of user information in their long-term operations. Using decision tree algorithms to perform data mining on user information can accurately analyze user needs and user satisfaction with services. Based on the research of decision tree algorithm, this paper conducts analysis and mining based on mobile user data in order to improve the level of mobile marketing services and implement customer refined management strategies.

Keywords: Decision Tree Algorithm, Data Mining, Marketing

After entering the information age, the mobile telecommunications market has ushered in an open and highly competitive new business collaboration environment. Telecom operators insist on customer demand-oriented when conducting marketing. While maintaining the development of traditional services, they also formulate targeted marketing methods through user data mining. A large amount of customer demand information is carried in customer data. How to mine valuable information from the massive information depends on data mining technology. From the perspective of application, data mining technology mainly realizes the classification, prediction, estimation, association grouping, and homogeneous grouping of target data. Common data mining methods include decision tree algorithm, KNN algorithm, SVM method, Bayesian method, etc. Different algorithms follow different principles. The specific collection is as follows: Table 1.
Table 1. Data mining methods and ideas

| Number | Data Mining Method   | Design Idea                                                                 |
|--------|----------------------|-----------------------------------------------------------------------------|
| 1      | Decision tree algorithm | Generates an easy-to-understand tree structure and classification rules by inductively learning training data. |
| 2      | KNN algorithm         | By calculating the distance between the current data and the other data in the data set, find the closest k data, and judge the category of the current data according to the majority of the k data. |
| 3      | SVM method            | Find a hyperplane in the data set that can separate the data set. At the same time, it is required that the vertical distance between this plane and the boundary of the data set should be the largest. This maximizes the distance between the classes in the data set. |
| 4      | Bayesian method       | First find the probability of the data under each possible classification condition, and then select the category with the highest probability is the category of the current data. |

1. The Application Status of Data Mining in Mobile Marketing

1.1. The Development History of Data Mining Technology

Data mining technology was first applied to the financial industry and telecommunications industry, and played an important role in this field. Financial operation requires accurate and effective data support, which requires data mining technology to have a complete theoretical foundation. With the rapid development of the mobile telecommunications industry, a large amount of user data has been accumulated in a short period of time. The use of data mining technology to find valuable information can provide guarantee for the sustainable development of the mobile telecommunications industry. For example, building customers through data mining technology and the churn prediction model develops an effective retention strategy for users who may churn.

In recent years, under the guidance of user-oriented strategies, on the one hand, it has promoted the innovative development of data mining technology. On the other hand, it has accelerated the integration of data mining technology and the telecommunications industry. Traditional single data mining models (for example, customer relationship management models, customer segmentation models) have become common in the mobile telecommunications industry, but after all, a single data mining model has a limit to the degree of user segmentation. So multiple models must be developed. Only the combined data mining technology can meet the future application needs of the mobile telecommunications industry.

1.2. Application Status of Data Mining Technology

At present, the application of data mining in the telecommunications industry is mainly concentrated in user segmentation, churn user prediction, customer relationship management, etc., and less involved in the field of mobile marketing. From a practical point of view, using data mining technology to understand the needs of customers in the mobile market, customer segmentation based on this, and formulating corresponding marketing strategies for customers with different needs can effectively
increase customer retention. In this context, this article applies data mining technology to analyze the basic attributes of mobile market customers to predict potential customers, fake customers and sticky customers, and provide corresponding products and services to different customers, so as to achieve better marketing effects.

2. Introduction and Implementation of Decision Tree Algorithm

2.1. Introduction to Decision Tree Algorithm

The decision tree algorithm analyzes the internal information of the known classification data set, thereby extracting a set of formal and simple classification rules, which are expressed as a tree structure. And each node corresponds to a decision point of a certain attribute, which leads to There are multiple classification branches, and each branch accurately classifies the data. When it reaches the leaf node of the decision tree, the data can be classified completely and accurately.

The core advantage of the decision tree algorithm is that it can learn from known historical data, thereby establishing a tree model that can reveal the internal information and rules of the data and has a high information density. According to this model, the target data can be classified. After a quick analysis of the classification data set, a simple, intuitive, clear and understandable tree structure model is established based on the decision tree algorithm. Any branch between the root node and the leaf node marks an accurate classification rule. In addition, the decision tree algorithm has strong scalability, not only can quickly process small data sets, but also can deal with large data sets.

The current common decision tree algorithms include ID3 algorithm, C4.5 algorithm, SLIQ algorithm, SPRINT algorithm and so on. Among them, the ID3 algorithm follows the idea that small decision trees are better than large decision trees, and uses a top-down recursive method to construct trees. At the same time, the ID3 algorithm also borrows the idea of information entropy in information theory, by calculating the information entropy of each attribute in the current internal node classification in the candidate attribute set. So it can obtain the attribute with the smallest information entropy and use this attribute as Classification basis. In addition, it can obtain classification results containing valuable information.

2.2. Implementation of Decision Tree Algorithm

First, we create a node to distinguish the attribute value of the data partition label. If the attribute value of the partition data label belongs to the same type, return to the node and select this type of label. If the candidate set is an empty set, return to the node and select. Most classes mark this node, otherwise the information entropy of each candidate attribute under this node needs to be solved. The minimum information entropy is selected by comparing the information entropy, and a branch is created for each value according to the attribute of the minimum information entropy, so as to achieve the purpose of expanding the decision tree. Then call its own method on each branch to create its own child nodes, and finally generate a decision tree.

In the implementation path of the ID3 algorithm, the key link is to first calculate the information entropy of each attribute in the candidate attribute set, and finally select the attribute with the smallest information entropy value as the classification attribute through information entropy comparison. In this regard, suppose there is a sample data set X, the size of the data set is represented by |X|, the X set contains a target attribute R, which contains m values, and the sample data set X is divided into m data according to the target attribute R Subset, \( P(X_i) \) represent the probability that each data in the X set belongs to the i-th subset. At this time, the expected information required to classify the data of the data sample set X is the formula (1):
\[ \text{Info} (X) = -\sum_{i=1}^{n} P(X_i) \log_2 P(X_i) \]  

(1)

However, the actual situation is that these branches contain a lot of data of other classes, so in order to get accurate classification, you need to get the entropy of the subset:

\[ \text{Info}_r(X) = \sum_{j=1}^{n} \frac{|X_j|}{|X|} \times \text{Info} (X) \]  

(2)

Information gain is the difference between the two demands:

\[ \text{Gain} (T) = \text{Info}(X) - \text{Info}_r(X) \]  

(3)

Gain \( (T) \) represents the information increment after dividing by the attribute \( T \), and the attribute with the highest increment is selected as the best attribute for splitting. According to formula (3), it can be seen that the entropy value of each node classified by the target attribute \( \text{Info}(X) \) is certain, and the smaller the value \( \text{Info}_r(X) \), the maximum value \( \text{Gain} (T) \). Therefore, in order to reduce the amount of calculation \( \text{Info}_r(X) \), only the calculated value is needed in the actual algorithm, and the minimum value is selected as the best attribute for classification.

3. Defects and Optimization of ID3 Algorithm Based on Mobile Marketing Applications

3.1. Application Defects of ID3 Algorithm in Mobile Marketing

First, since the ID3 algorithm selects the split attribute based on the size of the classification information entropy of the attribute, the algorithm will calculate the classification information entropy of the attribute at each internal node \( \text{Info}_i(X) = \sum_{i=1}^{n} P(X_i) \log P(X_i) \), which requires frequent calls to the system function Math.log for calculation. However, frequent calling of system functions greatly reduces the efficiency of the algorithm and causes a lot of waste of time.

Secondly, the ID3 algorithm is based on the highest information gain \( \text{Gain} (T) \) as the criterion for the selection of classification attributes. It can be seen from the formula \( \text{Gain} (T) = \text{Info}(X) - \text{Info}_r(X) \) that the higher the value \( \text{Gain} (T) \), the smaller the value \( \text{Info}_r(X) \) is required, and the attribute with more attribute values \( \text{Info}_r(X) \) will be smaller by calculation. The attribute \( T \) with the most attribute value will be considered as the best classification attribute and will be selected finally. However, in practical applications, the attributes with more attribute values are not the sticky attributes that are of interest to actual problems. For example, the customer attributes that operators are more interested in include "whether the user is the lowest consumer", "the user uses the service brand", etc. The information gain of the latter is significantly higher than that of the former, so the ID3 algorithm will choose the attribute of "users use service brand", but this will deviate from the original intention.

3.2. Application-oriented ID3 Algorithm Optimization

For the application scenario of mobile marketing, this section achieves the goal of simplifying the calculation by decomposing the information entropy formula of the classification attribute \( A \). Assuming that the label set \( E \) of the data target attribute value in the data set \( X \) has two values \( (e_1, e_2) \),
the size of each data set is \( m \) and \( n \) in turn, and the sum of the two is the size of the total data set \( |X| \). The attribute \( A \) has \( v \) values \( (a_1, a_2, \ldots, a_v) \), and when the attribute \( A = a_i \) is the attribute, there are \( m_i \) bars of data \( E = e_i \) and \( n_i \) bars of data \( E = e_2 \). Thus, the information entropy of \( A \) attribute:

\[
Info_A(X) = \sum_{i=1}^{v} \frac{m_i + n_i}{m + n} \cdot \log \frac{m_i + n_i}{m_i} \cdot P(X_i)
\]

(4)

\[
Info(X) = -\frac{m_i}{m_i + n_i} \log_2 \frac{m_i}{m_i + n_i} - \frac{n_i}{m_i + n_i} \log_2 \frac{n_i}{m_i + n_i}
\]

(5)

After integration and simplification, we get:

\[
Info_A(X) = \sum_{i=1}^{v} \frac{1}{(m + n) \ln 2} \left( -m_i \ln \frac{m_i}{m_i + n_i} - n_i \ln \frac{n_i}{m_i + n_i} \right)
\]

(6)

Knowing that \( \ln 2 \) is a constant, and \( m + n = |X| \), the above formula (6) is simplified:

\[
Info_A(X) = -\sum_{i=1}^{v} \left( m_i \ln \frac{m_i}{m_i + n_i} + n_i \ln \frac{n_i}{m_i + n_i} \right)
\]

(7)

Use the definitions of Taylor formula and McLaughlin formula, a large number of log functions are used in the formula for calculating information entropy. Let \( f(x) = \ln(1 + x) \), then we can get

\[
f(x) = \ln(1 + x) \approx x
\]

(8)

To further simplify the above formula (7), we get:

\[
Info_A(X) = \sum_{i=1}^{v} \frac{2m_i n_i}{m_i + n_i}
\]

(9)

After the above formula optimization processing, the simplified information entropy formula (9) of the classification attribute \( A \) is finally obtained. The calculation amount of this formula is greatly reduced, especially the log calculation is omitted. There is no need to frequently call the system function Math during the execution of the ID3 algorithm log, which significantly improves computational efficiency.

This chapter reveals the shortcomings of the ID3 algorithm in mobile marketing applications, that is, the amount of calculation is large and the selection of attributes is biased. By decomposing the information entropy formula of the classification attribute \( A \), an improved information entropy formula is finally obtained for user classification in marketing can be more effective in practice.

4. Conclusion

At this stage, data mining technology has been widely used in the mobile telecommunications industry, and has achieved fruitful applications in many aspects. This paper takes data mining
technology as a starting point, collects and analyzes common algorithms in data mining, and focuses on the idea and process of ID3 algorithm. At the same time, the ID3 algorithm is improved according to the characteristics of the user's own attributes in mobile marketing. The ID3 algorithm not only improves computing efficiency, but also closes the relationship between operators and customer groups, thereby guiding the formulation of more targeted marketing strategies.

References

[1] Zhang Meitu(2011). The application of data mining in the mobile communication market. Modern Telecommunications Technology, no.1, pp: 110-114.

[2] Xiao Mingwei(2012). Research on the application of data mining technology based on customer share of telecommunication groups. Computer and Telecommunications, no.6, pp: 38-39,42.

[3] Safaa O A, Jassim F S(2013). Evaluation of Different Data Mining Algorithms with KDD CUP 99 Data Set[J]. Journal of Babylon University, vol.21, no.8, pp: 2663-2681.

[4] Chen Ping(2009). Parallel algorithm design and performance analysis of decision tree in data mining grid. Journal of Beijing University of Posts and Telecommunications, no.1.

[5] Wang Yongmei(2011). Research on ID3 algorithm in decision tree. Journal of Anhui University (Natural Science Edition), no.3.

[6] Xue Yongning(2013). Research on Data Mining Technology. Chinese and Foreign Entrepreneurs, vol.32, pp: 236.

[7] Zhao Wei(2013). Research on Data Mining Technology of Computer Semi-structured Data Source. Computer CD Software and Application, no.8.