



Design of Financial Modeling System Based on Decision Tree Analysis

Shihui Du^(✉) and Xiaochen Guo

Weifang Engineering Vocational College, Weifang 262500, China
gjx1107752384@163.com

Abstract. The design of financial modeling system is based on decision tree analysis. The main objective of this analysis is to determine the best portfolio of investors and choose the best investment strategy. The process includes determining which factors are important for making decisions about investment strategies, how investors should make these decisions, what criteria they should use to choose different investment strategies, and how they will ultimately apply the strategies they choose. It also helps them understand why some investments perform better than others, so it enables them to identify which factors contribute most to this performance. The decision-making process includes three steps: 1. Data analysis: this step involves collecting data from different sources, such as the company's annual report or other documents, and then analyzing them using statistical techniques. 2. Decision making: according to the results obtained in step 1, you can take appropriate actions for next year's performance through this method. 3. Implementation of decision making: at this stage, you must take appropriate measures (such as setting new strategies or changing existing strategies) to implement the decision.

Keywords: Decision tree · Financial modeling · Systems analysis

1 Introduction

The quality of the company's financial situation must be the focus of attention of the enterprise itself, investors and creditors. A well run and financially healthy company can not only improve its reputation in the market, but also broaden its financing channels. On the contrary, any financial problems of the enterprise will restrict the long-term development ability of the enterprise and even directly lead to the bankruptcy of the enterprise [1]. Therefore, early detection of the company's financial crisis signal enables the company's operators to take effective measures to improve the business operation and prevent the crisis in the embryonic stage of the financial crisis.

The financial situation of the company will also affect the credit scale of the bank to the company. As the bond market in China is still underdeveloped, the debt financing of the company mainly comes from the loans of commercial banks. In order to control risks, commercial banks, as creditors, will reduce the scale of credit to companies with poor financial conditions. Reducing the credit scale of commercial banks will inevitably aggravate the financial difficulties of enterprises themselves. Therefore, the company's

managers must focus on the company's own financial risks and respond in a timely manner. Once the company encounters special treatment, it will have a great negative impact on the company's subsequent operation [2]. Among them, computer financial management includes financial modeling, which can effectively provide decision-making suggestions for managers by establishing financial data related models and optimization. Based on this, this paper studies the design of financial modeling system based on decision tree analysis.

2 Related Work

2.1 Decision Tree Modeling

First, to put it simply, the decision tree is a "if judgment tree structure", which follows the divide and rule strategy. Each node is a feature. The key to dividing data according to the feature is to find out how to select the feature as the node and divide the data, and how to define the loss function to evaluate the result. Each decision in the decision-making process is a "test" of a certain attribute, and the final decision result corresponds to the final decision result. In general, a decision tree includes a root node, several internal nodes and several leaf nodes. It is easy to know:

Each non leaf node represents a feature attribute test.

Each branch represents the output of this characteristic attribute in a certain value range.. Each leaf node stores a category.

The sample set contained in each node is divided into child nodes through attribute test, and the root node contains the full set of samples.

The construction of a decision tree is a recursive process. There are three situations that will lead to recursive returns: (1) all the samples contained in the current node belong to the same category. At this time, the node is directly marked as a leaf node and set as the corresponding category; (2) If the current attribute set is empty or all samples have the same value on all attributes and cannot be divided, the node is marked as a leaf node and its category is set as the category with the largest number of samples in the node; (3) The sample set contained in the current node is empty and cannot be divided. At this time, the node is also marked as a leaf node and its category is set as the category with the largest number of samples in the parent node [3]. The following Fig. 1 shows the flow of the decision tree algorithm.

It can be seen that the key to decision tree learning is how to select the partition attributes, and different partition attributes will lead to different branch structures, thus affecting the performance of the whole decision tree. The goal of attribute division is to make the divided child nodes as "pure" as possible, that is, they belong to the same category. According to the improvement of specific methods for quantifying purity, there are three kinds of decision tree algorithms: ID3, C4.5 and cart (most commonly used). Information entropy, gain rate and Gini index are used as measurement methods respectively.

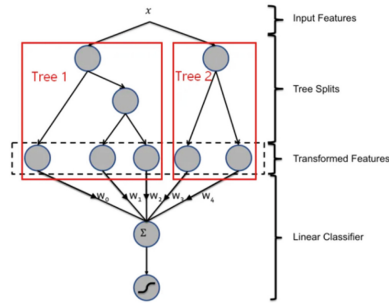


Fig. 1. Decision tree algorithm flow

Pruning is the main means for the decision tree algorithm to deal with overfitting. There are two pruning strategies:

Pre pruning: in the process of construction, it is first evaluated before considering whether to branch. Pre pruning makes many branches of the decision tree cut off, thus greatly reducing the training time cost and reducing the risk of over fitting. On the other hand, since pruning also cuts off the branches of the current node and the subsequent child nodes, the nature of “greedy” pre pruning prevents the expansion of branches, and to a certain extent, brings the risk of under fitting.

Post pruning: after constructing a complete decision tree, evaluate the necessity of branching from bottom to top. Post pruning usually retains more branches, so the performance of the decision tree adopting the post pruning strategy is often better than that of the pre pruning [4]. However, it traverses all nodes from bottom to top and calculates the performance. The training time cost is greatly improved compared with that of the pre pruning.

For continuous value processing: for continuous value attributes, it is not feasible if each value is taken as a branch, so it needs to be discretized. The commonly used method is dichotomy. The basic idea is: given the sample set D and the continuous attribute d , the dichotomy tries to find a partition point t to place the sample set D in the attribute α . It is divided into $< T$ and $> t$.

2.2 Financial Analysis

Financial analysis is an economic management activity based on accounting, statement data and other relevant data, which uses a series of special analysis techniques and methods to analyze and evaluate the profitability, operating capacity, solvency and growth capacity of enterprises and other economic organizations in the past and present related financing activities, investment activities, operating activities and distribution activities.

It is an economic application discipline that provides accurate information or basis for investors, creditors, operators and other organizations or individuals concerned with enterprises to understand the past of enterprises, evaluate the current situation of enterprises, predict the future of enterprises and make correct decisions [5].

The financial analysis model is to establish some models through some financial indicators to make complex financial information easy to understand. There are three main financial analysis models:

1. Enterprise strategy analysis model:

On the basis of clarifying the purpose of financial analysis, enterprise strategic analysis is the starting point of enterprise financial analysis. The essence of strategic analysis is to clarify the position of the enterprise in the industry and the competitive strategy that should be adopted through the analysis of the industry in which the enterprise is located or the industry that the enterprise intends to enter, so as to weigh the benefits and risks, understand and master the development potential of the enterprise, especially the potential in enterprise value creation or profit.

2. Accounting analysis model:

The essence of accounting analysis is to clarify the connotation and quality of accounting information, that is, to reveal its actual meaning from the surface of accounting data. The analysis includes not only the analysis of the connotation of each accounting statement and related accounting subjects, but also the analysis of accounting principles and policy changes, accounting method selection and changes, accounting quality and changes, etc.

3. Financial statement analysis model:

The analysis of financial statements is based on the financial statements, adopts scientific evaluation standards and applicable analysis methods, follows standard analysis procedures, and makes judgment, evaluation and prediction on the business situation and performance of the enterprise by comparing and analyzing the financial situation, operating results, cash flow and other important indicators of the enterprise [6].

3 Design of Financial Modeling System Based on Decision Tree Analysis

The most important thing about the design of financial modeling system based on decision tree analysis method studied in this paper is that because of this tool, the data architecture of the whole company can become standardized, and the next step is to build the big data platform of the enterprise. Moreover, it is written in Java and supports secondary development. It is an excel like designer. It is very easy to start, whether it is it or business: editing SQL optimization and data set reuse are all small cases, greatly reducing the threshold of report development. In the aspect of data security, which is most concerned in the enterprise, it supports multiple people to develop the same set of reports at the same time, and prevents editing conflicts through the template locking function [7]; Ensure data security through data analysis and permission control.

Whether in the database or in the DW/Bi design, it is necessary to do dimension modeling. However, due to different performance and objectives, the rules for dimension modeling are different. What I am talking about here is mainly dimension modeling in the DW/Bi design. The following is an example of the process of constructing a decision tree. The information gain is used to find the field with the largest amount of information in the data table, and a node of the financial analysis decision tree is established. Then, each branch of the tree is established according to the different values of the field, and then the lower node and branch of the tree are repeatedly established in each branch. Taking the traffic cost data set as an example, the classification decision tree of “overall traffic cost exceeding standard level” is constructed [8]. Classification label: Class P It refers to the overall over standard level of transportation expenses = “1”, and category n refers to the overall over standard level of transportation expenses = “0”. Similarly, a small amount of data is selected here to establish a branch of the tree and highlight the classification process.

4 Simulation Analysis

The financial data analysis model is established according to the decision tree ID3 algorithm, and the financial data set is tested. The prediction accuracy can be reached, and whether it can be reported as the budget of the next year. These problems are also completed through the decision tree algorithm.

In the process of financial data analysis, two-thirds of all data sets are used as data training sets, and the prediction model is established according to the decision tree ID3 classification algorithm to generate a decision tree and evaluate the accuracy of the established model. Then one-third of the whole data set is used as the data test set to test the accuracy of the model algorithm on the experimental platform. Data_ The pretreatment() function is mainly used to discretize the samples with continuous test attributes. Because ID3 decision tree algorithm cannot classify continuous attributes. If classification is enforced, each continuous value of an attribute will be treated as an attribute category [9]. In this case, the tree will be very large and the efficiency of building the tree will be very low. Even if the tree is constructed, because of the overfitting of continuous attributes, the accuracy of classification of unknown samples using the tree will be very poor. The test item code is shown in Fig. 2 below.

When different sample numbers are selected, when the two accuracy rates are in a stable or rising state, it means that this data set is well adapted to this kind of algorithm, and can be considered to provide auxiliary decision-making information for the actual financial work; Otherwise, it means that this data set is not suitable for this algorithm, can not be predicted by classification analysis algorithm, and can not guide the work. There are many factors affecting the evaluation results, such as the number of samples in the training set and the test set, the number of samples in a single tree, and the selection of algorithm parameters. With the increase of the number of samples, the decision tree structure will become more complex, and the accuracy may fluctuate [10]. Therefore, it is necessary to conduct multiple tests to determine the classification performance.

```

t0=clock;

[E correct_test correct_train

error_train_num]=M_ID3(handles.train,handles.test,clas.attribute_kind,long,the);

correct_test_num=sum(correct_test,2)*100;

set(handles.edit5,'string',correct_test_num/(handles.len_test));

set(handles.edit6,'string',1-error_train_num/handles.len_train);

set(handles.edit9,'string',length(E));

axes(handles.axes1)

plot(correct_train)

```

Fig. 2. Test item code

5 Conclusion

The design of financial modeling system is based on decision tree analysis. The main purpose of this type of analysis is to let users understand the relationship between different data sets and the relationship between them. This process first identifies all possible combinations or permutations of the variables in question and then proceeds through a series of steps designed to help you determine which combination is best suited to your particular situation. Decision trees are used as part of many different types of analysis, including those related to marketing, finance, sales forecasting, and customer relationship management.

References

1. Zhang, W.: Research on English score analysis system based on improved decision tree algorithm and fuzzy set. *J. Intell. Fuzzy Syst.* **39**(4), 5673–5685 (2020)
2. Zhang, X.Z.: Analysis and design of tourism CRM system based on decision tree. *Mod. Comput.* (2018)
3. Ming, Q., Li, R.: Analysis and Design of Personalized Learning System Based on Decision Tree Technology (2020)
4. Ming, Q.: The analysis and design of the personalized learning system based on decision tree. *Microcomput. Appl.* (2018)
5. Geng, X., Yang, D.: Intelligent Prediction Mathematical Model of Industrial Financial Fraud Based on Data Mining. Hindawi Limited (2021)
6. Jin, M., Wang, H., Zhang, Q., Luo, C.: Financial management and decision based on decision tree algorithm. *Wireless Pers. Commun.* **102**(4), 2869–2884 (2018). <https://doi.org/10.1007/s11277-018-5312-6>
7. Le, Y., Zizong, T., Panpan, W.U., et al.: Visual modeling of rice root growth based on B-spline curve (2022)
8. Huang, Z., Sun, Y., Gan, L., et al.: Durability Analysis of Building Exterior Thermal Insulation System in Hot Summer and Cold Winter Area Based on ANSYS (2022)
9. Varghese, V., Krishnan, V., Kumar, G.S.: Evaluating pedicle-screw instrumentation using decision-tree analysis based on pullout strength. *Asian Spine J.* **12**(4), 611–621 (2018)
10. Huang, Z., Liang, Y.: Research of data mining and web technology in university discipline construction decision support system based on MVC model. *Library Hi Tech* (2019)