Research on Decision Support Model Based on Decision Tree-Modified Back Propagation Neural Network

Junqing Liang\textsuperscript{1}, Xiaolin Lv\textsuperscript{2}, Jianshi Zhao\textsuperscript{3} and Yue Leng\textsuperscript{4}

School of Information and Control Engineering, Qingdao University of Technology, QingDao, ShanDong, 266520, China

\textsuperscript{*}Corresponding author’s e-mail: wannareset52@163.com

Abstract. In the context of today's global economic integration, various commercial network-based trading activities have become increasingly frequent. The gap in Chinese enterprises tends to be extreme. The introduction of the investment decision model helps to alleviate the extreme trend of the enterprise. The decision support model which based on decision tree and BP neural network has a good effect on financial investment. Both decision tree and BP neural network are the advantages of machine learning. This paper combines decision tree algorithm with BP neural network algorithm and uses the genetic algorithm improves the neural network model to form a financial decision model. Therefore, the model can make feasible and effective results for large data sources in a relatively short time, and the accuracy of the obtained results is high.

1. Introduction

Investment decision-making is a trading activity in which a company obtains the influence of other companies and obtains benefits under the market mechanism. In the investment process of the company’s decision-making organizations, it is divided into four aspects: making decision-making objectives, planning decision-making content, executing decision-making information, and decision-making evaluation. Once the evaluation cannot pass the decision, it will return to the starting point. On the basis of that, the decision-making process develops slowly. Financial investment decision support always contains a large number of influencing factors to interfere with its decision-making results, such as the decision-making company's own development dynamics, the investment company's own company's operating conditions and the value of the invested company's customer value will affect the decision results of investors. The general decision-making model only uses BP neural network to train all the features to obtain traditional predictions, which makes the prediction accuracy low, the model consumption is large, the decision time is long, the data source is too large, and the calculation amount is prominent, which also restricts the decision.

In the financial investment decision support model based on decision tree and improved BP neural network, the gain rate in the decision tree algorithm is computationally efficient and consumes relatively little time resources. The features can be sorted by the optimal structure of the decision tree and efficiently removed. Partial interference features improve the stability of the decision model; the neural network algorithm is easy to fall into the local optimal value, and the genetic algorithm is used to construct the appropriate population, combined with the fitness function to help the neural network jump out of the local optimum and find the global optimal which can improve the computational efficiency of the decision model and speed up the convergence of the model.
2. Decision Model Theory Construction

2.1. Decision tree algorithm
The decision tree algorithm is a common ML algorithm. It supports both classification and regression problems which is a nonlinear model. The feature selection in the model is based on the gain rate in the decision tree algorithm which can excluding the interfering features in the data, the gain rate is calculated as follows:

\[
\text{Gain\_ratio} = \frac{\text{Gain}(D, a)}{IV(a)}
\]

Where \( \text{Gain}(G, a) \) represents the information gain:

\[
\text{Gain}(D, a) = \text{Ent}(D) - \sum_{i=1}^{V} \left| D^i \right| / |D| \text{Ent}(D^i)
\]

\( \text{Ent}(D) \) represents information entropy:

\[
\text{Ent}(D) = -\sum_{k=1}^{x} P_k \log_2 P_k
\]

\( IV(a) = -\sum_{v=1}^{y} \left| D^v \right| / |D| \log_2 \left| D^v \right| / |D|
\]

Therefore, regardless of the number of values that can be attributed, the ratio is constrained by the ratio and the characteristic characteristics are affected.

2.2. BP neural network
The BP neural network has a hierarchical structure of three or more layers, each of which performs its own offset weighted sum and outputs through a transfer function, and the elements are serialized in a hierarchical feed forward topology. The gradient descent algorithm is generally used to update the weights to optimize performance and realize network learning. The number of input neurons is \( x \), the number of neurons in the hidden layer is \( D \), and the number of output neurons is \( Y \).

![BP neural network connection diagram](image)

Input layer to output layer weight is \( W \). The input of the \( h \)th hidden neuron is

\[
\sum_{i=1}^{x} W_{ih} X_i
\]

Similarly, the input of the \( j \)th output neuron is

\[
\sum_{h=1}^{D} V_{hj} d_h
\]

3. Data Processing

3.1. Sample source
Financial investment decision support model data sources are diverse, Yahoo Finance, Sina Finance and other financial websites, more officially provide financial information about stocks, futures, etc. In the sample source method, the model usually use python to crawl through web crawling or MATLAB. And the DataFeed toolbox is used for crawling.
For most of the company’s basic information, data can be extracted through the official website. For investment companies, company risks and other information needs further investigation, etc. Based on this, Tianyanchao.com provides a platform. It not only contains basic information about the company, legal representative information, but also information about the invested company, as well as company risk information. Therefore, the main data of the investment model comes from the query information of the sky-eye search website that uses Python to crawl through the web crawler. Information collection is based on the reliability and authenticity of market research. The company analyzes the existing system based on the expert system.

To avoid “overfitting” the decision model, sample data should be extracted as much as possible for training and testing.

3.2. Sample extraction
Before using the extracted data source, the data needs to be extracted and divided into training samples and test samples. Training samples are used for decision model training. And the test sample is used for the test evaluation of the model. And the sample extraction method is divided into three types: the leave method, the cross test method, and the self-help method.

Both the set-out method and the cross-validation method retain some samples as test sets. The training set used in the evaluation must be smaller than the data set, so there must be some deviation. Therefore, the improved self-help is used in the sample extraction of the decision model. The method, that is, all the data is put back into the sample of 10% as the test set, and the remaining 90% is used as the training set, and it is extracted ten times repeatedly. Under the premise of ensuring that the training set samples are sufficient, the deviation is reduced to some extent.

3.3. Feature selection
In order to simplify the calculation process, factors that influence the outcome of the decision-making investment are selected based on expert knowledge before feature selection. In order to avoid the subsequent deviation increase, expand the selection of feature sets.

| Table 1 Table of factors affecting investment model |
|--------------------------------------------------|
| Influencing factor | Possible value | Factor description |
|---------------------|----------------|--------------------|
| 1. Period factor    | Date           | The shortest return time that the investor can accept |
| 2. Capital injection cost | Natural number | Number of investor funds injected |
| 3. The risk of the invested company | High Medium Low | The risk of the invested company |
| 4. Invested company legal representative represents information value | Natural number | Value information provided by the top level of the investor |
| 5. Invested company judicial risk | High Medium Low | Invested company judicial risk |
| 6. Prospects for invested companies | L1, L2, L3, L4, L5 | Assessing the future development prospects of investors by combining market value |
| 7. Investment company's own capital turnover | Z1: rich, Z2: Ample, Z3: Moderate, Z4: Shortage, Z5: insufficient | The amount of funds the investor can invest in itself |
| 8. Invested company stock value | Natural number | Invested company stock value |
| 9. High net worth customer ratio | Decimal between 0-1 | Customer value owned by the investee |
| 10. High value of the invested company | Natural number | |
| 11. Investment company bears the | J1~J10 | From large to small, J1~J10 indicates |
Influencing factor | Possible value | Factor description
---|---|---
degree of risk | the size of the risk that the investor can bear.
12. Minimum revenue time | Date | The shortest date that the investor expects to be off-site than the return
13. Revenue cycle | Natural number | Revenue cycle
14. Industry sentiment index | Decimal between 0-1 |

According to the investment decision factors listed in Table 1 and the data set information, the calculation of the gain rate is based on each influencing factor, and the decision tree is constructed.

4. Model establishment

4.1. Genetic algorithm
The genetic algorithm is a random search method based on the law of "survival of the fittest, survival of the fittest" in the biological world [1]. According to the fitness function selected by the algorithm itself, the individuals are screened through operations such as “selection, crossover, and mutation”. According to the survival rule of the fittest, individuals with high fitness are selected, and vice versa.

4.2. Genetic algorithm improved neural network
(1) Initial population construction
The method of constructing the population mostly adopts the random sampling method and the uniform sampling method. The purpose of constructing the initial population is to obtain the objective function information. In the uniform distribution, the individual contains more objective function information than the random distribution, and the random distribution feasible solution is unevenly distributed. However, even distribution makes the individual lose randomness and reduces the accuracy of the algorithm. In this paper, pseudo-random sequences based on Halton sequences are used to generate individuals with low variability, and the uniformity is improved at the expense of partial randomness to improve the objective function information.

(2) Coding
In this paper, real number coding is used. Under the premise of the decision model and the amount of data is so large, the process of encoding and decoding is not needed, the efficiency of the model is improved, and the description and essence of the problem can be expressed intuitively.

(3) Fitness function
The fitness function is chosen to be the reciprocal of the absolute value of the expected output versus the true output difference. If the i-th data is output in the neural network, the output is \( y_i \), The raw output of the data is \( y_i \), Then the fitness function of the network on the data can be defined as

\[
\frac{1}{1 + |y_i - y'_i|}
\]

(4) Genetic operator
Selection operator: In the Darwinian "Evolution", the fitness rule of the survival of the fittest, the fitness function value is directly proportional to the individual survival probability, the greater the fitness function, the greater the survival probability, and vice versa. This article combines a conventional roulette strategy with an optimal preservation strategy. The roulette strategy is a replay-based proportional-based random sampling method. The individual should be determined by the proportion of the total fitness. Probability is expressed as \( P_i = \frac{S_i}{\sum S_i} \), The probability of selecting individual i is \( P_i \), \( S_i \) is its fitness function value. Combining the roulette strategy with the optimal preservation strategy, the individuals with the highest probability will not perform the crossover and mutation operations, replacing the individuals with the lowest fitness value after the cross-compilation operation. Improve the efficiency of model operations.

Crossover operator and Mutation operator: In order to simplify the model operation and improve the efficiency of the model operation, single point crossing and variation are adopted.
5. Design
Using BP neural network optimized by genetic algorithm, the BP neural network structure needs to be determined first, then the genetic algorithm optimizes the initial threshold and weight of the neural network; finally, the neural network prediction is performed. The BP-GA algorithm uses the individual to represent the initial weight threshold of the neural network. The absolute value of the error generated by the network when the individual is initialized is the fitness value of the individual, and the optimal weight threshold is selected for operations such as selection, mutation and intersection. The combination of decision tree algorithm, genetic algorithm and BP neural network algorithm not only accelerates the calculation speed, saves time, but also improves the accuracy of the decision model.

6. Model analysis
The financial investment decision support model based on decision tree-improved BP neural network proposed in this paper uses the gain rate to filter the unrelated interference factors and enters the selected feature attributes into the improved BP neural network. In the experimental data, the operating status of the invested companies of certain enterprises from 2008 to 2013 and the company’s own capital turnover are used as input data. Whether the investment subsidiaries can invest, income and return rate (including net profit rate) from 2014 to 2017 etc. etc. As the output data, the simulation prediction of the financial investment decision model is carried out.

Under the condition of limited data source, the improved BP neural network using genetic algorithm converges faster than the general neural network model. The general BP neural network model has a convergence iteration number of 300 times, and the improved model based on genetic algorithm has 224 times of convergence, and the model prediction accuracy is as high as 88.5%, which fluctuates within the target error. This model is fully applicable to financial investment decision support systems.

The investor can set a hard threshold as needed, which is the minimum income accepted and the range of fluctuations accepted. When it is impossible to make a corporate financial investment, consider the project investment or cooperative financing according to the income forecast of the model. The effect is as shown:

When financial investment is not possible, whether the net profit in the forecasting model can be converted to the direction of convergence can achieve the goal of maximizing revenue.
7. Conclusion

(1) This paper combines decision tree algorithm with genetic algorithm improved BP neural network to establish financial investment decision support model, use the gain rate calculation in decision tree algorithm to eliminate feature interference, eliminate the factors with low influence, and improve the accuracy of decision model.

(2) Genetic algorithm optimizes the initial weight threshold of BP neural network and reduces the risk of falling into local minimum in the traditional Stochastic Gradient Descent. Improve the convergence of the model to achieve the optimization model effect.

(3) The financial investment decision support model can be applied not only to corporate investment, but also to the project investment model. Refer to project return rates and reference predictions for cooperation

References

[1] Dong Xiaoshuai, Mao Zhengyuan. (2018)Research on solution of dynamic Vehicle Routing Problem based on improved genetic algorithm[J].Computer Engineering and Applications,54(19):49-55.

[2] Liu Xiang, Li Dongsheng, Hu Rui. (2018) Application of Improved Genetic Algorithm in Cooperative Jamming Resources Assignment[J]. Journal of Detection and Control, 40(5):69-75.

[3] Cao Leixin, Sun Hongbin. (2017) Stock Return Forecast Based on Grey Neural Network[J]. Computer and digital engineering, 45(01):24-28.

[4] Zhou Jia, Wei Renyong, Chen Chao, Hou Zheng. (2010) Improved Power Harmonic Detection Algorithm Based on Genetic Algorithm[J]. Science Association Forum (the second half of the month) 06:83-84.

[5] Hongtao Zheng, Fumin Pan, Ren Yang. (2012) Prediction of compressor characteristics by optimized BP neural networks based on genetic algorithm (GABP)[C]. 2012 International Conference on Maritime Technology. 2012:338-342

[6] Hossein Etemadi, Ahmad Ahmadpour, Seyed Mohammad Moshashaei. (2015) Earnings Per Share Forecast Using Extracted Rules from Trained Neural Network by Genetic Algorithm.《Computational Economics》, 46 (1) :55-63