ISSN 2616-5775 Vol. 5, Issue 3: 41-44, DOI: 10.25236/AJCIS.2022.050306

Trading Strategies Model Using the Apriori Algorithm and the Neural Network Optimized by the Genetic Algorithm

Ruiyao Yang

Statistical and Mathematics Institute, Xiamen University of Technology, Xiamen, Fujian, 361024, China

Abstract: *A neural network trading strategy optimized based on Apriori and the genetic algorithm is proposed to determine the trading strategies. Using the Apriori algorithm and the Neural Network optimized based on the Genetic Algorithm, we find all the frequent term sets of the fuzzy set and mine the fuzzy association rules from the frequent term set. Enter the price dataset, set minimum support and multiplier; output fuzzy association rule library, and iterate.*

Keywords: *Fuzzy, Apriori Algorithm, Genetic Algorithm, Neural Network*

1. Introduction

We need to develop a model that uses only the past daily price flow to date to determine whether the assets in its portfolio should be purchased, held, or sold daily.

The results show those fund managers or stock investors can give effective investment guidance to non-subjective fuzzy decision-making systems, and suggest trying to overcome their psychology and market benchmarks. The technical analysis index is blurred by triangular blurring and uses the Apriori algorithm and neural network to generate non-subjective class transaction rules. Design a fuzzy decision system with non-subjective class fuzzy trading rule library, product reasoning machine with Mamdani meaning, and Center average deblurring device. The structural parameters of the system are estimated using recurrence least squares with forgetting factors, and a non-subjective class trading strategy mode: neural network trading strategy based on the Apriori algorithm is proposed.

2. Model Establishment and Solution

2.1. The Apriori Algorithm Generates Fuzzy Trading Rules

We try to use the Apriori algorithm to find all frequent items of $x_{1,t}^{(m,n)}$ fuzzy sets and ed fuzzy sets (frequently occurring items), and mine fuzzy association rules from the frequent itemset (suggesting a strong relationship between the two items), the subjective association rules generated by which are no longer preset by domain experts.

The general procedure of the Apriori algorithm is as follows.

I. Input: The Price dataset $P\{p_0, p_1, \cdots, p_n\}$, Minimum support min_sup[1], Minimum size min_conf

II. Output: Fuzzy association rule library *R*

III. Iterate process:

(1)Dataset *P* is converted to a fuzzy fold-sign matrix by a fuzzy membership function:

$$[\mu A_i(P_t), \mu A_i(P_t), \mu A_i(P_t), \cdots, \mu B_l(P_t), \mu B_l(P_t), \mu B_l(P_t), \cdots] \quad (1)$$

Where A_i and $\mu A_i(P_t)$ is the *i*th fuzzy set of the rule IF part (predecessor) and its membership function, B_l and $\mu B_l(P_t)$ is the *j*th fuzzy set of the rule THEN part (posterior) and its membership function, the non-zero coordinates are extracted and marked to form the processed data set *D*.

(2)Identify all nonduplicate combinations (item set) C_1 of size 1 from the dataset *D*.

ISSN 2616-5775 Vol. 5, Issue 3: 41-44, DOI: 10.25236/AJCIS.2022.050306

(3)The records of the elements in C_1 were counted and the support formula (1) was calculated in *D*, by scanning through the dataset *D*.

(4)Determine whether the support of each element in C_1 satisfies the minimum support min_sup[1]. Satisfaction retains the element, otherwise, no action is done. The set of items with the minimum support constitutes a frequent itemset L_1, and the result is inserted into the empty list *L*.

(5)Find out all uninterrupted disorder combinations (item set) C_k of size *k* (*k*>1, *k*++, *k*=1,2...), such as item set {A} (size 1), {A→B} (size 2) ...

(6)Records of the elements in C_k were counted and supported in *D* by scanning and passing through C_k.

(7)Determine whether the support of each element in C_k satisfies the minimum support min_sup[1]. Satisfaction retains the element, otherwise, no action is done. The set of items with the minimum support constitutes a frequent itemset L_k, inserted into the list *L*.

(8)Determine whether L_k is a null set, if a null set, perform step (9), otherwise return to step (5).

(9)Similarly, find all the ordered rule combinations (item set) H_j. with size *j* (*j*> 1, j=1,2...), {A→B} (size = 2), A→BC, or {BC→A}, (size = 3) ...

(10)Determine whether the support of each element in H_j satisfies the minimum confidence min_conf. Satisfaction retains the element, otherwise, no action is done. The set of items that satisfy the minimum confidence constitutes the frequent itemset R_j into the list *R*.

(11)Determine whether H_j is a null set, if it is a null set, perform step (12), otherwise return to step (9).

(12)Returns the fuzzy association rule set *R*.

2.2. Algorithmic Steps for the Apriori

The specific implementation steps of the algorithm are as follows.

Step 1: Data processing. At time t, the membership functions of $x_{1,t}^{(m,n)} = \ln(\bar{P}_{t,m}/\bar{P}_{t,n})$ (m=1, n=5) to the corresponding fuzzy sets ("PS", "PM", "PB", "NZ", "N Z", "NS", "NM", "NB") represent the investor selects "commodity" 1 to 7, and the increase of the t + 1 stock as ed to the corresponding fuzzy sets ("BS", "BM", "BB", "B B", "SS", "SM", "SB", "N"). The resulting membership represents investors picking "goods" from 1 to 7. Next, for the "commodity" number on the fuzzy set label with membership greater than zero, the fuzzy set with membership less than zero is eliminated, and the resulting fuzzy set number list is defined as the investor "shopping list". Assuming that $x_{1,t}^{(m,n)}$ are calculated by historical data and fed into the corresponding blur, the result at time *t* is [0,0,0.9,0.9,0.1,0,0,0,0; 0,0,0,1,0,0,0,0,0] and the "shopping list" at time *t* according to the above numbering rule is [3,4,10].

Step 2: The Apriori algorithm looks for frequent sets. For selecting a frequent itemset, the support represents the proportion of records containing the item set in the selected dataset. The higher the support, the more likely it is to become a frequent set. Let the two sets of items be A. support=$P(A \cup B)$.

We, therefore, merge all the records of the "shopping list" generated in step 1 into a total dataset and can then define minimum support while retaining the set of items that satisfy the dataset with minimum support.

Step 3: The Apriori algorithm generates the association rules. For the generation of the association rules, we need to define the confidence from the frequent sets. The higher the confidence, the stronger the association, the easier it is to become a fuzzy rule with the association. The confidence of one of the rules A→B can be quantified as support $P(A \cup B)$ divided by support A. The formula can be expressed as Confidence=support($A \cup B$)/support(A).

To find the association fuzzy rules we need, we need to leverage the set of frequent items generated by the previous step to generate a list of possible rules, and then calculate the confidence of each rule. The rule is removed if the confidence does not meet the minimum requirements. As the method of frequent term set generation, we can use the Apriori principle to reduce the number of rules to be tested and improve the efficiency of the algorithm.

However, from the analysis method, as long as the training data is enough and sufficient, the association rules can reflect the market response to the trader behavior most likely. So, if we master the

ISSN 2616-5775 Vol. 5, Issue 3: 41-44, DOI: 10.25236/AJCIS.2022.050306

"buying" behavior of buying a stock trader, it is more incentive for "smart" investors actively looking for information in the market.

2.3. Generation of Fuzzy Transaction Rules Based on Neural Networks Optimized by Genetic Algorithms

Step 1: Data processing. At time t, the membership function of $x_{1,t}^{(m,n)}$ (m=1, n=5) into the corresponding fuzzy sets ("PS", "PM", "P M", "PB", "NZ", "NS", "NM", "NB") obtains the membership function of the t + 1 stock increase as ed, and the corresponding fuzzy sets ("BS", "BM", "B M", "BB", "SS", "SM", "S M", "SB", "N") to obtain the output signal matrix. Assuming that $x_{1,t}^{(m,n)}$, ed is calculated by historical data and replaced into the corresponding triangular blurring, a set of output-input data pairs are [0,0,0.9,0.0.1,0,0,0], the former acting as the input signal for the training sample and the latter training sample output signal and [0,0,1,0,0,0,0,0].

Step 2: BP Neural Network training

Since both the weights and the thresholds between the initial neurons of the BP neural network are randomly selected, it is easy to fall into local minima. To solve this problem, we use the Genetic Algorithm (GA) prediction model to optimize BP neural network, BP neural network, and organic integration, and use GA to connect the weights and threshold selection of random defects of BP neural network, which can not only promote BP neural network mapping, and the rapid convergence and strong learning ability of BP neural network.

- **Prediction of chaotic time series based on BP neural network**

Phase space reconstruction theory is the basis of chaotic time series prediction, and the theory of phase space reconstruction of chaotic time series is proposed by Packard et al. The state vector of a point in the state space can be represented as:

$$X_i = (x_i, x_{i+\tau}, \cdots, x_{i+(m-1)\tau})^T, i = 1,2,\cdots, M \quad (2)$$

Where $M = n - (m-1)\tau$ is the number of phase points in the reconstructed phase space, τ is the delayed time, and m is the embedding dimension.

For the typical three-layer BP neural network to predict the chaotic time series, the number of input layer neurons is better when m, the number of hidden layer neurons is p, the number of output layer neurons is 1, and BP neural network to complete the mapping $f: R^m \to R^1$, the mathematical expressions are:

$$x_{i+1} = f(X_i) = \left(\frac{\pi}{2} - \theta\right)\frac{1}{1+exp\left(-\sum_{j=1}^{p} v_j b_j + \gamma\right)} \quad (3)$$

Where v_j is the hidden layer to the output layer of the connection weights, γ is the threshold of the output layer, b_j is the output of the hidden layer node. BP neural network transfer function using Sigmoid function $f(x) = \frac{1}{(1+e^{-x})}$, we can come to:

$$b_j = \frac{1}{1+exp(-\sum_{i=1}^{m} w_{ij} x_i + \theta_j)} \quad (4)$$

Where w_{ij} is the connection weights of the input layer to the hidden layer, θ_j is the node of the hidden layer threshold. The connection weights of BP neural network, w_{ij}, v_j, θ_j, γ can be obtained by training the BP neural network, so x_{i+1}can be predicted. The formula is chaotic time series BP neural network prediction model, the number of neurons in hidden layer P general experience value is 2*m*+1.

Step 3: Data blur. When the trained network receives a new set of fuzzy signal input, a set of output multiplier data will be output. Finally, the output signal is converted according to the demulcent of the output signal into a non-fuzzy variable, as the output of the entire fuzzy system. All the above operations can be invoked the encapsulated neural network function in the neural network toolbox of MATLAB to generate a neural network that meets the control requirements as an implicit fuzzy rule inference machine for a fuzzy system.

ISSN 2616-5775 Vol. 5, Issue 3: 41-44, DOI: 10.25236/AJCIS.2022.050306

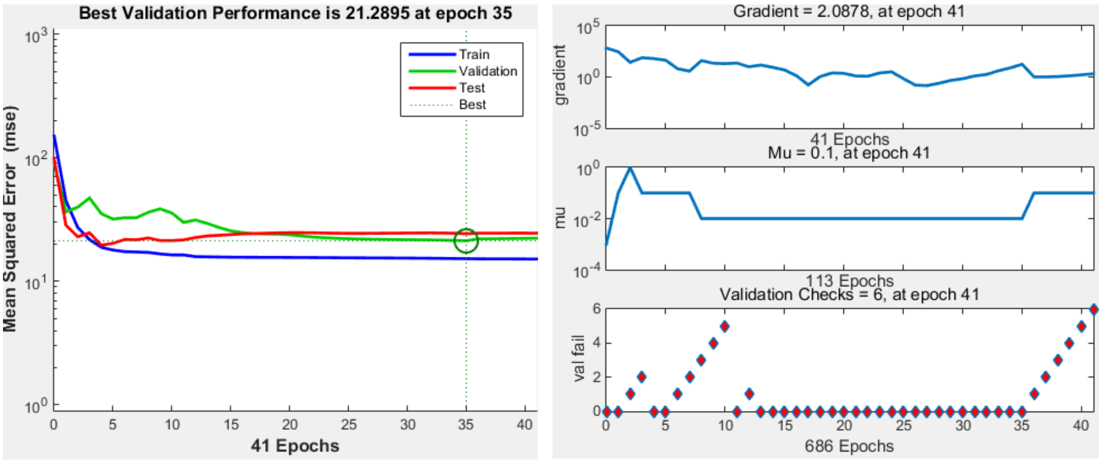

Figure 1: Results of analysis by bp neural network.

3. Model Evaluation

The fuzzy sets of fuzzy system theory are used to describe the technical analysis indexes of fuzzy subjectivity and expound the complex price model of financial assets with the dynamic price model of the fuzzy system.

We propose to utilize machine learning methods, such as Apriori algorithms and neural networks, to generate non-subjective fuzzy inference rules, instead of the fuzzy subjective rules set by domain experts, and to avoid the systematic bias caused by subjectivity of the artificially set rules.

The parameter estimation of the model simply uses the standard recursive least-squares algorithm with exponential forgetting to estimate the market trading intensity parameter, and it is well known that fast estimation methods and estimation accuracy are critical to strategy success due to unpredictable financial markets. Therefore, finding a faster and less error parameter estimation method becomes a key factor in improving the performance of fuzzy decision systems. The above deficiencies will also be the direction of our continued research in the future.

References

[1] TROW M., Problems in the Transition from Elite to Mass Higher Education [R]. Berkeley, CA: Carnegie Commission on Higher Education, 1973: 57.
[2] Guan Peijun, The mission of China's higher education in the new era [J]. China's higher education research, 2017(12): 17-19.
[3] THE., Teaching Excellence Framework (TEF) Results in 2017 [EB/OL]. https://www. timeshighereducation. com/news/teachingexcellence-framework -tef -results-2017.
[4] Wang Zhongkui, Hu Xiaotian, The way to realize the performance evaluation of undergraduate teaching quality: the experience and Enlightenment of Britain [J]. Research on education development, 2018, 38(17): 71-77+84.
[5] Wang bangquan, Research on quality evaluation of transnational higher education development in Australia [J]. Educational exploration, 2017(04): 115-120.
[6] Zhang Jing, Liao Xiangyang, Establishing a quality evaluation and monitoring system of higher education development guided by student development [J]. China's higher education, 2014 (01): 32-34+40.
[7] Zhang Nanxing, Wang Shu, Sun Jihong, Evaluation of comprehensive development level of higher education in China and Research on regional differences [J]. Education Research, 2014, 35 (05): 28-36.
[8] Wang Ning, Research on the quality evaluation index system of China's general higher education development [J]. Educational modernization, 2017, 4(18): 62-67.
[9] Ma Xing, Wang Nan, Construction of university teaching quality evaluation system based on big data [J]. Tsinghua university education research, 2018, 39(02): 38-43.