

# Neural Network Prediction of Stock Price Trend Based on RS with Entropy Discretization

Huang Haiqing<sup>1,a</sup>, Gan Xusheng<sup>2,b</sup>, Lei Lei<sup>3,c</sup>

<sup>1</sup> XiJing College, Xi'an, Shaanxi, 710123, China;

<sup>2</sup> Air Traffic Control and Navigation College, Air Force Engineering University, Xi'an, Shaanxi, 710051, China;

<sup>3</sup> School of Business Administration, Henan University of Economics and Law, Zhengzhou, Henan, 050046, China

<sup>a</sup>23492345@qq.com; <sup>b</sup>ganxusheng123@163.com; <sup>c</sup>lyx421@yahoo.com.cn

**Keywords:** BP Neural Network, Rough Set, Stock Price Trend, Discretization, Attribute Reduction.

**Abstract.** To accurately predict the stock price trend, an integration prediction method based on Rough Set (RS) and BP neural network are proposed. In the method, RS is firstly applied to reduce the features of stock price trend, and an entropy-based discretization algorithm is introduced to process the continuous attribute data, then, on this basis, BP neural network is used to establish the prediction model of stock price trend. The validation result indicates that, by RS attribute reduction, the prediction model of BP neural network for stock price trend can be simplified with performance improvement. The prediction result is better than those of traditional neural network and RBF neural network. This verifies its feasibility and effectiveness.

## 1. Introduction

The stock prediction is that, based on accurate survey statistics and stock market information, the scientific methods are used to predict the future prospects of stock market from the history, current situation and stock market laws. There are many factors that affect the stock market development. The uncertain interactions between these influence factors are very complex, which leads to great prediction deviations. With development of national economy and understanding of stock market in depth, it becomes increasingly important for stock investment and risk management to use a reasonable and effective method to accurately predict the stock price trend [1]. At present, the prediction of stock price trend mainly has the following problems:

### 1.1. Stock price trend has the mutability and variability.

The stock price trend often shows the mutability and variability because of a certain risk. This risk can be divided into systemic risk and non-systemic risks in a non-quantifiable form, which can cause the mutation of stock price trend or index. Sometimes the factors that affect the mutation of stock price trend cannot be observed, such as human manipulation and so on; sometimes although it can be observed, it is false information, such as fraud and spreading rumors and so on. These factors may cause the abnormal fluctuations and structure mutations of stock market, manifesting as the mutability and variability of stock price trend. This makes the stock price trend data show a non-normal distribution, resulting in low accuracy for stock price trend prediction.

### 1.2. There is a certain amount of redundant information in stock price trend data

The stock price trend data, obtained by trading after multi-party game, often contains a lot of valuable information such as investment intentions and behaviors and so on. At the same time, Stock trading data volume is larger, the relevant indices are more, making the stock price trend data contains a certain amount of redundant information. This affects the effective prediction of stock price trend.

The above two problems may lead to low prediction accuracy for stock price trend. Both the traditional prediction method and the single intelligent prediction method cannot solve these problems effectively. In this context, the hybrid prediction method has been taken more seriously. Therefore, a hybrid method based on BP neural network and Rough Set (RS) is proposed to predict the stock price trend. RS plays the roles in the method: 1. Reduce the input dimension of BP neural network to remove the redundant information in sample data;. The proposed method is to solve the problem of low prediction accuracy for stock price trend caused by the variability and mutability of stock price and redundant information, so as to provide the references for stock investment and scientific research.

## 2. BP Neural Network

Typical BP neural network has three layers including input layer, hidden layer and output layer. Each layer is connected each other, as shown in Fig. 1. In figure, the input nodes is  $x_i$ , the hidden nodes is  $h_j$ , the output nodes is  $y_l$ ; the connection weight between input nodes and hidden nodes is  $w_{ji}$ , the connection weight between hidden nodes and output nodes is  $w_{lj}$ .  $I$  is the number of input nodes,  $J$  is the hidden nodes,  $L$  is the number of output nodes.

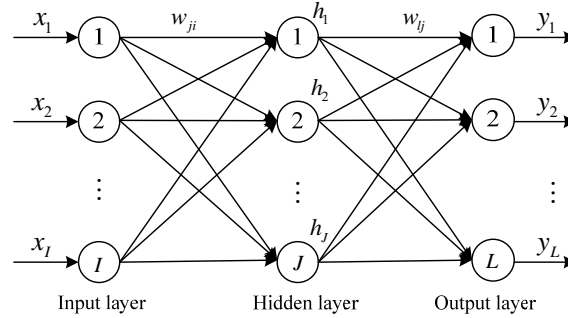


Fig. 1 Structure of BP neural network

The learning process of BP neural network [2] is as follows:

1. Forward propagation process of the mode, from input layer, through hidden layer to output layer;
2. Back propagation process that the connection weights are corrected using the error signal between actual network output and expectation output layer by layer, from output layer, through hidden layer to hidden layer;
3. Memory training process of the network repeated alternately by mode forward propagation process and error back propagation;
4. Learning convergence process that the global network error tends to the minimal value.

## 3. Rough Set

Rough Set (RS) theory is a soft computing method for dealing with uncertainties and inaccuracies after the theory of probability, fuzzy sets and evidence. The basic ideal is to derive the decision-making or classification rules by knowledge reduction under the premise of keeping the classification ability. The classification process: the objects that are not much difference are classified as one class. The basic relationship is indiscernibility. In recent years, RS have been paid more and more attention, and its effectiveness has been confirmed in many fields of science and engineering, which is one of the hotspots in the field of artificial intelligence.

RS is based on the classification mechanism. Knowledge is the ability to classify the object. The granularity of knowledge is thought as the reasons that the existing knowledge cannot determine for describing certain concepts, and discovery the knowledge implied in information systems in whole set approximation manner. Attribute reduction can reduce the overall number of attributes under the premise of keeping the classification ability, which is the basis of knowledge discovery. It has been the core of RS theory research content.

The minimum attribute reduction is to obtain the reduction with the least number of attributes in the whole reduction under the premise of keeping the information system classification ability unchanged. It has been shown that finding all reductions and minimum reduction is an NP-hard problem, so some researchers have proposed heuristic algorithms for attribute reduction, most of which realize the reduction by constructing the relationship between knowledge and information entropy. A heuristic attribute reduction algorithm based on mutual information is introduced [3]. The realization step is as follows:

Input: the decision table  $(U, C \cup D, V, f)$ , the initial attribute set  $T$  is empty;

Output: the attribute reduction result  $T$ .

1. Calculate the mutual information  $W(C;D)$  of the condition attribute set  $C$  and the decision attribute set  $D$ ;

2. Calculate the core attribute set of the condition attribute set  $C$  relative to the decision attributes  $D$ , namely  $C_0 = \text{CORE}_D(C)$ , and calculate  $W(C_0;D)$ . If  $W(C;D) = W(C_0;D)$ , then go to step 5; otherwise continue;

3. Let  $T = C_0$ , the condition attribute set  $C' = C - T$ , calculate the importance of each attribute according to the following formula, and choose the attribute  $a$  that makes  $SGF(a, T, D)$  maximum;

$$SGF(a, T, D) = \frac{W(T \cup \{a\}; D) - W(T; D)}{H(D|a)} \quad (1)$$

4.  $T = T \cup \{a_i\}$ ;

5. If  $W(C;D) = W(T;D)$ , it is terminated, otherwise turn to step 3.

#### 4. Entropy-Based Discretization Algorithm

RS is mostly used to analyze the discrete data. However, for many practical problems, the value of some condition attributes and decision attributes is usually continuous value, needing to be discretized before RS analysis. This is an important problem in the application of RS [26].

In essence, the discretization of condition attributes can be reduced to the problem that the condition attribute space is divided by the selected breakpoint, that is, the  $m$ -dimension space ( $m$  is the number of condition attributes) is divided into finite intervals. Obviously, the decision table discretized using the different division method may be different from the original decision table in compatibility. Assuming that an attribute has  $m$  attribute values, the number of candidate breakpoints that can be considered is  $m-1$ . With the increase in the number of attributes and the sample size, the number of candidate breakpoints doubles, so the efficiency of the breakpoint selection algorithm is very important for discretization.

For a certain continuous attribute, its value range can be divided into several intervals, and each interval contains one sample at least. So  $m$  samples can be divided into  $m$  intervals at most ( $O(m)$ ), that is, the continuous attribute variables can be converted into  $O(m)$  discrete variables. The study shows that the maximum entropy (or information) is achieved when the frequency-probability is distribution the maximum number of attribute values. Since each non-repetitive attribute value of continuous attributes corresponds to one discrete interval, it can be considered as no information (or entropy) loss in discretization process. If the discrete random variable is  $X \subseteq U$ , then its entropy can be defined as

$$H(X) = - \sum_{j=1}^{r(d)} p_j \log_2 p_j, \quad p_j = k_j / |X| \quad (2)$$

It can also be recorded as  $H(p)$ , where  $|X|$  is the number of instances, the number of instances for decision attribute  $j(j=1,2,\dots,r(d))$  is  $k_j$ . In general,  $H(p) \geq 0$ , the smaller  $H(p)$  indicates that the individual decision attribute values in the set  $X$  are dominant, so the degree of confusion is smaller, especially if and only if the decision attributes of instance in  $X$  are the same,  $H(p)=0$ . This ensures that the following discretization process doesn't change the compatibility of the decision table.

The key of discretization of continuous attribute is to reasonably determine the number and

position of discrete division points. Discretization should meet the following two requirements as much as possible: 1. The space dimension should be as small as possible after the discretization of continuous attribute. 2. The loss of information should be as little as possible after the discretization of continuous attribute [4][5].

According to above entropy properties of continuous attribute, an entropy-based discretization method of continuous attribute can be proposed, that is, the value range of each continuous attribute is first divided into several intervals and each interval corresponds to one non-repetition value, then two adjacent intervals are selected to merge in order to minimize the entropy difference before and after the merger. In this process, if there are more than one pair of adjacent intervals with the smallest entropy difference before and after the merger, a pair of them are randomly merged. Repeat this merger process until the stop point is obtained, and the division points for the defined intervals are stored.

It can be seen from the analysis that,  $H(p)$  is a concave function, which increases monotonically with the increase of  $k$ , and the increase rate of  $H(p)$  decreases when  $k$  approaches the maximum. On the function curve of the entropy shown in Figure 3, the maximum of  $H(p)$  is the point  $v_1$  and the corresponding  $k$  is also the maximum. If a chord  $l$  is drawn from the curve starting point  $v_2$  to the point  $v_1$ , all points on the concave function curve are above the chord  $l$ . According to the characteristic of concave function curve, it can be concluded that, at the farthest point  $v_0$  on the curve from the chord  $l$ , namely at the inflection point of the curve, the change of  $k$  is larger than that of entropy. It is also at this point that the best balance is achieved between loss of entropy and moderate interval number. In this way, the point  $v_0$  can be used as the best stop point for merging the adjacent intervals.

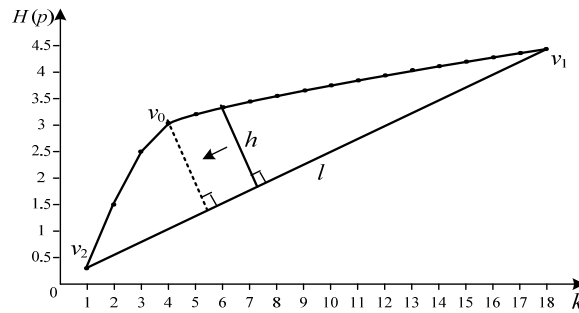


Fig. 2. Determination diagram of best stop point

Take a point on the function curve arbitrarily and from this point draw a vertical line segment  $h$  to the chord  $l$ . The distance from this point to the chord  $l$  is proportional to  $k_{\max}H(p) - H_{\max}(p)(k-1)$ , then the formula can be expressed as

$$h = k_{\max}H(p) - H_{\max}(p)(k-1) \quad (3)$$

where  $k_{\max}$  is the maximum number of intervals,  $H_{\max}(p)$  is the corresponding maximum entropy. It is easy to see from Figure 3 that,  $h$  at the point  $v_0$  is the maximum value, accordingly, the points  $v_0$  and corresponding interval numbers can be obtained.

## 5. Integrating RS with BP Neural Network

For modeling and predicting on stock price trend using BP neural network, while the input dimension is large, the network structure is very complex, which may result in long training time and low accuracy. Especially, the sample data collected from the complex system not only have a large dimension, but also contain many redundant variables, which may result in an unsatisfactory result for BP neural network modeling. The studies have shown that that, RS attribute reduction ability can be used to pre-process the sample data for modeling of BP neural network, which can not only reduce the redundant information, but also decrease the dimension of the input space by analyzing the internal relation between the sample data. The integration based on RS and BP neural

network is shown in Fig. 3.

## 6. Experiment Simulation

To validate the proposed method, the used data is from the famous stock market index: 03/15/2009 ~ 05/24/2014 from Nikkei 225 Index (Japan). To facilitate the modeling and testing, 15 features are used as the inputs of the prediction model to be established. And the output of the prediction model is the stock price trend  $T$ , which can be defined by the percent yield  $R$ , namely Drop:  $R \in (-\infty, -0.5\%]$ , Stability:  $R \in (-0.5\%, 0.5\%)$  and Rise:  $R \in [0.5\%, \infty)$ ,  $R$  can be calculated by

$$R_t = (P_t - P_{t-1}) / P_{t-1} \times 100\% \quad (4)$$

where  $P_t$  is the series of stock price.

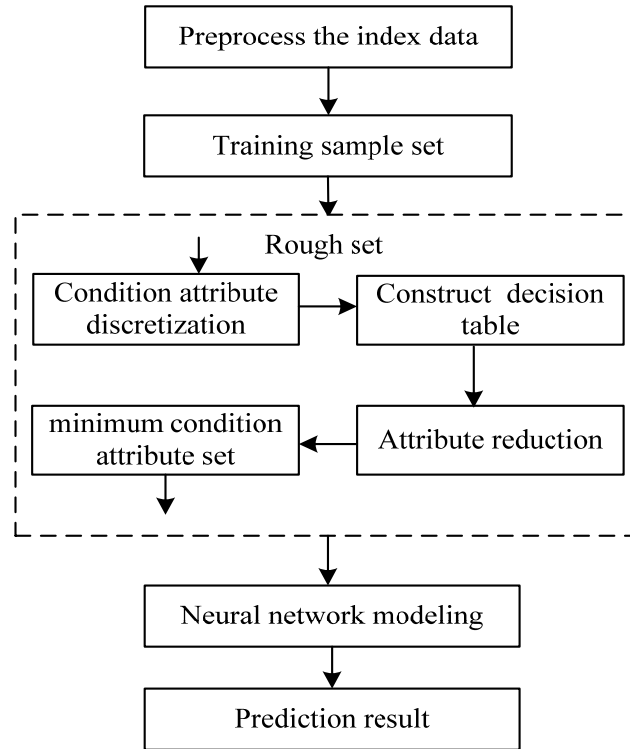


Fig. 3. Integration based on RS and BP neural network

According to the integration step in Fig. 3, 15 features of Nikkei 225 Index daily closing index can be taken as the condition attributes, and the following day variation movement  $T$  of Nikkei 225 Index can be taken as the decision attribute, building the initial decision table of dimension reduction for BP neural network input, as shown in Table 1, and the entropy-based discretization algorithm is applied to discretize the data of condition attributes in the decision table, as shown in Table 2. It can be seen from Fig. 3 that, the number of input nodes of BP neural network can be determined as 9 by RS attribute reduction. Compared with other BPNN and RBFNN models, the proposed RS-BPNN model has obvious advantages in network complexity, training and test accuracy. It shows that, using the attribute reduction ability adequately, RS can reduce the input dimensions of BPNN.

Table 1. Initial decision table of dimension reduction for input variables

Input	Sample ID				
	1	2	3	...	200
$x_1$	14.260	56.035	76.258	...	33.172
$x_2$	52.135	19.532	34.812	...	13.215
$x_3$	6.266	44.185	9.454	...	23.474
$x_4$	-83.183	215.156	47.550	...	-12.045
$x_5$	13.140	60.137	53.022	...	34.186
...	...	...	...	...	...
$x_{15}$	25.850	30.082	40.531	...	16.163
$y$	3	4	4	...	1

Table 2. Discretized decision table of dimension reduction for input variables

Input	Sample ID				
	1	2	3	...	200
$x_1$	1	5	7	...	3
$x_2$	5	2	3	...	1
$x_3$	1	2	1	...	1
$x_4$	3	8	2	...	5
$x_5$	1	6	5	...	3
...	...	...	...	...	...
$x_{15}$	3	3	4	...	2
$y$	3	4	4	...	1

Table 3. Performance comparison between different models for NIKKEI 225 Index

Assessment index	BPNN	RBFNN	RS-BPNN
Model structure	15-14-1	15-9-1	9-12-1
Training time (s)	16.2	9.5	12.5
RMSE	239.14	180.23	125.25
MAD	207.79	145.10	102.43
MAPE	3.95	3.34	3.32
DS%	60.74	61.95	64.27
CP%	63.78	60.90	65.97
CD%	57.89	63.44	66.84

## 7. Conclusion

For stock price prediction, an integration prediction method based on RS and BP neural network is proposed. To discretize the continuous attributes, an entropy-based discretization algorithm is used to preprocess the data in RS attribute reduction. The experiment result shows that the proposed method reduces the inputs number of the network, and has an improvement compared with traditional BP neural network in prediction accuracy. it provides an effective solution for the prediction of stock price trend.

## References

- [1] R. D. Edwards, J. Magee, W. H. C. Bassetti, "Technical analysis of stock trends," CRC Press, Boca Raton, 2007.
- [2] S. B. Ding, F. Wang, "Study on civil aviation safety forecasting method based on BP neural network," Journal of Civil Aviation University of China, vol. 24, No. 1, 2006, pp. 53-56.
- [3] Y. Yan, H. Z. Yang,, "Knowledge reduction algorithm based on mutual information," Journal of Tsinghua University, vol. 47, No. S2, 2007, pp. 1903-1906.
- [4] H. Xie, H. Z. Chang, D. X. Niu, "Discretization of continuous attributes in rough set theory based on information entropy," Chinese Journal of Computers, vol. 28, No. 9, 2005, pp. 1570-1574.
- [5] S. J. Hong, "Use of contextual information for feature ranking and discretization," IEEE Transactions on Knowledge and Data Engineering, vol. 9, No. 5, 1997, pp. 718-730.