Forecasting IPO Price using GA and ANN Simulation

Shi-hao Chou, Yen-sen Ni, William T. Lin
Information Management, Management Sciences, Banking & Finance
Hsing-Wu College, Tamkang University, Tamkang University
10/F, No. 7, Sec. 5 Xinyi Rd., Taipei 11049
Taiwan, ROC
892560441@mail.tku.edu.tw

Abstract: As the development of information technology, the economists are paying more attention to the price simulation, especially the stock price which has already become the reflection of economics. According to this point, this paper will take the trading data of Baidu stock which listed in Nasdaq as an example in order to take out the factor of influencing the stock price. It will settle the most suitable IPO price of new stock, making use of the training stylebook of other Nasdaq history index and in virtue of Genetic Algorithm and Artificial Neuron Network. This paper could offer other interrelated companies rational guidance.

Key-Words: ANN (Artificial Neuron Network), GA (Genetic Algorithm), BP (Back Propagation), IPO (Initial Public Offering)

1 Introduction

As the development of information technology, the economists are paying more attention to the price simulation, especially the stock price which has already become the reflection of economics. However, finance data is so complex and out-of-order that it becomes much harder to dope out. The new technology of AI can provide many new methods to solve these problems. ANN has more comprehensive ability of adapting, studying, mapping and so on. Besides, this fuzzy model is a non-linear one. Therefore, it possesses extensive ability. The method of stock forecasting which bases on ANN mainly uses ANN to train the stock price data, and then to dope by this model. Kryzanowski L pointed out in [3] as a tool that can obtain knowledge, ANN is designed into an organization that imitates obtaining knowledge of human’s brain. ANN studies by inputting the current sample data and changing weight in algorithms. Thus ANN can not only continuously backlog storage but also study according to the empirical knowledge model, and can re-study when the environment changes. The result indicates that ANN has the potential of ability in selecting stock. As a tool that indicates the financial market model, we can use it to predict the future price alignment. Jain and Nag [5] applied ANN in solving the problem of initial price of public negotiable securities. Based on the summary of financial derivatives, Montagna [6] compared ANN and path integral, which are two kinds of the option list price method with traditional calculating method, and he believed ANN was more...
suitable for applying.
GA makes use of the thought of nature choice and evolution to look for the optimal solutions. Speaking essentially, it has the ability to look for the optimal solutions in overall situation. The application is most on the financial market of is BP algorithms which bases on grads degree, but BP net converges slowly and partial superior etc. Sexton [8] used GA instead of the grads degree algorithms, and obtained the results which are better than BP algorithms.
According as traditional ANN arithmetic, this paper synthesizes many factors of influencing the stock price. In addition, it has introduced GA, which could make ANN model approach to the system model, to modulate fuzzy formula or studying weight of ANN.

2 Factors of influencing IPO price
In complete marketable launching condition, IPO price not only refers to financing situation of itself, but also the near-future, metaphase, long-dated investing environment of the whole stock market and the ratio of circulating capital stock in the whole. Hence, this paper use the Nasdaq index of one month, four month, twelve month before the Chinese stock listed in market, and the public company’s total capital stock of the first trading day, launching amount of the stock (data in the graph has been changed from CDR into ordinary stock), the ratio of circulating capital stock in the whole, retained profits one year before, EPS (Earnings Per Share), the ratio of EPS increasing as input variables. At last, we use the closing price as figure object for demonstration as shown in Table 1.

3 ANN model
In order to predict IPO price, we use BP net with many input but one output, as graph 1 shows:

\[
Y_t = f(\sum \omega_{ij}x_i) + \epsilon_t \quad (1)
\]

The ANN model can be denoted by formula (1), in which \(y\) is a dependent variable (it can be a vector, just as variable array), \(y_j(j=1,2,\ldots)\) is a in dependent variable., \(\epsilon\) is a variance.

\[
Y_t = f(\epsilon_{1t}, \epsilon_{2t}, \ldots) + \epsilon_t \quad (1)
\]

In formula (1), \(f\) means any nonlinear structure. ANN includes input layer, hidden layer and output layer. Input layer only receives data but not to conduct, then the neuron in the hidden layer will conduct these data, and then transfers the data to the output layer by transition function. Every node contacts each other by formula (2):

\[
Y_t = f(\sum \omega_{ij}x_i) + \epsilon_t \quad (2)
\]

In formula (2), \(\omega_{ij}\) is a parameter intervened i (input) and j (neuron), function \(f\) corresponds the transition function. This paper uses sigmoid transition function, as graph 2 shows.

\[
f(x) = \frac{1}{(1 + e^{-x})}
\]
As for the learning process of nerve network, is divided into two kinds theoretically: one kind dose not supervised study, namely and uncertain the righteousness output, the input arrays will be carried on a classification by network according to the problems which need to be resolved. The other kind is supervised learning which applies more extensively and widely. The network passes test importation and target exportation to reduce the mistake gradually along with the learning process.

This paper uses Back Propagation algorithm. Calculate the different signal of the last layer and transfer along the steepest route. The standard BP algorithm uses grads descending method to mapping variables of input and output, the method as follows:

The modification of weight is the minus grads from $\varepsilon_p$ to $\omega_{ki}$:

$$\Delta \omega_k = -\gamma \frac{\partial \varepsilon_p}{\partial \omega_{ki}} \frac{\partial \text{net}_k}{\partial \omega_{ki}}$$  \hspace{1cm} (3)

Similarly, the contribute rate of the summation that any nerve cell’s square margin from input to output is given by formula (4):

$$\lambda_{\omega} = \frac{\partial \varepsilon_p}{\partial y_p} \frac{\partial y_p}{\partial \text{net}_k} = (y_p - \tilde{y}_p) f'_k(\text{net}_k)$$  \hspace{1cm} (4)

For the hidden cell which is connect with nerve cell n, its contribute summation is given by formula (5):

$$\lambda_{hn} = \sum \lambda_{\omega} \omega_{np}$$  \hspace{1cm} (5)

4 GA model and implementation

4.1 ANN structure analysis of GA

GA assisting ANN, we can use it in variable optimization, weight optimization, controlling parameter optimization and so on. This paper attempts to ANN topology optimization, namely using GA evolves ANN, and make the error improved continuously through recalling the process, at last get the best route and weight, in this way, we can avoid the inefficient traditional methods.

The best ANN structure that this paper uses can be described by the dummy code as follows:

Begin GA //begin
    g := 0 //set the race generation 0
    InitPopu P(g) // initialization the generation
    EvalPopu P(g) // evaluate the adaptability
    While not done do // if the adaptability=99%, end programmed
        g ++ // add the generation
        Select P(g) from P(g-1) //choice operation
        CrossOver P(g) // cross operation
        Mutate P(g) // Mutate according definite probability
        Evalue P(g) // evaluate the new generation
    End While //end topology structure
End GA

In this process, ANN topology structure of
the two children needs to exchange network and weight in order to create new generation. This paper sets the cross probability 0.2, its cross mode is as follows:

Child 1: topology structure 9-7-1→(9-7 | 1) →(9-7-8-1)
Child 2: topology structure 9-6-8-1→ (9-6 | 8-1) →(9-6-1)

Combine the forepart of Child 1 and the back part of Child 2, come into being new structure (9-7-8-1); Combine the back part of Child 1 and the forepart of Child 2, come into being new structure (9-6-1), hence, the new topology structure are 9-7-8-1and 9-6-1. In order to avoid getting into local optimal solutions in searching process, we need to setup a random mutation function. This paper set the mutation probability 0.04, it denotes that in this probability, not to evolve as intrinsic process but to add or decrease the nerve cell in the hidden layers directly.

For the sake of judging whether going on searching, we need to evaluate current topology structure, this paper use two standards to evaluate: average variance $\varepsilon_1 < 0.01\%$ and the max variance $\varepsilon_2 < 0.5\%$, the calculate formulas are:

$$\varepsilon_1 = \frac{\Sigma (O_{\text{target}} - O_{\text{net}})}{(O_{\text{max}} - O_{\text{min}})}$$

$$\varepsilon_2 = \max (O_{\text{target}} - O_{\text{net}})$$

in which $O_{\text{target}}$ is objective output (namely the closing price in the first day); $O_{\text{net}}$ is the output of ANN (namely calculated price by ANN). When the two standards cause to conform, the searching ends.

In the evolvement mechanism, firstly, we elect a net topology structure and weight as the primal generation, copying, crossing and generating new generation. Set the member number of each generation 50, each group is trained by BP ANN, compared output with the objective result, take the adaptability as standard, elect 10 groups from 50 groups as the parent to generate child, repeat the operation until find the structure whose adaptability is 100%.

4.2 Implementation of GA and ANN

The optimal net structure that we searched by methods of GA and ANN will be studied integrally by BP arithmetic, and we use common method in statistics—Root Mean Square Error to observe the study effect. The method of Root mean square error is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i}^{N} \sum_{j}^{K} (T_{ij} - Y_{ij})^2}{N \times K}}$$

$T_{ij}$ mean the objective output of the $j_{th}$ cell in the $i_{th}$ example; $Y_{ij}$ means the calculated output of he $j_{th}$ cell in the $i_{th}$ example; $N$ means the number of examples; $K$ means the number of cells in the output layer. In this process, the study velocity is adjusted automatically from 0.01 to 0.5 according RMSE changes, until RMSE is convergent. Then we adjust the study velocity adown so that RMSE can approach optimal $S$, but to the Momentum Factor, as it can descend the frequency that the nerve cell weight changes, and make sure the whole net be not too fast to instability, hence this paper elect a bigger Momentum Factor whose value is 0.8, in order to get better study effect.
5 Finding and conclusion

It is supposed to say that IPO price is the one of the most enchanting problem which International Finance Corporation recognizes, because IPO price of making good winning means that the publisher could favorably publish according to the highest price that the investor could accept. In the other hand, it will be a failing price if publication is defeated or publish with too low price. In this process, many factors which included price accepted base line of investors even overstepped controllable scope of publisher and consignee. Actually, a transcendental objective IPO price does not exist. We made out comparatively reasonable price in virtue of GA & ANN and as a reference, we could work out more suitable IPO price. Through this emulated study, after 3000 times studies of 8 teams training datum, we could get comparatively perfect result when we were inputting data enchiridion of Baidu company to calculate. The predictable result of Baidu company’s IPO price is 110.78 dollars and the intraday closing price is 122.54 dollars, so we could say that there are little disparity between them and we would take out comparatively reasonable guidable price. (this paper making use of Java program and running in the environment of JDK1.4.2). This program passed through 2242 times periodic practice and then began to prognosticate with the example of Baidu. The IPO price of Baidu which we prognosticated is 117.6249 dollars and the IPO price of Baidu actually is 122.54 dollars. This result reveals that there are more running times of ANN, the prognosticative numerical value will be more close to actual numerical value, as graph 3 shows:

In this case, we can find that prognosticative numerical value will be very close to actual numerical value when the practice times reach 3000. It shows that ANN&GA will get a preferable result in the process of prognosticating IPO price.

Reference:


Table 1 : The factors that influencing the IPO price

<table>
<thead>
<tr>
<th>Stock</th>
<th>Nasdaq index (one month before)</th>
<th>Nasdaq index (four months before)</th>
<th>Nasdaq index (twelve months before)</th>
<th>total capital stock</th>
<th>launching amount</th>
<th>Ratio of circulating EPS (one year before)</th>
<th>ratio of EPS increasing retained profits</th>
<th>closing price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netease</td>
<td>-5.02%</td>
<td>-16.17%</td>
<td>13.2%</td>
<td>3099.4</td>
<td>450</td>
<td>29%</td>
<td>0.38</td>
<td>35.1</td>
</tr>
<tr>
<td>Sohu</td>
<td>-5.02%</td>
<td>-16.17%</td>
<td>13.2%</td>
<td>3570</td>
<td>450</td>
<td>25%</td>
<td>0.11</td>
<td>27.2</td>
</tr>
<tr>
<td>Tencent</td>
<td>-7.82%</td>
<td>-3.77%</td>
<td>4.32%</td>
<td>17000</td>
<td>5543</td>
<td>33%</td>
<td>0.49</td>
<td>58.0</td>
</tr>
<tr>
<td>NCTY</td>
<td>3.75%</td>
<td>3.20%</td>
<td>3.13%</td>
<td>2290</td>
<td>607.5</td>
<td>26%</td>
<td>0.37</td>
<td>16.5</td>
</tr>
<tr>
<td>SNDA</td>
<td>3.46%</td>
<td>-1.76%</td>
<td>1.68%</td>
<td>7094.7</td>
<td>866</td>
<td>24%</td>
<td>0.32</td>
<td>43.3</td>
</tr>
<tr>
<td>FMCN</td>
<td>-8.13%</td>
<td>9.31%</td>
<td>4.32%</td>
<td>3707</td>
<td>1010</td>
<td>27%</td>
<td>-0.24</td>
<td>24.6</td>
</tr>
<tr>
<td>Elong</td>
<td>4.17%</td>
<td>-7.83%</td>
<td>1.45%</td>
<td>3700</td>
<td>460</td>
<td>12.5%</td>
<td>0.12</td>
<td>11.9</td>
</tr>
<tr>
<td>baidu</td>
<td>-8.13%</td>
<td>9.31%</td>
<td>4.32%</td>
<td>3542</td>
<td>404</td>
<td>11.4%</td>
<td>-0.43</td>
<td>29.4</td>
</tr>
</tbody>
</table>