A Combination Prediction Model of Stock Composite Index Based on Artificial Intelligent Methods and Multi-Agent Simulation

Yongli LI
School of Management, Harbin Institute of Technology
Harbin 150001, P.R.China
E-mail: 0440004@fudan.edu.cn

Chong WU[*]
School of Management, Harbin Institute of Technology
Harbin 150001, P.R.China
E-mail: wuc_hit@126.com

Jiaming LIU
Beijing Institute of Information and Control
Beijing 100037, P.R.China
E-mail: JaminLiu1988@gmail.com

Peng LUO
School of Management, Harbin Institute of Technology
Harbin 150001, P.R.China
E-mail: luopeng_hit@126.com

Abstract

Predicting stock composite index is useful, which can raise the interest of both the investors and the corresponding researchers. This paper presented a new combination prediction model based on the technique of artificial intelligence and the principle of combination forecast. The principle of combination forecast, as a valid foundation of the new model, was strictly proved and carefully illustrated in this paper. Given the predicting rules, the new combination model was established by synthesizing three commonly used prediction models based on the principle of combination forecast. The comprehensive usage of qualitative forecast and quantitative forecast is also a feature of the new model. To valid the new model, comparison analysis and multi-agent simulation were both applied. Besides, the application of multi-agent simulation made the new model able to guide the investors' operations in a real stock market. According to the theoretical proof, the comparison analysis and the simulation experiment, the new combination prediction model tends to be a powerful and applicable tool in making the investment decisions.

Keywords: Combined forecast, Artificial intelligence, Decision analysis, Multi-Agent simulation, Stock composite index, Real world application.

---

1. Introduction

For many years, numerous researchers have been conducting work in the field of forecasting the stock market owing to its high risks and the attractive benefits. Accordingly, a variety of predicting models have been proposed and developed, including these famous artificial intelligence methods such as artificial neural network (ANN) and genetic algorithms (GA), the widely used machine learning method—support vector machine (SVM), and the two widespread methods based on the stochastic process, namely autoregressive moving average model (ARMA) and hidden markov model (HMM). A number of literatures have applied the above methods to make forecasting in the stock market, for example, Faria, et al (2009)[1] studied the principal index of the Brazilian stock market using ANN and the adaptive exponential smoothing method; Huang, et al (2005)[2] forecasted the weekly movement direction of NIKKEI 225 index with SVM; Hassan and Nath (2005)[3] proposed a new approach to predict the local stock market based on HMM. Also, the similar literatures can also be found in the newly published literatures such as Cao, et al (2005)[4], Du, et al (2013)[5], Ni, et al (2011)[6], and Golob, et al (2012)[7]. Therefore, the research on forecasting methods, especially on their applications in the stock market, is an ongoing one and also a useful one, which can still raise the interest of both the investors and the corresponding researchers.

There was a trend to make the hybrid model by adopting two or more of the above mentioned methods comprehensively, for example the so called "ARMA+ANN" fusion model (Luis, et al, 2008)[8], in which the results from ARMA are just the inputs of ANN. Many papers have achieved the conclusion that the performance of the hybrid model dominates the corresponding sole model, such as Hassan, et al (2007)[9], Ho and Oh (2010)[10], Asadi, et al (2012)[11] and so on. The trend prompted us that developing the combination model may be an effective way to lift the performance of the existing models. Accordingly, we aim, following the way of hybrid model, to explore a new way of combining the existing sole models in order to increase the prediction accuracy and enhance the ability of guide in a real world application.

From the literatures about the mentioned hybrid model, it can be found that the way of combining two models is "in series", which means that the output of

one sole model is just the input of the other sole model. The idea immediately encourages us to explore what will happen if combining these sole models "in parallel" rather than "in series". Accordingly, one question occurs that whether the combination model in parallel performs better than the one in series and what the theoretical basis is. It is the first question that we will work on in this paper. Contrast to the existing hybrid model without the valid and sound theoretical proof, the answer to the first question will lay a solid foundation for our new model on why the new model can increase the prediction accuracy. Secondly, the existing prediction models, no matter the sole ones or the hybrid ones, have no prediction rules or only have a few easy ones. As a result, those models can not deal with the complex situations such as the ones of the real stock market. The lack of sound prediction rules will increase the probability of wrong prediction results from those models. Therefore, the second question that we pay much attention to in the new model is how to draw up a series of useful rules so as to enhance the robust of prediction result from the new model. Compared to the existing models, the new combination model, if the prediction rules are sound, will be more credible and useful. Thirdly, both the sole models and the hybrid models can only give the result of prediction, and neither of them can make the operation guidance in a real world application. Taking the stock market as an example, although using the existing model can give a prediction result of the corresponding stock, we still can not know how much to buy or sell the stock and also can not understand how to utilize the result into making our benefits. Here comes the third question: how to make the new model able to give advice on the operations in a real stock market. It is obvious that, if the question can be solved, the new model will be more applicable than those existing ones. Through the comparison and contrast with the existing models, three questions are summarized and they become the three main points in the following paper. Besides, the three questions also show the advantages of the new model which dominates the old ones with some drawbacks.

Revolving around the above mentioned three questions, the rest of the paper is organized as follows: section 2 discusses and introduces how to choose the indicators for predicting a stock based on the principle of technical analysis, which is a necessary part for any prediction models since the indicators are the inputs of

the prediction model; section 3 presents the new combination model step by step, where section 3.1 demonstrates the principle of the new combination model with the theoretical proof in detail, which mainly answers the first question, section 3.2 shows the constitution of the new combination model, and section 3.3 explains the prediction rules to answer the second question; section 4 introduces the multi-agent simulation to make the new model more reasonable in the aspect of giving operation guidance in a real stock market, and also verifies the effectiveness of the new model through the comparison analysis; and then section 5 concludes.

2. Indicator Analysis and Indicator Selection

Generally speaking, the technical analysis is used to predict the stock index. From this viewpoint, the prediction model can be seen as one part of the technical analysis. The so called technical analysis used for predicting the stock trend, according to Bunn and Wright (1991)[12], is to use the stock prices (including the opening price and the closing price), the trading volume as well as other indicators calculated by the above listed figures. Besides, the technical analysis has been confirmed to be useful according to the research done by Kim (2003)[13] and Zhang and Qi (2005)[14].

Based on the above listed literatures, we summarize the widespread indicators used for the technical analysis in Table 1. These indicators are selected as the input variables for the forthcoming combination prediction model presented in this paper.

Following the Mathematical Expression given in Table 1, the vector $\mathbf{x}$ is introduced to denote the input variables, where $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)^T$.

Table 1. The widespread indicators used for the technical analysis.

Index	Explanation	Mathematical Expression
Composite index	It means the closing index or called the closing price (similarly hereinafter)	$y(t)$
Trading volume	The indictor reflects the whole trading volume of one stock during a trading day	$v(t)$
5-Day average	The average of the closing index within five days, so this is often used as a short-term indicator	$x_1 = \frac{1}{5}\sum_{i=1}^{5} y(t-i)$
20-Day average	The average of the closing index within twenty days, so this is often used as a mid-term indicator	$x_2 = \frac{1}{20}\sum_{i=1}^{20} y(t-i)$
60-Day average	The average of the closing index within sixty days, so this is often used as a long-term indicator	$x_3 = \frac{1}{60}\sum_{i=1}^{60} y(t-i)$
5-Day average trading volume	The average of the trading volume within five days, so it reflects the stock transaction in the short term. Besides, it is an important indicator to reflect the investor's enthusiasm, and it's of great significance for the composite index's prediction	$x_4 = \frac{1}{5}\sum_{i=1}^{5} v(t-i)$
DIF	Reflecting the difference between the short-term average and the long-term average. Besides, it also reflects the power of buyers and sellers in a stock market. If the buyer is more powerful, the stock's price will go up so that DIF is positive, and however if the seller is more powerful, the price will go down so that DIF is negative	$x_5 = x_1 - x_3$
DEA	The average of DIF within nine trading days, and it is the improvement of the above DIF, so it shares the similar meanings with DIF	$x_6 = \frac{1}{9}\sum_{i=1}^{9} x_5(t-i)$

3. The combination prediction model

3.1. *The principle of combination prediction model*

Contrast to the existing hybrid models with combining the sole models in series, the presented new combination prediction model will follow the way of "in parallel". On the one hand, in order to reduce the individual risk of the sole model or the hybrid model, we suggest exploiting a combination prediction model based on the majority voting rule; on the other hand, in order to lift the prediction accuracy, we try to fully play the role of every adopted model in the combination prediction model, and so we explore a new way to combine the sole models. However, why can the new combination prediction model achieve the above two goals? And what is its theoretical basis? The answers are included in the following lemma and its corollary.

[**Lemma 1**] If the results of different prediction models are independent and every sole model has the identical predication accuracy which is no less than 50%, then the prediction accuracy of the combination model is greater than that of any sole model when the majority voting rule is adopted and n is an odd number.

Proof. We denote θ_i as the predication accuracy of the *i-th* model, and according to the condition of the above lemma, we know that $\theta = \theta_i$ for any $i \in \{1, 2, \cdots, n\}$. Next, based on the result of each single model is independent with each other, the prediction accuracy of the combination model, denoted by $\tilde{\theta}$, can be expressed as

$$\tilde{\theta} = \sum\nolimits_{\{\sum_{i=1}^{n} \sigma_i > \frac{n}{2}\}} \prod\nolimits_{i=1}^{n} \theta_i^{\sigma_i} (1-\theta_i)^{1-\sigma_i}, \qquad (1)$$

where, σ_i is the indicator variable of the *i-th* model, that is, if the *i-th* model gives the right prediction, σ_i equals 1; otherwise, σ_i equals 0. Furthermore, the formula $\sum_{i=1}^{n} \sigma_i > n/2$ is the condition of summing according to the majority voting rule, because the combination model can give the right predication if and only if the above condition holds.

We further need to prove that $\tilde{\theta} \geq \theta$ under the condition of the above lemma. In fact, it is obvious that the formula (1) can be rewritten as

$$\tilde{\theta} = \theta^n + C_n^1 (1-\theta)^1 \theta^{n-1} + \cdots \\ + C_n^{(n-1)/2} (1-\theta)^{(n-1)/2} \theta^{(n+1)/2}, \qquad (2)$$

from which we can obtain

$$f(\theta) = \frac{\tilde{\theta}}{\theta} - 1 = \theta^{n-1} + C_n^1 (1-\theta)^1 \theta^{n-2} + \cdots \\ + C_n^{(n-1)/2} (1-\theta)^{(n-1)/2} \theta^{(n-1)/2} - 1, \qquad (3)$$

It is obvious that when θ equals 0.5 or 1, $f(\theta)$ in the above formula (3) equals 0. Then according to mathematical analysis, we can obtain that

$$\frac{\partial^2 f(\theta)}{\partial \theta^2} < 0, \text{ for any } \theta \in [0.5, 1], \qquad (4)$$

which indicates that $f(\theta) \geq 0$ for any $\theta \in [0.5, 1]$ because of the intermediate value theorem for continuous function. It infers from the above result that $\tilde{\theta} \geq \theta$, when θ is between 0.5 and 1. Thus, the lemma holds. $\qquad \square$

However, the conditions of the above lemma can't always be satisfied, especially the condition that every sole model has the identical predication accuracy. Fortunately, the function $\tilde{\theta}(\theta_i)$, expressed by formula (1), is a continuous one about θ_i, so according to the definition of the continuous function, we have that $\exists \; \delta, \varepsilon$, when $|\theta_i - \theta| < \delta$, the inequality $|\tilde{\theta}(\theta_i) - \tilde{\theta}(\theta)| < \varepsilon$ holds. From the above reasoning process, we can further obtain the corollary as follows.

[**Corollary 1**] For any $\theta_i \in (0.5, 1)$ where $i \in \{1, 2, \cdots, n\}$, $\exists \; \delta$, when $\max \theta_i - \min \theta_i < \delta$, the inequality $\tilde{\theta} \geq \max \theta_i$ holds.

The above corollary means that the prediction accuracy of the combination model can be higher than that of any sole model given some conditions as shown in the Lemma 1 and the Corollary 1. Therefore, the theoretical result explains why we choose the combination model to predict the stock index.

In order to easily understand the above lemma and corollary, we explain them by taking a combination model consisting of three sole models as an example, which is showed in Table 2. Table 2 demonstrates the correctness of the Lemma 1 and the Corollary 1. As the Table 2 shows, the former two cases reflect the results of the Lemma 1, and the other two cases exemplify the Corollary 1 with δ being 0.1 in the case 3 and 0.15 in the case 4.

Table 2. An exemplification of the Lemma 1 and the Corollary 1.

The accurate rate of predication	Case 1	Case 2	Case 3	Case 4
Model 1	50%	60%	60%	65%
Model 2	50%	60%	65%	70%
Model 3	50%	60%	70%	80%
Combination	50%	64.8%	71.9%	80.7%

From the example shown in Table 2, we can find that the Corollary 1 relaxes the condition of the Lemma 1, namely there is no need to require that every sole model has the identical predication accuracy, and what we need is that their predication accuracies are near to each other. The Corollary 1 enables the theorem of combining the sole models in parallel to be sound and reasonable.

Accordingly, the Lemma 1 and the Corollary 1 explain why the combination prediction model in parallel is needed rather than the sole ones or the hybrid ones. One thing should be paid attention to that the sole models which constitute the combination model must be independent with each other, as the condition of the Lemma 1 or the Corollary 1 shows. Fortunately, it is not a problem in practice, because the sole models are different with each other and also come from the distinguished model-making principles; as a result, they are independent naturally. The next part will reveal the above statement in detail.

3.2. Three sole models constituting the combination model

The presented combination model consists of three sole models which are ARMA+SVM, ARMA+ANN and HMM, respectively. Among the three sole models, the former two can provide qualitative and quantitative results simultaneously, while the last one is a qualitative model which only can predict a stock's three trends, namely the rise, the fall and the flat. Besides, the former two are the hybrid models sharing the identical principle which reflects the non-linear character of the stock.

In detail, the principle of AMRA+SVM and ARMA+ANN is to decompose a stock index as follows:

$$\text{stock index} = \text{linear explainable part} + \text{non-linear part} + \text{white noise}, \quad (5)$$

where, AMRA corresponds to the linear explainable part since it is a linear model, and SVM or ANN is a non-linear model so that either can be chosen as the non–linear explainable part. The last part (white noise), always assumed to meet Gaussian distribution, reflects the random interferences which are not able to be explained by the former two parts. The advantages of such hybrid models are validated by many papers, such as Vapnik. et al (1997)[15], Pai and Lin (2005)[16], Li, et al (2009)[17], Li, et al (2013)[18] and so on. Furthermore, the equivalent mathematical expression of the Equation (5) is

$$y_t = \bar{y}_t + \tilde{y}_t + w_t, \quad (6)$$

where, $\bar{y}_t$, $\tilde{y}_t$ and w_t respects the linear explainable part, the non–linear explainable part and the white noise, respectively. Understandably, $\bar{y}_t$ expresses ARMA, and $\tilde{y}_t$ expresses SVM or ANN in this paper. The detailed information of the three sole models is shown below.

(i) Sole Model 1: ARMA+SVM.

ARMA can be constructed as follows:

$$\bar{y}_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t \\ - \gamma_1 \varepsilon_{t-1} - \gamma_2 \varepsilon_{t-2} - \cdots - \gamma_q \varepsilon_{t-q}, \quad (7)$$

where, y_{t-j} ($j = 1, 2, \cdots, p$) is the real stock index at the time $t - j$, ε_{t-i} ($j = 0, 1, \cdots, q$) is the white noise at the time $t - i$, and the other variables in Formula (7) are the parameters which will be identified after fitting the model.

After the above step, we can have Δy_t defined to be

$$\Delta y_t = y_t - \bar{y}_t = \tilde{y}_t + w_t. \quad (8)$$

The Δy_t ($t = 1, 2, \cdots$) is one of the input variables of SVM since it includes the information which can not be explained by the linear part—ARMA. Then, we write the samples as follows,

$$(\mathbf{x}_t, \Delta y_t), \text{ with } t = 1, 2, \cdots, \quad (9)$$

where, $\mathbf{x}_t$ has been defined in section 2. These samples are the inputs of SVM which can also be called the learning samples.

Next we use the SVM and these samples to predict $\tilde{y}_t$. The detailed SVM regression method applied here is introduced as below.

First, we make the regression equation as

$$f(\mathbf{x}_t) = w\phi(\mathbf{x}_t) + b, \quad (10)$$

where, $\phi(\cdot)$ is a function satisfying the following equation $K(x,y) = \phi(x) \cdot \phi(y)$, and $K(x,y)$ is a Gauss function and is the SVM regression model's kernel function.

Second, to determine the value of w and b appearing in the Equation (10), it is necessary to introduce the following optimization problem:

$$\min_{w,b} \quad C\frac{1}{n}\sum_{t=1}^{n}L_\varepsilon(\Delta_t) + \frac{1}{2}\|w\|^2$$

$$s.t. \quad L_\varepsilon(\Delta_t) = \begin{cases} |f(\mathbf{x}_t) - \Delta_t| - \varepsilon, & |f(\mathbf{x}_t) - \Delta_t| \ge \varepsilon \\ 0, & |f(\mathbf{x}_t) - \Delta_t| < \varepsilon \end{cases},$$

(11)

where, C is the penalty coefficient and ε is the precision coefficient. The two values can be set in order to get the best predicting results by comparisons.

(ii) Sole Model 2: ARMA+ANN.

The first step is identical to the sole model 1 and shares the formulas from (7) to (9). After the first step, we can get the samples showed in Formula (9). Then, we predict the non–linear explainable ($\tilde{y}_t$) using the famous artificial intelligent technology—ANN.

The structure of ANN applied in this paper is showed in Fig.1 where six nodes (corresponding to $\mathbf{x}_t$) constitute the input layer and one node (corresponding to Δy_t) constitutes the output layer.
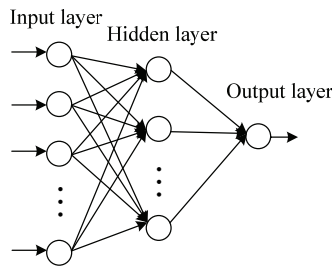

Fig.1 Structure Chart of Artificial Neural Network

The number of nodes in the hidden layer can be determined through numerous experiments in order to find the best one which makes the whole ANN owns the highest prediction accuracy. To train the above ANN, the Back-Propagation algorithm (BP) is needed by which we can get the two matrixes from the input layer to the hidden layer and from the hidden layer to the output layer, respectively. Then, we can get the prediction value of the non–linear explainable denoted

by $\tilde{y}_t$. The last step is the same with the sole model 1—ARMA+SVM.

(iii) Sole Model 3: HMM.

Essentially, HMM is a dual random process, which includes a random process of Markov Chain with a hidden state and a random process of generating the observed values. The topological structure of HMM is shown in fig.2.
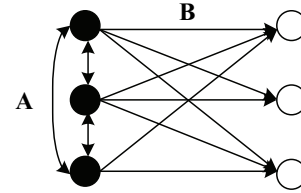

Fig.2 Topological Structure of the HMM

For generating the HMM and training it, three existing algorithms are useful, which are Forward–Backward algorithm solving the calculation issue of the given model, Viterbi algorithm aiming to solve the issue of model decoding, and Baum-Welch algorithm to solve the issue of model training.

In this paper, the HMM consists of three hidden nodes reflected by the three solid circles in Fig.2 and three output nodes depicted by the three empty circles in Fig.2. Given the series sg_t defined in the Formula (12), Baum-Welch algorithm introduced above can be applied to cope with the model training problem; As a result, the state-transmission matrix $\mathbf{A}$ among these hidden nodes and the phenotype matrix $\mathbf{B}$ from the hidden nodes to the output nodes can be achieved. Then, Forward–Backward algorithm will take effect to provide the prediction result denoted by g_t which reflects the trend of the stock index.

$$sg_t = \begin{cases} +1 & \text{if } (y_t - y_{t-1})/y_{t-1} > 0.5\%, \\ -1 & \text{if } (y_t - y_{t-1})/y_{t-1} < -0.5\%, \\ 0 & \text{others.} \end{cases}$$ (12)

Accordingly, the specific steps are listed below, and the programming codes are showed in the supplement materials.

Step 1: Calculating the series sg_t according to the Formula (12). In fact, the series can be seen as the learning samples of the HMM;

Step 2: Applying the Baum-Welch algorithm to obtain the matrix $\mathbf{A}$ and $\mathbf{B}$;

Step 3: Based on **A** and **B**, the trend of the predicted stock index can be achieved via Forward–Backward algorithm.

Following all the three steps, we can obtain the qualitative predicting results.

3.3. *Comprehensive predicting rules for the combination model*

The first two sole models can provide the quantitative results, and their results are denoted by $_1p_t$ and $_2p_t$ respectively, then their average gives the comprehensive quantitative result denoted by p_t here. Subsequently, the predicted volatility denoted by r_t can be defined as follows:

$$r_t = \frac{p_t - y_{t-1}}{y_{t-1}} = \frac{_1p_t + _2p_t - 2y_{t-1}}{2y_{t-1}} . \qquad (13)$$

Meanwhile, the third sole model introduced in section 3.2 can only provide the qualitative result which has been denoted as g_t.

Table 3. Comprehensive predicting rules and corresponding operations.

g_t	r_t	Decisions	Operations
	$\geq 2\%$	intensely bullish	buy at a large amount
	$[0.5\%, 2\%)$	bullish	buy at a proper amount
$+1$	$[-0.5\%, 0.5\%)$	slightly bullish	buy at a small amount
	$[-2\%, -0.5\%)$	not clear	wait and no operations
	$\leq -2\%$	not clear	wait and no operations
	$\geq 2\%$	slightly bullish	buy at a small amount
	$[0.5\%, 2\%)$	not clear	wait and no operations
0	$[-0.5\%, 0.5\%)$	not clear	wait and no operations
	$[-2\%, -0.5\%)$	not clear	wait and no operations
	$\leq -2\%$	Slightly bearish	sell at a small amount
	$\geq 2\%$	not clear	wait and no operations
	$[0.5\%, 2\%)$	not clear	wait and no operations
-1	$[-0.5\%, 0.5\%)$	slightly bearish	sell at a small amount
	$[-2\%, -0.5\%)$	bearish	sell at a proper amount
	$\leq -2\%$	intensely bearish	sell at a large amount

The comprehensive prediction rules are listed in Table 3 with the suggestions on how to operate the corresponding stock. By examining the above comprehensive decisions and operations, we are able to find that the presented model is a robust one because the suggestions are provided only when the qualitative model is accordance with the quantitative model. It is noted that the qualitative model will decide the direction of the operation, while the quantitative model will decide the strength of the operation. From this viewpoint, the comprehensive model not only would predict the results more correctly since the risk of system including three sole models is low, but also would make more moderate investment advices because the advice on investment is based on three models rather than one or two of them. Furthermore, the above Table 3 will also be useful in the designing the intelligence of the agents in the section of multi-Agent simulation.

4. Model verification and multi-Agent simulation

4.1. *Data select and basic results*

We selected the opening prices and closing prices of Shanghai Stock Exchange (SSE for short) Composite Index (Ticker code: 000001) from June 2nd, 2011 to November 10th, 2011 as the whole sample for training and testing the model. The whole sample is divided into two parts: the training sample (from June 2nd, 2011 to August 23rd, 2011) and the testing sample (from August 24th, 2011 to November 10th, 2011). The test sample number during the period is 100, since one trading day has two numbers, namely the opening price and the closing price.

Based on the section 2 of this paper, part of the original data is shown in Table 4, where the items listed

in Table 1 can be found. As shown in Table 4, one trading day has two rows to respond to, in which the first row is the opening price (index) and its corresponding information having been known and the second row presents the closing pricing and the other indicators which can be obtain at that time. It is noted that we use our combined model to predict both the

opening price and the closing price in one trading day because it is more useful for guiding the operations in a real stock market. That is to say, one person has more opportunities to achieve his benefits or avoid his loss when he can grasp more prediction results than when he can only predict the closing price as the existing models have done.

Table 4. Part of the original data

Trading Date	Opening or Closing Index	Opening or Closing Rate of Return	Trading volume	5-Day average	20-Day average	60-Day average	5-Day average trading volume	DIF	DEA
2011/06/02	2715.32	-1.0297	794443.63	2728	2815.7	2917	823193.625	-54.009	-46.974
2011/06/02	2705.18	-0.3734	968065.63	2721.7	2807.325	2912.15	844211.188	-55.393	-48.658
2011/06/03	2701.77	-0.1261	968065.63	2721.7	2807.325	2912.15	844211.188	-55.393	-48.658
2011/06/03	2728.02	0.9716	747106.88	2725.3	2800.525	2907.617	806425.625	-54.024	-49.731
2011/06/07	2720.97	-0.2584	747106.88	2725.3	2800.525	2907.617	806425.625	-54.024	-49.731
2011/06/07	2744.3	0.8574	717843.06	2732.9	2794.125	2903.325	802355.188	-51.037	-49.992
2011/06/08	2742.04	-0.0824	717843.06	2732.9	2794.125	2903.325	802355.188	-51.037	-49.992
2011/06/08	2750.29	0.3009	805516.5	2734.3	2787.125	2899.883	806604.813	-47.637	-49.521
2011/06/09	2743.54	-0.2454	805516.5	2734.3	2787.125	2899.883	806604.813	-47.637	-49.521
2011/06/09	2703.34	-1.4653	887054.19	2726.2	2778.1	2896.033	825113.625	-48.176	-49.252
2011/06/10	2696.13	-0.2667	887054.19	2726.2	2778.1	2896.033	825113.625	-48.176	-49.252
2011/06/10	2705.14	0.3342	796388.63	2726.3	2771.175	2892.167	790784	-47.906	-48.983

Table 5. The basic results of prediction

Prediction accuracy	Sole model			Combination model
	HMM	ARMA+SVM	ARMA+ANN	
Bullish	24.1%	62.1%	31.0%	42.2%
Flat	82.1%	67.9%	75.0%	82.5%
Bearish	69.8%	65.1%	74.4%	76.0%
Overall	68.9%	66.2%	67.1%	73.8%

Applying the above models, we can obtain the results listed in Table 5, where the prediction accuracy of each model is compared. Table 5 shows that the prediction accuracy of three sole models and the combination one in the three cases—"Bullish", "Flat" and "Bearish". Here, when the rate of return is more than 0.25%, the state of the stock trend is defined as "Bullish", when the rate of return lies between -0.25% and 0.25%, the state is defined as "Flat", and in other cases, the state is defined as "Bearish".

Following the explanations and steps shown in the section 3.2 of this paper, we can get the prediction results of ARMA+SVM, ARMA+ANN and HMM, respectively. In detail, as for ARMA+SVM and ARMA+ANN, we need to fit the ARMA part first, in which the rate of return is the dependent variable and these indictors denoted by v and $\mathbf{X}$ (see also Table 1 for the details) are the independent variables. Then we can get the non-linear part according to the Formula (6) and fit it by SVM or ANN. Here, we use the former two numbers to predict the next number in the usage of

SVM and we adopt the 30×20×1 network structure to establish the ANN. As for the HMM sole model, we can follow the steps listed in section 3.2 to get its prediction results. It is noted that the process of prediction is dynamic.

Although the two models ARMA+SVM and ARMA+ANN provide the quantitative outcomes, the corresponding qualitative results are also included so that we can judge the trend of a stock, thus it is no doubt that the two quantitative models are comparable with the qualitative model HMM. Through comparing the numbers listed in the Table 4, it is easily found that the combination model enjoys the highest overall prediction accuracy. It is interesting and important that the results are accord with the Lemma 1 and the Corollary 1 proved in the section 3. The more detailed prediction results of the three sole models are further presented in the following fig.3 to fig.5.
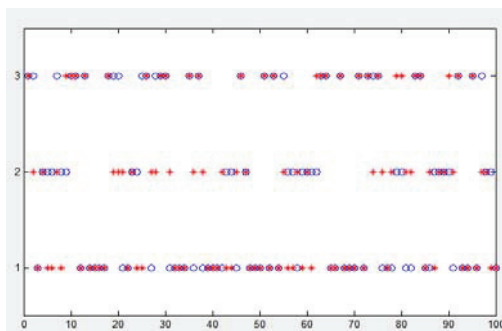

Fig.3 The prediction results of HMM

In the Fig.3, the horizontal axis shows the 100 sequential samples, and the vertical axis expresses the three states where "1" stands for Bearish, "2" for flat and "3" for Bullish. The bule circles show the real states of the stock on the corresponding sample, and the red stars show the prediction results, thus if and only if the two symbols coincide, the model gives the right forecast.

The horizontal axis in the Fig.4 or the Fig.5 is the same with that in the Fig.3, while the vertical axis in the Fig.4 or the Fig.5 shows the volume of stock fluctuation, which can be calculated by the Formula (13). Besides, the bule circles and the red stars have the similar meanings with those appearing in the Fig.3.
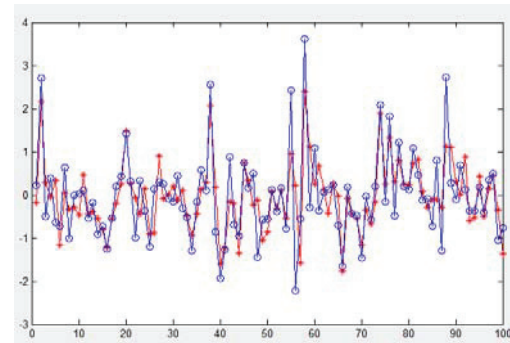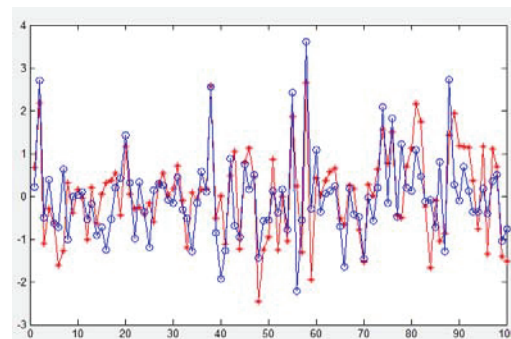

Fig.4 The prediction results of ARMA+SVM

Fig.5 The prediction results of ARMA+ANN

4.2. *Multi-Agent simulation*

The given comparisons only provide the information of which model has the highest prediction accuracy, and however it disables to guide the operations in the stock market. Therefore, it is necessary to make a multi-agent simulation to validate these models comprehensively by comparing the agents' profits.

We plan to design four types of agents which use the presented combination model, the qualitative model (based on the two models AMRA+SVM and ARMA+ANN), the quantitative model (based on HMM) and the random strategy respectively. Their descriptions and the corresponding intelligences are listed in Table 5.

Next, we set the trading constraints according to the real stock market of China as follows: (1) short-sell is not allowed; (2) T+1 trading, which means the bought stocks can not be sold at the same trading day; (3) transaction fees are 7‰ of turnover and are charged when buying. Meanwhile, the above designed four agents are allocated 10000 units of money initially.

Based on the whole intelligences and the trading constraints, the process of stimulation is illustrated in the Fig.6 and the final results are presented in the Fig.7 which have undergone 100 simulation cycles sharing the same periods with those showed in the Fig.3, the Fig.4 or the Fig.5.

Table 5 Four types of Agents and their intelligences

Agent type	Description	Intelligence
combined	carrying out their operations based on the presented combination model of this paper	According to the decisions and operations listed in the Table 3, (1) When the market is predicted to be intensively bullish, use all the remaining money to buy; (2) When the market is predicted to be bullish, use 50% of the remaining money to buy; (3) When the market is predicted to be slightly bullish, use 25% of the remaining money to buy; For the above three cases, if no currency exists at the moment, hold the stock. The situation of bearish is just opposite to the bullish counterpart. (4) When the market is predicted to be not clear, make no operations at all;
quantitative	carrying out their operations based on the average of two quantitative models (ARMA+ SVM and ARMA+ANN)	According to the average of the two models' results, (1) When the average is no less than 2%, the market is predicted to be intensively bullish, so use all the remaining money to buy; (2) When the average lies between 0.5% and 2%, the market is predicted to be bullish, so use 50% the remaining money to buy; (3) When the average lies between -0.5% and 0.5%, the market is predicted to be flat, so make no operations at all; (4) When the average lies between -2% and -0.5%, the market is predicted to be bearish, so sell 50% of the holding stocks; (5) When the average lies no more than -2%, the market is predicted to be intensively bearish, so sell all the holding stocks; If no remaining money or no holding stocks can be used, keep the current status.
qualitative	carrying out their operations based on the HMM	According to the results from the HMM, (1) When the market is predicted to be bullish, use all the remaining money to buy; (2) When the market is predicted to be flat, make no operations at all; (3) When the market is predicted to be bearish, sell all the holding stocks; If no remaining money or no holding stocks can be used, keep the current status.
random	carrying out their operations based on tossing a coin to decide to buy or to sell without consulting any prediction model	Randomly choose trading or no operating by generating two groups of random numbers where each group has three numbers with the same probability. (1) The first group decides the direction of operations, that is, to buy, to sell or to make no operations; (2) The second group decides the strength of operations, that is, 100%, 50% or 0. If no remaining money or no holding stocks can be used, keep the current status.

In the Fig.6 and the Fig.7, "1" and "2" correspond to the combined agent, "3" and "4" to the quantitative agent, "5" and "6" to the qualitative agent, and "7" and "8" to the random agent. Moreover, the blue bars express the amount of the currencies these agents have at the moment, and the red bars show the amount of the stocks these agents hold at the same simulation cycle as the blue ones. These agents carry out their operations according to their corresponding intelligences, and buy or sell their stocks based on the real price of the stock on that trading day. The Fig.7 illustrates the final profit of each agent in the form of currency after 100 trading days, from which we can find that the combined agent has the highest profitability. In all, the simulation analysis validates that the combined agent, who is guided by the combination model presented in this paper, can obtain satisfactory profits and is prior to the others. Thus, the newly proposed model is more applicable and effective than the other three.
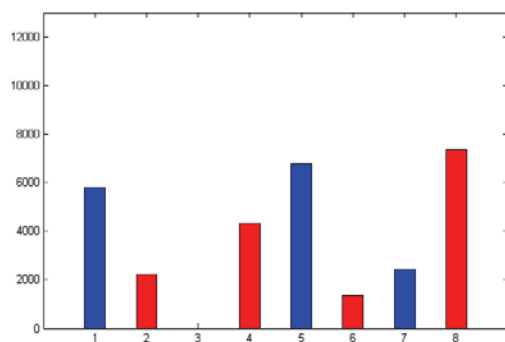
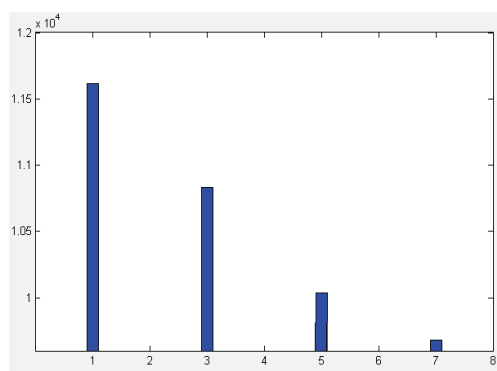Fig.6 The process of muti-agent stimulation

Fig.7 The profits of four kinds of agents

5. Conclusions

This paper presented a new combination forecast model, explained why to choose the method of combination forecast by proving an important principle, and validated the model based on a multi-agent simulation. The principle of combination forecast and the techniques of artificial intelligent are two cornerstones of this paper. In fact, the former interprets the rationality of establishing such a combination model, and the latter take effects not only in the sole models and the combination one, but also in the multi-agent simulation. Furthermore, the specialty of this paper can be summarized to be the application of multi-agent simulation to compare the combination model with the sole ones, which helps to demonstrate the functions of the newly proposed model comprehensively. Besides, to compare the prediction accuracy is commonly used when to validate several models, but it is not sound because it can not be applied directly to guide the operations in a stock market; However, the multi-agent

simulation can solve the problem well since the intelligences of agents are similar to the humankind and includes the guidance on when to buy or sell and how much of it in face of the real stock market.

Based on the whole model and the simulation analysis in this paper, the prediction accuracy of the combination model reaches up to 73.8%, and the profit of the combined agent is a little more than 16% which is higher than the others without use of the combination model. In all, the results are satisfactory. Therefore, we tend to believe that the model can benefit the investors in the real stock market.

References

1. Faria, E., Marcelo, P. A., Gonzalez, J.L., Cavalcante, J. and Marcio, P.A., Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods, *Expert Systems with Applications*. 36 (2009) 12506–12509.
2. Huang W., Nakamori, Y. and Wang, S.Y., Forecasting stock market movement direction with support vector machine, *Computers & Operations Research*. 32 (2005) 2513–2522.
3. Hassan, M.R. and Nath, B., Stock market forecasting using hidden Markov model: a new approach, *Proceedings of Intelligent Systems Design and Applications*. (2005) 192–196.
4. Cao Q., Leggio, K.B. and Schniederjans, M.J., A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market, *Computers & Operations Research*. 32 (2005) 2499–2512.
5. Du, X.F., Leung, S.C.H., Zhang, J.L. and Lai, K.K, Demand forecasting of perishable farm products using support vector machine, *International Journal of Systems Science*. 44(3) (2013) 556-567.
6. Ni, L.P., Ni, Z.W. and Gao, Y.Z., Stock trend prediction based on fractal feature selection and support vector machine, *Expert Systems with Applications*. 38 (2011) 5569–5576.
7. Golob, K., Bastic, M. and Psunder, I., Analysis of Impact Factors on the Real Estate Market: Case Slovenia, *Inzinerine Ekonomika-Engineering Economics*. 23(4) (2012) 357-367.
8. Luis A. Dı´az-Robles, Juan C. Ortega, Joshua S. Fu, Gregory D. Reed, Judith C. Chow, John G. Watson, Juan A. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile, *Atmospheric Environment*. 42(35) (2008) 8331–8340.
9. Hassan, M.R., A combination of hidden Markov model and fuzzy model for stock market forecasting, *Neurocomputing*. 72 (2009) 3439–3446.

10. Ho, C.T.B. and Oh, K.B., Selecting Internet company stocks using a combined DEA and AHP approach, *International Journal of Systems S cience*. 41(3) (2010) 325-336.

11. Asadi, S., Hadavandi, E., Mehmanpazir, F. and Nakhostin, M. M., Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction, *Knowledge-Based Systems*. **35** (2012) 245–258.

12. Bunn, D. and Wright, G., Interaction of judgemental and statistical foresting methods: issues and analysis, *Management Science*. 37(5) (1991) 501-518.

13. Kim, K., Financial time series forecasting using support vector machines, *Neurocomputing*. 55 (2003) 307-319.

14. Zhang, G.P. and Qi, M., Neural network foresting for seasonal and trend time series, *European Journal of Operational Research*. 160 (2005) 501-514.

15. Vapnik, V., Golowich, S., and Smola, A. (eds.), Support vector method for function approximation, regression estimation, and signal processing. Advances in Neural Information Processing Systems, 9th edn. (Cambridge, MIT, 1997).

16. Pai, P.F. and Lin, C.S., A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*. 33 (2007) 497-505.

17. Li, Y.L., Yuan, W.J. and Wang, K.S., Improvement of risk measure and its application to the risk management, in *Proceedings of 2010 International Conference on Management Science & Engineering*, (Melbourne, 2010), 1284- 1289.

18. Yongli L., Chong W., Xudong W. and Shitang W. A tree-network model for mining short message services seed users and its empirical analysis. *Knowledge-Based Systems*, 40 (2013) 50-57.