

## Research Article

# An Efficient Stock Recommendation Model Based on Big Order Net Inflow

Yang Yujun,<sup>1,2,3</sup> Li Jianping,<sup>1</sup> and Yang Yimei<sup>2</sup>

<sup>1</sup>School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

<sup>2</sup>Department of Computer Science and Technology, Huaihua University, Huaihua 418008, China

<sup>3</sup>Hunan Provincial Key Laboratory of Ecological Agriculture Intelligent Control Technology, Huaihua 418008, China

Correspondence should be addressed to Yang Yimei; [yym@hhtc.edu.cn](mailto:yym@hhtc.edu.cn)

Received 14 August 2015; Revised 10 December 2015; Accepted 15 December 2015

Academic Editor: David Bigaud

Copyright © 2016 Yang Yujun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In general, the stock trend is mainly driven by the big order transactions. Believing that the stock rise with a large volume is closely associated with the big order net inflow, we propose an efficient stock recommendation model based on big order net inflow in the paper. In order to compute the big order net inflow of stock, we use the M/G/1 queue system to measure all tick-by-tick transaction data. Based on an indicator of the big order net inflow of stock, we select some stocks with the higher value of the net inflow to constitute the prerecommended stock set for the target investor user. In order to recommend some stocks with which this style is familiar them to the target users, we divide lots of investors into several categories using fuzzy clustering method and we should do our best to choose stocks from the stock set once operated by those investors who are in the same category with the target user. The experiment results show that the recommended stocks have better gains during the several days after the recommended stock day and the proposed model can provide reliable investment guidance for the target investors and let them get more stock returns.

## 1. Introduction

In the area of stock recommendation method research [1], most of the research mainly focuses on the two areas: stock recommendation methods based on stock comment [2] and price forecasting [3]. The former is easy to understand and master for investors. However, in such a complicated stock market, the investors do not know which one to believe among the lots of stock comments with dubious authenticity and every choice has a great risk for them. The latter method [4] is difficulty for investors to understand and master since the application of the latter method is relatively complex and involves a lot of profound mathematical knowledge [5]. Given this situation, many scholars have done a lot of research on the stock recommendation [6].

Currently, the stock recommendation based on price forecasting relies [7] mainly on mathematical and statistical methods [8], time series model [9, 10], and machine learning model [11]. Sonsino and Shavit [12] have researched a stock

prediction and selection method based on unidentified historical data. M.-Y. Chen and B.-T. Chen [13] have proposed a stock price forecasting method based on the hybrid fuzzy time series and granular computing. Xin et al. [14] have given a strategy for filtering out users with similar demand characteristics by using collaborative recommendation algorithm with fuzzy clustering method, which shows excellent recommendation effect.

In present financial field, how to integrate multiple technologies [15], such as data mining, machine learning and herd psychology [11], and other nontraditional technologies, into stock recommendation has become a hot topic. Few papers use money net inflow as stocks recommendation techniques. Given this situation, we proposed an efficient stock recommendation model based on big order net inflow in the paper. At first, we divide lots of users into several categories utilizing collaborative filtering algorithm based on user fuzzy clustering [16]. We get some stocks from the stocks once operated by those users in same category and form a

prerecommended stock set. Then, we use a method based on M/G/1 [17] to compute the net inflow amount of big order for every stock in the prerecommended stock set. From the prerecommended stock set descending ordered by the value of big order net inflow, we choose some stocks with highest value in front of the set as the last recommendation stock set for the target user.

In general, we believe that the stock rise with a large trading volume is closely related to the purchase stock of big order. In order to analyze the big order net flow of stock, we need to observe the stock trading volume and turnover. The money net flow and the money flow [18] are different in concept. The big order refers to the amount of each transaction over one million yuan or the volume of each transaction over fifty thousand in a single transaction. So the big order net inflow refers to the amount of money of big order buy or sell of the same stocks within a day. Most of the time, the money flow is bigger than zero, and the money net flow is less than zero. In individual cases, the money net flow is bigger than zero, and the big order net flow is less than zero. Under this situation, we proposed an efficient stock recommendation model based on big order net inflow. The new recommendation model can measure the capital and the pulsation of the stock markets and consider investors preferences and behavior characteristics; it can improve the existing deficiencies of some current stock recommendation. In addition, the new recommendation model can analyze and filter the stock with less returns in the future and improve the investment gains of investors. The experiment results show that the recommended stocks have better gains during the several days after the recommended stock day and the proposed model can provide reliable investment guidance for the target investors and let them get more investment returns.

The rest of this paper is organized as follows. Section 2 briefly reviews the definitions and theorems of fuzzy clustering and the framework of the collaborative filtering algorithm based on user fuzzy clustering. Section 3 demonstrates the method for computing big order net inflow and the framework of the proposed model. Section 4 presents the simulation experiment and empirical analyses of the proposed model; finally some conclusions are given and some future works are pointed out in Section 5.

## 2. Theoretical and Modeling Framework

The concept of fuzzy set [16] in fuzzy cluster is put forward by Zadeh in 1965. Gath and Bar-On earlier applied that theory to compute the scoring of poly graphic sleep recordings in their study [19]. In this section, we will briefly review the definitions and theoretic of fuzzy cluster.

### 2.1. Fuzzy Clustering Theory

**Definition 1.** Defined set  $R = (r_{ij})_{m \times n}$ , a matrix with  $m$  rows and  $n$  columns; if  $0 \leq r_{ij} \leq 1$ , then  $R$  is called fuzzy matrix  $R$ . When  $r_{ij}$  is only 0 or 1, said  $R$  is a Boolean matrix. When elements  $r_{ii}$  of the diagonal are all 1 in fuzzy matrix  $R = (r_{ij})_{n \times n}$ , said  $R$  is reflexive fuzzy matrix.

**Definition 2.** If  $A$  is  $n$ -order square, defined  $A^2 = A \circ A$ ,  $A^3 = A^2 \circ A, \dots$ , and  $A^k = A^{k-1} \circ A$ .

**Definition 3.** Defined set  $R = (r_{ij})_{m \times n}$ , and called  $R^T = (r_{ij}^T)_{m \times n}$  is the transposed matrix of  $R$ , where  $r_{ij}^T = r_{ji}$ .

**Definition 4.** Defined set  $R = (r_{ij})_{m \times n}$ , for any  $\lambda \in [0, 1]$ , and said  $A_\lambda = (r_{ij}^{(\lambda)})_{m \times n}$  is the  $\lambda$ -cut matrix of the fuzzy matrix  $R$ . When  $r_{ij} \geq \lambda$ ,  $r_{ij}^{(\lambda)} = 1$  and  $r_{ij} < \lambda$ ,  $r_{ij}^{(\lambda)} = 0$ , said  $A_\lambda$  is a Boolean matrix, and said  $\lambda$  is a confidence level parameter or cut level parameter.

**Definition 5.** Suppose two limited discourse domains  $X = \{x_1, x_2, \dots, x_m\}$  and  $Y = \{y_1, y_2, \dots, y_n\}$ ; then  $X$  to  $Y$  fuzzy relation  $R$  is an  $m \times n$  order fuzzy matrix, and set  $R = (r_{ij})_{m \times n}$ , where  $r_{ij} = R(x_i, y_j) \in [0, 1]$  represents  $(x_i, y_j)$  relevance on fuzzy relation  $R$ .

**Theorem 6.** If  $R$  is fuzzy similar matrix, then, for any natural number  $k$ ,  $R^k$  is fuzzy similarity matrix.

**Theorem 7.** If  $R$  is an  $n$  order fuzzy similar matrix, then there is minimum natural number  $k$  ( $k \leq n$ ), for all natural number  $m$  ( $m \geq k$ ); constant  $R^m = R^k$ ,  $R^k$ , namely, fuzzy equivalent matrix ( $R^{2k} = R^k$ ). At this point, said  $R^k$  is the transitive closure of  $R$ , denoted by  $t(R) = R^k$ .

**Method 8.** If  $R$  is fuzzy similar matrix, then the following method will solve the transitive closure  $t(R) : R \rightarrow R^2 \rightarrow R^4 \rightarrow R^8 \rightarrow R^{16} \rightarrow \dots \rightarrow R^{2^k}$ , and the said method is the square self-synthesis method ( $S^3M$ ).

**2.2. Fuzzy Clustering Analysis.** The fuzzy clustering analysis is an analyzing clustering and classification method by establishing fuzzy similar relationship of objective things based on the objective characteristics, the degree of closeness, and similarity between objective things.

In Figure 1, the fuzzy clustering processing model can be divided into four stages, namely, the data preprocessing, the data standardization, the constructing fuzzy similar matrix (FSM), and the clustering and analysis.

**2.2.1. Data Preprocessing.** In China, it is well known that individual investors buy or sell stocks only through a securities company, not directly from the stock exchange. In order to trade stocks, individual investors firstly have to register as a member in a securities company. According to the National Security Act, investors must have a test in the risk tolerance when investors are registering. The securities company gives them a risk test paper with fifteen questions or more. Each question has four or five options for investors to choose. In order to store and process, we use an integer value instead of the chosen answer of investors in the user risk database. In this paper, we choose twelve answers among all answers to study. There are at least 1 million records about risk information of investors in the user risk database of any securities company because there

TABLE 1: The surveying contents of investor.

| $x_{ij}$  | Denoted contents  | Range  |
|-----------|---|--------|
| $x_{i1}$  | The age range of investor   | [1, 5] |
| $x_{i2}$  | The funds amount of investor that can be used to invest   | [1, 5] |
| $x_{i3}$  | The investment purposes of investor   | [1, 4] |
| $x_{i4}$  | The duration of your general investment   | [1, 4] |
| $x_{i5}$  | The maximum loss that you can accept within one year  | [1, 5] |
| $x_{i6}$  | The selection of four portfolio average returns with the best and worst of earnings in the next three years | [1, 4] |
| $x_{i7}$  | The annual income change of your family in next five years  | [1, 4] |
| $x_{i8}$  | The ratio of domestic consumption expenditure share of total income monthly in your family                  | [1, 5] |
| $x_{i9}$  | The highest risk products that you have invested before   | [1, 4] |
| $x_{i10}$ | The annual income of your family  | [1, 5] |
| $x_{i11}$ | The number of years since you invest  | [1, 4] |
| $x_{i12}$ | The type of your work   | [1, 5] |

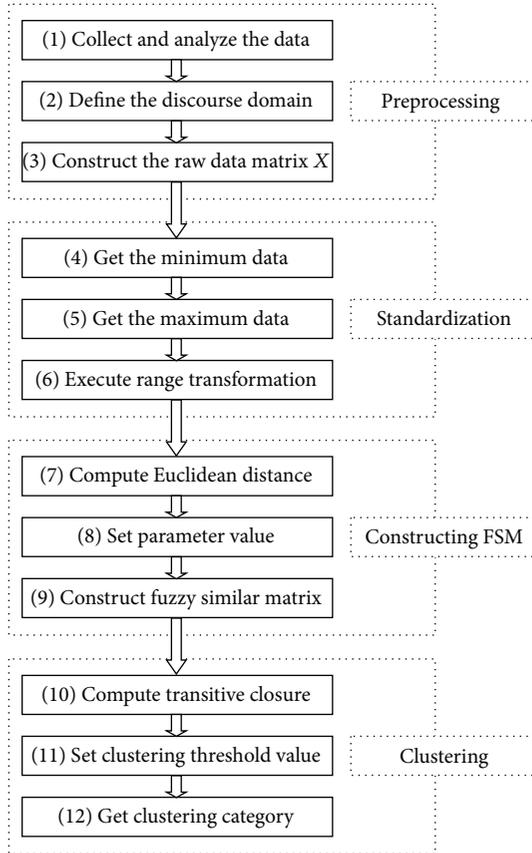


FIGURE 1: The structure of fuzzy clustering processing model.

are more than 200 million stocks investors in China. So we assume that an investor risk tolerance surveying database exists, which includes  $n$  investors risk tolerance surveying data. The characteristics of these data which include the following twelve contents reflect the investment style and risk tolerance of the investors. We define the set of investor risk tolerance surveying data as  $X = \{x_1, x_2, \dots, x_n\}$ , where  $x_i$  denotes the surveying data of the  $i$ th investor in the set  $X$ ,

which is constituted by the following contents vector in a fixed order  $x_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{ij}, \dots, x_{i12}\}$ , where  $j$  belongs to  $[1, 12]$ . Table 1 shows the detailed contents of the  $i$ th investor in the investor risk tolerance surveying database.

**2.2.2. Data Standardization.** We firstly compute all columns data of raw matrix  $X$ , get the minimum data and the maximum data of each column, and then compress each data of matrix  $X$  to  $[0, 1]$  using the following transformation formula. After the above standardizing process, we can get standardized matrix  $Y$  with standardizing data:

$$Y = \{y_{ij}\}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq 12,$$

$$y_{ij} = \frac{x_{ij} - \min_{1 \leq i \leq n} \{x_{ij}\}}{\max_{1 \leq i \leq n} \{x_{ij}\} - \min_{1 \leq i \leq n} \{x_{ij}\}}, \quad 1 \leq j \leq 12, \quad (1)$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1,12} \\ y_{21} & y_{22} & \cdots & y_{2,12} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{n,12} \end{bmatrix}.$$

**2.2.3. Constructing Fuzzy Similar Matrix.** After obtaining standardized matrix  $Y$  with standardizing data, we use the direct Euclidean distance method as the similarity coefficient method to determine the similarity coefficient among investors and construct the fuzzy similar matrix. Consider

$$r_{ij} = 1 - c \times d(x_i, x_j), \quad (2)$$

where  $c$  is a suitable choice of parameters so as to  $0 \leq r_{ij} \leq 1$  and the  $d(x_i, x_j)$  represents the distance between  $x_i$  and  $x_j$  as follows:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{12} (x_{ik} - x_{jk})^2}, \quad (3)$$

$$c = \frac{1}{1 + c'},$$

where  $c'$  belongs to  $[0, \infty)$  and  $c$  belongs to  $(0, 1]$ . We can choose any value for  $c'$ . If the  $c'$  value is too high, parameter  $c$  will be less, which will lead to increasing the accuracy of computing. Therefore, we chose an appropriate value for  $c'$ . In fact, we can choose zero value for  $c'$  which does not affect the experiment results;  $c$  is equal to 1.

Thus, we can get fuzzy similar matrix  $R$  of  $n$  by  $n$  between investors:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}. \quad (4)$$

**2.2.4. Clustering and Analysis.** Since fuzzy similar matrix  $R$  is a  $n$ -dimensional square matrix, we can carry out transitive closure  $t(R)$  using the square self-synthesis method:

$$\begin{aligned} t(R) : R &\longrightarrow R^2 \longrightarrow R^4 \longrightarrow R^8 \longrightarrow R^{16} \longrightarrow \cdots \\ &\longrightarrow R^{2^k}, \end{aligned} \quad (5)$$

where  $k \leq \lceil \log_2^n \rceil$ .

According to the actual situation, we have to choose an appropriate  $\lambda$  value between 0 and 1 for  $\lambda$ -cut matrix  $t(R)_\lambda$  of transitive closure  $t(R)$  of fuzzy matrix  $R$ . Having a classification of  $t(R)$  based on  $\lambda$ , we can get equivalence classification matrix  $t(R)_\lambda$  under the given  $\lambda$  value.

Thus, we will get clustering category for investors. One day, a new investor who becomes the target recommendation investor of stock recommendation system will be added to the investor risk tolerance surveying database. In order to determine which category the new investor belongs to, we utilize the fuzzy clustering method to subdivide the database into several groups based on the above  $\lambda$ -cut value.

### 2.3. Collaborative Filtering Algorithm

**2.3.1. Nearest Neighbors Choice.** Currently, the collaborative filtering algorithm is the most successful personalized recommendation algorithm. It can be classified into two categories: one is item-based collaborative filtering algorithm [20]; the other is user-based collaborative filtering algorithm. The former was first put forward by Sarwar et al. [21] that when we calculate the user similarity, first we calculate the similarity between items to select the most similar items and then predict the rating. The latter was first proposed by Goldberg et al. [22], which is according to the rating of target user's nearest neighbors to predict the target item rating. This algorithm can be computed offline, shorten the time of online calculation, and increase speed of the online recommendation.

In order to improve the accuracy of the nearest neighbor investor's choice, we will use the user-based collaborative filtering algorithm to compute the degree of similarity in behavior between other investors and the target investor in the same fuzzy cluster. Then, according to the stocks of the nearest investors and target investor, we can generate

optimized stock list from the stock set calculated by fuzzy cluster algorithm and choose the top- $k$  stock to recommend the target investor.

At first, we need to calculate the degree of similarity between investors based on their risk tolerance surveying data. At present, there are several methods to calculate the similarity between investors, such as cosine-based similarity, the adjusted-cosine similarity, and Person correlation-based similarity [23]. According to the degree of similarity between the target investor and other investors, we can generate neighbor set  $U = \{U_1, U_2, U_3, \dots, U_n\}$  for the target investor  $u$ . Then, we choose the top- $k$  investors as neighbor investor for the target investor based value of  $\text{sim}(u, U_i)$  ordered by descending.

**2.3.2. Constructing Stock Cluster Set.** In order to recommend some more accurate stocks to the target investor, we subdivide the stocks into several categories as follows based on the fuzzy clustering method in Section 2.2 in this paper. Then, we can utilize fuzzy clustering analysis method to construct stock cluster set. As we know, there are dozens of attributes in each stock. We choose six important attributes of stock to construct a stock attributes database. We think the six attributes of stock are the most important attributes for stock clustering. The six attributes are the daily average gains, the daily average amplitude, the days of price rise, the net profit in last year, the daily net amount of big order, and the days with net amount of big buying order. According to the six attributes of stock and the data format like Table 1 in this paper, we construct the stock attributes database. Then, we cluster the stocks in the stock attributes database and subdivide the stocks into several categories. We can effectively distinguish different stocks between poor stocks and good stocks. Such clustering results can reflect those stocks operational characteristics and be able to provide more accurate and effective recommendation information for target investor. In generally, the stocks in same cluster have similar trading dynamic characteristic. If we have a recommended stock for target investors, we will try to find some other stocks in its stock cluster set. This can improve the accuracy of recommended stocks and reduce the difficulty of the search for other recommended stocks. In order to reduce the computational complexity of clustering the stocks, we have to guarantee that the number of cluster is less than 10. The number of cluster will vary with the changes of recommended stocks. Generally, the number of cluster varied between five and seven in experimentation.

**2.3.3. Generating Recommended Stocks List.** At first, we get the stocks list of the target investor, and then we get the score of each stock in stocks list of target investor and the score of same stock like in stocks list of neighbor investors. If some stocks are not in the stocks list of neighbor investors, we can forecast their scores using the following formula:

$$f(u, j) = \frac{\sqrt{\sum_{k=1}^K \text{sim}(i, k)^2 * (S_k - \bar{S}_k)^2}}{\sqrt{\sum_{k=1}^K \text{sim}(i, k)^2}} + \bar{S}_u, \quad (6)$$

where  $\overline{S}_u$  represents the target investor  $u$ 's average scores for all stocks,  $S_k$  represents the  $k$ th neighbor investor scores of the target investor  $u$ ,  $\overline{S}_k$  represents the average scores of  $k$  neighbors investor for all stocks, and  $f(u, j)$  represents the  $j$ th stocks forecast score of target investor  $u$ .  $\text{sim}(i, k)$  is a modified similarity formula based on the cosine Similarity; it is defined as follows:

$$\text{sim}(i, k) = \frac{\sum_{v=1}^m (R_{vi} - \overline{R}_i)(R_{vk} - \overline{R}_k)}{\sqrt{\sum_{v=1}^m (R_{vi} - \overline{R}_i)^2} \sqrt{\sum_{v=1}^m (R_{vk} - \overline{R}_k)^2}}, \quad (7)$$

where  $\overline{R}_i$  is the  $i$ th stock average scores for all investors,  $\overline{R}_k$  is the  $k$ th stock average scores for all investors,  $R_{vk}$  represents the  $k$ th neighbor investor scores for the  $v$ th stock, and  $R_{vi}$  represents the  $i$ th neighbor investor scores for the  $v$ th stock.

Now we give a simple example for the process of generating recommendation stock list. Assuming that there are two stocks in the stocks list of the target investor, such as stock  $A$  and stock  $B$ , here we write them as a set  $\{A, B\}$ . Then we get the stocks set of four or more neighbors, such as set  $\{B, C, D\}$ ,  $\{B, C, E\}$ ,  $\{C, F\}$ , and  $\{B, C\}$ , and we can select the rated top- $k$  stocks as the target investor's extend stocks list from the above five stock set. When we set  $k$  as 2, we can get the top-2 stocks set  $\{B, C\}$ .

Secondly, in order to get the more large stock extend set of the target investor, such as stock set  $\{B, C, X, Y, Z\}$ , we need to revise the  $k$  value. If we recommend to the target investor only five stocks, we will stop revising the  $k$  value when the number of stock in the above stock set is greater than or equal to five.

Finally, we calculate net inflow amount of big order for each stock of the above stock extend set  $\{B, C, X, Y, Z\}$  recently, then we sort the five stocks in descending order according to the net inflow amount of big order and select top- $n$  stocks recommend to the target investor. If we set  $n$  as three, we get the recommend stock list  $\{B, X, Y\}$ . Because the process of judging money inflow is more complex, we propose a new method based on M/G/1 to compute the net inflow amount of big order and we can forecast the direction of stock price movements in the future.

**2.4. M/G/1 Queue System.** The M/G/1 queue system [24] has been extensively studied for the last three decades [25]. According to the circumstances of the securities transaction, here we review the single server queue system which behaves like the usual M/G/1 queue when the server is working. We assume that the server goes through cycles of idle and busy periods. The idle periods include two cases: one is when there is no work to do, and the other is when there is work to do but the server is on vacation. The busy periods are the times when the server is actually working on the customers of primary customers [26]. During the busy periods of the simple M/G/1 queue system, we assume that the customers arrival time follows Poisson distribution with parameter  $\lambda$ . Let  $\lambda$  be the customer arrival rate and let  $D$  be the distribution of the service times of customers arriving during busy periods; then the customers arrival time has a general distribution function  $B$ . Let  $S_i$  be the epoch at the end of the  $i$ th busy period and let  $T_i$  be the epoch beginning the  $i$ th busy period. Then,

$0 = T_i < S_i < T_{i+1} < S_{i+1} < \dots < T_n < S_n$ . We assume that the arrival process and the service times of the customers arriving in the interval  $(T_i, S_i)$  are independent of those arriving in  $(T_j, S_j)$  for  $i \neq j$ . Let  $w(t)$  be the work in the queue system at time  $t$ . Let  $W_{1i} = w(S_i)$  and  $W_{2i} = w(T_{i+1})$ , where  $i > 0$ . Then,  $W_{1i}$  is the work in the queue system at the end of the  $i$ th busy period and  $W_{2i}$  is the work in the queue system immediately after the beginning of the  $(i + 1)$ st busy period.

Let  $l(k)$  be the  $k$ th workload step in all workload process. Let  $L(k) = \min\{m : \sum_{i=1}^m (S_i - T_i) \geq k\}$  and  $E(k) = \sum_{i=1}^{L(k)-1} (T_{i+1} - S_i)$ .

If  $k < \sum_{i=1}^{L(k)} (S_i - T_i)$ , then  $l(k) = w(k + E(k))$ .

If  $k = \sum_{i=1}^{L(k)} (S_i - T_i)$ , then  $l(k^+) = w(k + E(k) + T_{L(k)+1} - S_{L(k)})$  and  $l(k^-) = w(k + E(k))$ .

Clearly, if  $k = \sum_{i=1}^{L(k)} (S_i - T_i)$ , then  $l(k^+) - l(k^-) = W_{2,L(k)} - W_{1,L(k)}$ . The  $l(k)$  process behaves like the work in a simple M/G/1 queue.

### 3. Proposed Method

This section consists of two subsections: Section 3.1 describes the method based on M/G/1 to compute the net inflow amount of big order and Section 3.2 describes the proposed model and method for stock recommendation.

#### 3.1. Compute the Net Inflow Method

**3.1.1. Funds Flow Theory.** The flow of funds is stock movement direction actively chosen by the funds in the stock market. From the amount of perspective to analyze the flow of funds, namely, observation volume and turnover, trading volume and turnover in the actual operation is directional, to buy or sell. For both the stock market trend analysis and the operation on individual stocks, the determination of the funds flow plays a vital role, and the process of the funds flow is more complicated, not easy to grasp. The funds flow can help investors see what others are doing in the end through the index (price) change fog. For example, the index (price) of a stock rise up to a point may be driven by 10 million funds or a billion of funds, both of which have a completely different significance for investors. In general, funds flow and the trend of stock index change are very similar, but in the following two cases funds flow measure has obvious significance. One is that the day's flow of funds and stock index change opposite. For example, the stock overall index is down throughout the day, but funds flow shows a positive net inflow of funds throughout the day. The other is that there is very big opposite between the funds flow and the stock index change. For example, the stock index rises highly throughout the day, but the actual net inflows are small or even negative which is called net outflows. When the funds flows and the stock index change is opposite, the funds flow can reflect stock actual movement direction more than the stock index change in the future.

**3.1.2. Funds Flow Concepts.** In order to more clearly describe the proposed method, we review some concept about funds flow as follows.

**Definition 9.** Funds inflow refers to the amount of active buying. It is active buying transactions where the buyer actively buys stock with price equal to or higher than the first selling price.

**Definition 10.** Funds outflow refers to the amount of active selling. It is active selling transactions where the seller actively sells stock with price equal to or less than the first buying price.

**Definition 11.** Funds net inflow refers to the amount of funds inflow minus funds outflow. If it is positive, the probability of stock price rise is higher than fall and vice versa if the net inflow is negative; then, the probability of stock price fall is higher than rise.

**Definition 12.** Big order refers to the buying or selling order with big amount. According to the amount of the order, we divide it into four categories: small order, general order, big order, and very big order (king order). Using the number of the order which measures it, we think that the number of big order is more than 5 million shares or more than 20 million yuan and the number of king order is more than 20 million shares or more than 100 million yuan.

**Definition 13.** Big order net amount refers to the difference between the amount of big buying order and the amount of big selling order. If it is positive, we call it the big order net inflow. If not, we call it the big order net outflow.

**Definition 14.** Tick-by-tick data refer to the single transaction during transactions. It reflects the true circumstances of the transaction process and is proprietary data of the Level-2.

**3.1.3. Method Based on M/G/1.** We review the classical M/G/1 queue system which has the following four characteristics [27].

- (1) The characteristic of the arrival process follows Poisson distribution with arrival rate parameter  $\lambda$ ; M indicates a Poisson process without memory.
- (2) Probability  $P$  of the service time follows general random distribution. Let  $t_i$  be the service time for the  $i$ th customer that is an independent and random value and has general distribution function  $B(t)$ :

$$P(t_i \geq t) = B(t), \quad i = 1, 2, 3, \dots \quad (8)$$

The mean value and variance of service time are given by

$$\begin{aligned} \bar{X} &= \frac{1}{u} = \int_0^{\infty} t dB(t), \\ \bar{X}^2 &= \int_0^{\infty} t^2 dB(t). \end{aligned} \quad (9)$$

- (3) The number of server desk is one; the arrival time and service time are independent of each other.
- (4) The system allows an infinite captain for the length of customers; the queue discipline is first come first serve (FCFS).

Assuming that interarrival time follows Poisson distribution with parameter  $\lambda$  in  $\{A(t), t \geq 0\}$ , let  $A(t)$  be the number of arrival customers in  $[0, t]$  time, let  $N(t)$  be the number of customers in queue system at time  $t$ , let  $X_n$  be the number of customers after the  $n$ th customer departure instant, let  $d_n$  be the departure time of the  $n$ th customer, and let  $a_n$  be the arrival time of the  $n$ th customer. If the number of customers is greater than zero at time  $d_n$ , then

$$X_{n+1} = X_n + A(d_{n+1}) - A(d_n) - 1 \quad (10)$$

or

$$d_{n+1} = d_n + t_n + 1. \quad (11)$$

At this time if  $X_n = 0$ , then

$$X_{n+1} = A(a_{n+1} + t_{n+1}) - A(a_{n+1}). \quad (12)$$

From (11) and (12) the following can be obtained:

$$P\{X_{n+1} = i \mid X_1, X_2, \dots, X_n\} = P\{X_{n+1} = i \mid X_n\}. \quad (13)$$

Transition probability matrix  $P$  of Markov Chain  $X_n$  is

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots \end{pmatrix}, \quad (14)$$

where chain  $X_n$  has Markov property and  $p_k$  is given by

$$p_k = \int_0^{\infty} \frac{e^{-\lambda t} (\lambda t)^k}{k!} dB(t), \quad (k = 0, 1, 2, \dots). \quad (15)$$

**3.2. The Framework of the Proposed Model.** In Figure 2, the process and data flow framework of the proposed model can be divided into two stages, namely, the user clustering stage and stock recommend stage. Generally, investors who have similar characteristics have similar investment interest. According to Figure 2, in order to obtain accurately the similar stocks for target users, we have to process user clustering and obtain some similar users. The selection of the clustering threshold value affects the number and size of the user category and then affects the accuracy of the stocks set which is selected from the same category user stocks, so it is critical to select the value of user clustering threshold. In the stock recommend stage, the computing of the big orders net for lots of stocks is the most important part. The stock trend is mainly driven by big order transactions.

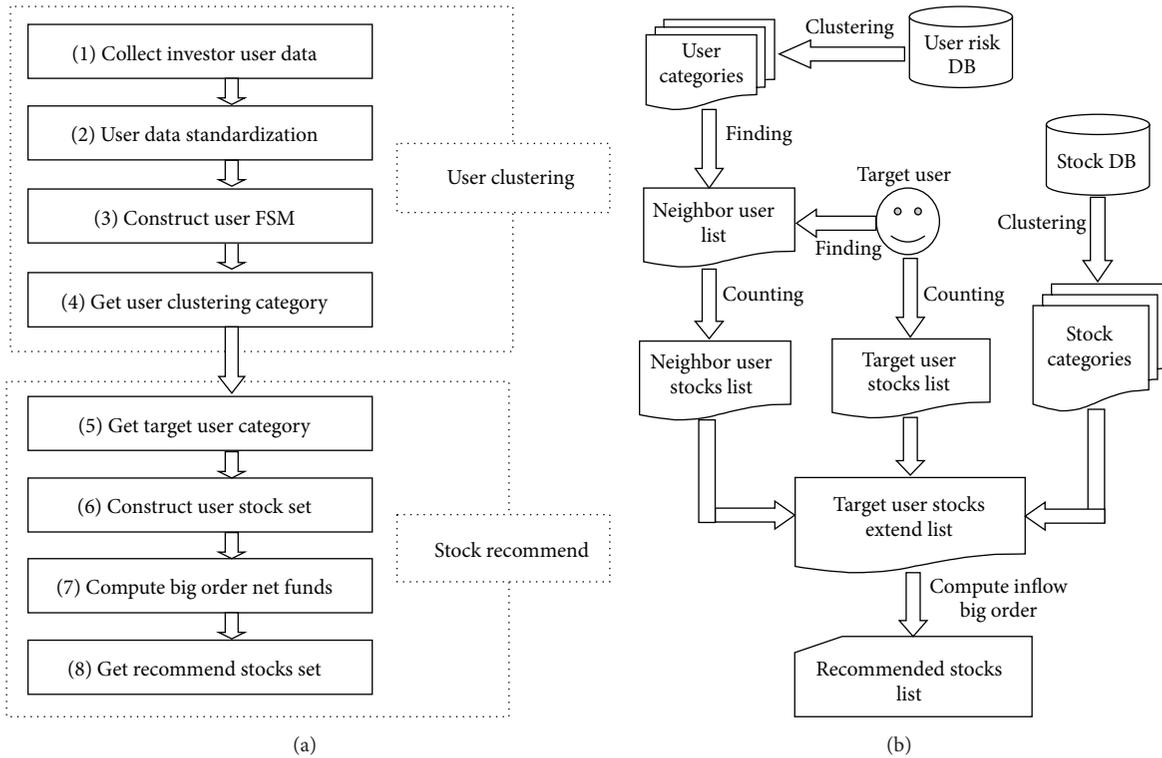


FIGURE 2: (a) The process framework of the proposed model. (b) The data flow framework of the proposed model.

It is generally believed that stocks rise with a large volume is closely associated with big orders net amount to buy, so the stocks are generally rising in price under the trend driven by big orders net inflow which is called big orders net buying. In contrast, the stocks are generally falling in price under the trend driven by big orders net outflow which is called big orders net selling. The big orders net amount includes the big orders net buying and the big orders net selling. The traditional model uses the funds inflow and funds outflow to predict stock trend that is unsuitable for some new special situation. For example, one day the funds inflow of a stock is far greater than zero, but the big orders net inflow of the stock is far smaller than zero. If it is that case, we can predict that this stock will go to falling trend over a period of time after that day. After observing many stocks, we found that it is indeed the case. Our proposed model can solve this problem by using big orders net amount that can avoid or reduce some forecast errors and can improve accuracy of recommend stocks trend in the future. Then, according to the indicated results of big orders net inflow, we select some optimal stocks recommend to target users to buy. Correspondingly, according to the indicated results of big orders net outflow, we select some optimal stocks recommend to target users to sell.

#### 4. Simulation Experiment

In this section, we study and compare the performance of the proposed model. In general, during the Shanghai



FIGURE 3: The research interval price trends graph of the Shanghai Composite Index (CSI).

and Shenzhen Composite Index (CSI) rise, the accuracy of the recommendation algorithm is higher, but during the CSI fall, the accuracy of recommendation algorithm is very low or even completely incorrect. Thus, in the course of falling, the experimental results can test a recommendation algorithm. In order to make the experiment results more objective and realistic, we use stock return to test whether the recommended stock has a good return from 10 to 30 days after that. For target investor, the higher the yields the investor gets, the better the effect of the recommendation model.

**4.1. Data Selection.** In order to examine whether the proposed model has made improvement in prediction accuracy, we select data at four different periods of the real stock market in China as the experiment data. The four periods include bottom period (2012/11/28–2012/12/04, see A in Figure 3), middle period (2012/12/17–2012/12/21, see B in Figure 3), top period (2013/02/04–2013/02/08, see C in Figure 3) of

TABLE 2: The users clustering results.

| Category number | Users numbers | Percentage (%) |
|-----------------|---------------|----------------|
| 1               | 912           | 35.63          |
| 2               | 138           | 5.39           |
| 3               | 576           | 22.50          |
| 4               | 482           | 18.83          |
| 5               | 314           | 12.27          |
| 6               | 48            | 1.88           |
| 7               | 90            | 3.52           |

a wave with rising trend, and middle period (2013/06/17–2013/06/21, see D in Figure 3) of a wave with falling trend. The experimental data include data of three parts that are the investor user data, the CSI data, and the tick-by-tick transaction data of stock. As the user's data are related to the user's personal privacy, we use the data which are removed from the user's privacy as the experimental data. The CSI and stocks data are stock market free data, but the tick-by-tick transaction data of stock are the Level-2 data and charging data.

*4.2. Data Processing.* In order to reduce the amount of computation and be suitable for parallel computing, we randomly select 2560 users from the investor user database to fuzzy clustering and divide those users into several categories according to the threshold of clustering. By the nature of the clustering, the value of the threshold can control the number of the user categories. The value of the threshold and the number of the user categories are inversely related. The threshold can be in the range from 0 to 1. If the threshold is set too high, we will get very few user categories. Contrarily, if the threshold is set too small, we will get much more user categories. For example, if the threshold is set to 1, we will get 2560 user categories. Clearly, if the threshold is set to 0, we will get one user categories. The complexity of clustering the users increases exponentially with the number of the user categories. In order to reduce the computational complexity of clustering the users, we have to choose an appropriate threshold. For that 2560 users data, we made several experiments to cluster that users data by constantly adjusting the value of the threshold. We can get seven user categories and get better distribution of user in Table 2 while the threshold was set to 0.6268. Certainly, it is allowed that the threshold is set higher or lower than 0.6268, but the distribution of user will become worse. Therefore, in order to achieve the purpose of the clustering users, we have to choose an appropriate clustering threshold. After making several experiments, we set the clustering threshold  $\lambda = 0.6268$  in this paper and then divide 2560 users into seven categories in Table 2.

After the users clustering, we use formula (7) based on the cosine similarity to calculate the similarity between the target user and the other users. We can find out the nearest neighbor users of the target user according to the similarity value. In the similarity calculation, here there are two cases: (1) the target user belongs to the clustered users; (2) the target user does not

TABLE 3: The nearest neighbors of the target user.

| No. | Similarity value | Users number |
|-----|------------------|--------------|
| 1   | 0.982            | 380          |
| 2   | 0.957            | 359          |
| 3   | 0.926            | 2478         |
| 4   | 0.903            | 1911         |
| 5   | 0.871            | 1653         |
| 6   | 0.819            | 406          |
| 7   | 0.735            | 141          |

belong to the clustered users. For case (1), the target user and his most nearest neighbors must belong to the same category of the clustered users; we only need to find out the  $k$  most nearest neighbors in the same category according to the value of the fuzzy similarity matrix. For case (2), we calculate the similarity value between the target user and the  $\log_2^n$  clustered users at most based on the Binary Search Algorithm [28], where  $n$  is the number of clustered users. In order to reduce the amount of calculation of the recommendation actions after that, we add the target user into the same category with the most clustered user. In order to make the experiment results more objective and realistic, we choose an investor user that does not belong to the clustered users as the study target in the experiment and the result of the nearest seven neighbors of the target user in Table 3.

With the complement classified for the target user, we can get the  $k$  clustered users and get the stock set from those users. Here we call stock set as a prerecommended stock set. We use a method based on  $M/G/1$  to compute the net inflow amount of big order for every stock in the prerecommended stock set. From the prerecommended stock set descending ordered by the value of big order net inflow, we choose the  $k$  stocks with highest value in front of the stock set as the last recommendation stock set for the target user. Due to limited paper space, we selected five stocks in each period of Figure 3 as the research stock in Tables 4, 5, 6, and 7. If the value of any stock in the stock set is smaller than 0, we have to set that the value of  $k$  is bigger and repeat the previous calculation steps. If the prerecommended stock set contains the all stock of the China Stock Market and any stock value of the big order net inflow is smaller than 0, we do not recommend any stock to the target user and recommend the target user to keep a wait state with holding money.

We do some research on some stocks that are recommended by the proposed model to analyze the maximum possible stocks return during the four different day intervals, that is 10 days, 30 days, 90 days, and 1 year, since that day when we recommend those stocks to the target user. By observing the experiment results, we found that those recommended stocks have better gains in Table 8, especially those stocks whose big order net inflow is higher proportion of trading volume of those days than the others. In addition, the stock price is the highest stock price without ex-rights or ex-dividend at the last trading day of the four corresponding day intervals.

TABLE 4: The recommended five stocks in the A period.

| Stock name (code)             | Date/mean  | Buy           | Sell          | Net amount (10 <sup>4</sup> share) | Percent (%)  | Net money (10 <sup>4</sup> yuan) |
|-------------------------------|------------|---------------|---------------|------------------------------------|--------------|----------------------------------|
| Jianghuai Automobile (600418) | 2012-11-28 | 65.87         | 162.00        | -96.11                             | -19.09       | -481.40                          |
|                               | 2012-11-29 | 77.98         | 22.48         | 55.50                              | 8.52         | 283.60                           |
|                               | 2012-11-30 | 476.20        | 216.60        | 259.50                             | 21.37        | 1359.30                          |
|                               | 2012-12-03 | 965.30        | 450.10        | 515.20                             | 18.93        | 2801.70                          |
|                               | 2012-12-04 | 1098.80       | 505.80        | 593.00                             | 24.59        | 3285.90                          |
|                               | Mean       | <b>536.83</b> | <b>271.40</b> | <b>265.42</b>                      | <b>10.86</b> | <b>1449.82</b>                   |
| Janus Precision (300083)      | 2012-11-28 | 0.00          | 0.00          | 0.00                               | 0.00         | 0.00                             |
|                               | 2012-11-29 | 6.49          | 0.00          | 6.49                               | 17.22        | 68.47                            |
|                               | 2012-11-30 | 0.00          | 0.00          | 0.00                               | 0.00         | 0.00                             |
|                               | 2012-12-03 | 3.35          | 0.00          | 3.35                               | 12.75        | 35.17                            |
|                               | 2012-12-04 | 3.26          | 4.00          | -0.74                              | -1.54        | -7.34                            |
|                               | Mean       | <b>2.62</b>   | <b>0.80</b>   | <b>1.82</b>                        | <b>5.69</b>  | <b>19.26</b>                     |
| Great Wall Motors (601633)    | 2012-11-28 | 58.62         | 51.76         | 6.86                               | 1.21         | 123.00                           |
|                               | 2012-11-29 | 93.53         | 109.20        | -15.70                             | -4.20        | -279.10                          |
|                               | 2012-11-30 | 125.40        | 99.61         | 25.77                              | 5.72         | 471.40                           |
|                               | 2012-12-03 | 259.20        | 216.80        | 42.35                              | 7.60         | 780.50                           |
|                               | 2012-12-04 | 301.40        | 236.00        | 65.43                              | 13.03        | 1196.20                          |
|                               | Mean       | <b>167.63</b> | <b>142.67</b> | <b>24.94</b>                       | <b>4.67</b>  | <b>458.40</b>                    |
| Huasun Group (000790)         | 2012-11-28 | 71.27         | 42.75         | 28.52                              | 8.24         | 179.00                           |
|                               | 2012-11-29 | 0.00          | 10.34         | -10.34                             | -4.72        | -62.21                           |
|                               | 2012-11-30 | 6.62          | 0.00          | 6.62                               | 2.84         | 40.48                            |
|                               | 2012-12-03 | 21.60         | 23.88         | -2.28                              | -0.81        | -14.04                           |
|                               | 2012-12-04 | 24.26         | 13.11         | 11.15                              | 4.54         | 63.76                            |
|                               | Mean       | <b>24.75</b>  | <b>18.02</b>  | <b>6.73</b>                        | <b>2.02</b>  | <b>41.40</b>                     |
| Faw Car (000800)              | 2012-11-28 | 9.79          | 41.20         | -31.41                             | -10.68       | -184.70                          |
|                               | 2012-11-29 | 23.03         | 24.87         | -1.84                              | -1.64        | -18.10                           |
|                               | 2012-11-30 | 30.49         | 40.12         | -9.64                              | -2.57        | -56.06                           |
|                               | 2012-12-03 | 241.80        | 245.50        | -3.73                              | -0.43        | -16.62                           |
|                               | 2012-12-04 | 314.90        | 141.30        | 173.60                             | 17.21        | 1075.10                          |
|                               | Mean       | <b>124.00</b> | <b>98.60</b>  | <b>25.40</b>                       | <b>0.38</b>  | <b>159.92</b>                    |

*4.3. Experimental Results Analysis.* In this section, we briefly analyze and explain the experimental results. After selecting an investor user that does not belong to the clustered users as the study target user in the experiment, we calculate the similarity value between the target user and the clustered users using formula (7) based on the cosine similarity. Because the nearest neighbors of the target user all belong to category 3 of clustered users, the target user belongs to category 3 according to the nature of the clustering that the target user and his nearest neighbors belong to the same cluster. We select the nearest seven neighbors of the target user to show you in Table 3. We use a method based on M/G/I to compute the net inflow amount of big order for every stock in the prerecommended stock set selected from those users. We select five stocks with higher value of big order net inflow during the period between 2012/11/28 and 2012/12/04 as the recommended stocks for the target user in Table 4. For simplicity, we sort the stocks using big order net inflow daily mean ratio from large to small and select five stocks to recommend the target user.

Table 4 shows the daily big order net flow, ratio, and money for the five recommended stocks. In order to ensure the data neat and objective, we calculate the daily big order net flow for every stock at continuous five trading days and compute the average of them. Combining the big order net money inflow, we use the big order net inflow ratio as first evaluation criterion for the stocks. From the data in Table 3, we find that the number of days of big order net money inflow is generally greater than the number of days of big order net money outflow. Although there are four days of big order net money outflow in the stock of Faw Car, the big order net money inflow of the fifth day is greater than the sum of big order net money outflow of the previous four days. For that reason we believe that the stock will rise in the future. Generally, the big order net inflow only affects the short-term price of stock and shows that some investors are upbeat about the stock in the near future. The short-term price of stock may be consistent with how much the big order net money inflow is, but the long-term price of stock may not be fully consistent with it. As we all know, the long-term price of stock

TABLE 5: The recommended five stocks in the B period.

| Stock name (code)                 | Date/mean  | Buy            | Sell           | Net amount (10 <sup>4</sup> share) | Percent (%)  | Net money (10 <sup>4</sup> yuan) |
|-----------------------------------|------------|----------------|----------------|------------------------------------|--------------|----------------------------------|
| Dahu Aquaculture (600257)         | 2012-12-17 | 1765.20        | 1012.20        | 753.00                             | 16.59        | 5306.90                          |
|                                   | 2012-12-18 | 1018.30        | 832.50         | 185.90                             | 5.12         | 1280.60                          |
|                                   | 2012-12-19 | 1207.50        | 723.90         | 483.60                             | 19.88        | 3346.80                          |
|                                   | 2012-12-20 | 1394.20        | 852.60         | 541.60                             | 20.45        | 3766.90                          |
|                                   | 2012-12-21 | 597.40         | 463.50         | 133.90                             | 6.17         | 956.00                           |
|                                   | Mean       | <b>1196.52</b> | <b>776.94</b>  | <b>419.60</b>                      | <b>13.64</b> | <b>2931.44</b>                   |
| Changan Automobile (000625)       | 2012-12-17 | 432.20         | 300.40         | 131.80                             | 7.11         | 720.60                           |
|                                   | 2012-12-18 | 427.10         | 840.50         | -413.40                            | -17.84       | -2185.60                         |
|                                   | 2012-12-19 | 5272.20        | 2329.90        | 2942.30                            | 31.97        | 17000.00                         |
|                                   | 2012-12-20 | 2238.60        | 1906.90        | 331.70                             | 6.30         | 1963.60                          |
|                                   | 2012-12-21 | 4297.70        | 2689.30        | 1608.40                            | 22.43        | 10000.00                         |
|                                   | Mean       | <b>2533.56</b> | <b>1613.40</b> | <b>920.16</b>                      | <b>9.99</b>  | <b>5499.72</b>                   |
| Xinning Modern Logistics (300013) | 2012-12-17 | 246.70         | 101.70         | 144.90                             | 43.29        | 1243.50                          |
|                                   | 2012-12-18 | 153.90         | 151.50         | 2.41                               | 0.28         | 21.33                            |
|                                   | 2012-12-19 | 20.56          | 29.87          | -9.32                              | -2.30        | -80.77                           |
|                                   | 2012-12-20 | 9.09           | 13.62          | -4.53                              | -0.96        | -40.18                           |
|                                   | 2012-12-21 | 10.73          | 8.35           | 2.39                               | 0.43         | 23.55                            |
|                                   | Mean       | <b>88.20</b>   | <b>61.01</b>   | <b>27.17</b>                       | <b>8.15</b>  | <b>233.49</b>                    |
| Robam Appliances (002508)         | 2012-12-17 | 141.70         | 79.90          | 61.83                              | 16.93        | 1150.70                          |
|                                   | 2012-12-18 | 12.43          | 24.04          | -11.61                             | -9.03        | -212.80                          |
|                                   | 2012-12-19 | 14.14          | 3.90           | 10.24                              | 15.52        | 189.80                           |
|                                   | 2012-12-20 | 1.70           | 14.35          | -12.65                             | -13.01       | -228.30                          |
|                                   | 2012-12-21 | 79.50          | 76.40          | 3.09                               | 1.08         | 60.14                            |
|                                   | Mean       | <b>49.89</b>   | <b>39.72</b>   | <b>10.18</b>                       | <b>2.30</b>  | <b>191.91</b>                    |
| Great Wall Motor (601633)         | 2012-12-17 | 139.40         | 147.70         | -8.27                              | -1.62        | -133.00                          |
|                                   | 2012-12-18 | 125.70         | 136.20         | -10.52                             | -1.95        | -218.80                          |
|                                   | 2012-12-19 | 97.63          | 55.14          | 42.50                              | 10.17        | 893.70                           |
|                                   | 2012-12-20 | 253.70         | 179.60         | 74.04                              | 14.04        | 1608.70                          |
|                                   | 2012-12-21 | 146.40         | 142.20         | 4.21                               | 1.06         | 90.46                            |
|                                   | Mean       | <b>152.57</b>  | <b>132.17</b>  | <b>20.39</b>                       | <b>4.34</b>  | <b>448.21</b>                    |

may be consistent with company's development, company's profitability, stock trader's goal and skill, and so on. We can see the above phenomenon from Table 8. Correspondingly, we can find that similar characteristics exist in Tables 5, 6, and 7.

Table 8 shows the recommended stocks maximum possible gains during the four different day intervals. In order to conveniently compare with the recommended stocks, we put the corresponding data of Shanghai Composite Index on the bottom of the data of the recommended stocks. Due to a similar trend of CSI 800 and Shanghai Composite Index, we chose Shanghai Composite Index as CSI in Table 8. If the highest price of the stock or CSI is equivalent in two or more continuous periods, this shows that the price of the stock or CSI is smaller than the previous price of that one and shows that the stock has been in decline during the following period. For example, the highest price of the CSI of the "highest price within 90 days" is equal to the highest price of CSI of the "highest price within 1 year"; they all are 2444.80 and we find that the running range of CSI is the interval from C to D in Figure 3. From Table 8, we find that the maximum possible

gains of most of the recommended stocks are bigger than the CSI. Without a doubt, the target investor cannot get the same gains with the maximum possible gains because the time of buying or selling stock is uncertain for the target investor. It is easy to understand that the recommended stocks can get better gains between A period and C period in Figure 3. But it is very difficult that the recommended stocks can get better gains between C period and D period or after D period since the CSI has been going down or fluctuant. They fully illustrate the effectiveness of the proposed model.

In order to show clearly the return of the recommended stocks, we choose the C and D period in Figure 3. For that is the start period and middle period of the CSI fall or fluctuant, the general stocks will fall or fluctuate with the CSI. If the recommended stocks can get better returns in the falling period, it clearly shows that the proposed method is excellent. We select two last stocks from the recommended five stocks in the C and D period. Figures 4 and 5 show the return comparing the two stocks with the CSI in the 20 weeks after recommending those stocks. Because the CSI market was closed for a holiday, Figure 4 shows only 18 trading weeks and

TABLE 6: The recommended five stocks in the C period.

| Stock name (code)               | Date/mean  | Buy           | Sell          | Net amount (10 <sup>4</sup> share) | Percent (%) | Net money (10 <sup>4</sup> yuan) |
|---------------------------------|------------|---------------|---------------|------------------------------------|-------------|----------------------------------|
| Irtouch Systems (300282)        | 2013-02-04 | 27.81         | 22.55         | 5.27                               | 9.26        | 78.49                            |
|                                 | 2013-02-05 | 6.39          | 5.54          | 0.85                               | 2.73        | 12.77                            |
|                                 | 2013-02-06 | 24.82         | 25.77         | -0.95                              | -1.43       | -13.60                           |
|                                 | 2013-02-07 | 9.18          | 2.36          | 6.82                               | 16.88       | 103.20                           |
|                                 | 2013-02-08 | 18.23         | 7.83          | 10.40                              | 19.58       | 161.00                           |
|                                 | Mean       | <b>17.29</b>  | <b>12.81</b>  | <b>4.48</b>                        | <b>9.40</b> | <b>68.37</b>                     |
| Transinfo Technology (002373)   | 2013-02-04 | 15.78         | 19.01         | -3.23                              | -5.41       | -36.05                           |
|                                 | 2013-02-05 | 3.28          | 7.37          | -4.09                              | -13.08      | -47.06                           |
|                                 | 2013-02-06 | 19.39         | 5.84          | 13.55                              | 23.65       | 159.30                           |
|                                 | 2013-02-07 | 47.76         | 39.35         | 8.41                               | 8.94        | 99.27                            |
|                                 | 2013-02-08 | 31.28         | 24.18         | 7.10                               | 9.52        | 85.24                            |
|                                 | Mean       | <b>23.50</b>  | <b>19.15</b>  | <b>4.35</b>                        | <b>4.72</b> | <b>52.14</b>                     |
| Microgate Technology (300319)   | 2013-02-04 | 15.84         | 23.28         | -7.44                              | -11.44      | -126.90                          |
|                                 | 2013-02-05 | 32.69         | 15.87         | 16.82                              | 20.43       | 293.30                           |
|                                 | 2013-02-06 | 14.57         | 14.34         | 0.23                               | 0.44        | 4.55                             |
|                                 | 2013-02-07 | 14.77         | 10.36         | 4.41                               | 8.45        | 77.36                            |
|                                 | 2013-02-08 | 22.16         | 19.66         | 2.50                               | 2.96        | 44.09                            |
|                                 | Mean       | <b>20.01</b>  | <b>16.70</b>  | <b>3.30</b>                        | <b>4.17</b> | <b>58.48</b>                     |
| Kingsun Optoelectronic (002638) | 2013-02-04 | 317.20        | 309.40        | 7.84                               | 1.08        | 84.20                            |
|                                 | 2013-02-05 | 251.90        | 244.50        | 7.39                               | 1.21        | 77.51                            |
|                                 | 2013-02-06 | 223.00        | 232.10        | -9.15                              | -1.84       | -94.10                           |
|                                 | 2013-02-07 | 229.90        | 181.00        | 48.91                              | 10.29       | 509.60                           |
|                                 | 2013-02-08 | 256.90        | 218.50        | 38.42                              | 6.88        | 408.50                           |
|                                 | Mean       | <b>255.78</b> | <b>237.10</b> | <b>18.68</b>                       | <b>3.52</b> | <b>197.14</b>                    |
| Victory Precision (002426)      | 2013-02-04 | 23.48         | 20.45         | 3.03                               | 2.85        | 19.94                            |
|                                 | 2013-02-05 | 25.84         | 31.34         | -5.50                              | -4.97       | -35.74                           |
|                                 | 2013-02-06 | 10.90         | 5.02          | 5.88                               | 8.48        | 38.82                            |
|                                 | 2013-02-07 | 3.58          | 10.97         | -7.39                              | -14.02      | -47.94                           |
|                                 | 2013-02-08 | 104.50        | 51.13         | 53.40                              | 24.99       | 358.40                           |
|                                 | Mean       | <b>33.66</b>  | <b>23.78</b>  | <b>9.88</b>                        | <b>3.47</b> | <b>66.70</b>                     |

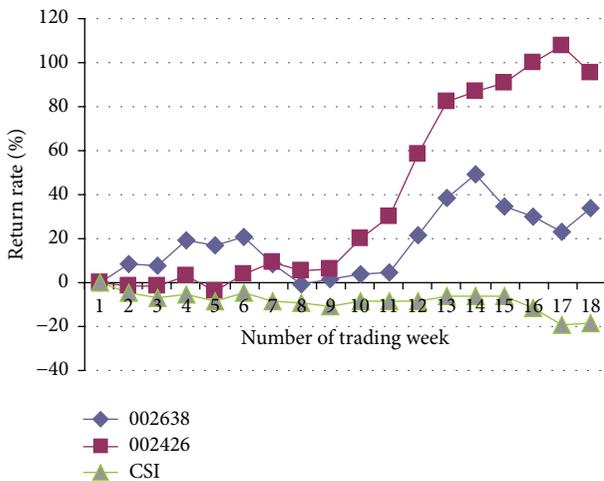


FIGURE 4: Return comparing of two stocks with the CSI in the C period.

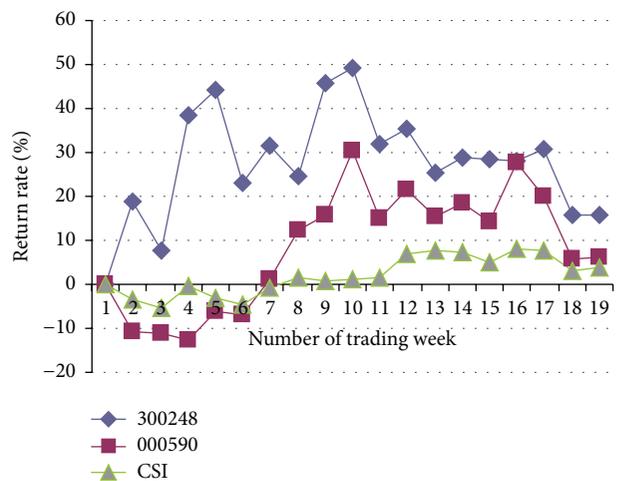


FIGURE 5: Return comparing of two stocks with the CSI in the D period.

TABLE 7: The recommended five stocks in the D period.

| Stock name (code)              | Date/mean  | Buy           | Sell          | Net amount (10 <sup>4</sup> share) | Percent (%) | Net money (10 <sup>4</sup> yuan) |
|--------------------------------|------------|---------------|---------------|------------------------------------|-------------|----------------------------------|
| Sunway Communicate (300136)    | 2013-06-17 | 182.30        | 162.30        | 20.01                              | 7.34        | 465.70                           |
|                                | 2013-06-18 | 96.59         | 89.67         | 6.92                               | 4.87        | 157.90                           |
|                                | 2013-06-19 | 51.76         | 41.94         | 9.82                               | 11.37       | 224.20                           |
|                                | 2013-06-20 | 164.20        | 131.00        | 33.15                              | 13.30       | 765.70                           |
|                                | 2013-06-21 | 136.10        | 137.20        | -1.18                              | -0.56       | -15.57                           |
|                                | Mean       | <b>126.19</b> | <b>112.42</b> | <b>13.74</b>                       | <b>7.26</b> | <b>319.59</b>                    |
| Huaфон Spandex (002064)        | 2013-06-17 | 1097.00       | 693.20        | 403.90                             | 21.44       | 2671.10                          |
|                                | 2013-06-18 | 829.70        | 759.60        | 70.12                              | 4.40        | 486.90                           |
|                                | 2013-06-19 | 356.00        | 305.80        | 50.22                              | 5.92        | 343.80                           |
|                                | 2013-06-20 | 282.10        | 287.60        | -5.49                              | -0.82       | -38.74                           |
|                                | 2013-06-21 | 168.40        | 151.50        | 16.88                              | 3.33        | 114.30                           |
|                                | Mean       | <b>546.64</b> | <b>439.54</b> | <b>107.13</b>                      | <b>6.85</b> | <b>715.47</b>                    |
| Tianshan Bio-Engineer (300313) | 2013-06-17 | 122.80        | 115.90        | 6.90                               | 2.38        | 96.30                            |
|                                | 2013-06-18 | 195.50        | 170.60        | 24.88                              | 6.01        | 351.50                           |
|                                | 2013-06-19 | 156.70        | 143.20        | 13.49                              | 3.83        | 199.30                           |
|                                | 2013-06-20 | 104.60        | 106.10        | -1.52                              | -0.64       | -18.87                           |
|                                | 2013-06-21 | 61.17         | 34.97         | 26.20                              | 14.15       | 354.60                           |
|                                | Mean       | <b>128.15</b> | <b>114.15</b> | <b>13.99</b>                       | <b>5.15</b> | <b>196.57</b>                    |
| Newcapec (300248)              | 2013-06-17 | 43.79         | 22.85         | 20.94                              | 12.56       | 264.70                           |
|                                | 2013-06-18 | 40.35         | 23.07         | 17.28                              | 11.14       | 219.90                           |
|                                | 2013-06-19 | 37.48         | 33.60         | 3.88                               | 2.58        | 51.34                            |
|                                | 2013-06-20 | 53.67         | 62.10         | -8.43                              | -4.35       | -100.70                          |
|                                | 2013-06-21 | 39.32         | 46.01         | -6.68                              | -4.83       | -79.32                           |
|                                | Mean       | <b>42.92</b>  | <b>37.53</b>  | <b>5.40</b>                        | <b>3.42</b> | <b>71.18</b>                     |
| Unisplendour Guhan (000590)    | 2013-06-17 | 134.40        | 149.50        | -15.15                             | -1.45       | -132.80                          |
|                                | 2013-06-18 | 315.30        | 153.30        | 162.00                             | 13.28       | 1677.50                          |
|                                | 2013-06-19 | 62.76         | 62.54         | 0.22                               | 0.04        | 5.04                             |
|                                | 2013-06-20 | 84.21         | 63.16         | 21.05                              | 4.03        | 212.60                           |
|                                | 2013-06-21 | 29.58         | 65.09         | -35.51                             | -6.31       | -353.50                          |
|                                | Mean       | <b>125.25</b> | <b>98.72</b>  | <b>26.52</b>                       | <b>1.92</b> | <b>281.77</b>                    |

Figure 5 shows only 19 trading weeks. Observing the return of the two stocks during the 20 weeks and comparing with the CSI over the same period, the two stocks get the better returns and it is far better than the CSI. The stock of Victory Precision (002426) gets the 107.96% best returns at the 17th trading week in Figure 4 and the stock of Newcapec (300248) gets the 49.21% best returns at the 10th trading week in Figure 5. Certainly, we cannot get the best return in practice, but the experienced investors will get far better returns than the CSI by purchasing the recommended stocks.

Figure 6 shows that different stock recommendation models will bring different stock returns. We compute the mean return rate of 18 trading weeks for 20 stocks in two different stocks sets which were recommended by the proposed method and the traditional method in the above four different periods and Figure 6 shows the result. Assuming that the target user bought the frontal stocks in the recommend stocks set, we find that the target user with appropriate operating will get far better stock return rate as the top line indicated in Figure 6.

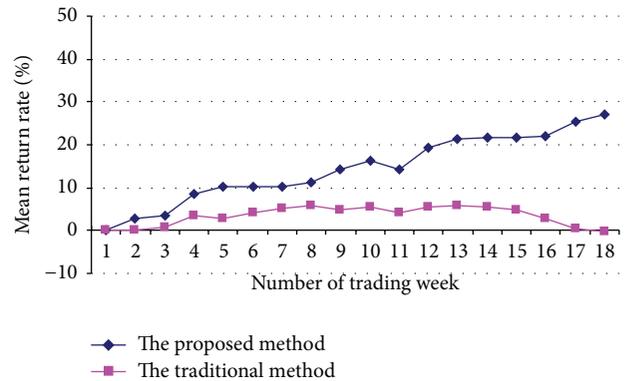


FIGURE 6: Mean return rate comparing of two different methods.

In Section 4.1, we selected data at four different periods of the real stock market in China. Each of the periods is very small and has only five trading days. What is the reason for validating the recommended stocks in experimentation

TABLE 8: The recommended stocks maximum possible gains during the four different day intervals.

| Recommend date/period  | Stock name (code)/CSI           | Recommend buy/open Price | Highest price within 10 days | Rise-up within 10 days | Highest price within 30 days | Rise-up within 30 days | Highest price within 90 days | Rise-up within 90 days | Highest price within 1 year | Rise-up within 1 year |
|------------------------|---------------------------------|--------------------------|------------------------------|------------------------|------------------------------|------------------------|------------------------------|------------------------|-----------------------------|-----------------------|
| 2012-12-05<br>A period | Jianghuai Automobile (600418)   | 5.56                     | 6.01                         | 8.09%                  | 6.98                         | 25.54%                 | 9.39                         | 68.88%                 | 10.37                       | 86.51%                |
|                        | Janus Precision (300083)        | 10.14                    | 11.60                        | 14.40%                 | 12.55                        | 23.77%                 | 17.35                        | 71.10%                 | 25.00                       | 146.55%               |
|                        | Great Wall Motor (601633)       | 18.43                    | 20.80                        | 12.86%                 | 23.91                        | 29.73%                 | 31.50                        | 70.92%                 | 52.85                       | 186.76%               |
|                        | Huasun Group (000790)           | 5.85                     | 6.67                         | 14.02%                 | 7.10                         | 21.37%                 | 8.55                         | 46.15%                 | 12.89                       | 120.34%               |
|                        | Faw Car (000800)                | 6.30                     | 7.13                         | 13.17%                 | 8.72                         | 38.41%                 | 8.79                         | 39.52%                 | 15.88                       | 152.06%               |
|                        | Shanghai Composite Index (CSI)  | <b>1973.11</b>           | <b>2152.50</b>               | <b>9.09%</b>           | <b>2296.11</b>               | <b>16.37%</b>          | <b>2444.80</b>               | <b>23.91%</b>          | <b>2444.80</b>              | <b>23.91%</b>         |
| 2012-12-24<br>B period | Dahu Aquaculture (600257)       | 6.83                     | 7.42                         | 8.64%                  | 7.42                         | 8.64%                  | 7.42                         | 8.64%                  | 9.36                        | 37.04%                |
|                        | Changan Automobile (000625)     | 6.17                     | 6.73                         | 9.08%                  | 7.62                         | 23.50%                 | 9.56                         | 54.94%                 | 13.19                       | 113.78%               |
|                        | Xinning Modern (300013)         | 9.20                     | 11.43                        | 24.24%                 | 11.43                        | 24.24%                 | 11.43                        | 24.24%                 | 19.97                       | 117.07%               |
|                        | Robam Appliances (002508)       | 18.62                    | 19.35                        | 3.92%                  | 25.00                        | 34.26%                 | 26.87                        | 44.31%                 | 42.79                       | 129.81%               |
|                        | Great Wall Motor (601633)       | 22.45                    | 23.91                        | 6.50%                  | 30.39                        | 35.37%                 | 34.18                        | 52.25%                 | 52.85                       | 135.41%               |
|                        | Shanghai Composite Index (CSI)  | <b>2150.65</b>           | <b>2296.11</b>               | <b>6.76%</b>           | <b>2335.81</b>               | <b>8.61%</b>           | <b>2444.80</b>               | <b>13.68%</b>          | <b>2444.80</b>              | <b>13.68%</b>         |
| 2013-02-18<br>C period | Irtouch Systems (300282)        | 15.55                    | 15.58                        | 0.19%                  | 16.61                        | 6.82%                  | 16.61                        | 6.82%                  | 32.69                       | 110.23%               |
|                        | Transinfo Technology (002373)   | 11.90                    | 12.22                        | 2.69%                  | 12.73                        | 6.97%                  | 14.01                        | 17.73%                 | 19.50                       | 63.87%                |
|                        | Microgate Technology (300319)   | 17.65                    | 18.62                        | 5.50%                  | 18.66                        | 5.72%                  | 18.66                        | 5.72%                  | 37.63                       | 113.20%               |
|                        | Kingsun Optoelectronic (002638) | 10.80                    | 12.65                        | 17.13%                 | 13.19                        | 22.13%                 | 15.42                        | 42.78%                 | 16.20                       | 50.00%                |
|                        | Victory Precision (002426)      | 6.78                     | 7.22                         | 6.49%                  | 7.22                         | 6.49%                  | 12.50                        | 84.37%                 | 20.46                       | 201.77%               |
|                        | Shanghai Composite Index (CSI)  | <b>2444.80</b>           | <b>2444.80</b>               | <b>0.00%</b>           | <b>2444.80</b>               | <b>0.00%</b>           | <b>2444.80</b>               | <b>0.00%</b>           | <b>2444.80</b>              | <b>0.00%</b>          |
| 2013-06-24<br>D period | Sunway Communicate (300136)     | 23.05                    | 27.20                        | 18.00%                 | 27.20                        | 18.00%                 | 27.20                        | 18.00%                 | 27.66                       | 20.00%                |
|                        | Huaton Spandex (002064)         | 6.64                     | 7.73                         | 16.42%                 | 8.30                         | 25.00%                 | 11.50                        | 73.19%                 | 11.50                       | 73.19%                |
|                        | Tianshan Bio-Engineer (300313)  | 13.41                    | 14.44                        | 7.68%                  | 14.44                        | 7.68%                  | 16.63                        | 24.01%                 | 23.10                       | 72.26%                |
|                        | Newcapec (300248)               | 11.99                    | 14.85                        | 23.85%                 | 17.90                        | 49.29%                 | 19.19                        | 60.05%                 | 31.32                       | 161.22%               |
|                        | Unisplendour Guhan (000590)     | 9.90                     | 9.38                         | -5.25%                 | 9.39                         | -5.15%                 | 13.50                        | 36.36%                 | 17.79                       | 79.70%                |
|                        | Shanghai Composite Index (CSI)  | <b>2068.86</b>           | <b>2068.86</b>               | <b>0.00%</b>           | <b>2092.87</b>               | <b>1.16%</b>           | <b>2270.27</b>               | <b>9.74%</b>           | <b>2270.27</b>              | <b>9.74%</b>          |

on larger period with one to eighteen weeks in Figures 4 and 5? As been well known, it is easy for buying stocks with big order, but it is very difficult for selling stocks with big order. The investors which can buy or sell stocks are generally institutional investors. The institutional investors mainly refer to a number of financial institutions, including banks, insurance companies, investment trust companies, credit unions, national bodies to establish a pension fund, and other organizations.

It is very difficult or impossible that the institutional investors can sell out the stocks within a few days after buying lots of stocks with big order. There may be two cases. The first case is that the institutional investors want to sell out all stocks in the next period of day. In general, selling out all the stocks required time is two to four weeks depending on the amount of their holding stocks which is generally more than the amount of buying stocks in those days we select data. The other case is that the institutional investors want to buy and not to sell some stocks in the next period of day. For example, this day when we select data is the starting period of buying stocks for the institutional investors. After some days, the amount of their buying stocks reached their desired amount and the institutional investors are no longer buying stocks. They want to sell out all the stocks when the stocks reach a certain range of price. In general, the time is twelve to sixteen weeks from buying stocks to selling out stocks for the institutional investors. In fact, nobody knows when they buy and when to sell out. In order to validate the proposal, we make the experimentation at the large period from one to eighteen weeks after our recommend date. As we see in Figures 4 and 5, the stock of Victory Precision (002426) and the stock of Kingsun Optoelectronic (002426) get, respectively, the best returns at the 17th trading week and at the 14th trading week in Figure 4 and the stock of Newcapec (300248) gets the better returns at the 4th trading week and it gets the best returns at the 10th trading week in Figure 5. Certainly, we cannot get the best return in practice, because nobody knows the highest price of the stocks in certain period and when they ought to sell out the stocks.

## 5. Conclusions and Future Works

In this paper, we proposed an efficient stock recommendation model based on big order net inflow. The proposed stock recommendation model based on big order net inflow can not only filter some low investment value stocks and improve the prediction accuracy in order to recommend some more investment value stocks to the target users and get more stock returns, but also improve the speed of compute by using some advanced algorithms, such as Binary Search Algorithm, and satisfy the investment desire of the investors in real life. From the experimental results, we found that the proposed stock recommendation model has a better performance than the model based on the money flow, and then it can filter the low or negative investment returns stocks and improve the investment gains for target investors.

In the future, the next work is to study how to improve the accuracy of recommendation model. We will reduce the

impact of the model on the influence of corporate events and rumors of trader and apply the model to more stock markets or future markets in the different country. We think that there are three aspects worth studying. Firstly, we can improve model performance with machine learning algorithms, time series analysis or latest technology of fuzzy cluster, and so on. For example, we may improve the accuracy thought training a large number of historical stocks data sets from different markets and countries using the machine learning algorithms. Secondly, we may improve the accuracy of the clustering set for stocks or users through selecting appropriate attributes of them or adding weights for every attribute. Finally, we may study how to find the big order transactions in real trading process, because some institutional investors may subdivide the big order into many random small orders by quantitative trading system software; it is very difficult for us.

## Conflict of Interests

All the authors of this paper declare that they have no conflict of interests in connection with the work submitted.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 61370073), the National High Technology Research and Development Program of China (no. 2007AA01Z423), Sichuan Province Science and Technology Support Program (no. 2013GZX0165), the Constructing Program of the Key Discipline in Huaihua University, the Scientific Research Fund of Hunan Provincial Education (nos. 12C0840 and 14C0886), the Scientific Research Fund of Huaihua University (nos. HHUY2012-15 and HHUY2011-17), the Science and Technology Plan Projects of Huaihua City, and the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province (no. ZNKZ2014-9).

## References

- [1] J. Duan, H. Liu, and J. Zeng, "Posterior probability model for stock return prediction based on analyst's recommendation behavior," *Knowledge-Based Systems*, vol. 50, pp. 151–158, 2013.
- [2] E. J. de Fortuny, T. de Smedt, D. Martens, and W. Daelemans, "Evaluating and understanding text-based stock price prediction models," *Information Processing & Management*, vol. 50, no. 2, pp. 426–441, 2014.
- [3] L. Huo, B. Jiang, T. Ning, and B. Yin, "A BP neural network predictor model for stock price," in *Intelligent Computing Methodologies*, D. S. Huang, K. H. Jo, and L. Wang, Eds., Lecture Notes in Computer Science, pp. 362–368, Springer, Berlin, Germany, 2014.
- [4] A. A. Adebisi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and artificial neural networks models for stock price prediction," *Journal of Applied Mathematics*, vol. 2014, Article ID 614342, 7 pages, 2014.
- [5] K. Miwa and K. Ueda, "Slow price reactions to analysts' recommendation revisions," *Quantitative Finance*, vol. 14, no. 6, pp. 993–1004, 2014.

- [6] T. Geva and J. Zahavi, "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news," *Decision Support Systems*, vol. 57, no. 1, pp. 212–223, 2014.
- [7] C. S. Han and C. Y. Wan, "Future stock price predicting system for use in enterprise, has future prediction compute server setting up weighted value and producing virtual total amount of market price information based on basis predictive model," KR2014120416-A, to Cs Co Ltd, Korea Advanced Institute of Science & Technology, 2014.
- [8] N. C. Brown, K. D. Wei, and R. Wermers, "Analyst recommendations, mutual fund herding, and overreaction in stock prices," *Management Science*, vol. 60, no. 1, pp. 1–20, 2014.
- [9] M.-Y. Chen, "A high-order fuzzy time series forecasting model for internet stock trading," *Future Generation Computer Systems*, vol. 37, pp. 461–467, 2014.
- [10] B. Q. Sun, H. F. Guo, H. R. Karimi, Y. Ge, and S. Xiong, "Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series," *Neurocomputing*, vol. 151, no. 3, pp. 1528–1536, 2015.
- [11] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, 2015.
- [12] D. Sonsino and T. Shavit, "Return prediction and stock selection from unidentified historical data," *Quantitative Finance*, vol. 14, no. 4, pp. 641–655, 2014.
- [13] M.-Y. Chen and B.-T. Chen, "A hybrid fuzzy time series model based on granular computing for stock price forecasting," *Information Sciences*, vol. 294, pp. 227–241, 2015.
- [14] Z. Xin, Z. Ma, and M. Gu, "Fuzzy clustering collaborative recommendation algorithms served for directional information recommendation," *Computer Science*, vol. 34, no. 9, pp. 128–130, 2007.
- [15] L. Chapple and J. E. Humphrey, "Does board gender diversity have a financial impact? Evidence using stock portfolio performance," *Journal of Business Ethics*, vol. 122, no. 4, pp. 709–723, 2014.
- [16] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [17] Q. Xu, "Continuous time M/G/1 queue with multiple vacations and server close-down time," *Journal of Computational Information Systems*, vol. 3, no. 2, pp. 753–757, 2007.
- [18] A. Frazzini and O. A. Lamont, "Dumb money: mutual fund flows and the cross-section of stock returns," *Journal of Financial Economics*, vol. 88, no. 2, pp. 299–322, 2008.
- [19] I. Gath and E. Bar-On, "Computerized method for scoring of polygraphic sleep recordings," *Computer Programs in Biomedicine*, vol. 11, no. 3, pp. 217–223, 1980.
- [20] M. Hong-Wei, Z. Guang-Wei, and L. Peng, "Survey of collaborative filtering algorithms," *Mini-Micro Systems*, vol. 7, pp. 1282–1288, 2009.
- [21] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International World Wide Web Conference*, pp. 285–295, May 2001.
- [22] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.
- [23] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem," *Information Sciences*, vol. 178, no. 1, pp. 37–51, 2008.
- [24] G. Choudhury, "Some aspects of M/G/1 queue with two different vacation times under multiple vacation policy," *Stochastic Analysis and Applications*, vol. 20, no. 5, pp. 901–909, 2002.
- [25] B. T. Doshi, "Conditional and unconditional distributions for M/G/1 type queues with server vacations," *Questa*, vol. 7, pp. 229–252, 1990.
- [26] H. W. Lee, "M/G/1 queue with exceptional first vacation," *Computers & Operations Research*, vol. 15, no. 5, pp. 441–445, 1988.
- [27] K. K. Leung, "On the additional delay in an M/G/1 queue with generalized vacations and exhaustive service," *Operations Research*, vol. 40, supplement 2, pp. S272–S283, 1992.
- [28] A. Hatamlou, "In search of optimal centroids on data clustering using a binary search algorithm," *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1756–1760, 2012.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

