

Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting

Esmail Hadavandi^a, Hassan Shavandi^{a,*}, Arash Ghanbari^b

^a Department of Industrial Engineering, Sharif University of Technology, P.O. Box 11365-9466, Tehran, Iran

^b Department of Industrial Engineering, College of Engineering, University of Tehran, P.O. Box 11155-4563, Tehran, Iran

ARTICLE INFO

Article history:

Received 8 December 2009

Received in revised form 8 May 2010

Accepted 11 May 2010

Available online 15 May 2010

Keywords:

Stock price forecasting

Genetic fuzzy systems

Self-organizing map (SOM)

Data clustering

Hybrid intelligence model

ABSTRACT

Stock market prediction is regarded as a challenging task in financial time-series forecasting. The central idea to successful stock market prediction is achieving best results using minimum required input data and the least complex stock market model. To achieve these purposes this article presents an integrated approach based on genetic fuzzy systems (GFS) and artificial neural networks (ANN) for constructing a stock price forecasting expert system. At first, we use stepwise regression analysis (SRA) to determine factors which have most influence on stock prices. At the next stage we divide our raw data into k clusters by means of self-organizing map (SOM) neural networks. Finally, all clusters will be fed into independent GFS models with the ability of rule base extraction and data base tuning. We evaluate capability of the proposed approach by applying it on stock price data gathered from IT and Airlines sectors, and compare the outcomes with previous stock price forecasting methods using mean absolute percentage error (MAPE). Results show that the proposed approach outperforms all previous methods, so it can be considered as a suitable tool for stock price forecasting problems.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction and literature review

1.1. Stock price forecasting

Forecasting is the process of making projections about future performance based on existing historic data. An accurate forecast aids in decision-making and planning for the future. Forecasts empower people to modify current variables in the present to predict the future to result in a favorable scenario.

Stock market prediction is regarded as a challenging task in financial time-series forecasting. This is primarily because of the uncertainties involved in the movement of the market. Many factors interact in the stock market including political events, general economic conditions, and traders' expectations. So, stock price time-series data is characterized by nonlinearities, discontinuities, and high-frequency multi-polynomial components and predicting market price movements is quite difficult.

Stock market forecasters focus on developing approaches to successfully forecast/predict index values or stock prices, aiming at high profits using well defined trading strategies. "The central idea to successful stock market prediction is achieving best results using minimum required input data and the least complex stock

market model" [3]. Considering this idea an obvious complexity of the problem paves the way for the importance of intelligent prediction paradigms [1].

Artificial intelligence techniques such as artificial neural networks (ANNs), fuzzy logic, and genetic algorithms (GAs) are popular research subjects, since they can deal with complex engineering problems which are difficult to solve by classical methods [28]. These techniques have been successfully used in the place of the complex mathematical systems for forecasting of time-series [8,26,33,40,41].

Each of AI-based techniques has advantages and disadvantages. One approach to deal with complex real-world problems is to integrate the use of several AI technologies in order to combine their different strengths and overcome a single technology's weakness to generate hybrid models that provides better results than the ones achieved with the use of each isolated technique. Using hybrid models or combining several models has become a common practice to improve forecasting accuracy and the literature on this topic has expanded dramatically [25].

Kuo et al. [29] proposed a genetic algorithm based fuzzy neural network (GFNN) to formulate the knowledge base of fuzzy inference rules which could measure the qualitative effect on the stock market. Next, the effect was further integrated with the technical indexes through the artificial neural network. An example based on the Taiwan stock market was utilized to assess the proposed intelligent system. Evaluation results indicated that the neural

* Corresponding author. Tel.: +98 21 6616 5711; fax: +98 21 6602 2702.

E-mail addresses: es.hadavandi@gmail.com (E. Hadavandi), shavandi@sharif.edu (H. Shavandi), arashghanbari@yahoo.com, arghanbari@ut.ac.ir (A. Ghanbari).

network considering both the quantitative and qualitative factors excels the neural network considering only the quantitative factors both in the clarity of buying–selling points and buying–selling performance.

Wang [39] proposed a hybrid model that uses a data mart to reduce the size of stock data and combined fuzzification techniques with the grey theory to develop a fuzzy grey prediction to predict stock price in Taiwan stock market. He concluded that the proposed model can effectively help stock dealers deal with day trading.

Chang and Liu [7] used a Takagi–Sugeno–Kang (TSK) type fuzzy rule based system (FRBS) for stock price prediction. They used simulated annealing (SA) for training the best parameters of fuzzy systems. They found that the forecasted results from TSK fuzzy rule based model were much better than those of back propagation network (BPN) or multiple regressions.

Hung [24] proposed a new application of fuzzy systems designed for a generalized auto-regression conditional heteroscedasticity (GARCH) model to forecast stock returns. The optimal parameters of the fuzzy membership functions and GARCH model were extracted using a GA to achieve a global optimal solution with a fast convergence rate for this fuzzy GARCH model estimation problem. The proposed model was also compared with the other methods, such as GARCH, EGARCH and outperformed them.

Majhi et al. [30] proposed a trigonometric functional link artificial neural network (FLANN) model for short (one-day) as well as long term (one month, two months) prediction of stock price of leading stock market indices: DJIA and S&P 500. They concluded that proposed model is an effective approach both computationally as well as performance wise to foresee the market levels both in short and medium terms future.

Atsalakis George and Valavanis Kimon [3] proposed a hybrid model that linked two Adaptive Neuro-Fuzzy Inference System (ANFIS) controllers to forecast next day's stock price trends of the Athens and the New York Stock Exchange (NYSE). The proposed system performed very well in trading simulation and comparisons with 13 other similar soft computing based approaches demonstrated solid and superior performance in terms of percentage of prediction accuracy of stock market trend.

Yudong and Lenan [42] proposed an improved bacterial chemotaxis optimization (IBCO) which was integrated into BPN to develop an efficient forecasting model for prediction of various stock indices. Performance comparison with the BPN model simulated indicated that the developed model offers less computational complexity, better prediction accuracy, and less training time.

One of the most popular approaches is the hybridization between fuzzy logic and GAs leading to genetic fuzzy systems (GFSs) [11]. A GFS is basically a fuzzy system augmented by a learning process based on evolutionary computation, which includes genetic algorithms, genetic programming, and evolutionary strategies, among other evolutionary algorithms (EAs) [15].

There are several researches which have used GFSs for forecasting problems [14,35], but there is not any research in the literature that uses a GFS with the ability of learning rule base and tuning data base of fuzzy system by use of genetic algorithm for stock price forecasting problem (a complete literature review on proposed techniques for stock market forecasting can be found in [4]).

This paper presents a hybrid artificial intelligence (AI) methodology for next day stock price prediction to extract useful patterns of information with a descriptive rule induction approach based on genetic fuzzy systems. Our method combines a data clustering technique and a GFS to learn rule base and tune data base of the fuzzy system. We test capability of the proposed method by applying it on stock price data of International Business Machines (IBM) and Dell Corporation from IT sector and British airlines and Ryanair

airlines from Airline sector that has been used in different studies as the case study.

1.2. Stock price forecasting from IT and airline sectors

In the following we present a brief review of using IT and Airlines sectors' stock price forecasting as the case study. Hassan and Nath [19] proposed a Hidden Markov Model (HMM) to generate one-day forecasts of stock prices in a novel way. Hassan et al. [20] proposed and developed a fusion model combining the HMM with an Artificial Neural Network and a Genetic Algorithm to achieve better forecasts. In their model, ANN was used to transform the input observation sequences of HMM and the GA was used to optimize the initial parameters of the HMM. This optimized HMM was then used to identify similar data pattern from the historical dataset. The comparison showed that forecasting ability of the fusion model is better than ARIMA model and HMM proposed in [19].

Hassan [21] proposed a novel combination of the HMM and the fuzzy models for forecasting stock market data. The model used HMM to identify data patterns and then used fuzzy logic to generate appropriate fuzzy rules and obtain a forecast value for next day stock price. The forecast accuracy of the combination HMM–fuzzy model was better when compared to the ARIMA and ANN and other HMM-based forecasting models.

2. Clustering-genetic fuzzy system

As mentioned before, this paper presents a hybrid artificial intelligence method called clustering-genetic fuzzy system (CGFS) for constructing an expert system (ES) for stock price prediction problems. There are three main stages in this research to construct CGFS. The first stage is variable selection stage; we use stepwise regression analysis (SRA) to choose the key variables that are to be considered in the model. In the second stage we use self-organization map (SOM) neural network to divide the data into sub-populations and reduce the complexity of the whole data space to something more homogeneous. And in the last stage we construct GFS for stock price prediction, to meet this purpose SOM outcomes will be fed into independent genetic fuzzy systems.

The procedure of making CGFS is shown in Fig. 1. Details of each stage are described in the following.

2.1. Variables selection by stepwise regression analysis

Variable selection is the process of selecting an optimum subset of input variables from the set of potentially useful variables which may be available in a given problem. Different researchers have applied variety of feature selection methods such as genetic algorithm [16], principal component analysis (PCA) [43], grey relation analysis [10] and stepwise regression analysis (SRA) [8] to select key factors in their forecasting systems. Among them, in recent years some researcher have used SRA for input variable selection in the field of stock market forecasting and they have obtained very promising results [7,17]. So, in this paper we adopt stepwise regression to analyze and select variables, and as its consequence improve the forecasting accuracy of the system. Stepwise regression method determines the set of independent factors that most closely determine the dependent variable. This task is carried out by means of the repetition of a variable selection. At each of these steps, a single variable is either entered or removed from the model. For each step, simple regression is performed using the previously included independent variables and one of the excluded variables.

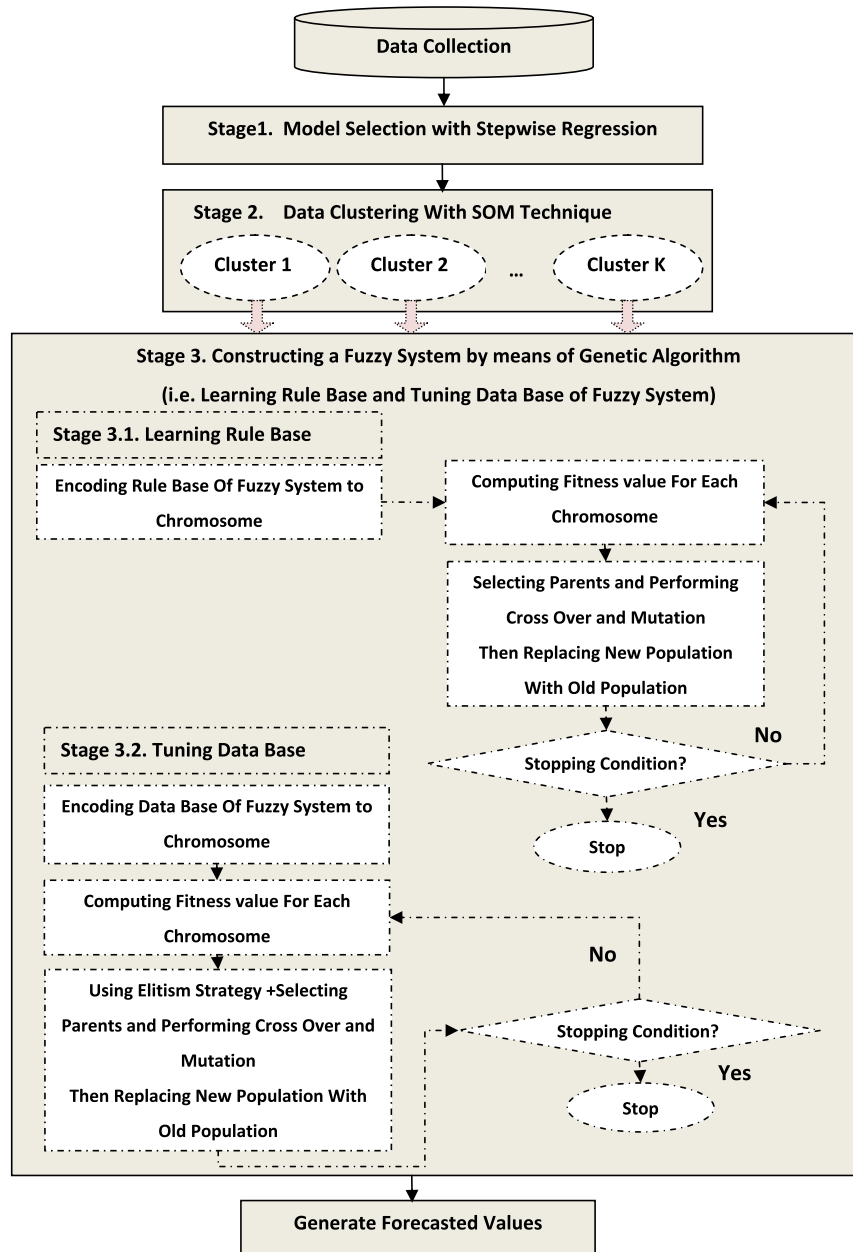


Fig. 1. The procedure of making CGFS model.

2.2. Data clustering by self-organizing map (SOM)

Clustering algorithms are classified to two groups: the agglomerative hierarchical algorithms [2] such as the centroid and Ward methods, and the nonhierarchical clustering [38], such as K-means and SOM neural networks. Each of these algorithms has their own advantages and disadvantages. Depending on the application, a particular type of clustering method should be chosen. Among clustering algorithms, because of the stable and flexible architecture of SOM neural networks, it has been used in a wide range of applications. Mangiameli et al. [31] made a comparison between the self-organizing map neural network clustering and the hierarchical clustering methods. A large number of data sets were used to test the performance of SOM and the hierarchical clustering methods. This research showed that SOM outperforms the hierarchical methods in clustering messy data and has better accuracy and robustness. There are also some researches in the field of fore-

casting which have used SOM neural networks for clustering data and they have obtained good results [9,23]. So, in this paper we use SOM neural networks for clustering datasets.

Self-organizing map is an unsupervised learning algorithm. This method was developed by Kohonen [27]. The SOM network consists of M neurons arranged in a 2-D rectangular or hexagonal grid. Each of the neurons i is assigned a weight vector, $w_i \in R^n$ (index $i = (p, q)$ for 2-D map). At each training step t , a training data $x^t \in R^n$ is randomly drawn from data set and calculates the Euclidean distances between x^t and all neurons. A winning neuron with weight of w_j can be found according to the minimum distance to x^t :

$$j = \arg \min_i \|x^t - w_i^t\|, \quad i \in \{1, 2, \dots, M\} \quad (1)$$

Then, the SOM adjusts the weight of the winner neuron and neighborhood neurons and moves closer to the input space according to:

$$w_i^{t+1} = w_i^t + \alpha^t \times h_{ji}^t \times [x^t - w_i^t] \quad (2)$$

where α^t and h_{ji}^t are the learning rate and neighborhood kernel at time t , respectively. Both α^t and h_{ji}^t decrease monotonically with time and within 0 and 1. The neighborhood kernel h_{ji}^t is a function of time and distance between neighbor neuron i and winning neuron. A widely applied neighborhood kernel can be written in terms of Gaussian function:

$$h_{ji}^t = \exp\left(-\frac{\|r_j - r_i\|^2}{2\sigma_t^2}\right) \quad (3)$$

where r_j and r_i are the position of winner neuron and neighbor neuron on map. σ_t is kernel width and decreasing with time. This process of weight-updating will be performed for a specified number of iterations. The Tanagra 1.4 was used for applying SOM clustering in this research [34].

2.3. Developing the genetic fuzzy system

Nowadays fuzzy rule based systems have been successfully applied to a wide range of real-world problems from different areas. In order to design an intelligent system of this kind for a concrete application, several tasks have to be performed. One of the most important and difficult ones is to derive an appropriate knowledge base (KB) about the problem. The KB stores the available knowledge in the form of fuzzy linguistic IF–THEN rules. It is composed of the rule base (RB), constituted by the collection of rules in their symbolic forms, and the data base (DB), which contains the linguistic term sets and the membership functions defining their meanings [6].

The difficulty presented by human experts to express their knowledge in the form of fuzzy rules has made researchers develop automatic techniques to perform this task. In this sense, a large amount of methods has been proposed to automatically generate fuzzy rules from numerical data. Usually, they use complex rule generation mechanisms such as neural networks [32] or genetic algorithms [11].

GAs have been demonstrated to be a powerful tool for automating the definition of the KB, since adaptive control, learning, and self-organization may be considered in a lot of cases as optimization or search processes. In particular, the application to the design, learning, and tuning of KBs has produced quite promising results. These approaches can be given the general name of genetic fuzzy systems [12].

In this paper we use a Mamdani-type fuzzy rule based system to deal with stock price forecasting problems. In a Mamdani-type fuzzy rule based system a common rule represented as follows:

If X_1 is A_1 and X_2 is A_2 THEN Y is C_1 , where X_1 , X_2 and Y are linguistic variables and A_1 , A_2 and C_1 are corresponding fuzzy sets. Evolutionary process that we use in this paper for evolving knowledge base of fuzzy rule based system consists of two general stages; stage 1 evolves rule base of fuzzy system and stage 2 tunes

data base of fuzzy system. In the following we will describe these two stages.

2.3.1. Evolving the rule base of fuzzy system by using genetic algorithm

About the appropriate number of linguistic variables, Herrera says “The usual way to define the DB involves choosing a number of linguistic terms for each linguistic variable (an odd number between 3 and 9, which is usually the same for all the variables) and setting the values of the system parameters by a uniform distribution of the linguistic terms into the variable universe of discourse” [22]. In the proposed model, a previously defined DB constituted by uniform fuzzy partitions with triangular membership functions crossing at height 0.5 is considered. The number of linguistic terms forming each of them can be specified by the GFS designer, and then Pittsburgh approach [36] is used for learning rule base. Each chromosome encodes a whole fuzzy rule set and the evolved rule base is the best individual of the last population. Pittsburgh approach can be decomposed in the following steps.

Step 1–Coding mechanism.

Many GFSs employ the decision table proposed by [37] as the common classical representation for the rule base of a fuzzy system. A fuzzy decision table represents a special case of a crisp relation (the ordinary type of relations we are familiar with) defined over the collections of fuzzy sets corresponding to the input and output variables. A chromosome is obtained from the decision table by going row-wise and coding each output fuzzy set as an integer number with start from 1 to number of output variable linguistic terms. Fuzzy decision table for an FRBS with two inputs (X_1 , X_2) and one output (Y) variable, with three fuzzy sets (A_{11} , A_{12} , A_{13} , A_{21} , A_{22} , A_{23}) related to each input variable and four fuzzy sets (B_1 , B_2 , B_3 , B_4) related to the output variable and applying this code to the fuzzy decision table are represented in Fig. 2.

Step 2–Generating the initial population.

Initial chromosomes (N_{pop}) are randomly generated; while the alleles are in the set $\{1, 2, \dots, N_B\}$ (N_B is the number of output variables' linguistic terms). All consequent labels have the same probability to be assigned to each gene.

Step 3–Calculating the fitness values.

As regards the fitness function, it is based on an application-specific measure usually employed in the design of GFSs, the mean squared error (MSE) over a training data set, which is represented by the following expression:

$$MSE(C_j) = \frac{1}{N} \sum_{i=1}^N (Y_i - P_i)^2 \quad (4)$$

where Y_i is the actual value and P_i is the output value of i th training data obtained from the FRBS using the RB coded in j th chromosome (C_j) and N is the number of training data.

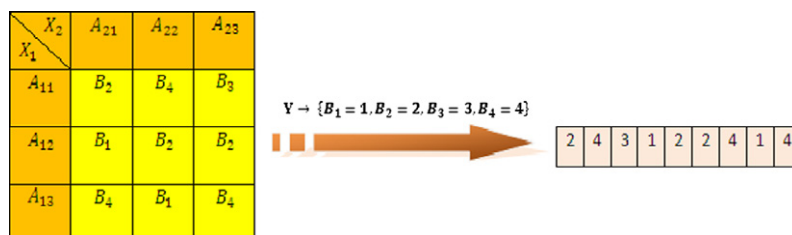


Fig. 2. Coding decision table as a chromosome.

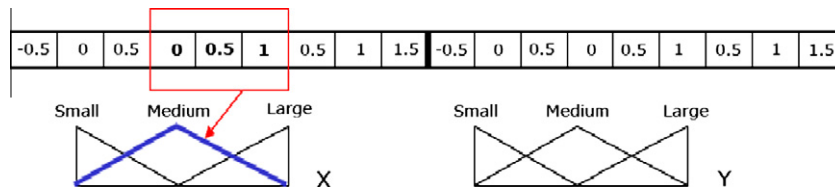


Fig. 3. Coding data base as chromosomes.

Step 4–Generate $(N_{pop} - 1)$ chromosome using the genetic operations.

We use binary tournament selection scheme for selection procedure. In binary tournament selection, two members of the population are selected at random and their fitness compared and the best one according to fitness value will be chosen to reproduce. Also we use one-point crossover and uniform mutation for genetic operators.

Step 5–Add the best rule set in the current population to the newly generated $(N_{pop} - 1)$ chromosome to form the next population.

Step 6–If the number of generations equals to the maximum generation number, then stop; otherwise go to *step 3*.

2.3.2. Genetic tuning of data base

After generation of rule base, we utilize the genetic tuning process that was proposed by Cordon and Herrera [13]. This tuning process slightly adjusts the shape of the membership functions of a preliminary DB defined. This approach can be decomposed in the following steps.

Step 1–Coding data base as chromosomes.

Each chromosome encodes a different data base definition. We use triangular membership functions for input and output variables' linguistic terms. Each triangular membership function is encoded by 3 real values, a primary fuzzy partition is represented as an array composed of $3N$ real values, with N being the number of linguistic terms for each variable (we take the same number of linguistic terms for each input and output linguistic variable). The complete DB for a problem, in which M linguistic variables are involved, is encoded into a fixed-length real coded chromosome C_j built by joining the partial representations of each one of variable fuzzy partitions as is shown in the following:

$$C_i^j = (a_{i1}^j, b_{i1}^j, c_{i1}^j, a_{i2}^j, b_{i2}^j, c_{i2}^j, \dots, a_{iN}^j, b_{iN}^j, c_{iN}^j)$$

$$C_j = C_1^j C_2^j C_3^j \dots C_{M-1}^j C_M^j$$

A sample coded data base with one input variable as well as one output variable was shown in Fig. 3. Each variable is defined by a fuzzy linguistic term such as small, medium, and large.

Step 2–Generating the initial population.

The initial population (N_{pop}) is created by using the initial DB definition. The first chromosome (C_1) is encoded directly from initial DB definition. The remaining individuals $(N_{pop} - 1)$ are generated by associating an interval of performance, $[c_h^l, c_h^r]$ to every gene c_h in C_1 , $h = 1, 2, \dots, 3NM$. Each interval of performance will be the interval of adjustment for the corresponding variable, $c_h \in [c_h^l, c_h^r]$.

If $t \bmod 3 = 1$, then c_t is the left value of the support of a triangular fuzzy number. The triangular fuzzy number is defined by the

three parameters (c_t, c_{t+1}, c_{t+2}) and the intervals of performance are as the following:

$$c_t \in [c_t^l, c_t^r] = \left[c_t - \frac{c_{t+1} - c_t}{2}, c_t + \frac{c_{t+1} - c_t}{2} \right]$$

$$c_{t+1} \in [c_{t+1}^l, c_{t+1}^r] = \left[c_{t+1} - \frac{c_{t+1} - c_t}{2}, c_{t+1} + \frac{c_{t+2} - c_{t+1}}{2} \right]$$

$$c_{t+2} \in [c_{t+2}^l, c_{t+2}^r] = \left[c_{t+2} - \frac{c_{t+2} - c_{t+1}}{2}, c_{t+2} + \frac{c_{t+3} - c_{t+2}}{2} \right]$$

Fig. 4 shows these intervals.

Step 3–Fitness value function.

We use MSE over a training data set as fitness function. This fitness function is applied on the chromosomes considering the tuned membership functions and the rule base extracted in previous phase.

Step 4–Selection and elitism.

The best 10% of the population are copied without changes in the elitism set. Elitism set ensures that the best chromosomes will not be destroyed during crossover and mutation. The selection process is then implemented. We use binary tournament selection scheme to selecting chromosomes for mating pool. The size of the mating pool equals ninety percent of the population size.

Step 5–Genetic operators.

We use *BLX-0.1* crossover [18] and uniform mutation in the proposed genetic tuning process.

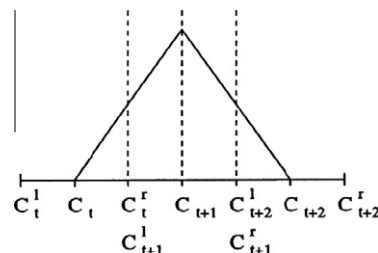


Fig. 4. Membership function and interval of performance for the tuning process.

Table 1
Variables description.

Variable	Description
Open price	The first price of a given stock in a daily trading
Close price	The price of the last transaction for a given stock at the end of a daily trading
High price	The highest price that was paid for a stock during a daily trading
Low price	The lowest price of a stock reached in a daily trading

Table 2
Training and test datasets.

Stock type	Stock name	Training data			Test data		
		From	To	#Observations	From	To	#Observations
IT sector	IBM corporation	10 Feb 2003	10 Sep 2004	400	13 Sep 2004	21 Jan 2005	91
	DELL Inc.	10 Feb 2003	10 Sep 2004	400	13 Sep 2004	21 Jan 2005	91
Airline sector	British airlines	17 Sep 2002	10 Sep 2004	503	11 Sep 2004	20 Jan 2005	91
	Ryanair airlines	6 May 2003	6 Dec 2004	400	7 Dec 2004	17 Mar 2005	71

Step 6–Replacement.

The current population is replaced by the newly generated offsprings, which forms the next generation by integrating the elitism set.

Step 7–Stopping criteria.

If the number of generations equals to the maximum generation number, then stop; otherwise go to step 3.

3. Experimental results

In this section we implement the proposed CGFS using the daily stock price of IBM and Dell Corporations from IT sector and British airlines and Ryanair airlines from Airline sector. All data set is collected from www.finance.yahoo.com.

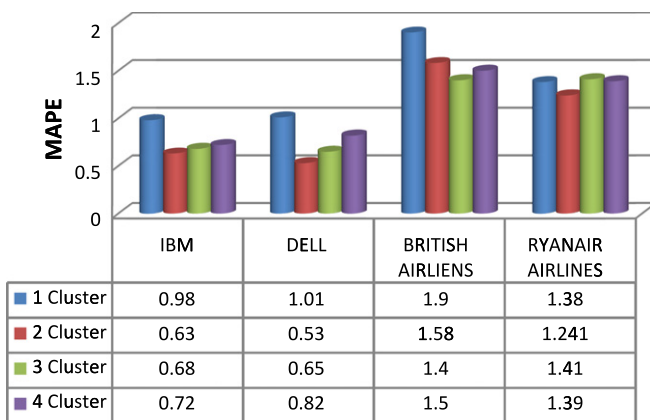
Four attributes: open, high, low and close price from the daily stock market are used to form the observation vector. The forecast variable is the next day's closing price. Description of the variables is shown in Table 1 and details about the training and test dataset

are given in Table 2. For the purpose of comparing the methods, we use the same train and test data sets which have been used by Hassan and Nath [19], Hassan et al. [20], Hassan [21].

3.1. Implementing clustering-genetic fuzzy system for stock price forecasting

In the first stage, we use SRA to eliminate low impact factors and choose the most influential ones out of mentioned factors. The criterion for adding or removing is determined by F -test statistic value and decreasing the sum of squared error. After the entrance of first variable to the model, the variable number is increased step by step; once it is removed from this model, it will never enter the model again. Before selecting variables, the critical point, level of significant and the values of F_e (F -to-enter) and F_r (F -to-remove) have to be determined first. Then the partial F value of each step has to be calculated and compared to F_e and F_r ; If $F > F_e$, it is considered to add variables to the model; otherwise, if $F < F_r$, the variables are removed from model [5]. The statistical software SPSS 17.0 was used to applying stepwise regression analysis in this research considering $F_e = 3.84$ and $F_r = 2.71$. The outcomes of this stage from IBM, British airlines and Ryanair airlines are open and close price and for Dell Corporation are low and close price. Selected variables are inputted into the SOM model and for each case different clusters are generated, where each cluster contains a portion of train data. We use trail-and-error process for determining the size of SOM map, after examination of different number of clusters such as 2, 3 and 4 we obtained that in our cases the suitable number of clusters (k) that provides minimum MAPE values for test data of IBM, Dell, British Airlines and Ryanair Airlines are 2,2,3,2, respectively, results are shown in Fig. 5. In the second stage we build a GFS for each cluster using related training data. Finally, in the testing phase, the data are first clustered and stock price forecasting is done by means of each cluster's test data. The suitable features of CGFSs and number of train and test observations for each cluster and each case after examination of different feature of parameters are shown in Table 3.

The proposed CGFS is applied for forecasting the next day's closing price of each case and the results are shown in Fig. 6.

**Fig. 5.** MAPE values of the forecasting results for different clusters for each case.**Table 3**
Suitable features of CGFS for different sectors and number of observations in each cluster.

CGFS-suitable features		Stock name								
		IT sector				Airline sector				
		IBM corporation		DELL Inc.		British airlines			Ryanair airlines	
		C ₁	C ₂	C ₁	C ₂	C ₁	C ₂	C ₃	C ₁	C ₂
Number of labels		3	3	3	3	3	3	3	3	3
Population size		50	80	100	100	60	60	80	80	80
Number of generations		120	200	1000	800	500	500	900	500	140
Crossover probability		0.6	0.87	0.86	0.86	0.87	0.87	0.87	0.87	0.86
Mutation rate		0.01	0.02	0.02	0.02	0.06	0.06	0.01	0.02	0.01
#Observations in clusters	Training data	226	174	287	113	167	160	176	202	198
	Test data	74	17	88	3	1	60	30	70	1

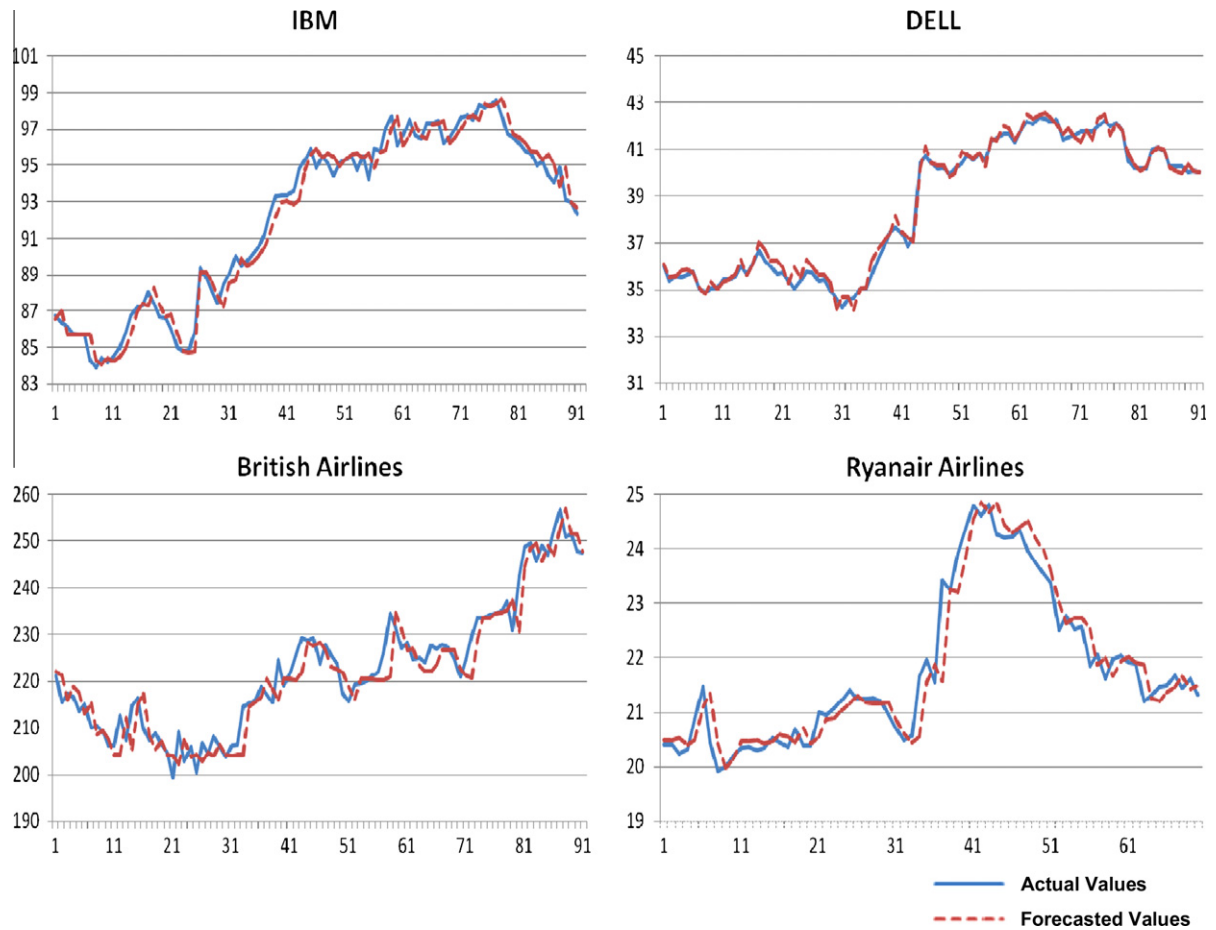


Fig. 6. The forecasting results of CGFS.

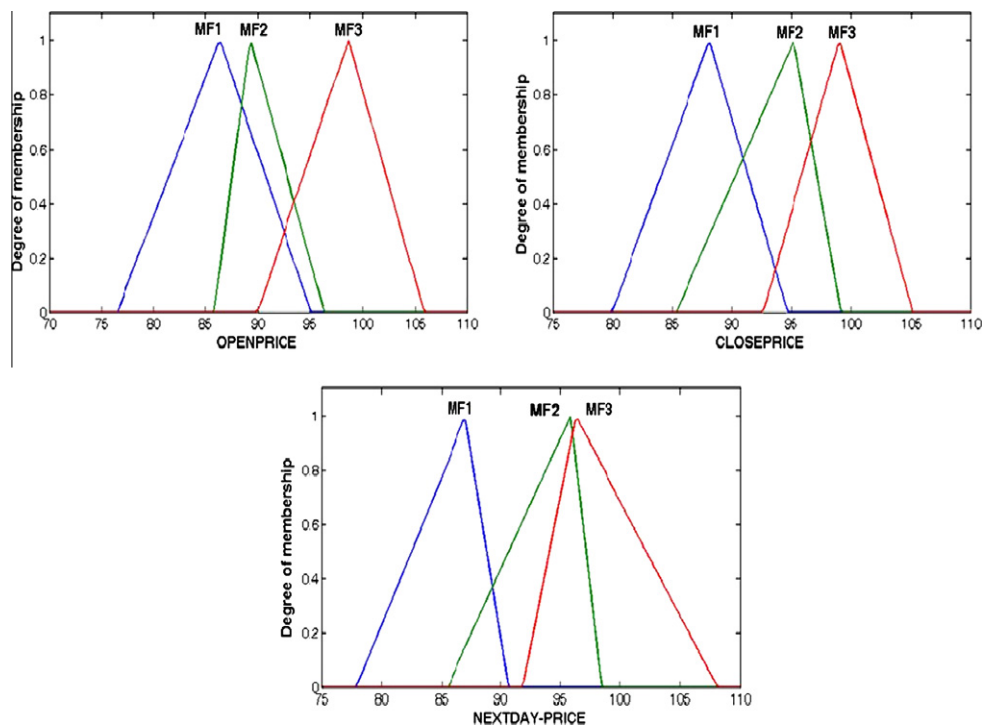


Fig. 7. The tuned membership functions of input and output variables in cluster 1's GFS of IBM.

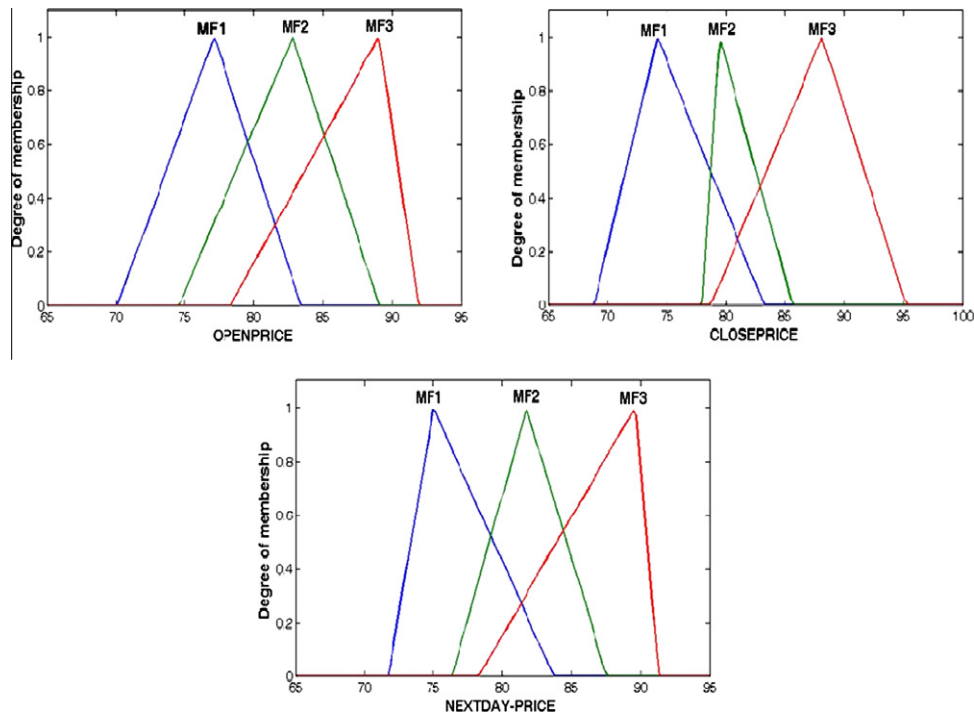


Fig. 8. The tuned membership functions of input and output variables in cluster 2's GFS of IBM.

	Close price		
	MF1	MF2	MF3
Open price	MF1	MF3	MF1
	MF2	MF2	MF3
	MF3	MF2	MF3

Fig. 9. Rule base of GFS for (a) cluster 1 (b) cluster 2 of IBM.

In the following the tuned membership functions of input and output variables in CGFS of cluster 1 and cluster 2 and rule base of them for IBM are shown in Figs. 7–9, respectively.

3.2. Performance analysis of CGFS

For the purpose of evaluating CGFS forecasting accuracy, we will compare outputs of this method with other methods that

are proposed for IT and Airlines sectors' stock price prediction. We perform this task by a common evaluation statistic called MAPE:

$$\text{MAPE} = 100 \times \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - P_i|}{Y_i} \quad (5)$$

where Y_i is the actual value and P_i is the forecasted value of i th test data obtained from CGFS and N is the number of test data. Summary of CGFS evaluations in comparison with the other methods is shown in Table 4.

Regarding to Table 4, our proposed approach has improved the forecasting accuracy of stock prices of all three cases. Namely, CGFS outperforms the rest of methods due to MAPE evaluation so this shows that it can be considered as a promising alternative for stock price prediction problems.

4. Conclusions

This paper presents a novel approach based on genetic fuzzy systems and SOM clustering (CGFS) for building a stock price forecasting expert system, with the aim of improving forecasting accuracy. At the first stage we used stepwise regression analysis to choose the key variables that are to be considered in the model. At the second stage, we categorize our data set into k clusters by

Table 4
CGFS evaluation vs. previous methods based on MAPE value.

Method	Stock name			
	IT sector		Airline sector	
	IBM corporation	DELL Inc.	British airlines	Ryanair airlines
HMM [19]	1.219	1.012	2.629	1.928
Hybrid of HMM, ANN and GA [20]	0.849	0.699	1.646	1.377
Hybrid of HMM and fuzzy logic [21]	0.779	0.405	1.529	1.356
ARIMA [21]	0.972	0.660	1.573	1.504
ANN [21]	0.972	0.660	2.283	1.504
CGFS (the proposed method)	0.63	0.534	1.413	1.241

means of SOM method, and at the third stage we fed each cluster into a robust genetic fuzzy approach to build the stock price forecasting expert system. CGFS approach has the following novel features:

- By using SOM clustering, we divide the data into sub-populations and reduce the complexity of the whole data space to something more homogeneous.
- GAs have been demonstrated to be a powerful tool for automating the definition of the fuzzy rule based systems. CGFS uses a genetic algorithm for extracting rule base of the fuzzy expert system.
- For the purpose of accuracy improvement, it tunes the data base of the fuzzy expert system using a unique genetic algorithm.
- We can be assured of working with optimum solutions, expressed in an easy, semantically understandable way of reasoning of the human being.

In order to evaluate the proposed approach we applied it on stock price data of IBM and Dell Corporations from IT sector and British airlines and Ryanair airlines from Airline sector which had been used in different papers as the case study. Results showed that forecasting accuracy of CGFS outperforms the rest of approaches regarding to MAPE evaluation, and CGFS can be used as a suitable forecasting tool to deal with stock price forecasting problems.

References

- [1] A. Abraham, B. Nath, P.K. Mahanti, Hybrid intelligent systems for stock market analysis, in: V.N. Alexandrov, J. Dongarra, B.A. Julianno, R.S. Renner, C.J.K. Tan (Eds.), *Computational Science*, Springer-Verlag, Germany, 2001, pp. 337–345.
- [2] M.R. Anderberg, *Cluster Analysis for Applications*, Academic Press, New York, 1973.
- [3] S. Atsalakis George, P. Valavanis Kimon, Forecasting stock market short-term trends using a neuro-fuzzy based methodology, *Expert Syst. Appl.* 36 (2009) 10696–10707.
- [4] S. Atsalakis George, P. Valavanis Kimon, Surveying stock market forecasting techniques – Part II: soft computing methods, *Expert Syst. Appl.* 36 (2009) 5932–5941.
- [5] J. Burkholder Thomas, L. Lieber Richard, Stepwise regression is an alternative to splines for fitting noisy data, *J. Biomech.* 29 (1996) 235–238.
- [6] J. Casillas, O. Cordón, F. Herrera, P. Villar, A hybrid learning process for the knowledge base of a fuzzy rule-based system, in: *Proceedings of the 2004 International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Perugia, Italy, 2004, pp. 2189–2196.
- [7] P.-C. Chang, C.-H. Liu, A TSK type fuzzy rule based system for stock price prediction, *Expert Syst. Appl.* 34 (2008) 135–144.
- [8] P.-C. Chang, C.-H. Liu, C.-Y. Fan, Data clustering and fuzzy neural network for sales forecasting: a case study in printed circuit board industry, *Knowl. Based Syst.* 22 (2009) 344–355.
- [9] P.-C. Chang, C.-H. Liu, Y.-W. Wang, A hybrid model by clustering and evolving fuzzy rules for sales decision supports in printed circuit board industry, *Decis. Support Syst.* 42 (2006) 1254–1269.
- [10] P.-C. Chang, Y.-W. Wang, T. Chi-Yang, Evolving neural network for printed circuit board sales forecasting, *Expert Syst. Appl.* 29 (2005) 83–92.
- [11] O. Cordón, F. Herrera, F. Hoffmann, L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, World Scientific, Singapore, 2001.
- [12] O. Cordón, F. Herrera, A general study on genetic fuzzy systems, in: J. Periaux, G. Winter, M. Galen, P. Cuesta (Eds.), *Genetic Algorithms in Engineering and Computer Science*, Wiley, 1995, pp. 33–57.
- [13] O. Cordón, F. Herrera, A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples, *Int. J. Approx. Reason.* 17 (1997) 369–407.
- [14] I.G. Damousis, P. Dokopoulos, A fuzzy expert system for the forecasting of wind speed and power generation in wind farms, in: *International Conference on Power Industry Computer Applications PICA*, Sydney, Australia, 2001, pp. 63–69.
- [15] A.E. Eiben, J.E. Smith, *Introduction to Evolutionary Computation*, Springer, Berlin, 2003.
- [16] M.E. ElAlami, A filter model for feature subset selection based on genetic algorithm, *Knowl. Based Syst.* 22 (2009) 356–362.
- [17] A. Esfahanipour, W. Aghamiri, Adapted neuro-fuzzy inference system on indirect approach TSK fuzzy rule base for stock market analysis, *Expert Syst. Appl.* 37 (2010) 4742–4748.
- [18] L.J. Eshelman, J.D. Schaffer, Real-coded genetic algorithms and interval-schemata, in: L.D. Whitley (Ed.), *Foundations of Genetic Algorithms*, Morgan Kaufmann Publishers, San Mateo, CA, 1993, pp. 187–202.
- [19] R. Hassan, B. Nath, Stock market forecasting using hidden markov model: a new approach, in: *5th International Conference on Intelligent System Design and Application*, Poland, 2005, pp. 192–196.
- [20] R. Hassan, B. Nath, M. Kirley, A fusion model of HMM, ANN and GA for stock market forecasting, *Expert Syst. Appl.* 33 (2007) 171–180.
- [21] R. Hassan, A combination of hidden Markov model and fuzzy model for stock market forecasting, *Neurocomputing* 72 (2009) 3439–3446.
- [22] F. Herrera, Genetic fuzzy systems: status, critical considerations and future directions, *Int. J. Comput. Int. Res.* (2005) 59–67.
- [23] S.-H. Hsu, J.P.-A. Hsieh, T.-C. Chih, K.-C. Hsu, A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression, *Expert Syst. Appl.* 36 (2009) 7947–7951.
- [24] J.-C. Hung, A fuzzy GARCH model applied to stock market scenario using a genetic algorithm, *Expert Syst. Appl.* 36 (2009) 11710–11717.
- [25] M. Khashei, M. Bijaria, G.A. Ardali, Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs), *Neurocomputing* 72 (2009) 956–967.
- [26] A. Keles, M. Kolcak, A. Keles, The adaptive neuro-fuzzy model for forecasting the domestic debt, *Knowl. Based Syst.* 21 (2008) 951–957.
- [27] T. Kohonen, The self-organizing map, *Neurocomputing* 21 (1998) 1–6.
- [28] A. Konar, *Computational Intelligence: Principles, Techniques*, Springer, Berlin, 2005.
- [29] R.J. Kuo, C.H. Chen, Y.C. Hwang, An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network, *Fuzzy Set. Syst.* 118 (2001) 21–45.
- [30] R. Majhi, G. Panda, G. Sahoo, Development and performance evaluation of FLANN based model for forecasting of stock markets, *Expert Syst. Appl.* 36 (2009) 6800–6808.
- [31] P. Mangiameli, S.K. Chen, D. West, A comparison of SOM neural network and hierarchical clustering methods, *Eur. J. Operational Res.* 93 (1996) 402–417.
- [32] D. Nauck, F. Klawonn, R. Kruse, *Foundations of Neuro-Fuzzy Systems*, Wiley, New York, 1997.
- [33] N. O'Connor, M.G. Madden, A neural network approach to predicting stock exchange movements using external factors, *Knowl. Based Syst.* 19 (2006) 371–378.
- [34] R. Rakotomalala, TANAGRA: a free software for research and academic purposes, *EGC'2005, RNTI-E-3 2* (2005) 697–702.
- [35] M. Shim, S. Seong, B. Ko, M. So, Application of evolutionary computation at LG electronics, in: *International Conference on Fuzzy Systems FUZZ-IEEE'99*, Seoul, South Korea, 1999, pp. 1802–1806.
- [36] S.F. Smith, A learning system based on genetic adaptive algorithms, PhD thesis, University of Pittsburgh, Dissertation Abstracts International PhD thesis, University of Pittsburgh, 1980.
- [37] P. Thrift, Fuzzy logic synthesis with genetic algorithms, in: *Fourth International Conference on Genetic Algorithms (ICGA'91)*, San Diego, USA, 1991, pp. 509–513.
- [38] J. Valente de Oliveira, W. Pedrycz, *Advances in Fuzzy Clustering and Its Applications*, John Wiley & Sons, 2007.
- [39] Y.-F. Wang, Predicting stock price using fuzzy grey prediction system, *Expert Syst. Appl.* 22 (2002) 33–39.
- [40] I.D. Wilson, S.D. Paris, J.A. Ware, D.H. Jenkins, Residential property price time series forecasting with neural networks, *Knowl. Based Syst.* 15 (2002) 335–341.
- [41] L. Yang, C.W. Dawson, M.R. Brown, M. Gell, Neural network and GA approaches for dwelling fire occurrence prediction, *Knowl. Based Syst.* 19 (2006) 213–219.
- [42] Z. Yudong, W. Lenan, Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network, *Expert Syst. Appl.* 36 (2009) 8849–8854.
- [43] Y.X. Zhang, Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis, *Talanta* 73 (2007) 68–75.