Information Sciences 424 (2018) 39–54

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Small and multi-peak nonlinear time series forecasting using a hybrid back propagation neural network

Xuefan Dong[a,b], Ying Lian[a,b], Yijun Liu[a,*]

[a] Institutes of Science and Development, CAS, No. 15 ZhongGuanCunBeiYiTiao Alley, Haidian District, Beijing 100190, PR China
[b] University of Chinese Academy of Sciences, No.19A Yuquanlu, Beijing 100049, PR China

ARTICLE INFO

Article history:
Received 23 March 2017
Revised 13 July 2017
Accepted 29 September 2017
Available online 29 September 2017

Keywords:
Public opinion
Nonlinear time series forecasting
Back propagation neural network
Particle swarm optimization
Equal division mechanism
Information entropy

ABSTRACT

Gushes of online public opinions may trigger unexpected incidents that significantly affect social security and stability. Number of posts published per time interval, which is a time series dataset featured with multiple small-scale peaks and nonlinearities, is a simple and direct indicator of how severe the situation is and how much attention has been attracted. Thus, it is of great interest and significance to be able to accurately forecast this type of time series datasets. In this paper, a hybrid Back Propagation Neural network (BPNN) model is proposed to predict the features of this kind of time series datasets. Specifically, a modified Particle Swarm Optimization (PSO) algorithm combined with an Information Entropy (IE) function is used to optimize the weights and thresholds of the network, and the Bayesian Regularization is applied during the training process. Two real online public opinion cases are investigated to verify the effectiveness of the proposed model. Results showed that the proposed model has better performance in accuracy and stability, compared with Levenberg–Marquardt (LM) based BPNN, PSO based BPNN, Bayesian Regularization (BR) based BPNN, Stochastic Gradient Descent (SGD) based BPNN and Least Squares Support Vector Machines (LS-SVM) models.

1. Introduction

In recent years, the internet has become one of the most important channels through which people accessing information and expressing their own attitudes and opinions about certain social events [40]. The online public opinion itself now has become a significant factor which will interfere with the development of social events. For instance, online public opinions have been proved to be an effective tool for the political candidates to win out in the 2016 United States Presidential Election [19]. Real-world social security can also be greatly affected by online public opinions [34]. The "Occupy Wall Street" movement in 2011 was mainly facilitated by Internet social media, among which Twitter played a crucial role in facilitating interactions and opinion sharing among the participants [4]. Rapid spreading speed, open access and high degree of connectivity among netizens make it easy to call on a huge number of supporters to drive a certain social movement [6]. Therefore, a deep understanding of the online public opinions has been a research focus for sociologists and data scientists in recent years. Previous studies on this issue mainly focused on six aspects: event analysis [18], text sentiment classification [41], online social network (OSN) analysis (Kim et al., 2015) [48], opinion leader identification [25,46], rumor spreading [35] and monitoring system [3,21]. All these studies shed light on understanding the online public opinions. However, in most cases,

* Corresponding author.
E-mail addresses: dongxuefan15@mails.ucas.ac.cn (X. Dong), lianying15@mails.ucas.ac.cn (Y. Lian), yijunliu@casipm.ac.cn (Y. Liu).

https://doi.org/10.1016/j.ins.2017.09.067
0020-0255/

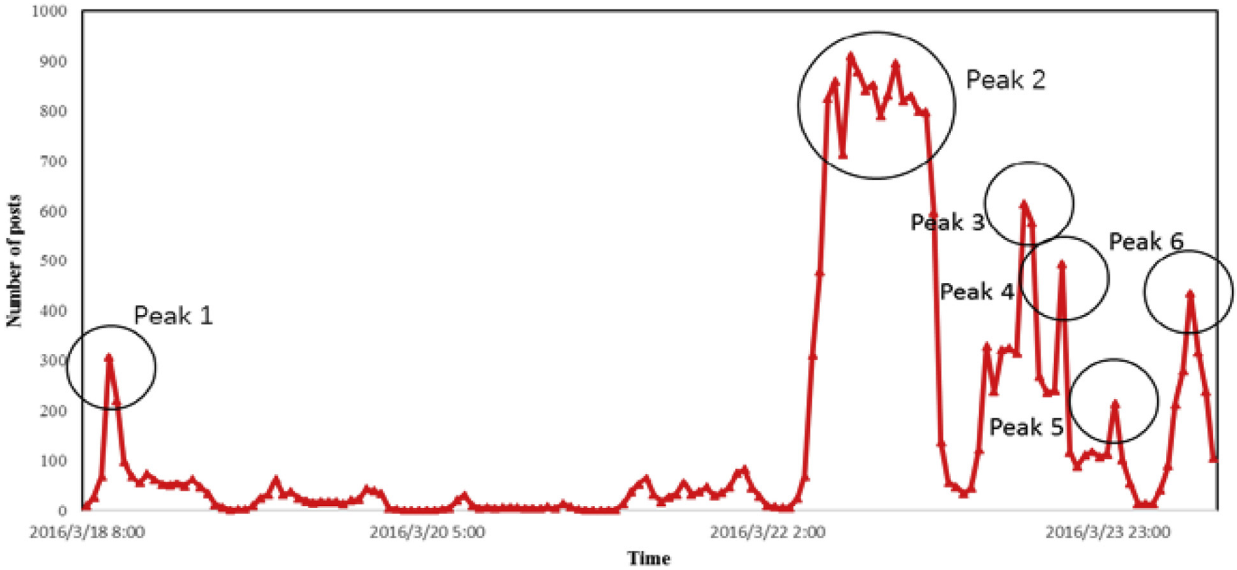


Fig. 1. Changes in number of posts published versus time in a real online public opinion case.
Notes: The online public opinion case shown above was caused by the "Illegal vaccine scandal in Shandong, China", happened on March 18, 2016 (For details, see http://www.bbc.com/news/world-asia-china-35859927*). Data are selected from the Weibo website (China), with time interval one hour. The timeline is from 08:00, March 18, 2016 to 13:00, March 24, 2016. All microblogs hashtagged "illegal vaccine; Shandong Province" are collected as the dataset. It can be noticed that there are six significant peaks in this graph. In addition, compared to other kinds of time series, the duration of this online public opinion case is relatively short.*

no timely guiding suggestions can be offered to the decision makers during the development of the event, especially in social and political areas. In order to formulate interventions in advance, it is curial for decision makers engaged in social and political management to know what is likely to happen as early and precise as possible.

As mentioned above, in most previous study programs related to Internet public opinions, focus had been mainly laid on the contents of those posts so as to know what have happened. In order to access the state more accurately and quickly, some techniques have been proposed. However, before analyzing the contents of those posts, it is necessary to keep an eye on the number of posts firstly as it can provide the most direct and simple information indicating the extent to which the situation is and how much attention has been attracted. In general, in most specific Internet public opinion cases, the social authorities or corporation managers will not pay attention on the contents until related posts have amounted to a quite large number. In addition, like in the case of the Occupy Wall Street movement, social group events tend to be caused only when a considerable number of participants are involved. Moreover, to a large extent, for both social and economic benefits, the earlier an accurate and comprehensive prediction on the number of posts is conducted, the more likely effective decisions on conducting intervention can be made. Indeed, a tremendous amount of attention has been attracted in previous studies [18]. In addition, most real data sets or real events employed in the past literature always involved a great number of participants (in other words, posts are very large in number). Therefore, it is very significant to propose an approach that can be adopted to accurately predict the number of posts related to certain Internet public opinion cases. However, as far as our knowledge is concerned, there is no paper focusing on such a point. A general form for the development trend of the Internet public opinions is shown as Fig. 1, which is featured with three main characteristics: nonlinearity, a small scale, and multiple peaks. It can be easily noticed from Fig. 1 that the data is relatively fewer compared to some other kinds of time series, such as financial time series. The reason lies in that focus is only laid on the number of posts related to the Internet public opinion case while simply ignoring the specific contents of respective posts. Hence, the number of posts at each time point can be set as one data. Moreover, as most specific Internet public opinion cases do not last long, only small data sets can be obtained. In addition, the nonlinear characteristic can be seen from the Fig. 1 as well, so we will not explain it in detail. With respect to the last characteristic, it can be noticed from Fig. 1 that there are six significant peaks. Such a phenomenon is mainly due to the uncertainty of posting behaviors and the random number of participants. Such features may lead the traditional neural network models to easily fall into the local optimum, thus tending to have a negative effect on the accuracy of the forecasted results, to a large extent.

As for methodologies used to forecast nonlinear time series, those models related to the neural network (NN) are commonly applied thanks to the NN's advantages of flexibly nonlinear transfer function and great self-learning ability [36]; thus, their promising applications in various areas, such as economics [27], finance [1], chemistry [29] have been illustrated. It has been revealed in several comparison research programs that compared to other models, neural network models have a generally better accuracy performance in nonlinear time series prediction (see the works of Lee et al. [20] and Yesilnacar and Topal [42]). Moreover, reviews in that area, mainly respecting the studies of Gooijer and Kumar [9] and Zhang et al. [45], have boosted our knowledge of that subject. However, certain inherent weaknesses of NN models have also been pointed out in previous studies. For example, the operation of complex criteria, by choosing the mean square empirical risk as the error function, is required during the training process [16]. In addition, Chau (2007) [49] suggested that the traditional ANN easily entraps into local optimum point. To overcome such limitations, various hybrid approaches have been proposed in the literature. Donate et al. [7] proposed hybrid forecasting methods by getting the neural networks integrated with the genetic algorithm while Chan et al. [2] and Kourentzes et al. [17] combined the Levenberg-Marquardt learning algorithm with neural network models. Moreover, the Stochastic Gradient Descent (SGD) is also a recently popular method for large scale optimization commonly applied as the training function for neural networks [32]. As a modified model of the traditional Gradient Descent (GD), it can to some extent increase the convergence speed, especially when the scale of samples is large [26]. However, due to the variance resulted from random sampling, it has slow convergence asymptotically, especially for searching a more accurate solution as stated in certain studies [[12],[15]]. Additionally, if there are multiple extremes on the error surface, SGD also tends to fall into the local optimum [39]. Additionally, the Support Vector Machine (SVM) method has been widely applied in certain areas related to artificial intelligence [23,24]. Compared to traditional BP neural network models, SVM has its advantage of being the only one required to search the optimal parameters; as a result, it can be more easily implemented [28]. In general, the SVM method is to find a kernel function from the input to output, in which the function is a hidden space putting the input to a high-dimensional feature space. Details about SVM can be found in the work of Sun et al. [33] and Hwang et al. [10]. Furthermore, the particle swarm optimization (PSO) algorithm is also a common option for training the neural networks thanks to its strengths of simple rules and a high searching speed [5,14,43]. Particularly, Kennedy et al. [14] applied the particle swarm optimization algorithm as a training function to optimize the structure and weights of feed forward neural networks. It is a great attempt for the intersection of PSO and neural networks in which details of such a hybrid method are discussed clearly. Das et al. [5] applied the particle swarm optimization algorithm to optimize various variables of neural network method, including but not limited to the number of layers, input and hidden neurons. The simulated results show that the model performs better than all other neural network approaches. Yolcu et al. [43] proposed a novel neural network model and employed the modified particle swarm optimization method as the training function. Empirical results show that the model has a considerable forecasting ability. Overall, a key contribution by all the illustrated studies and unique works mentioned above is to adopt hybrid methods of neural networks in an attempt to predict the nonlinear time series. However, there is no paper in which attention has been focused on the performance test of neural network models in forecasting the number of times that Internet public opinions are posted, a dataset featured with three characteristics: nonlinearity, a small scale and multiple peaks.

Therefore, in this paper, a hybrid Back Propagation Neural Network (BPNN) model is proposed to forecast the possible time series dataset of number of posts published in an online public opinion event. This model can be divided into three stages. In stage I, each one of the time series datasets collected is divided into two halves. While the first half is fed as the input to the model, the second half is used as the validation dataset to test the output predicted by the model. To calculate the initial connection weights and thresholds of the input, a modified PSO algorithm that has an equal division mechanism integrated with the Information Entropy (IE) function is applied. In stage II, the BPNN is designed with Bayesian Regularization to figure out the inherent pattern in the input weight set. In stage III, through the trained neural network, the forecasted time series dataset can be generated. In addition, two more real online public opinion cases are introduced, and the performance of this proposed model is compared with that of several other commonly used approaches.

The remainder of this paper is organized as follows: In Section 2, some basic concepts will be briefly reviewed and demonstrated. In Section 3, full method of the proposed model is presented. Results are presented and discussed in Section 4. In Section 5, a conclusion is drawn from the current study.

2. Brief review of basic concepts

2.1. Back propagation neural network (BPNN)

During the past decade, the BPNN has been cited as the most commonly used method for forecasting [38]. BPNN is a feedback network firstly introduced by Rumelhart et al. [30]. A general structure of BPNN model is shown as Fig. 2, which consists of three layers: the input, hidden and output layers, among which the hidden layer conveys important information between the input layer and the output layer. Particularly, the BPNN always consists of one or more hidden layers, thus allowing the networks to model complex functions. It is composed of two main processes: the dissemination of positive information and error back-propagation. According to Wang et al. [38], the mathematical relationships between those three layers can be formulated as follows:

Input layer-hidden layer:

$$y_j = f_I\left(\mu_j + \sum_{m=t-n}^{t-1} \mu_{jm} y_m\right)(0 \le \mu_j, \mu_{jm} \le 1) \tag{1}$$

Hidden layer-output layer:

$$y_t = f_o\left(\lambda_o + \sum_{j=1}^{I} \lambda_{oj} y_j\right)(0 \le \lambda_o, \lambda_{oj} \le 1) \tag{2}$$

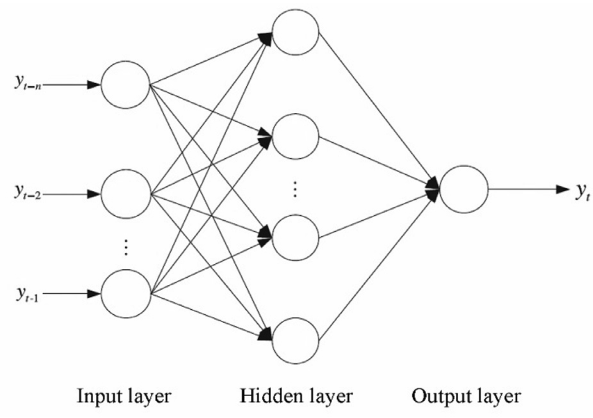


Fig. 2. The structure of BPNN.

where: y_m and y_j represent the input from the input layer and the hidden layer respectively; y_t represents the predicted value of point t; μ_{jm} and λ_{oj} are the network weights for the output and hidden layers; μ_j and λ_o are the threshold values of the hidden and output layers; n and I are the numbers of the nodes of the input layer and the hidden layer; f_I and f_o denote the activation functions of the hidden and output layers respectively. In most cases, the logistic and hyperbolic functions are commonly used as the hidden layer activation functions f_I while the linear function is frequently applied as the output layer activation function f_o [38].

2.2. Particle swarm optimization (PSO)

The particle swarm optimization (PSO) algorithm was firstly proposed by Kennedy and Eberhart [13] to imitate the preying behavior of birds. PSO is an interaction-based optimization algorithm similar to the genetic algorithm (GA). The system of PSO is initialized to a set of random solutions, named particles, and the optimal solution is achieved by searching particles with the iteration in the solution space. Due to the fact that the PSO algorithm has such advantages as a simple principle, a high searching speed, and implementation simplicity. However, some studies also suggested that PSO is easy to trap in local optimum for multi-peak search problems [8]. In order to overcome such a weakness, the initial number of population always needs to be increased. The dynamical rules of PSO are shown as in formula (3).

$$\begin{aligned} V_{it+1} &= V_{it} + c_1 \times Rand() \times (P_{it\,best} - X_{it}) + c_2 \times Rand() \times (G_{tbest} - X_{it}) \\ X_{it+1} &= X_{it} + V_{it}(i = 1, 2, \ldots, N) \end{aligned} \tag{3}$$

Where: N is the initial number of particles, V_{it} measures the velocity of the ith particle at the point of t, X_{it} measures the position information of the ith particle at Point t, P_{itbest} measures the position information the ith particle's best individual performance, G_{tbest} represents the position information of the best particle, *Rand* () is the random values from the interval [0,1], and c_1 and c_2 are two real numbers, with the range of $(0, \infty)$.

2.3. Information entropy (IE)

The information entropy (IE) is mainly used to measure the confusion degree in terms of probability distribution for information. In the past decade, the IE has been applied in various areas, such as industry [22], mathematics [44], finance [47], etc. A general form of IE can be formulated as in (4).

$$S = -\sum_{i=1}^{N} p(x_i)\log_2(p(x_i)) \tag{4}$$

Where: S measures the entropy value, $p(x_i)$ is the probability of the ith situation ($\sum_{i=1}^{N} p(x_i) = 1, 0 < p(x_i) < 1$). In addition, the information entropy has the following characteristics: (1) Non-negativity: $S \geq 0$; (2) Extremum: the IE reaches the maximum when $p(x_i) = 1/N$; (3) Concavity: the IE is a symmetric concave function; (4) Additivity: For independent situations $p(x_1)$, $p(x_2)$,..., $p(x_n)$, $S(\sum_{i=1}^{N} p(x_i)) = \sum_{i=1}^{n} S(p(x_i))$.

3. Method

3.1. The modified PSO

As mentioned above, in order to prevent trapping local optimum within the traditional PSO algorithm, the particles need to be increased in number. However, the algorithm will become less efficient but more complex in such a practice.

In addition, due to the fact that the time series of Internet public opinions posting number always have several peaks, the traditional PSO algorithm cannot be applied to do forecasting effectively. That problem can be solved by use of an equal division mechanism. In that mechanism, the solution space is divided into several sub-spaces in which a number of particles and an equal number of iterations are initialized. The specific algorithm processing steps are shown as follows:

Step 1: The solution space U is equally divided into some sub-spaces: u_1, u_2,..., u_n, and the number of iterations is set as T.
Step 1: Initialize particles with the number N in each sub-space where the velocity and position of each particle are set as V_n and X_n respectively.
Step 3: Calculate the fitness value of each particle and find the current global extremum of each sub-space, which is set as G_{besty}.
Step 4: Update the velocity and position of each particle in each sub-space based on the rule of traditional PSO (Eq. 3).
Step 5: Calculate the personal extremum of each particle p_{bestn} and the global extremum G_{bestn} in each sub-space.

So far as the IE used in the present paper is concerned, the traditional IE is transformed as

$$S(t) = -\sum_{i=1}^{N} p(x_{it})\log_2(p(x_{it}))$$
$$p(x_{it}) = \frac{h_{it}}{\sum\limits_{i=1}^{N} h_{it}} \tag{5}$$

where: $S(t)$ measures the entropy value in the tth generation, $p(x_{it})$ is the weight of the ith particle in the tth generation ($\sum_{i=1}^{N} p(x_{it}) = 1, 0 < p(x_{it}) < 1$), and h_{it} represents the fitness value of the ith particle in the tth generation. A larger $S(t)$ indicates a more significant similarity within the fitness values of particles while a smaller $S(t)$ indicates a more dispersed situation. Moreover, in order to maintain a high degree of efficiency during the whole process of the algorithm, the value of IE should be continuously kept in different appropriate intervals within different stages. In particular, the IE should be relatively higher so as to prevent trapping into local optimum during the early stages while during the latter stages, the IE should be relatively lower so as to prevent the dispersive distribution of particles, which may lead to the failure of gaining the optimal solution. Therefore, a time-varying adjustment mechanism is proposed in the present paper, which is shown as follows:

Step 1: Define the interval of IE as $[S(t)_{\min}, S(t)_{\max}]$.

$$S(t)_{\max} = \begin{cases} a + \alpha_1 t, & \alpha_1 t \le 1 \\ 1, & \alpha_1 t > 1 \end{cases}$$
$$S(t)_{\min} = \alpha_2 t \tag{6}$$

Where: a, α_1, α_2 are coefficient constants ($0 < a$, α_1, $\alpha_2 < 1$), which is to be used to determine the ideal interval of IE.
Step 2: If the calculated IE is too large or too small, indicating a non-ideal state of the distribution of particles' fitness values, the velocity of particles will be adjusted as

$$V_{it+1} = C \times V_{it} + c_1 \times Rand() \times (P_{itbest} - X_i) + c_2 \times Rand() \times (G_{tbest} - X_{it}) \tag{7}$$

Where: C is the inertia weight coefficient, calculated as:

$$C = \begin{cases} 1 + \beta_1[S(t) - S(t)_{\max}], & S(t) > S(t)_{\max} \\ 1, & S(t)_{\min} \le S(t) \le S(t)_{\max} \\ 1 - \beta_2[S(t) - S(t)_{\max}], & S(t) < S(t)_{\min} \end{cases} \tag{8}$$

where: β_1 and β_2 are two coefficient constants.
In addition, in the present study the fitness function of the modified PSO algorithm is the average mean squared error between expected and predicted values.

3.2. The modified PSO-BR-BPNN

As discussed above, the modified PSO algorithm is featured with such advantages as global convergence ability, preventing any trapping into local optimum, high efficiency in gaining optimal solutions, and easy implementation. Therefore, to utilize the modified PSO for BPNN seems to be an appropriate attempt in the nonlinear time series forecasting of the posting number of Internet public opinions.

Particularly, Wang et al. [38] put forward an excellent idea in order to make it possible for the traditional BPNN to forecast more accurately by applying the adaptive differential evolution (ADE) algorithm to select the initial weights and thresholds and using the Levenberg–Marquardt (LM) method as the training function. Such a technique will be adopted in the present paper. To be specific, the initial weight and threshold value between the input layer and the hidden layer, and the initial weight and threshold value between the hidden layer and the output layer are updated by use of the velocity and position features of particles within the modified PSO. In addition, the Bayesian Regularization is set as a training function for the BPNN in the present paper. Its excellent ability in coping with complexity issues makes it possible for the created networks to have a great performance in generalizing and predicting the nonlinear time series featured with such characteristics as a small scale and multiple peaks. The Bayesian Regularization can be regarded as a penalty term aimed to limit the interval for the connection weights of the networks. Such a method can be implemented by using the 'trainbr' function of the Matlab, and the details can be seen from the works of Kocadağlı and Aşıkgil [16] and Tipping and Lawrence [37], who have deeply discussed the mechanism of Bayesian learning, and Faundez-Zanuy (2010) [50], who has proposed a good demonstration of the integration of the BPNN and the Bayesian Regularization. Moreover, the Bayesian Regularization function requires either the mean squared error (MSE) or the sum of squared error (SSE) as the fitness function. The former is chosen in the present paper, shown as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(output_i - target_i)^2 \tag{9}$$

Where N is the length of learning sample, *output* and *target* measure the predicted values and the actual values respectively.

The specific processing steps of the Modified PSO-BR-BPNN is presented as follows:

Step 1: Divide the solution space U equally into several sub-spaces u_1, u_2,..., u_n, and set the number of iterations as T. Specify the lower and the upper values for the position and velocity of the particles.
Step 2: Initialize particles with the number N in each sub-space, in which the ith particle in jth sub-space is represented by $particle_{ij}$ ($i=1,2,\ldots,N;\ j=1,2,\ldots,n$). The velocity and position for the ith particle within the jth sub-space are randomly set as V_{ij} and X_{ij}.
Step 3: Calculate the fitness value of each particle and find the current global extremum of each sub-space, which is to be represented as P_{bestij} and $G^t{}_{bestj}$ respectively ($t=1,2,\ldots,T$).
Step 4: Compare the global optimums of each sub-space G_{bestj}, and gain the optimal solution G_{best}.
Step 5: Judge whether the G_{best} reaches the accuracy requirement λ or the number of iterations reaches the maximal number. If *yes*, carry out Step 9; otherwise, carry out Step 6.
Step 6: Update the velocity and position for the particles in each sub-space based on Eq. (3). Set $T=T+1$.
Step 7: Calculate the IE value of the particles $S(t)$ based on Eq. (5), and judge whether the IE value is involved in the ideal interval of $[S(t)_{\min}, S(t)_{\max}]$. If *yes*, carry out Step 3; otherwise, carry out Step 8.
Step 8: Update the velocity and position of the particles in each sub-space based on Eqs. (7) and (8). Carry out Step 7.
Step 9: Set the optimal solution G_{best} as the initial weights and thresholds of BPNN.
Step 10: Train the network with the Bayesian Regularization and then create the best-fitting network for making predictions. Then, the final combined forecasts can be achieved.

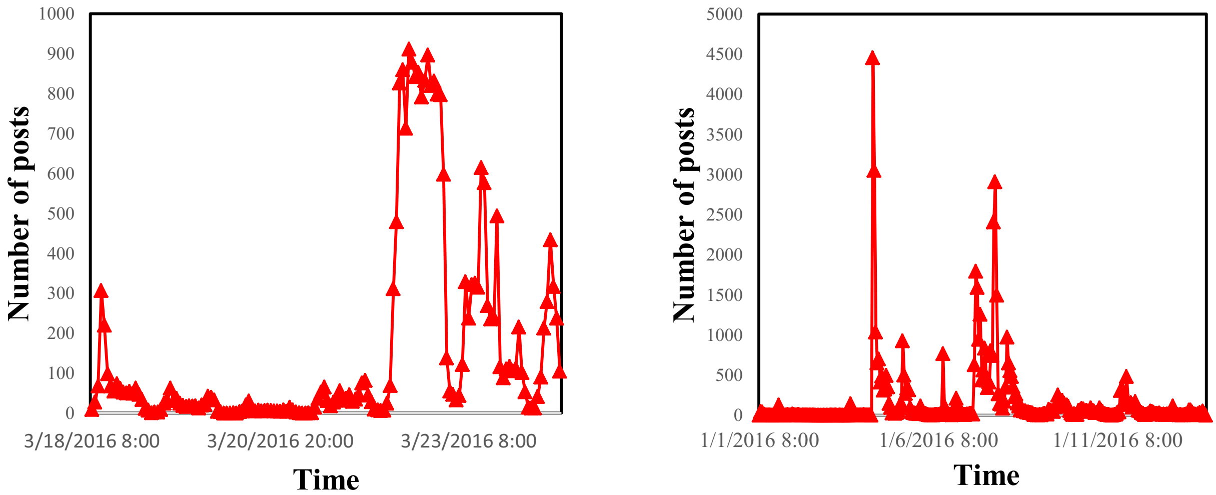
In the present paper, each particle corresponds to the initial connection weights and thresholds of BPNN. The solution space is within the range of [−1, 1]. The training set $Y_{tj}=[y_1,y_2,\ldots,y_{Ntj}]^{\mathrm{T}}$ is set as the input, where Y_{tj} is the set of past observations and N_{tj} is the size of training set. The BPNN learning process includes two phases: In Phase I, the modified PSO algorithm is used to calculate out the optimal or approximate optimal connection weights and thresholds of the network; and then in Phase II, the backpropagation learning rule and the Bayesian Regularization is to be employed to make adjustments to the final weights and thresholds. Finally, the combined forecasts $Y' = [y'_{Ntj+1}, y'_{Ntj+2},\ldots,y'_{Ntj+Nto}]^{\mathrm{T}}$ can be obtained, which are set as the output, where N_{to} is the size of testing set.

4. Experiment results

In the present paper, two Internet public opinion cases that happened recently are considered to testify the feasibility and effectiveness of the proposed model. The reasons for choosing those two cases are as follows: (1) both those two cases are recent events (both of which occurred in 2016), thus having ensured the timeliness of the data; (2) the changes of the number of posts in both those two cases are featured with a small scale, nonlinearity and multiple peaks, thus having guaranteed the consistency of the data; (3) although showing similar characteristics, the development of posts number in two cases is totally different in shape, thus having guaranteed the diversification of the data and made the results more convincing.

4.1. Data description

Case 1 is as shown in Fig. 1, and it was incurred by the "Illegal vaccine scandal in Shandong, China", which happened in March 2016. The data are excerpted from the Weibo website (China). Specifically, by using the web crawler tool, all microblogs hashtagged "illegal vaccine; Shandong Province" are collected as the dataset. The reason for choosing the Weibo website as our data source is due to its significant role in the human social interaction domain (according to the website's earnings during the second quarter of 2016, the website's active users have amounted to approximately 2.82 billion people). The time interval of the data is 150 hours (from 08:00, March 18, 2016 to 13:00, March 24, 2016), and the number



a. Development process of the Internet public opinion Case 1

b. Development process of the Internet public opinion Case 2

Fig. 3. Development process of the Internet public opinion cases.

of posts is 23,012. Case 2 was caused by the failure of the Circuit-Breaker mechanism policy authorized by China Securities Regulatory Commission (CSRC) in the Chinese stock market, which happened in January 2016. The data are excerpted from the Weibo website (China). The microblogs hashtagged "circuit breaker mechanism" and "circuit breakers" were collected as the dataset. The time interval of the data is 300 hours (from 08:00, January 1, 2016 to 20:00, January 13, 2016), and the number of posts is 28,445. The hashtags used to collect posts are some of the most related topic-specific tokens for each case. Having tried several combinations of different hashtags, we finally chose the one with the most posts. After the data collection process, we count the number of posts at each time point for each case. Then, two different time series have been obtained as shown in Fig. 3a and b respectively.

In the present paper, five popular models, LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, and LS-SVM, that were commonly applied to forecast nonlinear time series from previous studies are chosen to compare with the proposed model in the application of Internet public opinion posting number predictions. In specific, the training functions of LM-BPNN and PSO-BPNN models are both set as the Levenberg–Marquardt. The training function of SGD-BPNN is set as the Stochastic Gradient Descent. In addition, within the PSO-BPNN model, the traditional PSO algorithm is used to optimize the initial weights and thresholds. In addition, the kernel function of LS-SVM is set as the RBF function. Besides, the training function of BR-BPNN model is set as the Bayesian Regularization.

When conducting the analysis, each posting number dataset was divided into three subsets, namely, the training set, the validation set and the test set. The former two sets are used to train the network and control the model complexity respectively while the performances were evaluated over the test data. However, there is no theoretical guideline for determining the exact percentage of each set. In the present paper, the ratio of 60%:10%:30% for the training set, the validation set and the test set is applied for LM-BPNN, PSO-BPNN, and SGD-BPNN while the ratio of 70%:30% for the training set and the test set is applied for BR-BPNN, LS-SVM, and the model we proposed. The reason for choosing such a ratio (the percentage for the training set is relatively smaller than in other relevant studies) lies in the realistic demand for Internet public opinion supervision, that is, maximum predictions with least time and data. It was suggested in certain papers that a larger magnitude of the test set may have negative effects on the forecasting quality of all models [16]; however, with a similar percentage of test set implemented in every model of each case study, no unfair comparison analysis will be resulted in. In addition, according to Saini [31], the Bayesian framework allows the parameter to be selected by using only the training data. In other words, there is no requirement on the use of separated training and validation data. That is why the validation data set was taken into consideration within the BR-BPNN and the proposed method.

4.2. Performance assessments

In the present paper, following common practice [16,38], several methods are employed to measure the accuracy of the models, including the root mean squared error (RMSE), the mean absolute percentage error(MAPE), and the mean absolute

Table 1
The parameters setting for the proposed method.

Parameter	Value	Parameter	Value
Number of sub solution spaces	4	λ	0.19 [38]
Number of particles in each sub-space	20	Number of hidden layers	1
Number of iterations	200	Number of output layers	1
c_1 and c_2	2	Number of hidden neurons	4 [38]
w	0.4	Number of output neurons	1
a	0.1	Number of delays	2
α_1	0.15	Range of connection weights and thresholds	[−1, 1]
α_2	0.4	Activation function of the hidden layer	*logsig*
β_1	1.3	Activation function of the output layer	*purelin*
β_2	0.7		

Table 2
Performance comparison of the MPSO-BR-BPNN with other forecasting models.

		Case 1			Case 2		
Model		RMSE	MAE	MAPE	RMSE	MAE	MAPE
LM-BPNN	Training	123.18	91.47	**50.57%**	115.26	121.38	367.75%
	Test	133.46	93.15	**67.96%**	119.33	79.40	607.18%
PSO-BPNN	Training	114.96	94.40	90.13%	137.19	99.69	177.69%
	Test	141.21	87.16	155.00%	107.52	76.64	173.54%
BR-BPNN	Training	112.81	78.49	83.62%	102.64	82.86	105.21%
	Test	121.36	73.88	103.94%	113.98	63.77	187.05%
SGD-BPNN	Training	114.17	85.37	104.77%	116.49	85.63	94.23%
	Test	113.12	88.83	83.27%	100.50	62.42	109.77%
LS-SVM	Training	110.66	80.80	101.37%	96.65	58.90	103.28%
	Test	120.92	87.10	93.25%	98.72	60.25	99.63%
MPSO-BR-BPNN	Training	**105.81**	**74.17**	73.56%	**90.03**	**54.70**	**92.99%**
	Test	**90.86**	**60.48**	77.90%	**68.33**	**36.63**	**80.14%**

error (MAE). The expressions for RMSE, MAPE and MSE are respectively illustrated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (output_i - target_i)^2}{N}} \tag{10}$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{output_i - target_i}{target_i} \right| \tag{11}$$

$$MAE = \frac{\sum_{i=1}^{N} |output_i - target_i|}{N} \tag{12}$$

where $output_i$ and $target_i$ are the predicted and actual observations, respectively and N is the total number of observations.

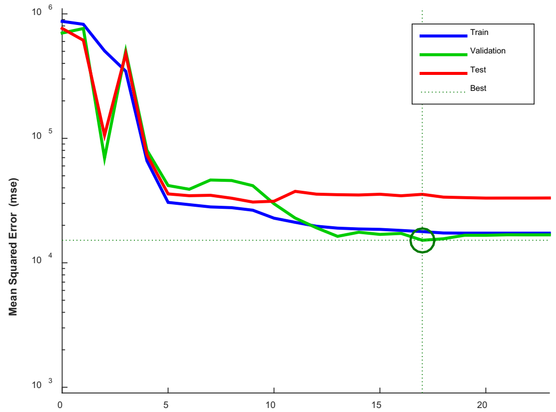
4.3. Parameter settings

There are some parameters for the proposed method to be specified, shown as Table 1.

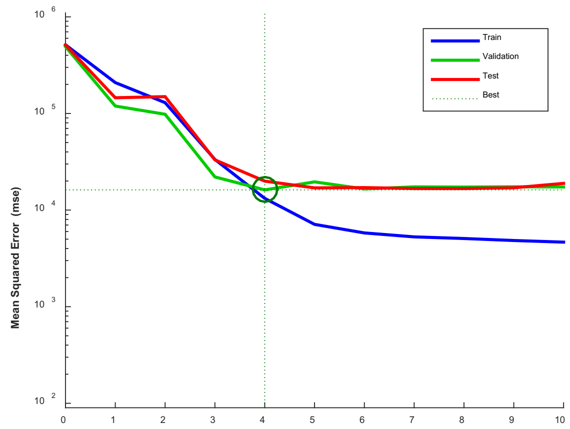
In addition, to set the number of hidden neurons as four (based on the work of Wang et al. [38]) were not meant to be optimal for our research. However, according to Jin et al. [11], the performance results will not be affected because only the weights and thresholds of BPNN are optimized. Moreover, within the RBF function of the LS-SVM method, following previous practice [33], the regularization constant and the bandwidth are set as 500 and 0.5 respectively (13).

4.4. Results analysis

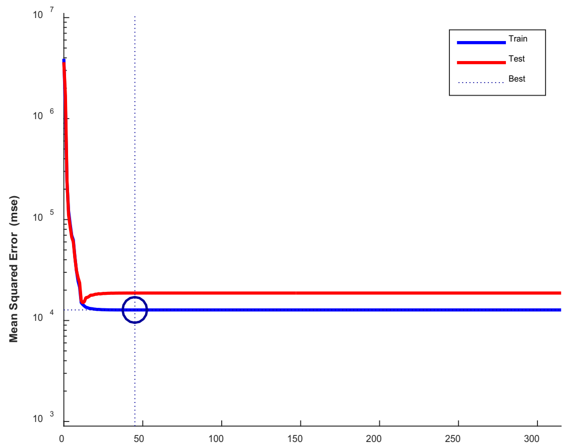
With each model run for 50 times, the best ones were taken as the final solutions. An overview of performance evaluation results by considering RMSE, MAPE and MAE is shown as in Table 2, and it is indicated that the proposed MPSO-BR-BPNN model can be adopted to provide significantly better forecasts than adopting any other model in forecasting small-scale multi-peak nonlinear time series. Specifically, the MPSO-BR-BPNN model exhibits the best in terms of RMSE and MAE for both training and test sets of all cases (with respect to Case 1, RMSE for training sets and that for test sets are 105.81 and 90.86 respectively while MAE for training sets and that for test sets are 74.17 and 60.48, respectively; with respect to Case 2, RMSE for training sets and that for test sets are 90.03 and 68.33 respectively while MAE for training sets and that for test sets are 54.70 and 36.63). In addition, in terms of MAPE, compared with LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN



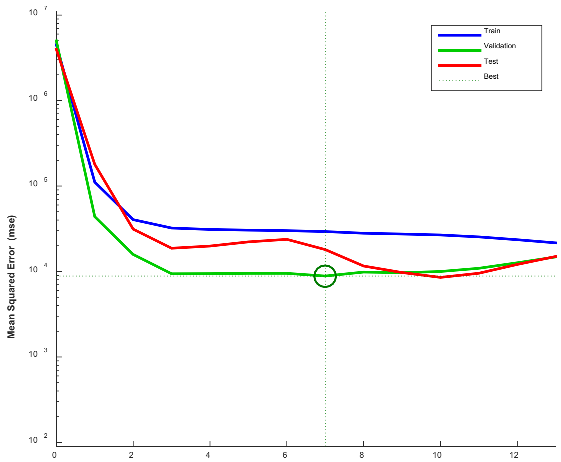
a. the training process of LM-BPNN for Case 1 (17 epochs)



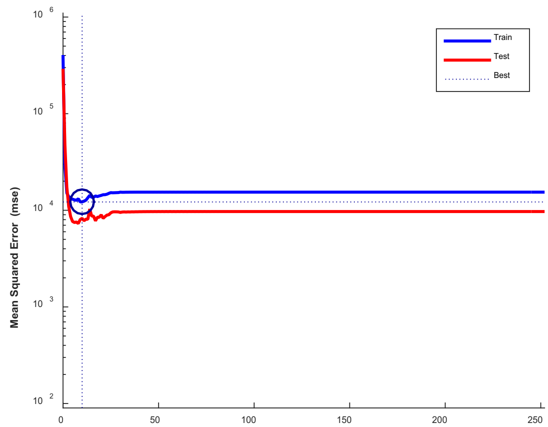
b. the training process of PSO-BPNN for Case 1 (4 epochs)



c. the training process of BR-BPNN for Case 1 (45 epochs)

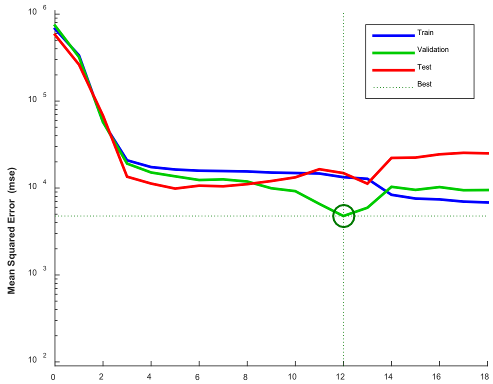


d. the training process of SGD-BPNN for Case 1 (7 epochs)

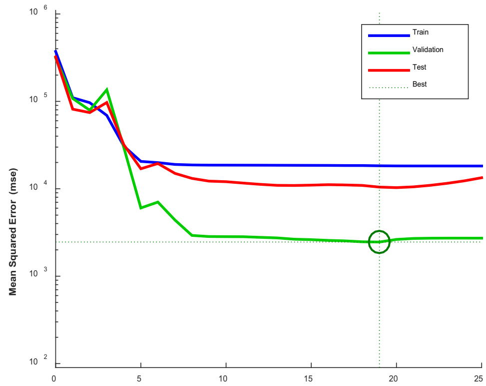


e. the training process of MPSO-BR-BPNN for Case 1 (10 epochs)

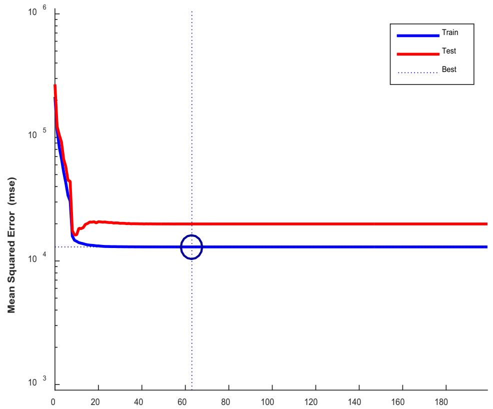
Fig. 4. The training processes of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, and MPSO-BR-BPNN for Case 1.



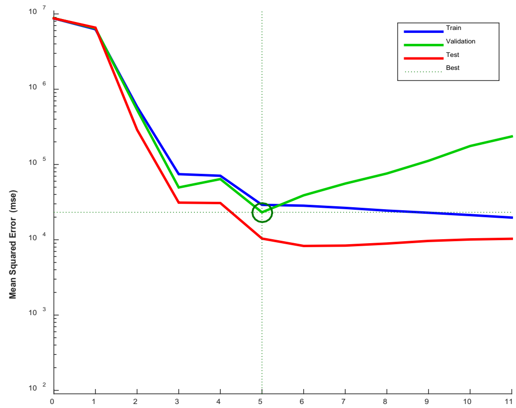
a. the training process of LM-BPNN for Case 2 (12 epochs)



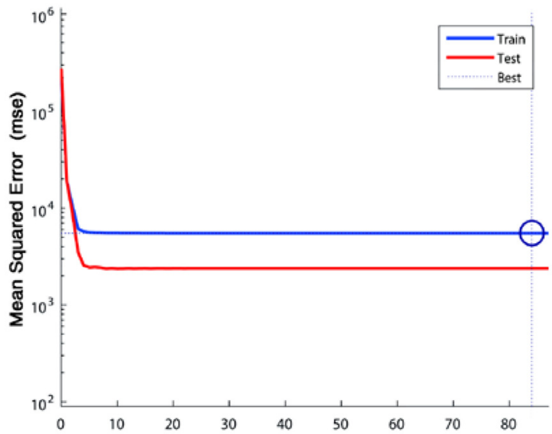
b. the training process of PSO-BPNN for Case 2 (19 epochs)



c. the training process of BR-BPNN for Case 2 (62 epochs)

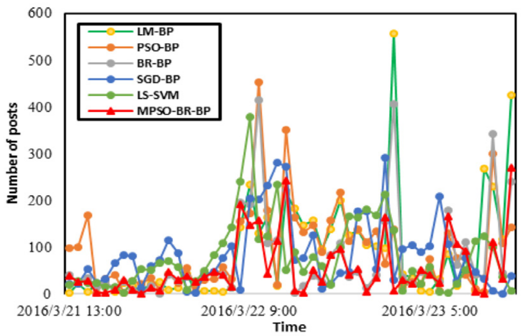


d. the training process of SGD-BPNN for Case 2 (5 epochs)

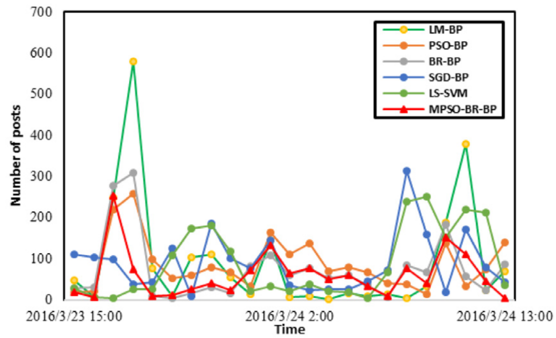


e. the training process of MPSO-BR-BPNN for Case 2 (84 epochs)

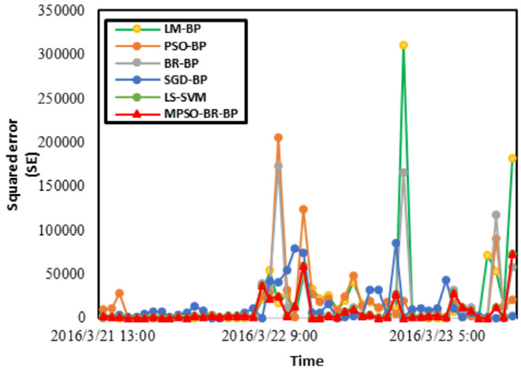
Fig. 5. The training processes of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, and MPSO-BR-BPNN for Case 2.



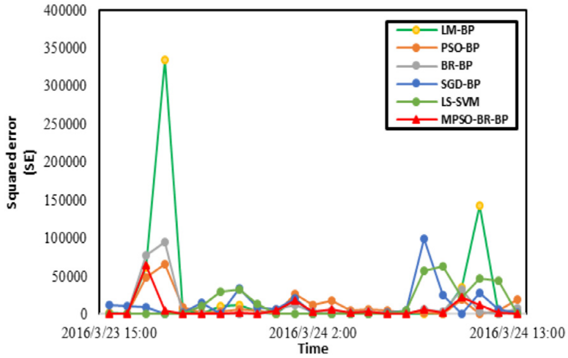
a. Results from comparisons in terms of AE for Case 1 (training and validation sets)



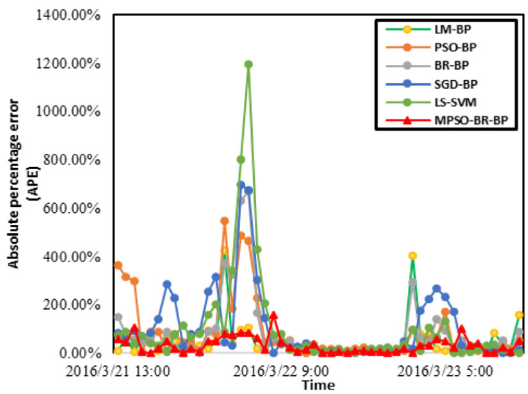
b. Results from comparisons in terms of AE for Case 1 (test set)



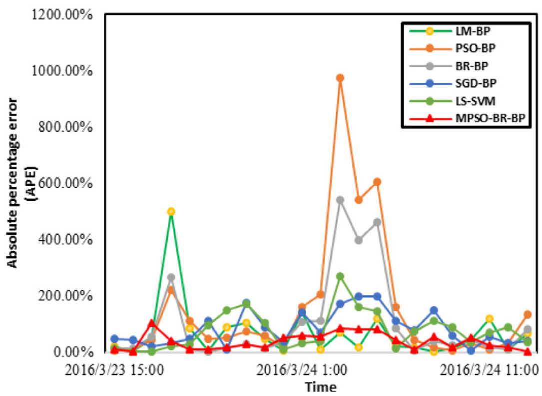
c. Results from comparisons in terms of SE for Case 1 (training and validation sets)



d. Results from comparisons in terms of SE for Case 1 (test set)

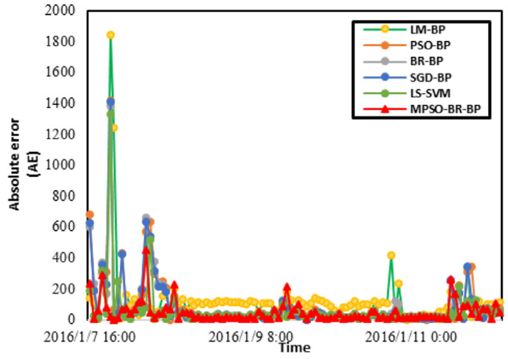


e. Results from comparisons in terms of APE for Case 1 (training and validation sets)

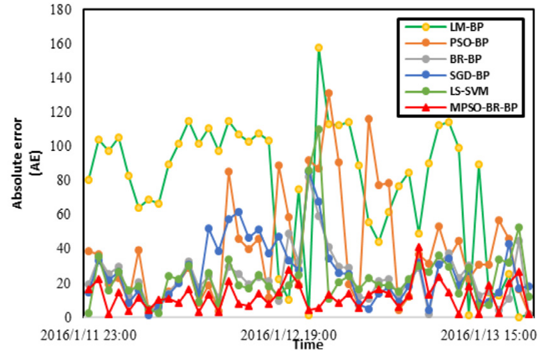


f. Results from comparisons in terms of APE for Case 1 (test set)

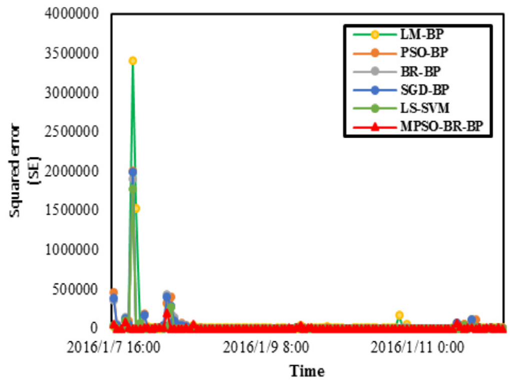
Fig. 6. Results from comparisons between LM-BPNN, PSO-BPNN, BR-BPNN SGD-BPNN, LS-SVM, and MPSO-BR-BPNN for Case 1.



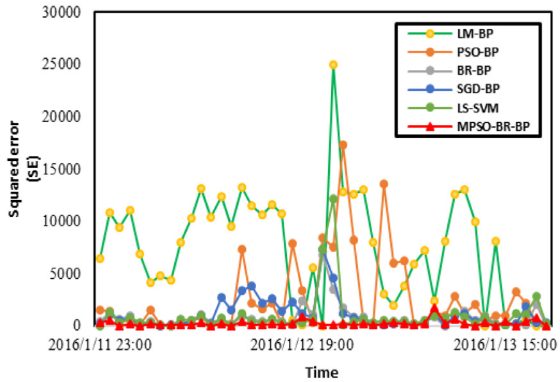
a. Results from comparisons in terms of AE for Case 2 (training and validation sets)



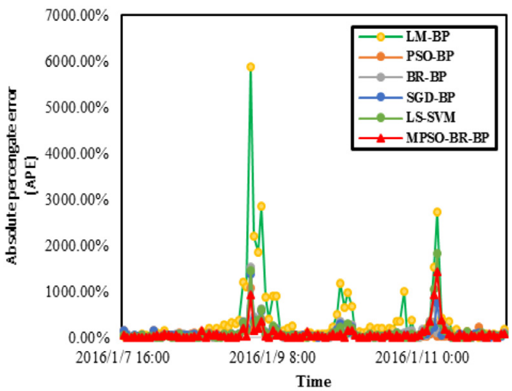
b. Results from comparisons in terms of AE for Case 2 (test set)



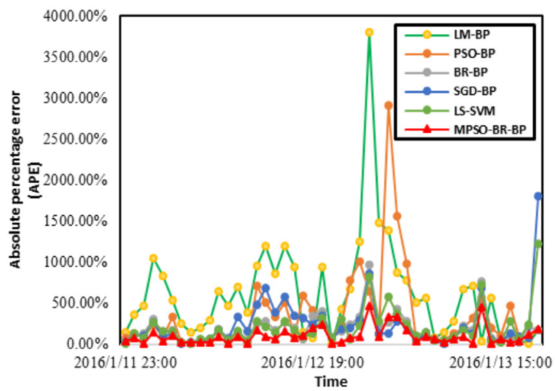
c. Results from comparisons in terms of SE for Case 2 (training and validation sets)



d. Results from comparisons in terms of SE for Case 2 (Test set)

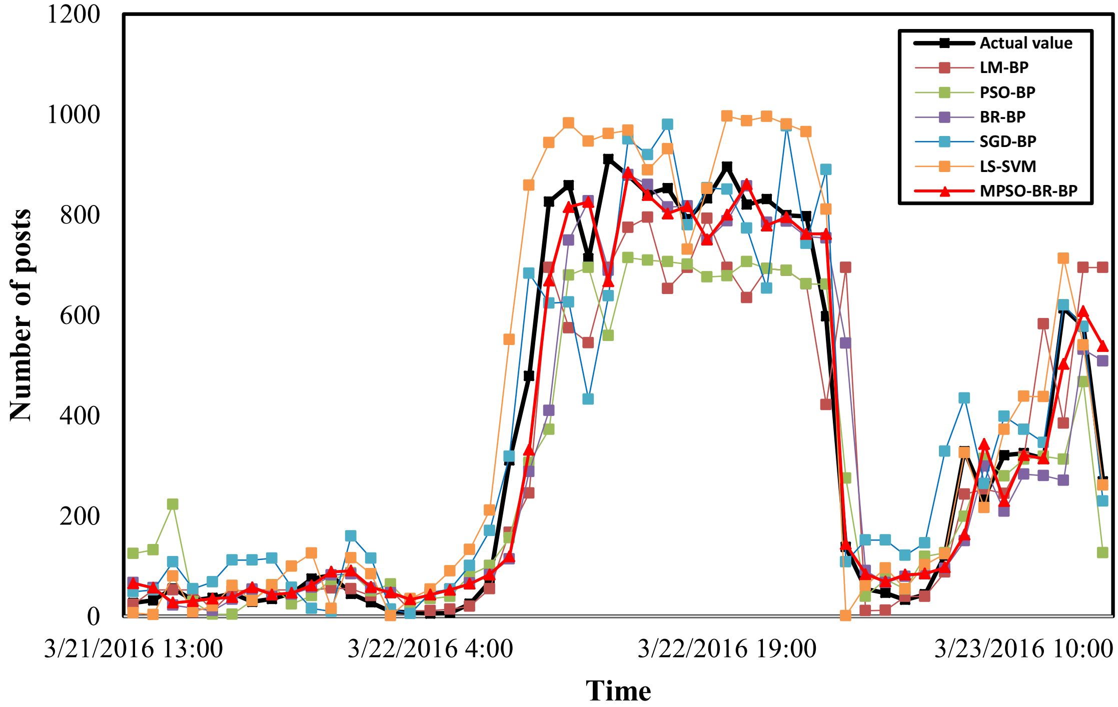


e. Results from comparisons in terms of APE for Case 2 (training and validation sets)

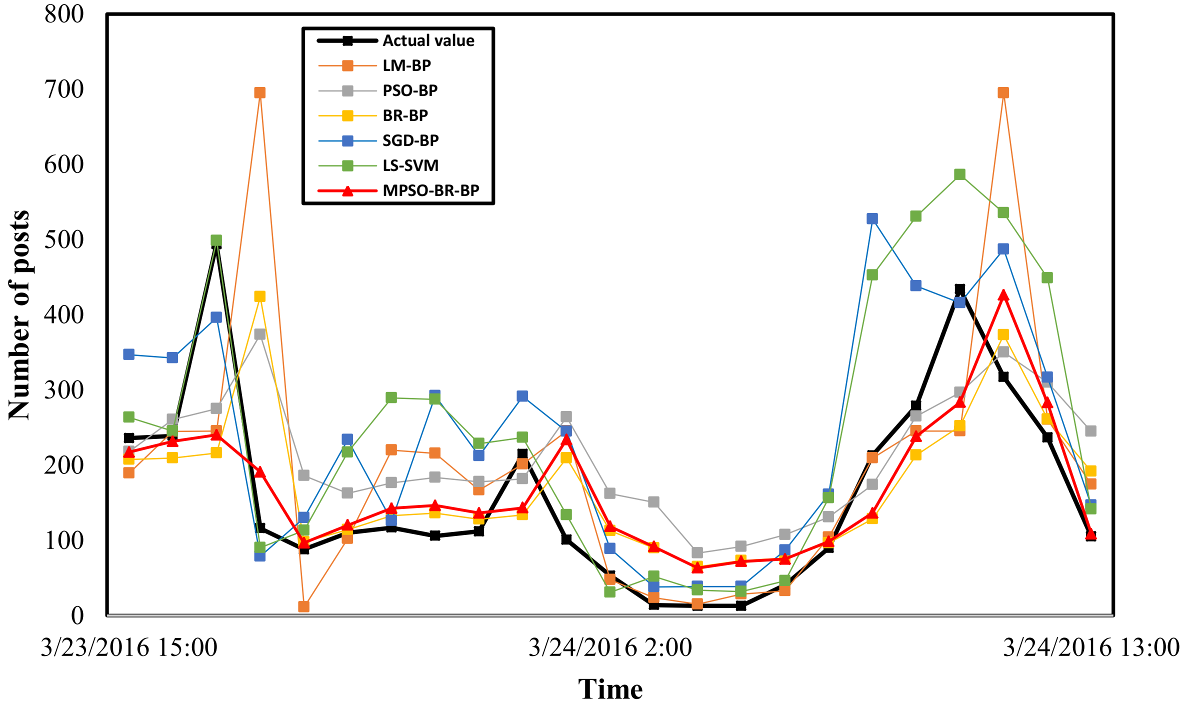


f. Results from comparisons in terms of APE for Case 2 (test set)

Fig. 7. Results from comparisons between LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM, and MPSO-BR-BPNN for Case 2.

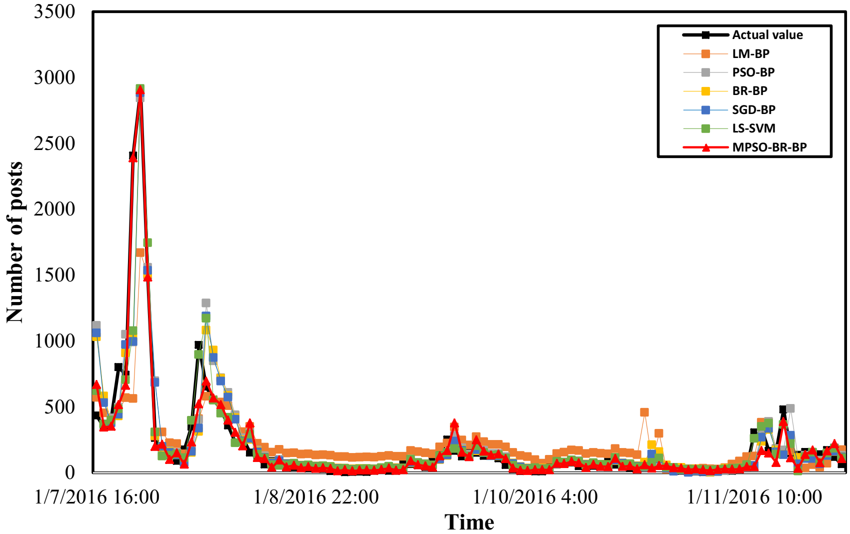


a. Fitting to training and validation data of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM and MPSO-BR-BPNN for Case 1

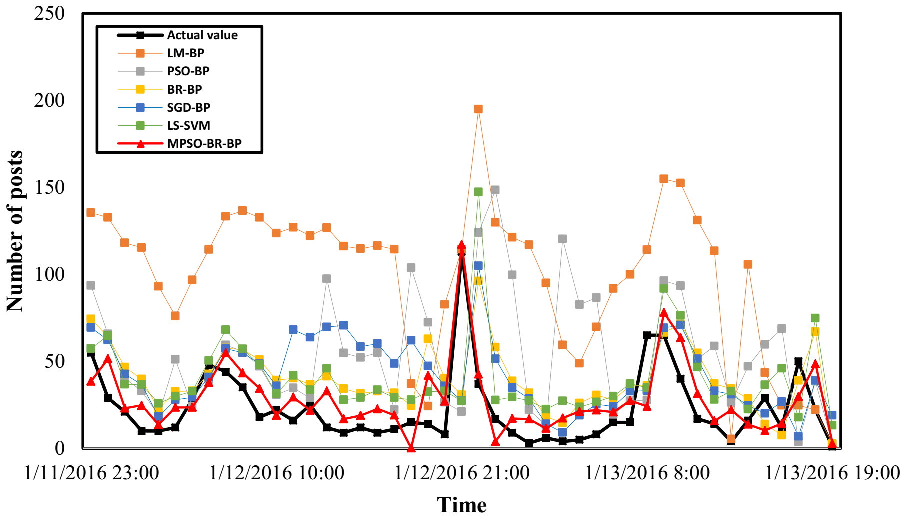


b. Fitting to test data of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM and MPSO-BR-BPNN for Case 1

Fig. 8. Fitting to training, validation and test data of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM and MPSO-BR-BPNN for Case 1.



a. Fitting to training and validation data of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM and MPSO-BR-BPNN for Case 2



b. Fitting to test data of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM and MPSO-BR-BPNN for Case 2

Fig. 9. Fitting to training, validation and test data of LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM and MPSO-BR-BPNN for Case 2.

and LS-SVM models, the proposed model performs the best for both training sets and test sets of Case 2 with the values of 92.99% and 80.14%, respectively. Except the proposed model, it can be also observed from Table 2 that the performance of the BR-BPNN model is relatively better as well. This proves the excellent ability of Bayesian Regularization in training networks for forecasting small and noisy nonlinear time series. Moreover, such an important finding has also empirically justified the curial role of the proposed modified PSO algorithm in the optimization of BPNN initial weights and thresholds.

The training processes of LM-BPNNM, PSO-BPNN, BR-BPNN, SGD-BPNN, and MPSO-BR-BPNN models for Case 1 and Case 2 are shown as Figs. 4 and 5, respectively. In addition, all comparison results of each single assessment at each time node for Case 1 and Case 2 can be classified into two parts: the training and validation sets (or the training set) and the test set, as respectively displayed in Figs. 6 and 7. It can be noticed that with the higher-order accuracy performance improved, the proposed model also shows a better numerical stability when being applied to forecasting small- and multi-peak nonlinear time series than other three models. Moreover, all fitting curve results obtained by adopting LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN, LS-SVM and the proposed MPSO-BR-BPNN models for Case 1 and Case 2 can be classified into two parts: the training and validation sets (or the training set) and the test set as shown in Figs. 8 and 9, respectively. It can be found out that all models can be effectively used to predict a rough description of the trend, to a large extent. However, LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN and LS-SVM are of a hysteretic nature, especially in any situation where there are fewer training data (Case 1). In comparison, the MPSO-BR-BPNN model shows a surprising accuracy strength in forecasting global peak values. In addition, it also has a better performance in capturing other wave peaks within the time series. The main reason for those two kinds of strengths lies in that the equal division and information entropy techniques and the partly contribution of Bayesian Regularization mechanism are added.

Actually, such two advantages are really important for predicting the number of times Internet public opinions are posted because in general, a more accurate forecast of the peak value and a better ability of identifying other wave peaks could always allow the decision makers to take more effective intervention measures for purpose of relieving the irritation or fear of the public and discover potential secondary Internet public opinions so as to prevent any potential social threat.

5. Conclusion

In this paper, a hybrid BPNN model (MPSO-BR-BPNN model) has been proposed to forecast the number of posts published per time interval during an online public opinion event, which is most likely a time series dataset featured with multiple small-scale peaks and nonlinearities. In this model, the Bayes Regularization is introduced in the training function, and a modified PSO algorithm, consisting of equal division mechanism and information entropy techniques, is applied to optimize the initial weights and thresholds. Two real online public opinion cases are calculated to compare the predicting performance of the proposed model with five other popular models. According to experimental results, the following conclusions can be drawn,

(1) The proposed MPSO-BR-BPNN model showed a significantly better forecasting performance than other five models, i.e. LM-BPNN, PSO-BPNN, BR-BPNN, SGD-BPNN and LS-SVM, in terms of RMSE, MAPE and MAE. In addition, this overmatch even grows with more training data.
(2) Besides its higher accuracy performance, the proposed model also exhibits a better numerical stability compared to other five models.
(3) Although all tested models can predict a rough contour of the number of posts published, three of these models showed hysteretic nature to some extent, especially with insufficient training data. In contrast, the MPSO-BR-BPNN model showed great accuracy in forecasting the peak values and capturing the posting peaks, which is really important for the decision makers to take effective interventions during the development of the online public opinion event.

As for the contributions of this paper, the following issues could be considered:

(1) This paper is the first research attempt focusing on the application of neural networks in predicting the posting number of online public opinions.
(2) A modified PSO algorithm by taking equal division and information entropy techniques into consideration is proposed to optimize the initial weights and thresholds of the neural networks. Actually, this algorithm can be applied in other models as well and some improvements may be achieved.
(3) The MPSO-BR-BPNN model proposed in this paper showed better accuracy and stability in predicting the posting number of online public opinion events, compared with traditional models. This model could also be used to forecast time series data in other areas that have similar characteristics.

However, the underlying formation mechanism of online public opinions is more complex than a simple nonlinear time series, as it is driven by various factors such as emotional and behavioral interactions of netizens, economic and social backgrounds, behavior of the government, etc. Thus, we just carried out a relatively general and experimental analysis focusing on the area from a forecasting perspective. For future studies, more attention could be paid to further improve the model or combine it with other successful models, and to forecast other characteristics of the online public opinions other than the time series dataset of number of posts.

Acknowledgments

Thankfulness shall be expressed to the reviewers for their useful discussions and comments on this manuscript. The corresponding author of the present paper is supported by the National Natural Science Foundation of China under Grant Nos. 71573247, 91024010, 91324009 and also by the Youth Innovation Promotion Association Projectunder Grant No: 2014139.

References

[1] D. Allen, M. Mcaleer, S. Peiris, A. Singh, Nonlinear time series and neural-network models of exchange rates between the us dollar and major currencies, Econ. Inst. Res. Pap. 4 (1) (2016) 7.
[2] K.Y. Chan, T.S. Dillon, J. Singh, E. Chang, Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm, IEEE Trans. Intell. Transp. Syst. 13 (2) (2012) 644–654.
[3] X.Y. Cheng, L.L. Zhu, Q. Zhu, J. Wang, X.Y. Cheng, L.L. Zhu, et al., The framework of network public opinion monitoring and analyzing system based on semantic content identification, J. Convergence Inf. Technol. 5 (10) (2010) 48–55.

[4] M.D. Conover, E. Ferrara, F. Menczer, A. Flammini, The digital evolution of occupy wall street, Plos One 8 (5) (2013) e64679.
[5] G. Das, P.K. Pattnaik, S.K. Padhy, Artificial neural network trained by particle swarm optimization for non-linear channel equalization, Expert Syst. Appl. 41 (7) (2014) 3491–3496.
[6] K.M. Deluca, S. Lawson, Y. Sun, Occupy wall street on the public screens of social media: the many framings of the birth of a protest movement, Commun. Culture Critique 5 (4) (2012) 483–509.
[7] J.P. Donate, X. Li, G.G. Sánchez, A.S.D. Miguel, Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm, Neural Comput. Appl. 22 (1) (2013) 1–10.
[8] R.C. Eberhart, Y. Shi, Comparing inertia weights and constriction factors in particle swarm optimization, in: Proceedings of the 2000 Congress on Evolutionary Computation, vol. 1, IEEE, 2000, pp. 84–88.
[9] J.G.D. Gooijer, K. Kumar, Some recent developments in non-linear time series modelling, testing, and forecasting, Int. J. Forecasting 8 (2) (1992) 135–156.
[10] S.H. Hwang, D.H. Ham, J.H. Kim, Forecasting performance of LS-SVM for nonlinear hydrological time series, KSCE J. Civ. Eng. 16 (5) (2012) 870–882.
[11] C. Jin, S.W. Jin, L.N. Qin, Attribute selection method based on a hybrid BPNN and PSO algorithms, Appl. Soft Comput. 12 (8) (2012) 2147–2155.
[12] R. Johnson, T. Zhang, Accelerating stochastic gradient descent using predictive variance reduction, Adv. Neural Inf. Process. Syst. (2013) 315–323.
[13] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the IEEE International Conference on Neural Networks, 1995, vol. 4, IEEE, 1995, pp. 1942–1948.
[14] J. Kennedy, R. Eberhart, Y. Shi, Swarm Intelligence, Morgan Kaufmann Publishers, San Francisco, 2001.
[15] S. Klein, J.P.W. Pluim, M. Staring, M.A. Viergever, Adaptive stochastic gradient descent optimisation for image registration, Int. J. Comput. Vis. 81 (3) (2009) 227–239.
[16] O. Kocadağlı, B. Aşıkgil, Nonlinear time series forecasting with Bayesian neural networks, Expert Syst. Appl. 41 (15) (2014) 6596–6610.
[17] N. Kourentzes, D.K. Barrow, S.F. Crone, Neural network ensemble operators for time series forecasting, Expert Syst. Appl. 41 (9) (2014) 4235–4244.
[18] K.H. Kwon, C.C. Bang, M. Egnoto, H.R. Rao, Social media rumors as improvised public opinion: semantic network analyses of Twitter discourses during Korean saber rattling 2013, Asian J. Commun. 26 (3) (2016) 201–222.
[19] J. Lee, Y.S. Lim, Gendered campaign tweets: the cases of Hillary Clinton and Donald Trump, Pub. Relat. Rev. 42 (5) (2016) 849–855.
[20] T.H. Lee, H. White, C.W.J. Granger, Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative tests, J. Econ. 56 (3) (1993) 269–290.
[21] T. Li, J.P. Pan, J. He, A Network public opinion analysis system based on semantic web, in: Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, vol. 3, IEEE, 2010, pp. 550–554.
[22] L. Liu, J. Zhou, X. An, Y. Zhang, L. Yang, Using fuzzy theory and information entropy for water quality assessment in three gorges region, China, Expert Syst. Appl. 37 (3) (2010) 2517–2521.
[23] M. Liu, L. Wang, L. Nie, J. Dai, D. Ji, Event graph based contradiction recognition from big data collection, Neurocomputing 181 (2016) 64–75.
[24] M. Liu, L. Jiang, H. Hu, Automatic extraction and visualization of semantic relations between medical entities from medicine instructions, Multimed. Tools Appl. 76 (8) (2017) 10555–10573.
[25] N. Ma, Y. Liu, Superedgerank algorithm and its application in identifying opinion leader of online public opinion supernetwork, Expert Syst. Appl. 41 (4) (2014) 1357–1368.
[26] N. Meuleau, M. Dorigo, Ant colony optimization and stochastic gradient descent, Artif. Life 8 (2) (2014) 103–121.
[27] P. Melin, O. Castillo, A. Mancilla, M. Lopez, Simulation and forecasting complex economic time series using neural network models, J. Intell. Syst. 14 (2-3) (2006) 193–212.
[28] D. Niu, Y. Wang, C. Duan, M. Xing, A new short-term power load forecasting model based on chaotic time series and SVM, J. Univ. Comput. Sci. 215 (13) (2009) 2726–2745.
[29] A. Raich, X. Wu, H.F. Lin, A. Cinar, A comparative study of neural networks and nonlinear time series techniques for dynamic modeling of chemical processes, in: Dynamics & Control of Chemical Reactors Distillation Columns & Batch Processes, vol. 143, 1993, pp. 63–68.
[30] D.E. Rumelhart, G.E. Hinton, J.L. McClelland, A general framework for parallel distributed processing, in: D.E. Rumelhart, J.L. McClelland (Eds.), *Parallel Distributed processing: Explorations in the Microstructure of Cognition*. Vol. 1: Foundations, MIT Press, Cambridge, MA, 1986, pp. 45–76.
[31] L.M. Saini, Peak load forecasting using Bayesian regularization, resilient and adaptive backpropagation learning based artificial neural networks, Electr. Power Syst. Res. 78 (7) (2008) 1302–1310.
[32] M. Soltiz, D. Kudithipudi, C. Merkel, G.S. Rose, R.E. Pino, Memristor-based neural logic blocks for nonlinearly separable functions, IEEE Trans. Comput. 62 (8) (2013) 1597–1606.
[33] J. Sun, C. Zheng, Y. Zhou, Y. Bai, J. Luo, Nonlinear noise reduction of chaotic time series based on multidimensional recurrent LS-SVM, Neurocomputing 71 (16–18) (2008) 3675–3679.
[34] N. Thanthry, M.S. Ali, R. Pendse, Security, Internet connectivity and aircraft data networks, IEEE Aerosp. Electron. Syst. Mag. 21 (11) (2006) 3–7.
[35] R.Y. Tian, Y.J. Liu, Isolation, insertion, and reconstruction: three strategies to intervene in rumor spread based on supernetwork model, Decis. Support Syst. 67 (2) (2014) 121–130.
[36] J.L. Ticknor, A Bayesian regularized artificial neural network for stock market forecasting, Expert Syst. Appl. 40 (14) (2013) 5501–5506.
[37] M.E. Tipping, N.D. Lawrence, Variational inference for Student-*t* models: Robust Bayesian interpolation and generalised component analysis, Neurocomputing 69 (1-3) (2005) 123–141.
[38] L. Wang, Y. Zeng, T. Chen, Back propagation neural network with adaptive differential evolution algorithm for time series forecasting, Expert Syst. Appl. 42 (2) (2015) 855–863.
[39] Z. Wang, N. Djuric, K. Crammer, S. Vucetic, Trading representability for scalability: adaptive multi-hyperplane machine for nonlinear classification, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, California, USA, 2011, pp. 24–32.
[40] D.J. Watts, P.S. Dodds, Influentials, networks, and public opinion formation, J. Consumer Res. 34 (4) (2007) 441–458.
[41] R. Xia, C. Zong, S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Inf. Sci. 181 (6) (2011) 1138–1152.
[42] E. Yesilnacar, T. Topal, Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey), Eng. Geol. 79 (3-4) (2005) 251–266.
[43] U. Yolcu, E. Egrioglu, C.H. Aladag, A new linear & nonlinear artificial neural network model for time series forecasting, Decis. Support Syst. 54 (3) (2013) 1340–1347.
[44] H.C. Yuan, F.L. Xiong, X.Y. Huai, A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy, Comput. Electron. Agricult. 40 (1-3) (2003) 57–64.
[45] G.P. Zhang, B.E. Patuwo, M.Y. Hu, A simulation study of artificial neural networks for nonlinear time-series forecasting, Comput. Oper. Res. 28 (4) (2001) 381–396.
[46] W. Zhang, X. Li, H. He, X. Wang, Identifying network public opinion leaders based on Markov logic networks, Sci. World J. 2014 (5) (2014) 435–444.
[47] R. Zhou, X. Wang, X. Dong, Z. Zong, Portfolio selection model with the measures of information entropy-incremental entropy-skewness, Adv. Inf. Sci. Serv. Sci. 5 (8) (2013) 853–864.
[48] H. Kim, K. Beznosov, E. Yoneki, A study on the influential neighbors to maximize information diffusion in online social networks, Computational Social Networks 2 (1) (2015) 3.
[49] K.W. Chau, Application of a pso-based neural network in analysis of outcomes of construction claims, Automation in Construction 16 (5) (2007) 642–646.
[50] Faundez-Zanuy, M. (2010). Nonlinear predictive models computation in ADPCM schemes 1. Signal Processing Conference, 2000, European(pp.1-4).