

Railway freight volume forecast using an ensemble model with optimised deep belief network

Fenling Feng¹, Wan Li¹ ✉, Qiwei Jiang¹

¹School of Traffic and Transportation Engineering, Central South University, Changsha, People's Republic of China

✉ E-mail: lingleeabd@126.com

ISSN 1751-956X

Received on 8th November 2017

Revised 8th April 2018

Accepted on 17th April 2018

E-First on 21st May 2018

doi: 10.1049/iet-its.2017.0369

www.ietdl.org

Abstract: Forecasting freight traffic contributes to the improvement of traffic facilities and making industrial policy, so it is significant to predict freight volume accurately. Extensive works had proved that ensemble model performed better than single model, so an ensemble model, combining seasonal autoregressive integrated moving average (SARIMA) with deep belief network (DBN), is proposed here. SARIMA, a linear model, is used to find the regularities of railway freight traffic. DBN, a non-linear model, is taken to mine the complex relationships between indexes and railway freight. In order to decide appropriate architecture of DBN, including the number of network layers and neurons in each hidden layer, Gaussian particle swarm optimisation algorithm is designed to decide appropriate architecture of DBN, including the number of network layers and neurons in each hidden layer. Besides, Spearman rank correlation analysis is used for selecting indexes related to freight volume. Experimental results show that, compared with SARIMA, DBN, back propagation neural network, Elman neural network, and radial basis function neural network, the proposed ensemble model obtains best performance, and the mean absolute error is 5.5159 million t and the mean absolute percentage error is 1.9657%.

1 Introduction

Along with the development of world economy, logistics freight volume is an important indicator of development degree for a country or an area. As the most flourishing developing country that has a significant influence on the current global economy, Chinese logistics freight traffic has become a global concern [1]. As the main artery of the national economy, the railway has the advantages of large transport capacity, low transportation cost, low carbon, and environmental protection. It has become the main way of the long distance freight transportation in China, and plays an important role in China's economic development. Currently, the growth rate of the macroeconomy in China is low, and the price and volume of the staple freight are declining [2]. Freight volume is also one of the main bases for determining the scale of logistic infrastructure and making industrial policies. However, there are two difficulties to predict freight volume accurately: (i) freight volume is a typical non-stationary time series; (ii) it is influenced by many factors such as foreign trade and economic cooperation, industry. Thus, it is hard to forecast railway freight volume.

Nevertheless, many scholars tried to develop different forecast models to overcome these difficulties. These models included: gray model, Markov method, regression model, and neural network. Wang and Chen [3] proposed a combined model with optimal weights which combined elasticity coefficient method with GM(1,1) and forecasted passenger and freight traffic volumes from 2011 to 2015 of all provinces in China. Ma and Chen [4] and Lu and Song [5] used Markov model improved by gray model and other methods to build predicting models for forecasting civil aviation freight volume and railway freight volume, respectively. The gray system theory is simple and has fast operating speed, which can give a good performance for short-term forecasting, but it is not ideal for the changing system. Markov model assumes that past states are independent of the state of the future. Obviously, monthly railway freight volume that exhibits certain fluctuation is affected by the state of national economy. In addition, there is a certain relationship between the current freight volume and the past freight volume, thus they are not suitable to predict monthly railway freight volume. Zia [6], Yang and Yu [7, 8], and Gopal and Prasanta [9] took various regression models to forecast freight volume of aviation, port, and railway, respectively. Recently,

artificial neural network (ANN) had achieved great success in predicting freight or cargo. Sun and Wei [10] used back propagation neural network (BPNN) to predict railway freight turnover. Yang *et al.* [11] proposed a hybrid model which embraced wavelet neural network, Markov chain, and volatility model for short-term travel time prediction. Tang *et al.* [12] and Ma *et al.* [13] improved fuzzy neural network and convolution neural network to forecast traffic speed, respectively. Regression model is one of the linear models, while there is a non-linear relationship between index and railway freight volume. So, regression model is also not suitable for predicting monthly railway freight traffic. Owing to good self-learning, self-adapting, and generalisation ability, ANN got a wide range of applications, such as traffic speed prediction [14–16].

However, there are still problems when ANNs are applied to resolve some real-world problems: (i) how to decide the structure of them? (ii) how to decide the initial values of connection weights? (iii) how to find a suitable learning rate during training [17]? To solve the first and third problem, Takashi and Shinsuke [17] and Shao and Jiang [18] applied successfully particle swarm optimal (PSO) algorithm to decide the structure, learning rate, and momentum of neural network. As for the second problem, Hinton and Osindero [19] proposed a deep belief network (DBN) and corresponding efficient algorithm. This algorithm becomes the main framework of deep learning since then. DBN, made up of multiple restricted Boltzmann machines (RBMs), uses a layer-by-layer unsupervised learning to pre-training the initial weights of the networks, then the global supervised learning is used for fine-tuning it.

In recent years, the application of ensemble model has become more and more popular. Different models have both different advantages and disadvantages, so some scholars combined linear models with non-linear models recently. Ratnadip [20] proposed using ANN to optimise weights of each model in ensemble models, which consisted of Box–Jenkins, feedforward ANN (FANN), Elman ANN (EANN), and support vector machines (SVM). What is more, DBN was used to process data and mine different data features. Bai *et al.* [21] predicted daily reservoir inflow using ensemble models, which consisted of three DBNs. In order to increase predictive accuracy, this paper attempts to use ensemble model for forecasting railway freight traffic.

This paper has the following three innovations: (i) DBN, one of deep learning methods, is used to predict freight volume; (ii) put forward an ensemble model composed of SARIMA and DBN; (iii) Gaussian particle swarm optimisation (GPSO) is used to optimise the architecture of DBN (GPSO-DBN). SARIMA, a linear model, is good at predicting cyclical time series, and railway freight volume shows cyclical changes yearly. So, it is suitable to forecast time series of railway freight volume. DBN, a non-linear model, has more extensive adaptability and mapping capability which can fit any continuous function with arbitrarily desired accuracy to solve many complex problems. Besides, the DBN is a non-parametric data-driven model and it does not need restrictive assumptions on the underlying process from which data are generated [18]. The relationships between railway freight traffic and indexes are complicated; therefore, DBN is used for mining the non-linear relationships between them. Combining SARIMA with DBN can give full play to the advantages of each model and make up for shortcomings of single model, thereby predictive accuracy is improved. Given the previous studies focused on forecasting annual freight volume and they could not guide monthly management of transportation facilities and staff, this paper concentrates on predicting monthly freight traffic. In order to optimise the predictive performance of DBN, GPSO is used to find the appropriate architecture of network (the number of hidden layers and the neurons of each hidden layer). Moreover, Spearman rank correlation analysis is used to obtain indexes that have close relationships with freight traffic.

The remainder of the paper is organised as follows. Section 2 briefly describes the methods this paper has used, including Spearman rank correlation analysis, SARIMA, GPSO, and DBN. Our proposed ensemble model and experimental procedure are described in Section 3. Section 4 reports and discusses the empirical results, followed by conclusions in Section 5.

2 Methods and models

The railway freight volume has two characteristics, which include periodic change and certain fluctuation. Besides, it has close relationships with many industrial and economy indexes. If only one linear or non-linear model is used to predict it, then the model cannot capture these features at the same time, which will affect the predictive accuracy. SARIMA is good at processing data with periodicity, and DBN is adept at mining the relationships between independent variables and dependent variables. Therefore, combining these two models cannot only effectively capture the periodicity and fluctuation characteristics of railway freight volume, but also make full use of the relationships between indexes and railway freight volume.

2.1 Spearman rank correlation analysis

The Spearman rank correlation coefficient ρ is a non-parametric measure of the strength of dependence between two time series or variables whose relation may be non-linear as long as it is monotonic. The Spearman correlation coefficient differs from the Pearson's product-moment correlation coefficient in that the values are converted to ranks before computing the coefficient [22]. Through this method, the relational degrees between indexes and railway freight traffic can be obtained. The correlation degree range is $[-1, 1]$, the closer to 1 the relational degree is, the more positive relevant two time series are; the closer to 0 the relational degree is, the less relevant two time series are; the closer to -1, the more negative relevant two time series are.

2.2 Seasonal autoregressive integrated moving average model

An ARIMA, commonly known as a Box-Jenkins model, assumes that each future value of a time series is a linear function of several past observations and white noise terms. It employs an ordinary differencing process of order to make a time series stationary. For seasonal time series forecasting, the basic ARIMA model is extended to SARIMA, represented as SARIMA [23].

2.3 Gaussian particle swarm optimisation algorithm

The PSO is a meta-heuristic algorithm that was proposed by Kennedy and Eberhart in 1995 [24]. In beginning, the position and velocity of a group of particles are initialised randomly. Each particle i is defined by its position vector X_i^0 and a random velocity V_i^0 . At the following iteration, the particle moves according to its velocity and be evaluated according to the fitness function $f(X_i^k)$ which is related to the problem to be optimised. The value of the fitness function is compared with the best value attained before. The best value ever obtained for each particle is stored as pbest which actually is the personal optimum, and the best value among all pbest is stored as gbest which actually is the global optimum. Then velocity of the particle is updated by:

$$V_i^{k+1} = \lambda * V_i^k + c_1 * \text{rand} * (\text{pbest} - X_i^k) + c_2 * \text{rand} * (\text{gbest} - X_i^k) \quad (1)$$

where λ is the inertia weight coefficient; rand generates a random number between 0 and 1.

Eberhart proposed the 'adaptation' concept by introducing an inertial weight λ into the algorithm. The inertial weight decreases rapidly in the beginning, then after some number of iterations, the weight decreases slowly by using the equation below:

$$\lambda = \begin{cases} q_0 - ((q_0 - q_1)/f_1) * g, & 1 \leq f \leq f_1 \\ (q_1) * e^{(q_1 - f)/k}, & f_1 \leq f \leq f_2 \end{cases} \quad (2)$$

Where q_0 is the initial inertial weight, q_1 the inertial weight used in the linear reduction, and k the coefficient to adjust the non-linear reduction [25]. The total number of iterations is represented as f_2 and f_1 is the total number of generations for the linear weight decreasing strategy.

This adaptive particle swarm optimisation still retains the problem of the initial parameters setting. If these parameters are not properly set, the convergence speed becomes slow, especially when the search is close to the global optima [26, 27]. Gaussian random variables were used rather than uniformly distributed random variables to escape from the local minimum [28]. The particles' new velocities and positions are calculated with the introduction of Gaussian random variables and the velocity and position equations become

$$V_i^{k+1} = \lambda * V_i^k + c_1 * G(\mu, \sigma^2) * (\text{pbest} - X_i^k) + c_2 * G(\mu, \sigma^2) * (\text{gbest} - X_i^k) \quad (3)$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \quad (4)$$

Where the Gaussian distribution is given by

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \quad (5)$$

The Ziggurat algorithm is a rejection-based method used for sampling the Gaussian random variables [29].

2.4 Deep belief network

The main framework of deep learning includes: (i) the DBN based on the RBM; (ii) the stacked auto-encoder [30], (iii) the convolutional neural network (CNN) [31]. The DBN, an FANN with many hidden layers, has been one of the most common and effective approaches among all deep learning methods. DBN is a stack of RBMs. After an RBM has been trained, the output can be applied as the input data for training the higher level RBM (Fig. 1).

An RBM is an undirected graphical model in which visible variables v are connected to stochastic hidden units h using undirected weighted connections [32]. They are restricted that there

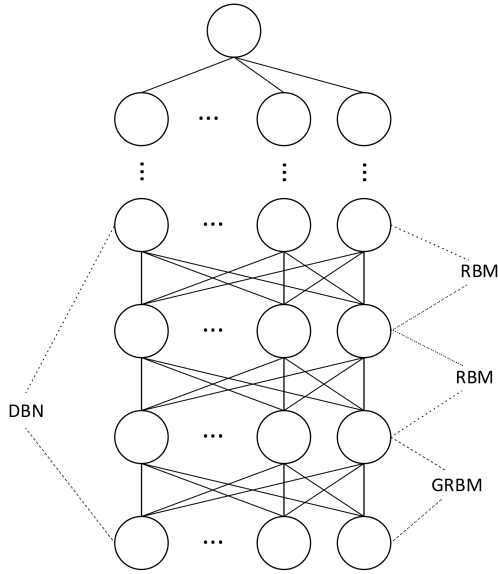


Fig. 1 Structure of DBN

are no connections within hidden variables or visible variables. The model defines a probability distribution over v , h via an energy function E

$$E(v, h|\theta) = - \sum_{i=1}^V \sum_{j=1}^H w_{ij} v_i h_j - \sum_{i=1}^V b_i v_i - \sum_{j=1}^H a_j h_j \quad (6)$$

where $\theta = (w, b, a)$ is the parameter set, w means the weight between the units of hidden layers and visible layers, b and a are the corresponding bias, and V and H stand for the number of the v and the h , respectively.

Hinton and Osindero [19] proposed a fast greedy learning method, which gave a way to train the DBN with one layer at a time, and solved the issue of the θ initialisation successfully. The nodes of visible layer v_i and nodes of hidden layer h_j are binary state nodes, i.e. $v_i, h_j \in \{0, 1\}$. The conditional probability distributions P is as follows:

$$P(h_j = 1|v) = \text{Sigm}\left(\sum_{i=1}^V w_{ij} v_i + a_j\right) \quad (7)$$

$$P(v_i = 1|h) = \text{Sigm}\left(\sum_{j=1}^H w_{ij} h_j + b_i\right) \quad (8)$$

Where $\text{Sigm}(\cdot)$ represents a log-sigmoid function. The goal of training networks is to train the probabilities in (7) of visible units as approximations, which means making $P(v)$ maximisation. To solve this problem, Hinton [33] proposed the contrastive divergence (CD) algorithm:

- i. initialise v using the input data, and compute h according to the conditional probability distributions (7);
- ii. obtain reconstruction state v' based on (8) using h , and repeat (7) to update the hidden nodes using v' , getting h' . The weight update can be calculated as follows:

$$\Delta w_{ij} = \varepsilon (E[v_i h_j] - E[v'_i h'_j]) \quad (9)$$

where ε is the learning rate, and $E[\cdot]$ represents the expectation of the training data.

Then stack several RBMs together into a DBN, and the probability of generating visible nodes $P(v)$ could be derived as

$$P(v) = \sum P(h|\theta)P(v|h, \theta) \quad (10)$$

After θ learned from an RBM, $P(v|h, \theta)$ is kept. In addition, $P(h|\theta)$ could be replaced in turns by continuous RBM which treats hidden layer of previous RBM as visible data. By this way, it can improve a variational lower bound on the probability of the training data as introduced in [19]. This process of feature learning is continued until achieve maximum of $P(v)$, or setting trained number of hidden layers for the DBN.

Considering the real data normalised into $[0, 1]$ for training DBNs, Cho and Ilin [34] replaced the binary visible neurons with the Gaussian ones, and established GRBM (Fig. 1). Therefore, the first layer RBM of DBN is replaced by GRBM. Equations (6)–(9) can be replaced by

$$E(v, h|\theta) = - \sum_{i=1}^V \sum_{j=1}^H w_{ij} \frac{v_i}{\sigma_i} h_j - \sum_{i=1}^V \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^H a_j h_j \quad (11)$$

$$P(h_j = 1|v) = \text{Sigm}\left(\sum_{i=1}^V w_{ij} \frac{v_i}{\sigma_i} + a_j\right) \quad (12)$$

$$P(v_i = 1|h) = N\left(b_i + \sigma_i \sum_{j=1}^H w_{ij} h_j, \sigma_i^2\right) \quad (13)$$

$$\Delta w_{ij} = \varepsilon \left(E\left[\frac{v_i}{\sigma_i} h_j\right] - E\left[\frac{v'_i}{\sigma_i} h'_j\right] \right) \quad (14)$$

where $\sigma(\sigma = 1)$ is the standard deviation, and $N(b_m, \sigma)$ represents a Gaussian probability density function.

After getting the initial weights according to the CD algorithm, the back propagation algorithm and the batch gradient descent are used to fine-tune the weights. Finally, the model obtained is applied to forecast freight traffic.

3 Proposed model and experimental procedure

3.1 Optimised DBN model

After the DBN is constructed, the next step is to optimise its architecture. Extensive experiments by Yoshua Bengio's group suggested that several hidden layers were better than one [35]. However, the architecture of DBN was hard to decide, there were researches using PSO to find appropriate numbers of hidden nodes [17, 18]. So, in this paper, GPSO is used to search better architecture and parameters of DBN. In [17, 18], DBN with four or five layers was applied. In view of these, in this paper, GPSO is used to optimise architecture of DBN. We select the DBN containing two or three hidden layers, i.e. two or three RBMs. As shown in Fig. 1, the hidden layer of RBM i has n_i ($i=1,2,3$) neurons, then a particle is designed to be a three-dimensional vector $X_i(n_1, n_2, n_3)$. Designing population of the particles X_i with enough size S ($i=1,2,3,\dots,S$) and using the GPSO algorithm, the optimised DBN is able to forecast railway freight volume perfectly. The flowchart of the optimisation process is shown in Fig. 2.

A summarised GPSO algorithm used to decide the structure of DBN is shown in Table 1. Error of training is mean squared error (MSE) between output values of DBN and real values. Also, e represents the convergence error of this algorithm.

3.2 Ensemble technique

It is not hard to integrate two submodels, while the troubling matter is how to assign the weights of forecast results of two submodels. In this paper, the weights of each model are assigned by the fitting errors of the submodels. The fitting errors contain mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The weights are assigned by the following formulas [20]

$$\mu_i = \frac{\exp(\gamma_i)}{\sum_{j=1}^T \exp(\gamma_j)}, \quad i = 1, 2, \dots, T \quad (15)$$

$$\gamma_i = (\text{MAE}_i + \text{RMSE}_i + \text{MAPE}_i)^{-1}, \quad i = 1, 2, \dots, T \quad (16)$$

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |y_n - \hat{y}_n| \quad (17)$$

$$\text{RMSE} = \frac{1}{N} \sqrt{\sum_{n=1}^N (y_n - \hat{y}_n)^2} \quad (18)$$

$$\text{MAPE} = \frac{1}{N} \sum_{n=1}^N \left| \frac{y_n - \hat{y}_n}{y_n} \right| \times 100 \quad (19)$$

Here, (15) is so-called SOFTMAX distribution [36], μ_i represents the weight of i th submodel. y_n and \hat{y}_n are the actual and forecast observations, respectively, and N is the total number of observations. By (15)–(19), the differences in magnitudes of the error measures are regularised through using the exponential function. Each weight in a specific sample forecasting trial is non-negative and unbiased.

3.3 Experimental procedure

After having addressed the constituents separately, the ensemble procedure can be concluded as follows: (Fig. 3)

Step 1 Correlation analysis: Collect original freight traffic data, including railway freight volume and indexes data. Then Spearman rank relation degrees between them are computed.

Step 2 Submodels fitting and training: Training data is fitted using SARIMA and parameters p, d, q, P, D, Q are decided according to the value of Akaike information criterion (AIC) and Bayesian information criterion (BIC).

Similarly, training data are used for training DBN optimised by GPSO. The input of DBN is indexes affecting railway freight volume and output is monthly railway freight volume. The parameters of DBN mentioned in Section 3.1 can be obtained by GPSO.

Step 3 Calculating weights: Take fitting errors of SARIMA and optimised DBN into (15) and (16), then the weights of two submodels μ_1 and μ_2 are obtained.

Step 4 Predicting: After parameters of SARIMA being set up, its predictive result $x_i^j (i = 1, 2, 3, \dots, 12)$ is obtained.

Taking test indexes data as input of optimised DBN predictive result $x_i^j (i = 1, 2, 3, \dots, 12)$ is gotten. Finally, predictive result of ensemble model $R^j (1, 2, 3, \dots, 12)$ is calculated by $R^j = \mu_1 \cdot x_1^j + \mu_2 \cdot x_2^j$.

4 Experiments and results

4.1 Data pre-processing

4.1.1 Data normalisation: The time series, concluding railway freight volume and indexes related to the former, are normalised by the following formula

$$x_i' = \frac{x_i}{x_{\max}} \quad (20)$$

where x_i' is the normalised value, and x_{\max} represents the maximum value of the data series.

4.1.2 Correlation analysis: The data used in this paper are obtained from the website of the National Bureau of Statistics of China. Data include railway freight volume and indexes from January 2011 to December 2015. Spearman rank correlation analysis is performed on these indexes, and results are shown in Table 2.

As can be seen from Table 2, there are 19 indexes related to railway freight volume. These indexes are divided into six types,

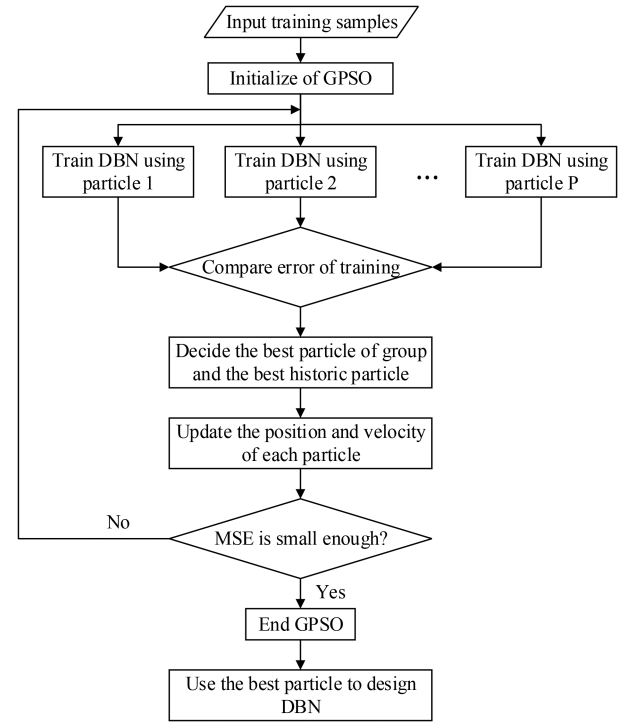


Fig. 2 Flowchart of the optimised DBN using GPSO

Table 1 Procedure of optimised DBN

Input the training samples and prepare a DBN.
Decide the population size of particles S and limitation of iteration number l .
Initialise the start position X_i^0 and the start velocity V_i^0 of each particle.
While iteration number less than l and training error less than ϵ .
Evaluate each particle using MSE, and find the best position of a particle g_{best} from its history and the best particle position of the swarm p_{best} .
Update the positions and velocities of all the particles by Eq. (3) and Eq. (4).
End

containing industry, energy, investment in fixed assets, foreign trade and economic cooperation, domestic trade, postal, and telecommunication services. Among them, indexes of industry type include: (i) output of microcomputer equipment (OME, sets); (ii) output of lead (OLE, t); (iii) output of electronic computer (OEC, sets); (iv) output of passenger car (OPC, cars); (v) output of wine (OW, kl); (vi) output of steel bar (OSB, t); (vii) output of metal cutting machine tool (OMCMT, sets); (viii) output of wire rod (OWR, t); (ix) output of household washing machine (OHWM, sets); (x) output of iron ore (OIO, t); (xi) output of pig iron (OPI, t); (xii) output of liquor (OLI, kl). Indexes of energy type include: (i) output of coke (OC, t); (ii) thermal power (TP, kW h). Indexes of investment in fixed assets type include the coal mining and washing industry (CMWI, yuan). Indexes of foreign trade and economic cooperation type include the total value of imports (TVI, dollars). Indexes of domestic trade type include retail value of petroleum and related products (RVPRP, yuan). Indexes of postal and telecommunication services type include: (i) issue of magazines (IM, copies); (ii) total business volume of postal and telecommunication (TBVPT, yuan). The words following abbreviation are the indexes' units of measure.

Among these indexes, the indexes in industry type are the most. As railway is good at transporting goods that are long distance and large quantities, it has become one of the main ways to transport industrial products. In these industrial indexes, OME, OLE, OEC, OPC, OW, OMCMT, OHWM, and OIO have positive correlations with railway freight volume, and the correlation degrees are 0.35,

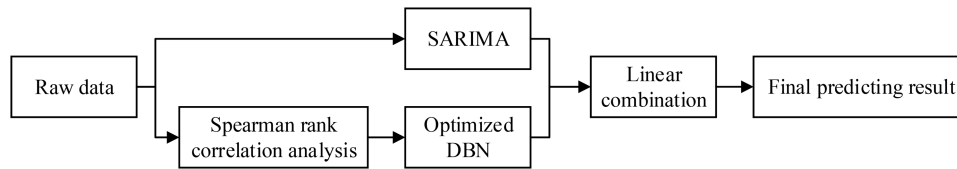


Fig. 3 Experimental procedure

Table 2 Spearman rank correlation analysis for indexes

Types	Index	Degree
industry	OME	0.35
	OLE	0.28
	OEC	0.26
	OPC	0.20
	OW	0.18
	OSB	-0.11
	OMCMT	0.11
	OWR	-0.10
	OHWM	0.07
	OIO	0.07
	OPI	-0.06
	OLI	-0.05
	OC	0.16
energy	TP	0.04
investment in fixed assets	CMWI	0.16
foreign trade and economic cooperation	TVI	0.46
domestic trade	RVPRP	-0.06
postal and telecommunication services	IM	0.48
	TBVPT	-0.47

0.28, 0.26, 0.20, 0.18, 0.11, 0.07, and 0.07, respectively. Besides, OSB, OWR, OPI, and OLI have negative correlations with railway freight volume, and the correlation degrees are -0.11, -0.10, -0.06, and -0.05, respectively. These four indexes being negatively correlated with railway freight volume is because that the output of these industrial products increased first and then decreased in fluctuation from 2011 to 2015, and railway freight volume showed a downward trend in fluctuation.

As shown in Table 2, indexes in energy type, investment in fixed assets type, and domestic trade type are related to coal and oil. Railway is one of the main ways to transport coal and oil, so railway freight volume is related to these indexes to a certain extent. Both OC and TP in energy type have positive correlations with railway freight volume and the correlation degrees are 0.16 and 0.04, respectively. The CMWI in investment in fixed assets type has positive correlation with railway freight volume and the correlation degree is 0.16. RVPRP in domestic trade type has negative correlation with railway freight volume and the correlation degree is -0.06. This index having a negative correlation with railway freight volume is because it increased in fluctuation from 2011 to 2014, and began to decrease in 2015, while railway freight volume had been decreasing in fluctuation.

In addition, TVI in foreign trade and economic cooperation type in Table 2 has positive correlation with railway freight volume and the correlation degree is 0.46. Railway freight volume is mainly affected by national economy. TVI reflects the state of the national economy. Therefore, there is a high correlation between railway freight volume and TVI.

Of all the indexes, the two indexes in Table 2 belong to postal and telecommunication services (IM and TBVPT) have the highest correlation degrees with the railway freight volume, which are 0.48 and -0.47. Railway is one of the main ways of transporting paper magazines; therefore, there is a high positive correlation between them. TBVPT has been gradually increased in recent years and the main transportation way of postal services is highway. Therefore, compared with other transporting ways, highway transportation has a negative impact on railway freight transportation. Given this,

Table 3 Experimental parameters

Name	Abbreviation	Value
the number of neurons in input layer	—	19
the number of neurons in output layer	—	1
the number of RBMs	—	2 or 3
iteration number of each RBM	—	50
the number of neurons in the first and second hidden layer	$n_i (1 \leq n_i \leq 30, i = 1, 2)$	given by GPSO
the number of neurons in the third hidden layer	$n_3 (0 \leq n_3 \leq 30)$	given by GPSO
learning rate of DBN	α_1	0.1
momentum of DBN	β_1	0.1
BP algorithm convergence error	ϵ	0.0001
iteration times of BP algorithm	—	200
population size of GPSO	S	10
iteration times of GPSO	I	30
initial inertia weight of GPSO	q_0	0.9
linear reduction inertia weight of GPSO	q_1	0.4
coefficient of non-linear reduction inertia weight of GPSO	k	10
iteration times of inertia weight linear reduction in GPSO	f_1	15
the number of matrix of Zigurat algorithm	—	128

there is a negative correlation between railway freight and total business volume of postal and telecommunication.

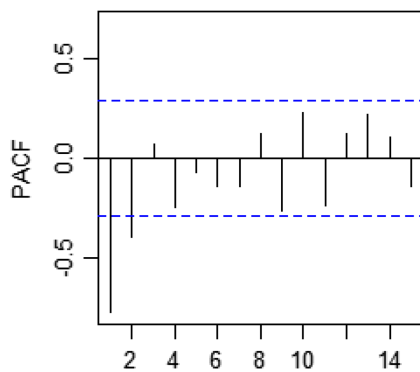
4.2 Prediction results and analysis

4.2.1 Parameters setting: The parameters of predicting experiments are shown in Table 3. As there are 19 indexes having correlation with railway freight volume, they are taken as the input of DBN and the number of neurons in input layer is 19. The output of DBN is monthly railway freight volume, thus, the number of neurons in output is 1. Before doing experiments, it is unclear how many RBMs can get better performance. Therefore, the number of RBMs is 2 or 3. The number of iteration for each RBM is 50. The number of neurons in hidden layer is decided by GPSO. The number of neurons in the first and second hidden layer range from 1 to 30 and the number of neurons in the third hidden layer range from 0 to 30. As the third hidden layer may not exist, the number of neurons in the third hidden layer may be equal to 0. The learning rate and momentum of DBN are 0.1 and 0.1, respectively. It should be noted that here the learning rate and momentum of DBN represent the learning rates and momentums of RBM and BP algorithm. The convergence error and iteration times of BP algorithm are 0.0001 and 200, respectively. The population size, iteration times, initial inertia weight, linear reduction inertia weight, coefficient of non-linear reduction inertia weight, and iteration times of inertia weight linear reduction in GPSO are 10, 30, 0.9, 0.4, 10, and 15, respectively. Besides, the number of matrix of Zigurat algorithm is 128.

The training samples of experiments are indexes data and railway freight volume from January 2011 to December 2014, total 4 years. The test samples are indexes data and railway freight volume from January 2015 to December 2015, total 1 year.

Table 4 Parameters of SARIMA candidate models

Models	AIC	BIC
(1, 1, 2)	826.79	834.19
(2, 1, 2)	820.21	829.47
(0, 2, 2)	818.2	823.69
(1, 2, 2)	815.93	823.25
(2, 2, 2)	817.88	827.02
(0, 2, 3)	818.7	826.02
(1, 2, 3)	816.65	825.79
(2, 2, 3)	809.97	820.94
$(0, 2, 2) \times (1, 0, 0)^{12}$	791.62	798.94
$(0, 2, 3) \times (1, 0, 0)^{12}$	790.76	799.91
$(1, 2, 2) \times (1, 0, 0)^{12}$	794.06	803.2
$(1, 2, 3) \times (1, 0, 0)^{12}$	792.66	803.63

**Fig. 4** PACF graph of the second-order differentiated series of railway freight volume

4.2.2 Selecting SARIMA model: Owing to the decreasing trend of railway freight volume from January 2011 to December 2014, it needs to be differentiated. After railway freight volume was differentiated with the first order and second order, the stationary time series was gotten. As autocorrelation function (ACF) graph and partial autocorrelation (PACF) graph of the differentiated time series do not show obvious seasonality, no more seasonal difference is needed. Finally, some candidate models are obtained and its parameters are analysed, which are shown in Table 4. We can see that the AIC and BIC of $(0, 2, 2) \times (1, 0, 0)^{12}$ are 791.62 and 798.94, respectively, and the AIC and BIC of $(0, 2, 3) \times (1, 0, 0)^{12}$ are 790.76 and 799.91, respectively. These parameters are smaller than other candidate models. The values of ACF and PCAF of $(0, 2, 3) \times (1, 0, 0)^{12}$ fall into the confidence interval, while the ACF and PACF values of $(0, 2, 2) \times (1, 0, 0)^{12}$ do not completely fall into the confidence interval. Besides, the AIC and BIC of $(1, 2, 3) \times (1, 0, 0)^{12}$ are 792.66 and 803.63. Both of them are slightly smaller than the values of $(0, 2, 3) \times (1, 0, 0)^{12}$ and the values of ACF and PCAF of $(1, 2, 3) \times (1, 0, 0)^{12}$ fall into the confidence interval. The PACF graph of the second-order differentiated series of railway freight volume is shown in Fig. 4. According to this figure, the value of p can be 1, but the AIC and BIC of $(0, 2, 3) \times (1, 0, 0)^{12}$ are smaller than these of $(1, 2, 3) \times (1, 0, 0)^{12}$. Therefore, this paper chooses the SARIMA model $(0, 2, 3) \times (1, 0, 0)^{12}$ to predict the railway freight volume.

4.2.3 Predicting and discuss: The fitting results of SARIMA and training results of DBN are shown in Fig. 5. As can be seen from Fig. 5, both of these two models fit well. According to (15), weights assigned to SARIMA and DBN are 0.5129 and 0.4871, respectively.

For neural networks, computation time is one of key points that is often compared with other models. However, in this paper, the DBN is used to predict and what is more important is the predictive precision. Therefore, the computation times of all models for predicting are not considered. However, studies had shown that

DBN's computation time is indeed longer than other methods, such as SVM [18].

In order to exhibit predicting accuracy of the model proposed in this paper, the multiple hidden layers BPNN, single hidden layer BPNN, multiple hidden layers Elman neural network (ENN), single hidden layer ENN, and radial basis function neural network (RBF) are compared. The convergence error, maximum number of iteration, and learning rate of BPNN and ENN, containing multiple hidden layers and single hidden layer, are 0.0001, 1000, and 0.01, respectively. All activation function in BPNN is log-sigmoid function. All activation function in ENN is tan-sigmoid function. Owing to the small amount of data in this paper, we use the batch gradient descent algorithm to train DBN, BPNN, and ENN. The convergence error and the value of spread of RBF are 0.0001 and 0.8, respectively. In this paper, the convergence error represents the MSE.

The predictive results are shown in Table 5. GPSO-BPNN5 means BPNN has three hidden layers and is optimised by GPSO. BPNN3 indicates that BPNN only has one hidden layer. Similarly, GPSO-ENN5 means that ENN has three hidden layers and is optimised by GPSO. ENN3 indicate ENN only has one hidden layer. Structures in Table 5 represent the structure parameters of models when optimal results are obtained. p , d , q , P , D , and Q in SARIMA is 0, 2, 3, 1, 0, and 0, respectively. The structure of DBN is 19–13–12–22–1, where the number of values represents the number of network layers and the values represent the number of the neurons in each layer. In other words, 13 neurons in the first hidden layer, 12 neurons in the second hidden layer, and 22 neurons in the third hidden layer are the best combination. The interpretation of other models' structures are the same as above.

The predictive errors of eight models are shown in Table 5. SARIMA-DBN, the proposed model in this paper, achieves the best performance, and its MAE and MAPE are 5.5159 million t and 1.9657%, which are the smallest predictive error among all models. It indicates that the proposed ensemble model has better generalisation ability and higher predictive accuracy. Thus, it is appropriate to predict railway freight volume.

The MAE and MAPE of SARIMA, which is one of the submodels of the proposed ensemble model, are 7.2023 million t and 2.5967%, respectively. The MAE and MAPE of DBN, which is also one of the submodels of the proposed ensemble model, are 7.1320 million t and 2.5592%, respectively. Compared to SARIMA and DBN, the MAE and MAPE of SARIMA-DBN are smaller than those two models, which means the prediction accuracy of the ensemble model is better than single model.

The predictive results of SARIMA-DBN and submodels are shown in Fig. 6. The results of SARIMA and DBN have deviations from real values. After results of these two submodels being combined, the final result is closer to actual value, so better prediction results are obtained. The predicting errors of SARIMA are close to that of DBN. Compared to DBN, SARIMA is simpler and its computation cost is less. Since it is expert in fitting and predicting time series with periodic change, it can achieve almost the same performance as DBN.

The MAE and MAPE of five layers BPNN optimised by GPSO are 10.1534 million t and 3.6069%, respectively. The MAE and MAPE of three layers BPNN are 12.0858 million t and 4.3590%, respectively. The MAE and MAPE of five layers ENN optimised by GPSO are 16.5125 million t and 5.9524%, respectively. The MAE and MAPE of three layers ENN are 10.5750 million t and 3.7828%, respectively. The MAE and MAPE of RBF are 9.4767 million t and 3.4141%, respectively. The proposed ensemble model SARIMA-DBN and both of two submodels in SARIMA-DBN have smaller predicting errors than other neural networks, which include five layers BPNN optimised by GPSO, three layers BPNN, five layers ENN optimised by GPSO, and three layers ENN. This indicates that on the one hand, railway freight volume is indeed a periodic time series and SARIMA is suitable for forecasting railway freight volume, on the other hand, railway freight volume is affected by many indexes and it is reasonable to use the indexes for forecasting railway freight volume. DBN is appropriate for mining the relationships between the indexes and the railway freight volume.

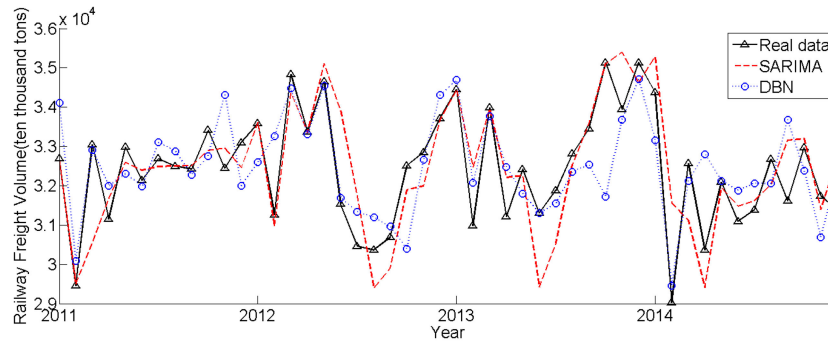


Fig. 5 Fitting results of SARIMA and DBN

Table 5 Forecasting results

Models	Structures	Predicting errors	
		MAE, million t	MAPE, %
SARIMA-DBN	-	5.5159	1.9657
SARIMA	$(0, 2, 3) \times (1, 0, 0)^{12}$	7.2023	2.5967
DBN	19-13-12-22-1	7.1320	2.5592
GPSO-BPNN5	19-20-29-18-1	10.1534	3.6069
BPNN3	19-13-1	12.0858	4.3590
GPSO-ENN5	19-21-30-27-1	16.5125	5.9524
ENN3	19-11-1	10.5750	3.7828
RBF	19-30-1	9.4767	3.4141

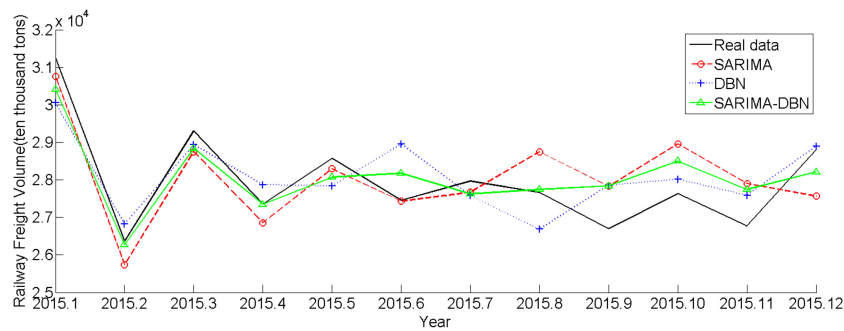


Fig. 6 Forecast results of SARIMA, DBN, and SARIMA-DBN

Table 6 Maximum and minimum AEs and APEs of predictive results

	Models	Maximum value	Minimum value
AE, million t	SARIMA-DBN	10.2392	0.0231
	SARIMA	11.7427	0.1610
	DBN	13.3252	0.8528
APE, %	SARIMA-DBN	4.3123	0.0095
	SARIMA	4.7762	0.0659
	DBN	5.4564	0.3329

The SARIMA-DBN performs better than other models. Although five layers BPNN and five layers ENN are optimised by GPSO, the predictive accuracy is lower than that of SARIMA-DBN. This may be resulted from the weights and thresholds in the network locating in a local minimum.

Besides, the maximum and minimum absolute errors (AEs) and absolute percentage errors (APEs) of predictive results for SARIMA, DBN, and SARIMA-DBN are shown in Table 6. Compared with SARIMA and DBN, whether it is AE or APE, the proposed SARIMA-DBN has smaller error. Also, the SARIMA has smaller maximum and minimum values than DBN, whether they are AEs or APEs. The AEs of predictive results for SARIMA, DBN, and SARIMA-DBN are shown in Fig. 7. The APEs of predictive results for SARIMA, DBN, and SARIMA-DBN are shown in Fig. 8. From Figs. 7 and 8, we can see the forecasting errors for every month. In some months, the errors of combined results decrease sharply, such as February 2015, this is because the

predictive result of one of submodel is larger than the real value and the other one is smaller than the real value. In some months, the errors of combined results are the average of two submodels' results, such as January 2015, this is because the predictive results of both submodels are larger or smaller than the real value. In this case, the result of the ensemble model is larger than that of one of the submodels. However, it is inevitable for ensemble model. It is reasonable to consider the whole performance of the ensemble model rather than some of the results.

5 Conclusions

Freight traffic forecasting is a dynamic research area, having fundamental importance in numerous practical fields. Improving accuracy of freight traffic is a challenging task that has been gaining continuous research attentions in the past decades. However, they almost focused on forecasting annual freight

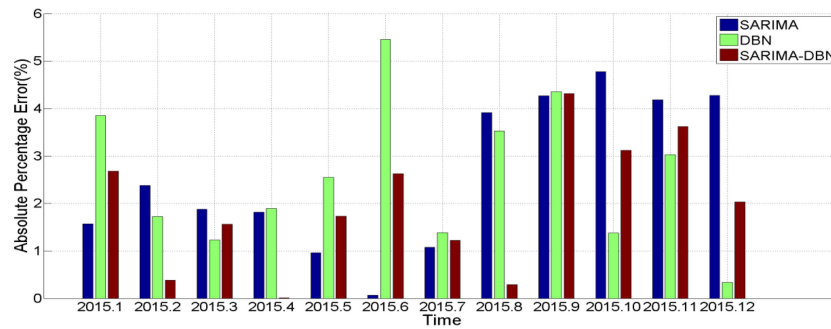


Fig. 7 AEs of predictive results

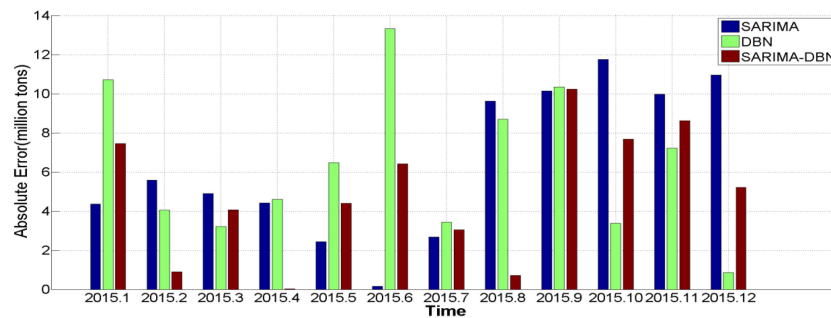


Fig. 8 APEs of predictive results

volume, there was scarcely research on forecasting monthly freight traffic. Obviously, this work is more challenging. This may be due to monthly railway freight traffic fluctuating largely. Railway freight traffic belongs to periodic time series, which means it shows regularity. Besides, railway freight traffic is affected by many indexes. Given these two reasons, an ensemble model SARIMA-DBN is proposed to predict railway freight volume accurately. In order to get the appropriate number of hidden layers and neurons in DBN, GPSO is used to optimise the structure of DBN to get better predictive results.

The predictive results indicate that compared with SARIMA, DBN, five layers BPNN, three layers BPNN, five layers ENN, three layers ENN, and RBF, the proposed SARIMA-DBN obtains the best performance. The future studies will focus on two points: (i) if indexes that are more relevant to railway freight volume are found, the predictive accuracy will be further improved; (ii) combining other deep learning methods and time series models may get better performance for predicting railway freight volume, such as deep auto-encoder and Markov chain.

6 Acknowledgment

This research is supported by China Railway Science and Technology Research Development Program (2015F024).

7 References

- [1] Zhou, C., Tao, J.: 'Adaptive combination forecasting model for China's logistics freight volume based on an improved PSO-BP neural network', *Kybernetes*, 2015, **44**, (4), pp. 646–666
- [2] Feng, F., Xu, Y., Tang, Z.: 'Research on the charge rate of railway value-guaranteed transportation based on competitive and cooperative relationships', *Adv. Mech. Eng.*, 2018, **10**, (1), pp. 1–11
- [3] Wang, Y., Chen, X.: 'Forecast of passenger and freight traffic volume based on elasticity coefficient method and grey model'. Proc. 13th COTA Int. Conf. of Transportation Professionals (CICTP), Shenzhen, China, August 2013, pp. 136–147
- [4] Ma, J., Chen, Z.: 'Forecast of civil aviation freight volume using unbiased grey-fuzzy-Markov chain method'. Proc. 1st Int. Conf. on Information Management, Xi'an, China, 23–24 November 2013, pp. 528–531
- [5] Lu, H., Song, R.: 'Forecast of railway freight ton-kilometers based on the ubgpm-markov model'. Proc. 4th Int. Conf. on Intelligent System and Applied Material (GSAM), Taiyuan, China, August 2014, pp. 2199–2202
- [6] Zia, W.: 'Simultaneous modeling of passenger and cargo demand at an airport', *Transp. Res. Record*, 2013, **2336**, pp. 63–74
- [7] Yang, Y., Yu, C.: 'Prediction models based on multivariate statistical methods and their applications for predicting railway freight volume', *Neurocomputing*, 2015, **158**, pp. 210–215

- [8] Yang, Y., Yu, C.: 'Development of the regional freight transportation demand prediction models based on the regression analysis methods', *Neurocomputing*, 2015, **158**, pp. 42–47
- [9] Gopal, R.P., Prasanta, K.S.: 'Estimation of freight demand at Mumbai port using regression and time series models', *KSCCE J. Civ. Eng.*, 2016, **20**, (5), pp. 2022–2032
- [10] Sun, J., Wei, C.: 'Railway freight turnover forecast based on the BP neural network'. Proc. 3rd Int. Conf. on Materials Engineering for Advanced Technologies (ICMEAT), Brisbane, Australia, January 2014, pp. 837–840
- [11] Yang, H., Zou, Y., Wang, Z., et al.: 'A hybrid method for short-term freeway travel time prediction based on wavelet neural network and Markov chain', *Can. J. Civ. Eng.*, 2018, **45**, (2), pp. 77–86
- [12] Tang, J., Liu, F., Zou, Y., et al.: 'An improved fuzzy neural network for traffic speed prediction considering periodic characteristic', *IEEE Trans. Intell. Transp. Syst.*, 2017, **18**, (9), pp. 2340–2350
- [13] Ma, X., Dai, Z., He, Z., et al.: 'Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction', *Sensors*, 2017, **17**, (4), p. 818
- [14] Jia, Y., Wu, J., Moshe, B., et al.: 'Rainfall-integrated traffic speed prediction using deep learning method', *IET Intell. Transp. Syst.*, 2017, **11**, (9), pp. 531–536
- [15] Yarah, B., Hazem, M.A., Hoda, T., et al.: 'Speed prediction from mobile sensors using cellular phone-based traffic data', *IET Intell. Transp. Syst.*, 2017, **11**, (7), pp. 387–396
- [16] Tang, J., Liu, F., Zhang, W., et al.: 'Lane-changes prediction based on adaptive fuzzy neural network', *Expert Syst. Appl.*, 2018, **91**, pp. 452–463
- [17] Takashi, K., Shinsuke, K.: 'Time series forecasting using a deep belief network with restricted Boltzmann machines', *Neurocomputing*, 2014, **137**, pp. 47–56
- [18] Shao, H., Jiang, H.: 'Rolling bearing fault diagnosis using an optimization deep belief network', *Meas. Sci. Technol.*, 2015, **26**, (11), p. 115002
- [19] Hinton, G., Osindero, S.: 'A fast learning algorithm for deep belief nets', *Neural Comput.*, 2006, **18**, (7), pp. 1527–1554
- [20] Ratnadip, A.: 'A neural network based linear ensemble framework for time series forecasting', *Neurocomputing*, 2015, **157**, pp. 231–242
- [21] Bai, Y., Chen, Z., Xie, J., et al.: 'Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models', *J. Hydrol.*, 2016, **532**, pp. 193–206
- [22] Lohninger, H.: 'Teach/Me data analysis' (Springer, Berlin, Heidelberg, 2002), p. 258
- [23] Hamzaebi, C.: 'Improving artificial neural networks' performance in seasonal time series forecasting', *Inf. Sci.*, 2008, **178**, (23), pp. 4550–4559
- [24] Kennedy, J., Eberhart, R.: 'Particle swarm optimization'. Proc. IEEE Conf. on Neural Networks, Perth, Australia, November 1995, pp. 1942–1948
- [25] Zhang, J., Zhang, J.: 'A hybrid particle swarm optimization-backpropagation algorithm for feedforward neural network training', *Appl. Math. Comput.*, 2007, **185**, pp. 1026–1037
- [26] Venu, G., Ganesh, K.: 'Comparison of particle swarm optimization and back propagation as training algorithms for neural networks'. Proc. IEEE Swarm Intelligence Symp. 2003, Indianapolis, IN, USA, April 2003, pp. 110–117
- [27] Settles, M., Rodebaugh, B., Soule, T.: 'Comparison of genetic algorithm and particle swarm optimizer when evolving are current neural network'. Proc. Genetic and Evolutionary Computation, Chicago, IL, USA, July 2003, pp. 148–149

- [28] Melo, H., Watada, J.: 'Gaussian-PSO with fuzzy reasoning based on structural learning for training a neural network', *Neurocomputing*, 2016, **172**, pp. 405–412
- [29] George, M., Wai, W.: 'The Ziggurat method for generating random variables', *J. Stat. Softw.*, 2000, **5**, (8), pp. 1–7
- [30] Hinton, G.E., Salakhutdinov, R.R.: 'Reducing the dimensionality of data with neural networks', *Science*, 2006, **313**, (5786), pp. 504–507
- [31] Lecun, Y., Bottou, L.: 'Gradient-based learning applied to document recognition', *Proc. IEEE*, 1998, **86**, (11), pp. 2278–2324
- [32] Hinton, G.E., Teh, Y.W.: 'Rate-coded restricted Boltzmann machines for face recognition'. Proc. 14th Neural Information Processing Systems, Colorado, USA, November 2000, pp. 872–878
- [33] Hinton, G.E.: 'Training products of experts by minimizing contrastive divergence', *Neural Comput.*, 2002, **14**, (8), pp. 1771–1800
- [34] Cho, K., Ilin, A.: 'Improved learning of Gaussian-Bernoulli restricted Boltzmann machines'. Proc. 21st Conf. on Artificial Neural Networks, Espoo, Finland, June 2011, pp. 10–17
- [35] Le, R.N., Bengio, Y.: 'Representational power of restricted Boltzmann machines and deep belief networks', *Neural Comput.*, 2008, **20**, (6), pp. 1631–1649
- [36] Kuncheva, L.I.: '*Combining pattern classifiers: methods and algorithms*' (Wiley-Interscience, New York, USA, 2005)