ISSN 2395-8618

# Traffic Accidents Forecasting using Singular Value Decomposition and an Autoregressive Neural Network Based on PSO

Lida Barba and Nibaldo Rodríguez

***Abstract*—In this paper, we propose a strategy to improve the forecasting of traffic accidents in Concepción, Chile. The forecasting strategy consists of four stages: embedding, decomposition, estimation and recomposition. At the first stage, the Hankel matrix is used to embed the original time series. At the second stage, the Singular Value Decomposition (SVD) technique is applied. SVD extracts the singular values and the singular vectors, which are used to obtain the components of low and high frequency. At the third stage, the estimation is implemented with an Autoregressive Neural Network (ANN) based on Particle Swarm Optimization (PSO). The final stage is recomposition, where the forecasted value is obtained. The results are compared with the values given by the conventional forecasting process. Our strategy shows high accuracy and is superior to the conventional process.**

***Index Terms*—Autoregressive neural network, particle swarm optimization, singular value decomposition.**

## I. Introduction

FORECASTING of time series with neural networks has been widely implemented due to its capability of approximation and universal generalization [1], [2] in diverse areas of knowledge [3], [4]. Conventionally, the neural networks show difference and improvement through the adequate selection of transfer and activation functions [5], [6], the variation in the input dimension and the time delay [7], also changing the number of hidden nodes [8], others researchers propose modifications in the learning algorithms [9], is common also the use of explanatory variables [10], there are works that implement hybrid solutions reaching good performance [11], [12], whereas the decomposition, disaggregation or aggregation of the time series before the forecasting have demonstrated to be an effective strategy [13], [14]. The combination ANN-PSO has improved the forecasting over some classical algorithms [15], [16], [17]

Based on these arguments, in this work we propose a strategy of improving traffic accidents forecasting based on the decomposition of a time series in components of low and high frequency from the singular values of the Hankel matrix. The strategy is applied in four stages, embedding with Hankel matrix, decomposing with SVD, estimation with ANN-PSO, and recomposing with simple addition. The time series are the traffic accidents of Concepción - Chile, sinister number and injured, from year 2000 to 2012, with weekly sampling.

The paper is structured as follows. Section II describes the time series forecasting strategy. Section III presents the forecasting accuracy metrics. Section IV presents the results and discussion. Section V gives conclusions.

Manuscript received on December 24, 2014, accepted for publication on April 20, 2015, published on June 15, 2015.

Lida Barba is with the Pontificia Universidad Católica de Valparaíso, Chile and Universidad Nacional de Chimborazo, Ecuador (e-mail: lbarba@unach.edu.ec).

Nibaldo Rodríguez is with Pontificia Universidad Católica de Valparaíso, Chile (e-mail: nibaldo.rodriguez@ucv.cl).

## II. Time series forecasting strategy

Our forecasting strategy is presented in Figure 1. It consists of four stages: embedding, decomposition, estimation, and recomposition. Embedding means to map the time series in a Hankel matrix, decomposition is developed with SVD, the singular values are used to extract the components of low and high frequency, the estimation of the found components is based on an ANN based on PSO, and the recomposition is developed with the simple addition of the ANNs outputs.

The original time series is represented with $x$, $H$ is the Hankel matrix, $S$, $V$, and $U$ are the matrix elements obtained with SVD, $C_L$ is the component of low frequency, $C_H$ is the component of high frequency, $\hat{C_L}$, and $\hat{C_H}$ are the estimated components, $\hat{x}$ is the forecasted time series, and $er$ is the error computed between $x$ and $\hat{x}$.

### A. Embedding the time series

The time series is embedded in the Hankel matrix, the process is illustrated as follows:

$$H_{M\times L} = \begin{bmatrix} x_1 & x_2 & \dots & x_L \\ x_2 & x_3 & \dots & x_{L+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_M & x_{M+1} & \dots & x_N \end{bmatrix} \quad (1)$$

where $H$ is a matrix of order $M \times L$, $x_1 \dots x_N$, are the original values of the time series, of length $N$. The value of $L$ is computed as

$$L = N - M + 1. \quad (2)$$

ISSN 2395-8618

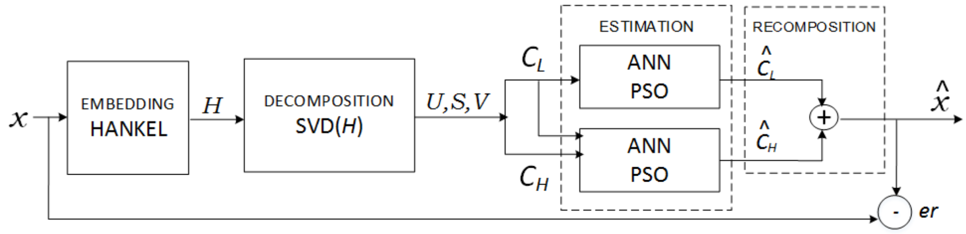


Fig. 1. Time series forecasting strategy

### B. Singular Value Decomposition

Let $H$ be an $M \times n$ real matrix, then there exist an $M \times M$ orthogonal matrix $U$, an $n \times n$ orthogonal matrix $V$, and and $M \times n$ diagonal matrix $S$ with diagonal entries $s_1 \geq s_2 \geq \ldots \geq s_p$, with $p = min(M,n)$, such that $U^T HV = S$. Moreover, the numbers $s_1, s_2, \ldots, s_p$ are uniquely determined by $H$ [18].

$$H = U \times S \times V^T \tag{3}$$

The extraction of the components is done by means of the singular values $s_i$, the orthogonal matrix $U$, and the orthogonal matrix $V$, for each singular value is obtained one matrix $A_i$, with $i = 1 \ldots m$:

$$A_i = s(i) \times U(:,i) \times V(:,i)^T \tag{4}$$

Therefore, the matrix $A_i$ contains the $i-th$ component, the extraction process is:

$$C_i = \left[ \begin{array}{cc} A_i(1,:) & A_i(2,n:m)^T \end{array} \right] \tag{5}$$

where $C_i$ is the $i-th$ component, the elements of $C_i$ are located in the first row and last column of $A_i$.

The optimal number of components ($M$) is given by the maximum peak of differential energy $\Delta E$ of each pair of sequential components, and it computation is

$$\Delta E_i = E_i - E_{i+1} \tag{6}$$

where $E_i$ is the energy of the $i-th$ component, and $i = 1, \ldots, M-1$. The energy of each singular is computed with

$$E_i = s_i^2 / (\sum_{i=1}^{M} s_i^2) \tag{7}$$

where $s_i$ is the $i-th$ singular value of the Hankel matrix obtained before. The first component extracted is the component $C_L$ and the second is the component $C_H$ (if $M = 2$). When the optimal number of components $M > 2$, the component $C_H$ is computed with the summation of the components from 2 to $M-th$, as follows

$$C_H = \sum_{i=2}^{M} C_i \tag{8}$$

### C. Estimation of components with an Autoregressive Neural Network based on PSO

The ANN is based on the algorithm PSO, it performs the estimation to obtain $\hat{C}_L$, and $\hat{C}_H$. The ANN inputs are the lagged terms of $C_L$ and $C_H$. The ANN has a common structure of three layers [19], at the hidden layer the sigmoid transfer function is applied, and at the output layer the estimated value is obtained. The ANN output is

$$\hat{x} = \phi(net) \times b \tag{9}$$

where $\hat{x}$ is the estimated value, $net$ is the output of the hidden layer, $b$ is the vector that contains the weights on the connections from the hidden layer to the output layer, $net$ is computed with

$$net = x \times w \tag{10}$$

where $x$ is the data input matrix, with order $N \times P$, $N$ is the sample length, and $P$ is the number of input variables (lags terms), $w$ is the weight matrix of order $P \times N_h$, with $N_h$ hidden units. The sigmoid transfer function is applied at hidden layer with

$$\phi(net) = 1/(1 + e^{-net}) \tag{11}$$

The weights of the ANN connections, $w$ and $b$ are adjusted with PSO learning algorithm. In the swarm the $N_p$ particles has a position vector $X_i = (X_{i1}, X_{i2}, \ldots, X_{iD})$, and a velocity vector $V_i = (V_{i1}, V_{i2}, \ldots, V_{iD})$, each particle is considered a potential solution in a $D$-dimensional search space. During each iteration the particles are accelerated toward the previous best position denoted by $p_{id}$ and toward the global best position denoted by $p_{gd}$. The swarm has $N_p$x$D$ values and is initialized randomly, $D$ is computed with $P \times N_h + N_h$; the process finish when the lowest error is obtained based on the fitness function evaluation, or when the maximum number of iterations is reached [20], [21].

$$V_{id}^{l+1} = I^l \times V_{id}^l + c_1 \times rd_1(p_{id}^l + X_{id}^l) + c_2 \times rd_2(p_{gd}^l + X_{id}^l) \tag{12}$$

$$X_{id}^{l+1} = X_{id}^l + V_{id}^{l+1} \tag{13}$$

$$I^l = I_{max}^l - \frac{I_{max}^l - I_{min}^l}{iter_{max}} \times l, \tag{14}$$

where $i = 1, \ldots, N_p$, $d = 1, \ldots, D$; $I$ denotes the inertia weight, $c_1$ and $c_2$ are learning factors, $rd_1$ and $rd_2$ are positive random numbers in the range $[0,1]$ under normal distribution,

ISSN 2395-8618

$l$ is the $l$th iteration. Inertia weight has linear decreasing, in equation 14, $I_{max}$ is the maximum value of inertia, $I_{min}$ is the lowest, and $iter_{max}$ is total of iterations.

The particle $X_{id}$ represents the optimal solution of the set of weights in the neural network, therefore $X_i d$ contains the ANN connections weights $w$ and $b$.

### D. Recomposing the time series

The recomposition of the time series is done with the addition of the estimated components, then the forecasted time series is obtained using

$$\hat{x} = \hat{C}_L + \hat{C}_H. \tag{15}$$

## III. Forecasting accuracy metrics

The number of lags for the ANN is determined with the metric: Generalized Cross Validation (GCV), this determines the best number based on the accuracy of the forecasting probing a determined range of values. The evaluation of the forecasting is computed with the metrics: Mean Absolute Percentage Error (MAPE), Coefficient of determination $R^2$, Root Mean Squared Error (RMSE), and Relative Error (RE).

$$RMSE = \sqrt{\frac{1}{N_v}\sum_{i=1}^{N_v}(x_i - \hat{x}_i)^2} \tag{16}$$

$$GCV = \frac{RMSE}{(1 - K/N_v)^2} \tag{17}$$

$$MAPE = \left[\frac{1}{N_v}\sum_{i=1}^{N_v}|(x_i - \hat{x}_i)/x_i|\right] \times 100 \tag{18}$$

$$R^2 = 1 - \frac{\sigma^2(er)}{\sigma^2(x)} \tag{19}$$

$$RE = \sum_{i=1}^{N_v}(x_i - \hat{x}_i)/x_i \tag{20}$$

where $N_v$ is the validation (testing) sample size, $x_i$ is the $i$-th observed value, $\hat{x}_i$ is the $i$-th estimated value, and $K$ is the number of lagged values.

## IV. Results and discussion

The applied data are available from the CONASET web site [22], and they represent the number of accidents and the injured of Concepción-Chile, from year 2000 to 2012 with weekly sampling. The training data set contains the 70% of the sample, consequently the testing data set contains the 30% In the next subsections are evaluated the time series forecasting strategy presented in Fig. 1.

### A. Embedding and Decomposition

The time series is embed in a Hankel matrix, the optimal number of components $M$ was determined using and initial number $M = N/2$ components. Once obtained the number the components, the differential energy $\Delta E$, of each component was computed, this is shown in the Fig. 2, the maximum peak represents the optimal $M$. The embedding and the decomposing is executed again with the optimal $M$, for time series number of accidents the optimal was $M = 6$, and for the time series injured people the optimal found was $M = 4$. The component of low frequency extracted and estimated for the time series number of accidents is shown in the Fig. 4a, while the component of high frequency the same time series is shown in the Fig. 4b. The component of low frequency extracted and estimated for the time series injured people is shown in the Fig. 5a, while the component of high frequency the same time series is shown in the Fig. 5b.

### B. Estimation and Recomposition

The calibration of the number of lags of the ANN was determined with the GCV metric, for the two time series was found an optimal $ANN(K, N_h, 1)$, with $K = 7$ inputs for the time series number of accidents and $K = 5$ for the time series injured people as shown the Fig. 3, and the number of hidden nodes was assigned in $N_h = 6$ for the two time series, this value was computed with the natural logarithm of the training data set length (normally used in our experiments).

The PSO learning algorithm was applied to determine the weights of the ANN, after trial and error they were configured with a swarm of $N_p \times D$ dimension, $N_p = 40$ particles, and $D = Np \times N_h + N_h$, inertia weight parameter $I$ has linear decreasing with a maximum value of 1 and a minimum value of 0.2, the acceleration factors $c_1$ and $c_2$ were fixed in 1.05 and 2.95 respectively, the $iter_m ax$ is 2500.

The evaluation performed at the testing stage for the time series number of accidents is presented in the Fig. 6 and Table I. The observed values vs. the estimated values are illustrated in the Fig. 6a, reaching a good accuracy, while the relative error is presented in the Fig. 6b, which shows that the 98.5% of the points present an error lower than the ±10%.

The evaluation performed at the testing stage for the time series number of injured people is presented in the Fig. 7 and Table II. The observed values vs. the estimated values are illustrated in the Fig. 7a, reaching a good accuracy, while the relative error is presented in the Fig. 7b, which shows that the 95.54% of the points present an error lower than the ±10%.

TABLE I
Number of Accidents Forecasting

| | SVD-ANN-PSO | ANN-PSO |
|---|---|---|
| *Components* | 6 | — |
| *RMSE* | 0.0211 | 0.087 |
| *MAPE* | 3.17% | 14.48% |
| $R^2$ | 98.29% | 70.71% |
| $RE \pm 10\%$ | 98.5% | 48.76% |

The results presented in Table I show that the major accuracy of the forecasting of the time series number of accidents is achieved with the model SVD-ANN-PSO(7,6,1), with a $RMSE$ of 0.0211, and a $MAPE$ of 3.17%, the 98.5% of the points have an relative error lower than the ±10%.

ISSN 2395-8618

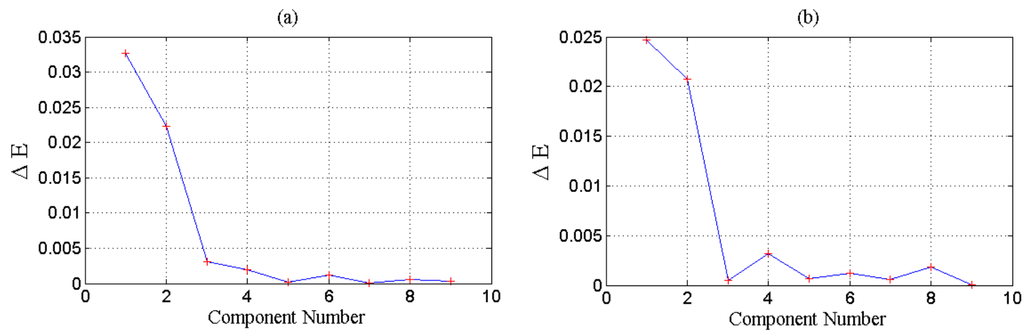


Fig. 2. $\Delta E$ : (a) Number of Accidents, (b) Injured people

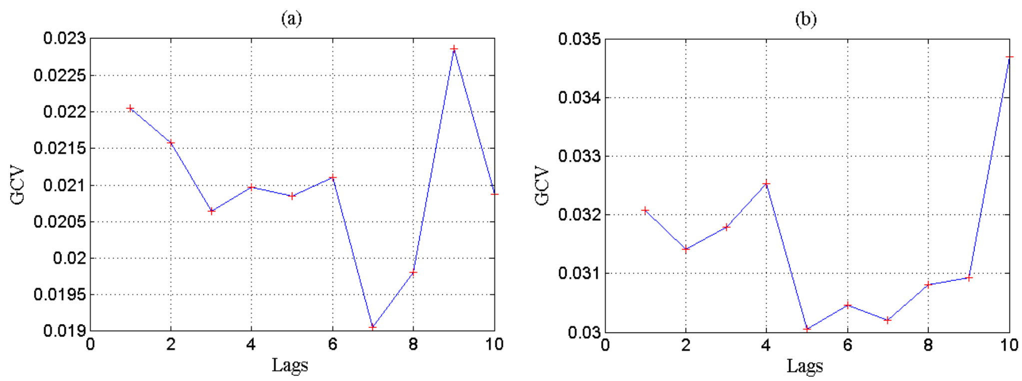


Fig. 3. Lags calibration (a) Number of accidents (b) Injured people

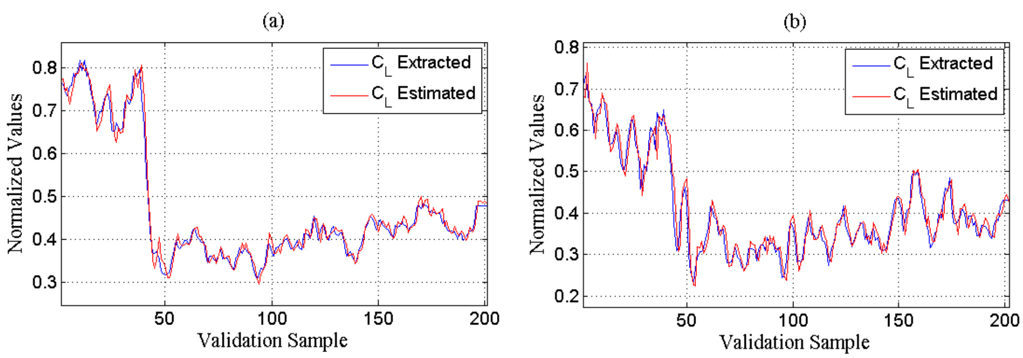


Fig. 4. Number of Accidents components (a) $C_L$, (b) $C_H$

TABLE II
INJURED PEOPLE FORECASTING

| | SVD-ANN-PSO | ANN-PSO |
|---|---|---|
| *Components* | 4 | — |
| *RMSE* | 0.0172 | 0.101 |
| *MAPE* | 3.58% | 21.42% |
| $R^2$ | 98.5% | 46.08% |
| $RE \pm 10\%$ | 95.54% | 40.59% |

The results presented in Table II show that the major accuracy of the forecasting of the time series injured people is achieved with the model SVD-ANN-PSO(5,6,1), with a *RMSE* of 0.0172, and a *MAPE* of 3.58%, the 95.54% of the points have an relative error lower than the ±10%.

## V. Conclusions

The proposed forecasting strategy is based on the time series decomposition using the singular values of the Hankel matrix. The strategy consists of four stages: embedding, decomposition, estimation, and recomposition. The embedding consists in mapping the time series in a Hankel matrix. The decomposition is based on SVD technique, SVD extracts the components of low and high frequency of the time series, the estimation is executed with an ANN based on PSO, while the recomposition is made with the single addition of the estimated components.

For evaluation of this strategy, we implemented a conventional ANN based on PSO. The best result was obtained

ISSN 2395-8618

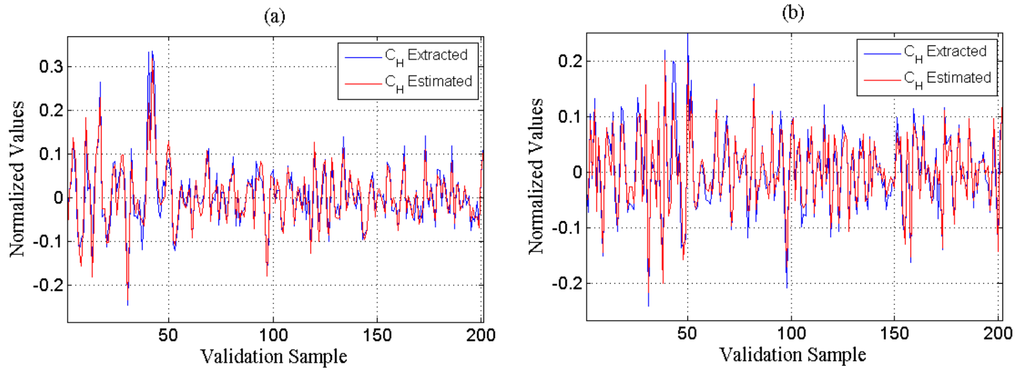


Fig. 5. Injured people components(a) $C_L$, (b) $C_H$

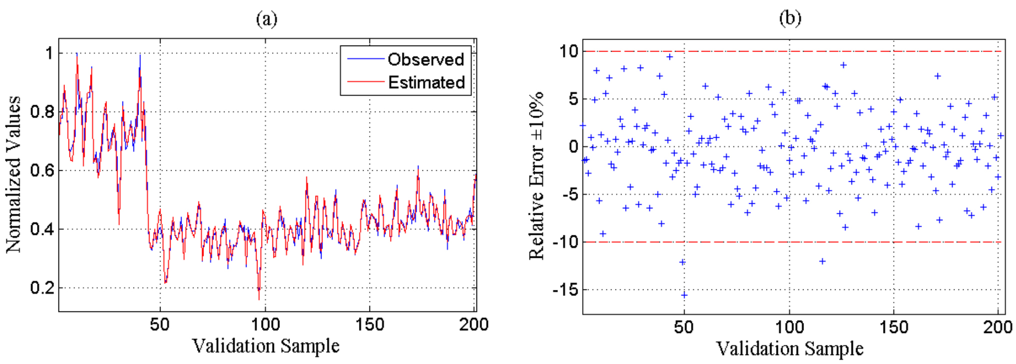


Fig. 6. SVD-ANN-PSO(7,6,1) (a) Observed vs. Estimated (b) Relative Error

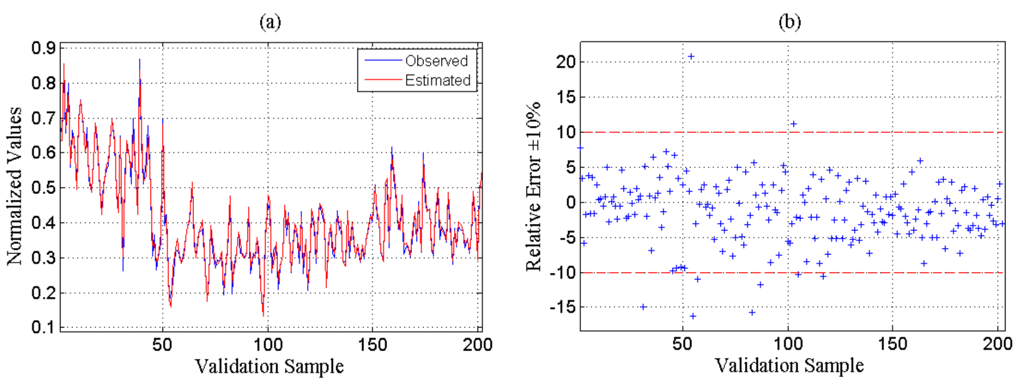


Fig. 7. SVD-ANN-PSO(5,6,1) (a) Observed vs. Estimated (b) Relative Error

with the proposed strategy for the two time series analyzed, SVD-ANN-PSO shows superiority regards the conventional implementation. For the time series number of accidents, SVD-ANN-PSO reaches an *RMSE* of 0.0211, and a *MAPE* of 3.17%, in front of the conventional ANN-PSO that reaches an *RMSE* of 0.087, and a *MAPE* of 14.48%. For traffic accidents reaches an *RMSE* of 0.0172, and a *MAPE* of 3.58%, in front of the conventional ANN-PSO that reaches an *RMSE* of 0.101, and a *MAPE* of 21.42%.

In the future, this strategy will be evaluated with data of traffic accidents of other regions of Chile, other countries, and with time series of other engineering fields.

## ACKNOWLEDGEMENTS

This research was partially supported by the Chilean National Science Fund through the project Fondecyt-Regular 1131105 and by the VRIEA of the Pontificia Universidad Católica de Valparaíso.

## REFERENCES

[1] Hornik K., Stinchcombe X., White H.: Multilayer feedforward networks are universal approximators. Neural Networks. 2(5), 359–366 (1989)

[2] Svozil D., Kvasnicka V., Pospichal J.: Introduction to multi-layer feed-forward neural networks. Chemometrics and Intelligent Laboratory Systems. 39(1), 43–62 (1997)

ISSN 2395-8618

[3] Chattopadhyay G., Chattopadhyay S.:Autoregressive forecast of monthly total ozone concentration: A neurocomputing approach. Computers & Geosciences. 35(9), 1925–1932 (2009)
[4] Maali Y., Al-Jumaily A.: Multi Neural Networks Investigation based Sleep Apnea Prediction. Procedia Computer Science. 24, 97–102 (2013)
[5] Rojas I., Pomares H., Bernier J.L., Ortega J., Pino B., Pelayo F.J., Prieto A.: Time series analysis using normalized PG-RBF network with regression weights. Neurocomputing. 42(1–4), 267–285 (2002)
[6] Roh S.B., Oh S.K., Pedrycz W.: Design of fuzzy radial basis function-based polynomial neural networks. Fuzzy Sets and Systems. 185(1), 15–37 (2011)
[7] Liu F., Ng G.S., Quek C.: RLDDE: A novel reinforcement learning-based dimension and delay estimator for neural networks in time series prediction. Neurocomputing. 70(7–9), 1331–1341 (2007)
[8] Scarselli F., Chung A.: Universal Approximation Using Feedforward Neural Networks: A Survey of Some Existing Methods, and Some New Results. Neural Networks. 11(1), 15–37 (1998)
[9] Gheyas I.A., Smith L.S.: A novel neural network ensemble architecture for time series forecasting. Neurocomputing. 74(18), 3855–3864 (2011)
[10] Gao D., Kinouchi Y., Ito K., Zhao X.: Neural networks for event extraction from time series: a back propagation algorithm approach. Future Generation Computer Systems. 21(7), 1096–1105 (2005)
[11] Khashei M., Bijari M., Ali G.: Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks (PNNs). Computers & Industrial Engineering. 63(1), 37–45 (2012)
[12] Jeong K., Koo C., Hong T.: An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network). Energy. 71, 71–79 (2014)
[13] Wei Y., Chen M.C.: Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C: Emerging Technologies. 21(1), 148–162 (2012)
[14] Shoaib M., Shamseldin A.Y., Melville B.W.: Comparative study of different wavelet based neural network models for rainfall–runoff modeling. Journal of Hydrology. 515, 47–58 (2014)
[15] Zhou J., Duan Z., Li Y., Deng J., Yu D.: PSO-based neural network optimization and its utilization in a boring machine. Journal of Materials Processing Technology. 178(13), 19–23 (2006)
[16] Mohandes M.A.: Modeling global solar radiation using particle swarm optimization PSO. Solar Energy. 86(11), 3137–3145 (2012)
[17] de Mingo López L.F., Blas N.G., Arteta A. The optimal combination: Grammatical swarm, particle swarm optimization and neural networks. Journal of Computational Science. 3(12), 46–55, (2012)
[18] Shores, T.S.: Applied Linear Algebra and Matrix Analysis. Springer, 291–293, (2007)
[19] Freeman J.A., Skapura D.M.: Neural Networks, Algorithms, Applications, and Programming Techniques. Addison-Wesley, California (1991)
[20] Eberhart R.C., Shi Y., Kennedy J.: Swarm Intelligence. Morgan Kaufmann, San Francisco CA (2001)
[21] Yang X.S.: Chapter 7. Particle Swarm Optimization: Nature-Inspired Optimization Algorithms. Elsevier. 99–110 (2014)
[22] National Commission of Transit Security, http://www.conaset.cl