Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (17): 3557-3562, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3557.3562

Logistics Demand Forecasting using KPCA-based Lssvr with Two-order Oscillating Particle Swarm Algorithm

[1]Li-Yan Geng and [2]Xi-Kui Lv
[1]School of Economics and Management, Shijiazhuang Tiedao University, Shijiazhuang, 050043, China
[2]Transportation Institute, Shijiazhuang Tiedao University, Shijiazhuang, 050043, China

Abstract: Logistics demand forecasting is an important step in the process of logistics system planning and development. The existing models for logistics demand forecasting often encounter problems of low forecasting accuracy and slow convergence speed. Consider these problems, a hybrid model called KPCA-LSSVR-TOOPSO was proposed to improve the forecasting accuracy and accelerate the convergence speed. The hybrid model integrated the Kernel Principal Component Analysis (KPCA), the Two-order Oscillating Particle Swarm Optimization (TOOPSO) and least squares support vector regression. First, the nonlinear features of the influential factors of logistics demand were extracted by KPCA. Then, the kernel principal components were put into LSSVR and a LSSVR model was established for logistics demand forecasting. Finally, TOOPSO algorithm was used to optimize the parameters in LSSVR. Empirical results from the China logistics demand indicate that the proposed model decreases the dimension of the modeling data. The minimum and maximum relative prediction errors of the proposed model are only -0.33548 and 5.3270%, respectively. The NRMSE, NMAE and MPE of the model are 0.1890, 0.1367 and 0.0182, which are smaller than ones of the other three models. The convergence speed of the proposed model is the fastest among the four models.

Key words: Logistics demand forecasting, least squares support vector regression, two-order oscillating particle swarm optimization algorithm

INTRODUCTION

Logistics system is a complex dynamic system and logistics demand is affected by many factors, such as the social economic factor, the social environmental factor, the national policy factor and so on. These factors have a complex impact on the logistics demand, which leads to the nonlinear relationship between the logistics demand and its influential factors. The traditional logistics demand forecasting methods, such as the regression analysis (Fite *et al*., 2001), the time series analysis (Adrangi *et al*., 2001), the grey forecasting model (He, 2008) and the rough set theory (Feng *et al*., 2010), can't obtain satisfactory results. In recent years, artificial neural network (ANN) is introduced into logistics demand forecasting (Geng *et al*., 2007). As a nonparametric method, ANN can approximate the nonlinear relationship between logistics demand and its influential factors well. However, ANN often meets with some problems in practical applications including an over-fitting problem and a local minimum point, which may have bad influence on forecasting accuracy of logistics demand.

Support Vector Machines (SVM), proposed by Vapnik (1995), is one of the latest machine learning methods, which is based on structural risk minimization. SVM solves the problems in ANN and achieves better logistics demand forecasting results relative to ANN (Hu and Lu, 2008). Least squares support vector regression (LSSVR) is the modified form to SVM for regression problems that simplified inequality constraints to equation constraints, which improves the computational efficiency (Suykens and Vandewalle, 1999). The selection of the parameters in LSSVR has direct impact on LSSVR performance. As a commonly used parameter selection method of LSSVR, cross validation method needs trials and tests, which will waste time and have difficult to obtain the optimal parameters. Two-order oscillating particle swarm optimization (TOOPSO) is an improved PSO algorithm. By introducing a two-order oscillating evolutionary equation, TOOPSO could adjust the global and local search capability of the algorithm and avoid the local optimization (Hu, 2007). The parameters of LSSVR will be optimized using TOOPSO algorithm in this study.

Corresponding Author: Li-Yan Geng, School of Economics and Management, Shijiazhuang Tiedao University, Shijiazhuang, 050043, China

If all the influential factors are used to forecasting logistics demand, the established forecasting model could become complex and then limit the popularization and application of the model. Liang *et al.* (2012), combined Kernel principal component analysis (KPCA) with LSSVR, to forecast logistics demand and the results showed that the combined model simplified the model structure and outperformed the LSSVR in logistics demand.

In this study, KPCA-LSSVR-TOOPSO model is proposed. The kernel principal components extracted from the influential factors of logistics demand using KPCA are selected as the input variables of LSSVR and TOOPSO algorithm is used to optimize the parameters of LSSVR.

THEORY

Kernel principal component analysis: Given the observed data x_j, R^m, for j = 1, 2,..., N, where N is the number of observed data. Kernel Principal Component Analysis (KPCA) maps the observed data into the high-dimensional feature space F using the nonlinear mapping function $\Phi(x)$ and the corresponding covariance matrix can be computed as:

$$C^F = \frac{1}{N}\sum_{j=1}^{N}\Phi(x_j)\Phi(x_j)^T \quad (1)$$

where, $\Phi(x_j)$, for j = 1, 2,..., N, is assumed to be mean centered and variance scaled. Then the principal components can be obtained by the following equation:

$$\lambda v = C^F v = \frac{1}{N}\sum_{j=1}^{N}\Phi(x_j, v)\Phi(x_j)^T \quad (2)$$

where, eigenvalues $\lambda = 0$ and $v \in F\backslash\{0\}$. All solutions v with $\lambda = 0$ must lie in the span of $\Phi(x_1),\ldots, \Phi(x_N)$, that is, there exist coefficients such that:

$$v = \sum_{j=1}^{N}\beta_j\Phi(x_j) \quad (3)$$

We can get the following equation by left multiplying $\Phi(x_j)$ with (2) for all j = 1, 2,..., N:

$$\lambda\,(\Phi(x_j)v) = \Phi(x_j).C^F v \quad (4)$$

Define a kernel matrix K of order N×N with elements as:

$$k_{ij} = \Phi(x_i)\times\Phi(x_j) = k(x_i, x_j) \quad (5)$$

By introducing a kernel function $k(x_i, x_j)$, we can compute inner products in F space without performing nonlinear mappings. The function that satisfies the mercer's condition should be as the kernel function. Some of widely used kernel functions are Gaussian kernel, polynomial kernel and sigmoid kernel. Before applying KPCA, the mean centering of $\Phi(x_j)$ in feature space should be performed. The centered kernel matrix $\bar{K}$ can be obtained from:

$$\bar{K} = K - l_N K - l_N + l_N K l_N \quad (6)$$

where, l_N is a N×N order matrix with elements as $l_{ij} = 1/N$ for i,j = 1,2,...,N. Then, the eigenvalue decomposition to $\bar{K}$ is written as:

$$N\lambda ß = \bar{K}ß \quad (7)$$

where, b_j (j = 1, 2,..., N) are the eigenvectors corresponding the eigenvalues l_j (j = 1, 2,..., N). The solution (l_j, b_j) should satisfy $l_j\,(b_j, b_j) = 1$. Then, the pth kernel principal component s_l of a test vector x is obtained by projecting $\Phi(x)$ onto the direction of the pth eigenvector:

$$s_p = v_p\Phi(x_i) = \sum_{i=1}^{N}\beta_{i,p}\,k(x_i, x_j) \quad (8)$$

METHODS

KPCA-based LSSVR model: The basic idea of KPCA-based LSSVR model is that the kernel principal components extracted using KPCA are selected as the input variables of LSSVR to reduce the dimension of modeling data.

Suppose there are a set of training data $L = \{(s_l, y_l)|l = 1,2,\ldots n\}$, where n is the number of data for training. The input vector s_l, R^d is the d-dimensional kernel principal components and y_l, R is the corresponding 1-dimensional output value. The input data are mapped into a hyperspace by the nonlinear mapping function $\Phi(s_l)$. The primal constrained optimization problem of the LSSVR model is obtained as below:

$$\min_{\omega,e} J(\omega, e) = \frac{1}{2}\|\omega\| + \frac{1}{2}\gamma\sum_{l=1}^{n}e_l^2 \quad (9)$$

$$y_l = \omega^T\varphi(s_l) + b + e_l,\ l = 1,\ldots,n \quad (10)$$

where, ω, b are the weight vector and bias constant value, respectively. e_l is the error variable and γ is the regularized

parameter. Solving this optimization problem in dual space, leads to transforms the primal problem into the following set of linear equations based on the Karush-Kuhn-Tucker (KKT) condition:

$$\begin{bmatrix} 0 & 1^T \\ 1 & K+I/\gamma \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix}=\begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (11)$$

where, $\alpha = [\alpha_1, \alpha_2,\ldots, \alpha_n]^T$, $1 = [1,\ldots,1]^T$, $Y = [y_1, y_2,\ldots, y_n]^T$ and I is an unit matrix of order n. K is the kernel function matrix in which the kernel function is defined as $k(s_l,s) = \Phi(s_l)^T\Phi(s)$ for l = 1,2,…,n. Here, the Gaussian kernel function with the form $k(s_l,s) = \exp(-\|s_l-s\|^2/\sigma^2)$ is considered, where σ^2 is the kernel parameter. Finally, the LSSVR model is given by:

$$y_l = \sum_{l=1}^{n} \alpha_l \exp(-\| s_l - s\|^2 / \sigma^2) + b \quad (12)$$

where, α_l(l = 1,2,…,n) are the Lagrange multipliers.

Optimal parameters of KPCA-based LSSVR optimized by TOOPSO: According to Eq. 11 and 12, there are two parameters needed to be decided: The kernel parameter σ^2 and the regularized parameter γ. Usually, cross validation method is used to select the parameters (γ, σ^2) of LSSVR. But this method is an experimental method. The parameters obtained may not be the optimal, which will have bad impact on forecasting performance of LSSVR.

To improve the forecasting ability of LSSVR, TOOPSO algorithm is used to choose the optimal parameters (γ, σ^2) of LSSVR in this paper. TOOPSO algorithm introduces two oscillating factors into the evolutionary equation to adjust the influence of the acceleration on the velocity, which can overcome the premature problem validly and increase the evolutionary speed. Let t denote the number of current iterations and t_{max} denote the number of maximal iterations. If $t<0.5t_{max}$, oscillating factors ξ_1 and ξ_2 and are taken as:

$$\xi_1 < (2\sqrt{c_1 r_1} - 1)/c_1 r_1,\ \xi_2 < (2\sqrt{c_2 r_2} - 1)/c_2 r_2 \quad (13)$$

If $t = 0.5t_{max}$, oscillating factors ξ_1 and ξ_2 are taken as:

$$\xi_1 \geq (2\sqrt{c_1 r_1} - 1)/c_1 r_1,\ \xi_2 \geq (2\sqrt{c_2 r_2} - 1)/c_2 r_2 \quad (14)$$

where, c_1 and c_2 are acceleration factors which control the maximum step size. r_1, r_2 are random numbers between zero and one. After the parameters (γ, σ^2) being obtained through the input and output data, the well trained LSSVR model can be applied to forecasting.

The procedures of the KPCA-LSSVR-TOOPSO model for forecasting are summarized as below.

Step 1: Preprocess the data. The whole dataset are normalized using the mean and standard deviation of each variable

Step 2: Extract the nonlinear feature. Choose the kernel function and compute the kernel matrix. Apply eigenvalue decomposition to the kernel matrix and extract nonlinear principal components according to Eq. 8

Step 3: Generate initial M sets of particles composed of the parameters (γ, σ^2). Set the parameters containing acceleration factors c_1 and c_2 and the number of maximal iterations t_{max}, the maximal and minimal inertia weight w_{max} and w_{min}

Step 4: The Fitness function is defined as follows:

$$E = \frac{1}{n}\sum_{l=1}^{n}(y_l - \hat{y}_l)^2 \quad (15)$$

where, l is the number of training samples. y_1 and $\hat{y}_2$ stand for the actual values and forecasting results from training samples, respectively.

Step 5: Compute the fitness value of each particle according to Eq. 15. Update the optimal particle position as the position corresponding to the minimum fitness value of the individual particle. Update the global optimal particle position as the position corresponding to the minimum fitness value of all the particles. The inertia weight *w* is automatically linearly decreased with the evolutionary process (Zhao *et al*., 2006)

Step 6: If the stopping criteria are met, the evolutionary process is terminated. The global optimum particle position corresponds to the optimal parameters (γ, σ^2), otherwise, go back to step

Step 7: The LSSVR model is established by the obtained optimal parameters (γ, σ^2) for forecasting logistics demand. Then the forecasts are transformed into the primal forecasts

EMPIRICAL RESEARCH

Data description: The data used in empirical research is China logistics data from 1991 to 2011, which comes from National Bureau of Statistics of China. Here the total social logistics costs (x_0) is selected as the reflection of the logistics demand.

According to the availability and operability of data, seventeen indicators are selected as the influential factors

Table 1: Results of KPCA

Components	Eigenvalue	Contribution rate (%)	Cumulative contribution rate (%)
1	0.0054	84.59	84.59
2	0.0005	7.66	92.26
3	0.0003	4.93	97.19
4	0.0001	1.38	98.57
5	0.0001	1.01	99.58
6	0.0000	0.20	99.77
7	0.0000	0.15	99.92
8	0.0000	0.04	99.96
9	0.0000	0.02	99.98
10	0.0000	0.01	99.99
11	0.0000	0.01	100.00
12	0.0000	0.00	100.00
13	0.0000	0.00	100.00
14	0.0000	0.00	100.00
15	0.0000	0.00	100.00
16	0.0000	0.00	100.00
17	0.0000	0.00	100.00

of the logistics demand. They are the gross domestic product (x_1) the total output value of the primary industry (x_2), the total output value of the second industry (x_3), the total output value of tertiary industry (x_4), the total investment in fixed assets (x_5), the business volume of postal and telecommunication services (x_6), the total value of imports and exports(x_7), the total retail sales of the consumer goods (x_8), the household consumption expenditure (x_9), the total freight traffic (x_{10}), the total freight ton-kilometers (x_{11}), the number of employed persons in logistics industry (x_{12}), the possession of civil trucks vehicles (x_{13}), the number of national owned railway freight cars (x_{14}), the possession of civil cargo vessels (x_{15}), the total population (x_{16}) and the retail price index (x_{17}).

Empirical process and results analysis: In KPCA, the kernel function should be selected first. In this paper, the Gaussian kernel function is used which is as the same to the kernel function in LSSVR. Results from the KPCA to the influential factors of logistics demand are shown in Table 1. It is clear that KPCA obtains good feature extraction effectiveness. The first three principal components extract 97.19% information of the original data and we select the first three principal components obtained as the input variables of LSSVR model.

Whole data is divided into two parts: The period from 1991 to 2005 is selected as the training samples for training model and the period from 2006 o 2011 is as the testing samples for testing the forecasting performance of the model.

The original parameters of TOOPSO algorithm are set as follows: M = 10, $c_1 = 0.2$, $c_2 = 1.8$, $w_{max} = 0.9$, $w_{min} = 0.1$, $t_{max} = 30$. To reduce the stochastic influence on parameters, the parameters of LSSVR is optimized

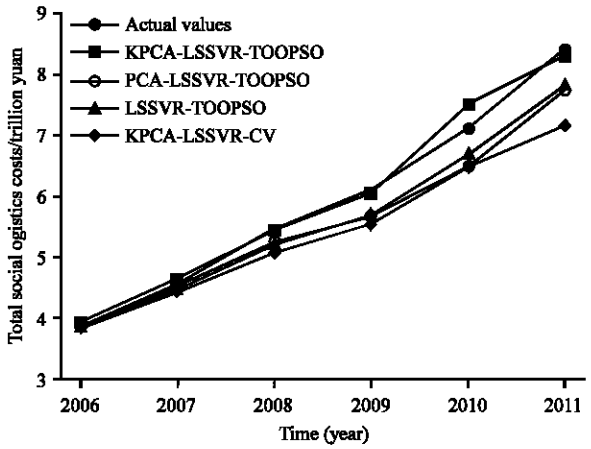

Fig. 1: Forecasts curves of the four models

continuously by TOOPSO algorithm for ten times, the optimal parameters obtained are used to establish LSSVR model.

The forecasting results of the KPCA-LSSVR-TOOPSO model are compared with those of the three models: PCA-LSSVR-TOOPSO, LSSVR-TOOPSO and KPCA-LSSVR-CV. In PCA-LSSVR-TOOPSO model, the twelve linear principal components extracted from the original influential factors using PCA are selected as the input variables of the LSSVR and the parameters of the LSSVR are optimized by the TOOPSO algorithm. In LSSVR-TOOPSO model, all of the original seventeen influential factors are directly put into the LSSVR and the parameters of the LSSVR are optimized by the TOOPSO algorithm. In KPCA-LSSVR-CV model, the three nonlinear principal components extracted from the original influential factors by KPCA are as the input variables of the LSSVR and the parameters of the LSSVR are optimized by five-fold cross validation method. The forecasting results are given in Fig. 1 and Table 2.

Based on Fig. 1 and Table 2, the KPCA-LSSVM-TOOPSO model forecasts the growing trends of the total social logistics costs better than the other three models. Apart from relative error of PCA-LSSVR-TOOPSO of 2007, the smaller relative errors are founded in KPCA-LSSVR-TOOPSO. And the minimum and maximum relative errors are only -0.33548 and 5.3270%, which is the smallest among the four models. This shows that as a whole, KPCA-LSSVR-TOOPSO model obtains higher forecasting accuracy than the other three models.

Three statistics are used to measure the forecasting performance of the four models. They are the Normalized Root Mean Squared Error (NRMSE), the Normalized Mean

Table 2: Forecasting results of the four models

	Actual values	KPCA-LSSVR-TOOPSO		PCA-LSSVR-TOOPSO		LSSVR-TOOPSO		KPCA-LSSVR-CV	
Year	(BY)	Forecasts (BY)	Relative error (%)	Forecasts (BY)	Relative error (%)	Forecasts (BY)	Relative error (%)	Forecasts (BY)	Relative error (%)
2006	3841	3895	1.4042	3854	0.3378	3827	0.3810	3808	0.8603
2007	4541	4621	1.7726	4504	0.7952	4462	1.7216	4419	2.6782
2008	5454	5436	-0.3354	5242	3.8948	5180	5.0345	5045	7.4972
2009	6083	6032	-0.8284	5655	7.0367	5680	6.6229	5536	8.9838
2010	7098	7477	5.3270	6478	8.7456	6661	6.1614	6472	8.8293
2011	8400	8296	-1.2411	7733	7.9354	7798	7.1643	7159	14.7698

Table 3: Forecasting performance of the four models

Models	NMSE	NMAE	MPE	Time
KPCA-LSSVR-TOOPSO	0.1890	0.1367	0.0182	5.10
PCA-LSSVR-TOOPSO	0.4772	0.3943	0.0479	7.33
LSSVR-TOOPSO	0.4140	0.3609	0.0451	7.97
KPCA-LSSVR-CV	0.7203	0.5938	0.0727	82.42

Absolute Error (NMAE) and the mean percentage error (MPE). The three statistics are expressed as follows:

$$NMSE = \sqrt{\sum_{q=1}^{P}(\hat{y}_q - y_q)^2 / \sum_{q=1}^{P}(y_{q-1} - y_q)^2} \quad (13)$$

$$NMAE = \sum_{q=1}^{P}\left|\hat{y}_q - y_q\right| / \sum_{q=1}^{P}\left|y_{q-1} - y_q\right| \quad (14)$$

$$MPE = P^{-1}\sum_{q=1}^{P}\left|(\hat{y}_q - y_q)/y_q\right| \quad (15)$$

where, y_q and $\hat{y}$ are the actual logistics costs and the forecasted logistics costs of different models, respectively. P is the number of the forecasted logistics costs. At the same time, the searching time for the optimal parameters of different model (TIME) is used to evaluate the convergence performance of the four models. The results are reported in Table 3.

From Table 3, it is obvious that the NMSE, NMAE, MPE of KPCA-LSSVR-TOOPSO model are smaller than ones of the other three models. Compare with PCA-LSSVR-TOOPSO, LSSVR-TOOPSO and KPCA-LSSVR-CV, KPCA-LSSVR-TOOPSO has the fastest convergence speed. And the convergence speed of the TOOPSO based three models is much faster than that under the five-fold cross validation method. Therefore, it is concluded that KPCA-LSSVR-TOOPSO model has a better performance than the other three models on logistics demand forecasting.

CONCLUSIONS

In this study, a KPCA-LSSVR-TOOPSO model is proposed, in which LSSVR model is combined with KPCA for forecasting logistics demand and TOOPSO algorithm is adopted to search the optimized parameters of LSSVR. An example on China logistics demand is taken to illustrate the effectiveness of the proposed model. Empirical results indicate that KPCA-LSSVR-TOOPSO model greatly simplifies the structure of the forecasting model. Moreover, it provides a higher logistics demand forecasting accuracy and a faster convergence speed relative to the PCA-LSSVR-TOOPSO model, LSSVR-TOOPSO model and KPCA-LSSVR-CV model. Further works can focus on using kernel independent component analysis (KICA) to reduce the dimension of the modeling data and choose other improved PSO algorithms to optimize the parameters needed in LSSVR.

ACKNOWLEDGMENTS

This study was supported in part by the National Natural Science Foundation of China "The research of spatial railway line selection based on a large proportion of the true three-dimensional geological entity modeling" (No. 51278316).

REFERENCES

Adrangi, B., A. Chatrath and K. Raffiee, 2001. The demand for US air transport service: A chaos and nonlinearity investigation. Transp. Res. E, 37: 337-353.

Feng, Y., Z.Y. Zhang, G.S. Xu and P.N. Wen, 2010. The forecasting of logistics demand in China based on rough set theory. Logistics Technol., 29: 60-62.

Fite, J.T., G.D. Taylor, J.S. Usher, J.R. English and J.N. Roberts, 2001. Forecasting freight demand using economic indices. Int. J. Phys. Distrib. Logistics Manage., 34: 299-308.

Geng, Y., S.D. Ju and Y.N. Chen, 2007. Analysis and forecast of logistics demand based on BP neural network. Logistics Technol., 26: 35-37.

He, G.H., 2008. Forecast of regional logistics requirements and application of grey prediction model. J. Beijing Jiaotong Univ., 7: 33-37.

Hu, J., 2007. A two-order particle swarm optimization model. J. Comput. Res. Dev., 44: 1825-1831.

Hu, Y.Z. and H.Y. Lu, 2008. Study on logistics demand forecast model based on support vector regression. Logistics Technol., 27: 66-68.

Liang, Y.G., L.Y. Geng and Z.F. Zhang, 2012. Forecast of regional logistics demand based on KPCA-LSSVM. Railway Transp. Econ., 34: 63-67.

Suykens, J.A.K. and J. Vandewalle, 1999. Least squares support vector machine classifiers. Neural Process. Lett., 9: 293-300.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. 1st Edn., Springer-Verlag, New York, USA.

Zhao, B., C.X. Guo, B.R. Bai and Y.J. Cao, 2006. An improved particle swarm optimization algorithm for unit commitment. Int. J. Elect. Power Energy Syst., 28: 482-490.