Comparison of Different Independent Component Analysis Algorithms for Sales Forecasting

Wensheng Dai, Jui-Yu Wu and Chi-Jie Lu*

***Abstract*—**Sales forecasting is one of the most important issues in managing information technology (IT) chain store sales since an IT chain store has many branches. Integrating feature extraction method and prediction tool, such as support vector regression (SVR), is a useful method for constructing an effective sales forecasting scheme. Independent component analysis (ICA) is a novel feature extraction technique and has been widely applied to deal with various forecasting problems. But, up to now, only the basic ICA method (i.e. temporal ICA model) was applied to sale forecasting problem. In this paper, we utilize three different ICA methods including spatial ICA (sICA), temporal ICA (tICA) and spatiotemporal ICA (stICA) to extract features from the sales data and compare their performance in sales forecasting of IT chain store. Experimental results from a real sales data show that the sales forecasting scheme by integrating stICA and SVR outperforms the comparison models in terms of forecasting error. The stICA is a promising tool for extracting effective features from branch sales data and the extracted features can improve the prediction performance of SVR for sales forecasting.

***Keywords*—**Sales forecasting, spatial ICA, spatiotemporal ICA, temporal ICA.

I. Introduction

INDEPENDENT component analysis (ICA) is one of the most used methods for blind source separation (BSS) which is to separate the source from the received signals without any prior knowledge of the source signal[1]. The goal of ICA is to recover independent sources when given only sensor observations that are unknown mixtures of the unobserved independent source signals. It has been investigated extensively in image processing, time series forecasting and statistical process control [1-4].

For time series forecasting problems, the first important step is usually to use feature extraction to reveal the underlying/interesting information that can't be found directly from the observed data. The performance of predictors can be improved by using the features as inputs [4-8]. Therefore, the two-stage forecasting scheme by integrating feature extraction method and prediction tool is a well-known method in literature [5-7]. The basic ICA is usually used as a novel feature extraction technique to find independent sources (i.e. features) for time series forecasting [1]. The independent sources called independent components (ICs) can be used to represent hidden information of the observable data. The basic ICA has been widely applied in different time series forecasting problems, such as stock price forecasting, exchange rate forecasting and product demand forecasting [9-11].

The basic ICA was originally developed to deal with the problems similar to the "cocktail party" problem in which many people are speaking at once. It assumed the extracted ICs are independent in time (independence of the voices) [12]. Thus, the basic ICA is also called temporal ICA (tICA). However, for some application data such as biological time-series and functional magnetic resonance imaging (fMRI) data, it is more realistic assumed that the ICs are independent in space (independent of the images or voxel) [13-14]. This ICA model is called spatial ICA (sICA). Besides, spatiotemporal ICA (stICA) based on the assumption that there exist small dependences between different spatial source data and between different temporal source data is also proposed [13-14]. In other words, stICA maximizes the degree of independence over space as well as over time, without necessarily producing independence in either space or time [13-14]. In short, there are three different ICA algorithms. tICA seeks a set of ICs which are strictly independent in time. On the contrary, sICA tries to find a set of ICs which are strictly independent in space. stICA seeks a set of ICs which are not strictly independent over time nor space.

Sales forecasting is one of the most crucial challenges for managing the information technology (IT) chain store sales since an IT chain store has many branches. By predicting consumer demand before selling, sales forecasting helps to determine the appropriate number of products to keep in inventory, thereby preventing over- or under-stocking. Because of the IT chain store's volatile environment, with rapid changes to product specifications, intense competition, and rapidly eroding prices, constructing an effective sales forecasting model is a challenging task.

The sales of a branch of an IT chain store may be affected by other neighboring branches of the same IT chain store. Therefore, to forecast sales of a branch, the historical sales data of this branch and its neighboring branches will be good

Wensheng Dai is with the Financial School, Renmin University of China, Beijing, People's Republic of China.

Jui-Yu Wu is with the Department of Business Administration, Lunghwa University of Science and Technology, Taiwan.

Chi-Jie Lu is with the Department of Industrial Management, Chien Hsin University of Science and Technology, Taoyuan County 32097, Taiwan, ROC (corresponding author's; e-mail: jerrylu@uch.edu.tw; chijie.lu@gmail.com).

predictor variables. The historical sales data of the branches of an IT chain store are highly correlated in space or time or both. Thus, three different ICA algorithms are used in this study to extract features from the branch sales data of an IT chain store. The feature extraction performance of the three different ICA algorithms is compared by using the two-stage forecasting scheme.

In this study, we propose a sales forecasting model for the branch of an IT chain store by integrating ICA algorithms and support vector regression (SVR). SVR based on statistical learning theory is an effective neural network algorithm and has been receiving increasing attention for solving nonlinear regression estimation problems. The SVR is derived from the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error [15]. Due to the advantages of the generalization capability in obtaining a unique solution, SVR can lead to great potential and superior performance in practical applications. It has been successfully applied in time series forecasting problem [4,5,11].

In the proposed sales forecasting scheme, we first use three different ICA algorithms (i.e., tICA, sICA and stICA) on the predictor variables to estimate ICs. The ICs can be used to represent underlying/hidden information of the predictor variables. The ICs are then used as the input variables of the SVR for building the prediction model. In order to evaluate the performance of the three different ICA algorithms, a real branch sales data of an IT chain store is used as the illustrative example.

The rest of this paper is organized as follows. Section 2 gives brief overviews of temporal ICA, spatial ICA and Spatiotemporal ICA and SVR. The sales forecasting scheme is described in Section 3. Section 4 presents the experimental results and this paper is concluded in Section 5.

II. Methodology

A. Temporal, Spatial and Spatiotemporal ICA

Let $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_\mathrm{m}]^\mathrm{T}$ be an input matrix of size $m\times n, m\le n$, consisting of observed mixture signals x_i of size $1\times n, i=1,2,\ldots,m.$ Suppose that the singular value decomposition (SVD) of $\boldsymbol{X}$ is given by $\boldsymbol{X}=\boldsymbol{UDV}^T$, where $\boldsymbol{U}\in R^{m\times k}, k\le m$ corresponds to eigenarrays, $\boldsymbol{V}\in R^{n\times k}$ is associated with eigensequences, and $\boldsymbol{D}$ is a diagonal matrix containing singular values. Following the notations in Stone et al. [14], it is defined $\widetilde{\boldsymbol{X}}=\boldsymbol{UDV}^T=\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^T$, where $\widetilde{\boldsymbol{U}}=\boldsymbol{UD}^{1/2}$ and $\widetilde{\boldsymbol{V}}=\boldsymbol{VD}^{1/2}$.

For temporal ICA (tICA), it embodies the assumption that $\widetilde{\boldsymbol{V}}$ can be decomposed: $\widetilde{\boldsymbol{V}}=\boldsymbol{PA}_P$, where $\boldsymbol{A}_P$ is an $k\times k$ mixing matrix, and $\boldsymbol{P}$ is an $n\times k$ matrix of k statistically independent temporal signals. tICA can be used to obtain the decomposition $\boldsymbol{p}_T=\widetilde{\boldsymbol{V}}\boldsymbol{W}_P$. $\boldsymbol{W}_P$ is a permuted version of $\boldsymbol{A}_P^{-1}$. The vector $\boldsymbol{p}_T$ is a set of extracted temporal signals and is a scale version of exactly one column vector in matrix $\boldsymbol{P}$. This is achieved by maximizing the entropy $h_T=H(\boldsymbol{Y})$ of $\boldsymbol{Y}=\tau(\boldsymbol{P}_T)$, where τ is approximates the cdf of the temporal source signals.

For spatial ICA (sICA), it is assumed that $\widetilde{\boldsymbol{U}}$ can be decomposed as $\widetilde{\boldsymbol{U}}=\boldsymbol{SA}_S$, where $\boldsymbol{A}_S$ is an $k\times k$ mixing matrix and $\boldsymbol{S}$ is an $m\times k$ matrix of k statistically independent spatial signals. sICA can be applied to generate the decomposition $\boldsymbol{y}_S=\widetilde{\boldsymbol{U}}\boldsymbol{W}_S$, where $\boldsymbol{W}_S$ is a permuted version of $\boldsymbol{A}_S^{-1}$. The vector $\boldsymbol{y}_S$ is a scale version of exactly one column vector in matrix $\boldsymbol{S}$ and is a set of extracted spatial signals. This is achieved by maximizing the entropy $h_S=H(\boldsymbol{Z})$ of $\boldsymbol{Z}=\delta(\boldsymbol{y}_S)$, where δ is approximates the cdf of the spatial source signals.

In spatiotemporal ICA (stICA), it is trying to find the decomposition $\widetilde{\boldsymbol{X}}=\boldsymbol{S\Lambda P}^T$, where $\boldsymbol{S}$ is an $m\times k$ matrix with a set of k statistically independent spatial signals, $\boldsymbol{P}$ is an $n\times k$ matrix of k mutually independent temporal signals, and $\boldsymbol{\Lambda}$ is a diagonal scaling matrix and is required to ensure that $\boldsymbol{S}$ and $\boldsymbol{P}$ have amplitudes appropriate to their respective cdfs δ and τ. Under the condition of $\widetilde{\boldsymbol{X}}=\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^T$, there exist two $k\times k$ un-mixing matrices $\boldsymbol{W}_P$ and $\boldsymbol{W}_S$, such that $\boldsymbol{P}=\widetilde{\boldsymbol{V}}\boldsymbol{W}_P$ and $\boldsymbol{S}=\widetilde{\boldsymbol{U}}\boldsymbol{W}_S$. Then, if $\boldsymbol{W}_S\boldsymbol{\Lambda}\boldsymbol{W}_T=\boldsymbol{I}$, the following relation holds: $\widetilde{\boldsymbol{X}}=\boldsymbol{S\Lambda P}^T=\widetilde{\boldsymbol{U}}\boldsymbol{W}_S\boldsymbol{\Lambda}(\widetilde{\boldsymbol{V}}\boldsymbol{W}_P)^T=\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^T$. We can estimate the $\boldsymbol{W}_P$ and $\boldsymbol{W}_S$ by maximizing an objective function associated with spatial and temporal entropies at the same time. That is, the objective function for stICA has the following form: $h_{ST}(\boldsymbol{W}_S,\boldsymbol{\Lambda})=\alpha H(\boldsymbol{Z})+(1-\alpha)H(\boldsymbol{Y})$, where α (0.5 is used in this study) defines the relative weighting for spatial entropy and temporal entropy. More details on tICA, sICA and stICA can be found in [12-14].

B. Support Vector Regression

Support vector regression (SVR) is an artificial intelligent forecasting tool based on statistical learning theory and structural risk minimization principle [15]. The SVR model can be expressed as the following equation [15]:

$$f(x)=(\boldsymbol{z}\emptyset(x))+b \tag{1}$$

where $\mathbf{z}$ is weight vector, b is bias and $\emptyset(x)$ is a kernel function which use a non-linear function to transform the non-linear input to be linear mode in a high dimension feature space.

Traditional regression gets the coefficients through minimizing the square error which can be considered as empirical risk based on loss function. Vapnik [15] introduced so-called ε-insensitivity loss function to SVR. Considering empirical risk and structure risk synchronously, the SVR model can be constructed to minimize the following programming:

$$\text{Min:}\frac{(\boldsymbol{z}^T\boldsymbol{z})}{2}+C\sum_i(\varphi_i+\varphi_i^*)$$

$$\text{Subject to:}\begin{cases} y_i-\boldsymbol{z}^Tx_i-b\le\varepsilon+\varphi_i \\ \boldsymbol{z}^Tx_i+b-y_i\le\varepsilon+\varphi_i^* \\ \varphi_i,\varphi_i^*\ge 0\end{cases} \tag{2}$$

where $i=1,\ldots,n$ is the number of training data; $(\varphi_i+\varphi_i^*)$ is the empirical risk; ε defined the region of ε-insensitivity, when the predicted value falls into the band area, the loss is zero. Contrarily, if the predicted value falls out the band area, the loss is equal to the difference between the predicted value and the margin; $\frac{\boldsymbol{z}^T\boldsymbol{z}}{2}$ is the structure risk preventing over-learning and lack of applied universality; C is modifying coefficient representing the trade-off between empirical risk and structure risk.

After selecting proper modifying coefficient (C), width of band area (ε) and kernel function ($\emptyset(x)$), the optimum of each parameter can be resolved though Lagrange function. The performance of SVR is mainly affected by the setting of parameters C and ε [4,5]. There are no general rules for the choice of C and ε. This study uses exponentially growing sequences of C and ε to identify good parameters. The parameter set of C and ε which generate the minimum forecasting mean square error (MSE) is considered as the best parameter set.

III. Sales Forecasting Scheme

This study uses a two-stage sales forecasting scheme. In this scheme, we use different ICA algorithms as feature extraction method and utilize support vector regression as prediction tool. The schematic representation of the two-stage sales forecasting scheme is illustrated in Fig. 1.

As shown in Fig. 1, the first step of the ICA-SVR scheme is data scaling. In this step, the original datasets and prediction variables are scaled into the range of [-1.0, 1.0] by utilizing min-max normalization method. The min-max normalization method converts a value x of variable X to x' in the range [-1.0, 1.0] by computing: $x' = -1 + \frac{2(x-minX)}{(maxX-minX)}$, where $maxX$ and $minX$ are the maximum and minimum values for attribute/variable X.

Then, the three different ICA algorithms including tICA, sICA and stICA are used in the scaled data to estimate ICs. In the third step, the ICs contained hidden information of the prediction variables are used as input variables to construct SVR sales forecasting model.

Since this study uses three ICA algorithms to extract features, based on the two-stage scheme, four sales forecasting methods including tICA-SVR, sICA-SVR, t-stICA-SVR and s-stICA-SVR are presented in this study. For the tICA-SVR and sICA-SVR methods, the tICA and sICA are used as feature extraction method, respectively. As stICA generates two different sets of ICs which are respectively are can be used to represent the temporal and spatial ICs, t-stICA-SVR corresponding to tICA and s-stICA-SVR corresponding to sICA are generated.

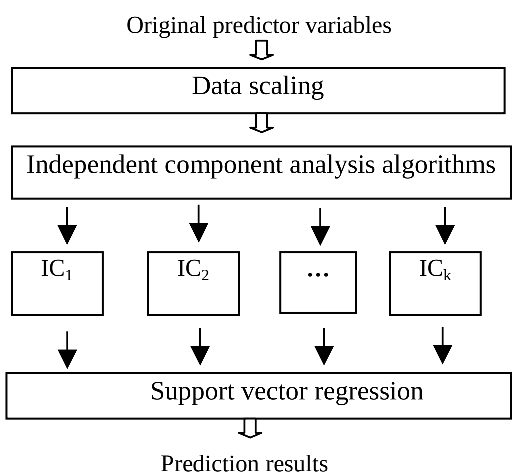

Fig. 1. The two-stage sales forecasting scheme.

IV. Experimental Results

For evaluating the performance of the three different ICA algorithms for sales forecasting for IT chain store, a real weekly branch sales dataset of an IT chain store is used in this study. This data contains 12 branches. There are totally 96 data points in each branch. The first 70 data points (72.9% of the total sample points) are used as the training sample and the remaining 26 data points (27.1% of the total sample points) are used as testing sample. Fig. 2 shows the sales data of the target branch. Fig. 3 shows the sales data of the 11 neighboring branches. From Fig.s 2 and 3, it can be seen that sales trend between the target branch and 11 neighboring branches are similar. The historical data of the 11 neighboring branches can be used as predictor variables for forecasting sales of the target branch. Therefore, the previous week's sales volume (T-1) of the target branch and the 11 neighboring branches are used as 12 predictor variables in this study. The input matrices $\mathbf{X_{tr}}$ of size 12×70 and $\mathbf{X_{te}}$ of size 12×26 are then generated for training stage and testing stage, respectively.

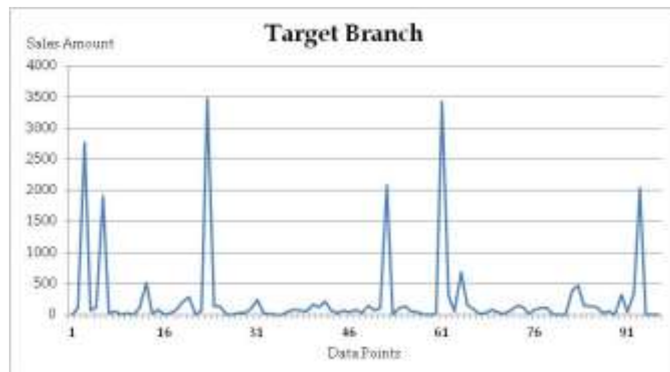

Fig. 2. The sales data of the target branch.

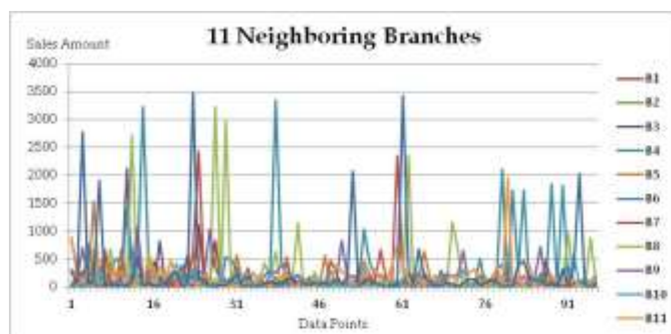

Fig. 3. The sales data of the 11 neighboring branches.

The prediction results of the four ICA-SVR sales forecasting scheme including tICA-SVR, sICA-SVR, t-stICA-SVR and s-stICA-SVR methods are compared to the SVR model without using ICA for feature extraction (called the single SVR model). The prediction performance is evaluated using the following statistical metrics, namely, the root mean square error (RMSE), mean absolute difference (MAD) and mean absolute percentage error (MAPE). RMSE, MAD and MAPE are measures of the deviation between actual and predicted values. The smaller values of RMSE, MAD and MAPE, the closer are the predicted time series values to that of the actual value.

TABLE I
THE MODEL SELECTION RESULTS OF THE SINGLE SVR MODEL

C	ε	Training MSE	Testing MSE
2^{9}	2^{-5}	0.0521	0.0688
	2^{-7}	0.0537	0.0678
	2^{-9}	0.0552	0.0667
	2^{-11}	0.0547	0.0664
2^{11}	2^{-5}	0.0712	0.0582
	2^{-7}	**0.0399**	**0.0492**
	2^{-9}	0.0407	0.0523
	2^{-11}	0.0407	0.0567
2^{13}	2^{-5}	0.0539	0.0673
	2^{-7}	0.0555	0.0655
	2^{-9}	0.0561	0.0676
	2^{-11}	0.0572	0.0672

TABLE II
THE MODEL SELECTION RESULTS OF THE tICA-SVR MODEL

C	ε	Training MSE	Testing MSE
2^{7}	2^{-3}	0.0401	0.0541
	2^{-5}	0.0391	0.0523
	2^{-7}	0.0380	0.0515
	2^{-9}	0.0377	0.0495
2^{9}	2^{-3}	0.0396	0.0537
	2^{-5}	0.0388	0.0529
	2^{-7}	0.0379	0.0541
	2^{-9}	0.0379	0.0592
2^{11}	2^{-3}	0.0390	0.0609
	2^{-5}	**0.0374**	**0.0455**
	2^{-7}	0.0376	0.0753
	2^{-9}	0.0395	0.0508

In the modeling of single SVR model, the scaled values of the 12 predictor variables are directly used as inputs. In selecting the parameters for modeling SVR, the parameter set ($C=2^{11}$, $\varepsilon=2^{-7}$) is used as the start point of grid search for searching the best parameters. The testing results of the SVR model with combinations of different parameter sets are summarized in Table I. From Table I, it can be found that the parameter set ($C=2^{11}$, $\varepsilon=2^{-7}$) gives the best forecasting result (minimum testing MSE) and is the best parameter set for single SVR model.

For the tICA-SVR model, first, the original predictor variables are scaled and then passed to tICA model to estimate ICs, i.e. features. The ICs are then used for building SVR forecasting model. As the same process with above single SVR, the parameter set ($C=2^{9}$, $\varepsilon=2^{-7}$) is used as the start point of grid search. Table II summaries the testing results of the tICA-SVR model with combinations of different parameter sets. Table II shows that the parameter set ($C=2^{11}$, $\varepsilon=2^{-5}$) gives the best forecasting result and is the best parameter set for the tICA-SVR model.

Using the similar process, the sICA-SVR, t-stICA-SVR and s-stICA-SVR models uses, respectively, the sICA and stICA for generating spatial ICs, temporal ICs of stICA and spatial ICs of stICA from the scaled predictor variables and then utilizes the features as inputs of the SVR models. Tables III to V shows the model selection results of the sICA-SVR, t-stICA-SVR and s-stICA-SVR models, respectively. From the tables, it can be found that the best parameter sets for sICA-SVR, t-stICA-SVR and s-stICA-SVR models are ($C=2^{9}$, $\varepsilon=2^{-9}$), ($C=2^{11}$, $\varepsilon=2^{-7}$) and ($C=2^{11}$, $\varepsilon=2^{-9}$), respectively.

TABLE III
THE MODEL SELECTION RESULTS OF THE sICA-SVR MODEL

C	ε	Training MSE	Testing MSE
2^{9}	2^{-7}	0.0410	0.0639
	2^{-9}	**0.0390**	**0.0478**
	2^{-11}	0.0395	0.0791
	2^{-13}	0.0415	0.0533
2^{11}	2^{-7}	0.0416	0.0564
	2^{-9}	0.0407	0.0555
	2^{-11}	0.0398	0.0568
	2^{-13}	0.0398	0.0622
2^{13}	2^{-7}	0.0421	0.0568
	2^{-9}	0.0411	0.0549
	2^{-11}	0.0399	0.0541
	2^{-13}	0.0396	0.0520

TABLE IV
THE MODEL SELECTION RESULTS OF THE t-stICA-SVR MODEL

C	ε	Training MSE	Testing MSE
2^{7}	2^{-3}	0.0324	0.0438
	2^{-5}	0.0316	0.0423
	2^{-7}	0.0307	0.0417
	2^{-9}	0.0305	0.0401
2^{9}	2^{-3}	0.0320	0.0434
	2^{-5}	0.0314	0.0428
	2^{-7}	0.0306	0.0438
	2^{-9}	0.0306	0.0479
2^{11}	2^{-3}	0.0304	0.0610
	2^{-5}	0.0315	0.0493
	2^{-7}	**0.0303**	**0.0369**
	2^{-9}	0.0320	0.0411

TABLE V
THE MODEL SELECTION RESULTS OF THE s-stICA-SVR MODEL

C	ε	Training MSE	Testing MSE
2^{9}	2^{-7}	0.0343	0.0459
	2^{-9}	0.0334	0.0453
	2^{-11}	0.0325	0.0463
	2^{-13}	0.0322	0.0506
2^{11}	2^{-7}	0.0320	0.0463
	2^{-9}	**0.0314**	**0.0421**
	2^{-11}	0.0326	0.0440
	2^{-13}	0.0324	0.0423
2^{13}	2^{-7}	0.0334	0.0521
	2^{-9}	0.0320	0.0389
	2^{-11}	0.0321	0.0644
	2^{-13}	0.0338	0.0435

The forecasting results using the tICA-SVR, sICA-SVR, t-stICA-SVR, s-stICA-SVR and single SVR models are computed and listed in Table VI. Table VI depicts that the RMSE, MAD and MAPE of the t-stICA-SVR model are, respectively, 20.369, 8.140 and 0.06%. It can be observed that these values are smaller than those of the tICA-SVR, sICA-SVR, s-stICA-SVR and single SVR models. Therefore, the t-stICA-SVR model can generate the best forecasting result.

Moreover, it also can be seen from the table that s-stICA-SVR outperform the tICA-SVR, sICA-SVR and single SVR models. Since the t-stICA-SVR and s-stICA-SVR can provide better forecasting results than the tICA-SVR and sICA-SVR models, stICA can estimate more effective ICs and improve sales forecasting performance for IT chain store. It indicates that stICA is a promising tool for extracting effective features from branch sales data and the extracted features can improve the prediction performance of SVR for sales forecasting. Besides, from the Table VI, we find that temporal

ICs are more suitable for forecasting branch sales since the forecasting performance of t-stICA-SVR and tICA-SVR models are better than that of s-stICA-SVR and sICA-SVR models, respectively.

TABLE VI
THE SALES FORECASTING RESULTS USING THE tICA-SVR, sICA-SVR, t-stICA-SVR, s-stICA-SVR AND SINGLE SVR MODELS

Models	RMSE	MAD	MAPE
tICA-SVR	70.596	17.551	0.13%
sICA-SVR	104.946	50.635	0.21%
t-stICA-SVR	**20.369**	**8.140**	**0.06%**
s-stICA-SVR	33.757	10.922	0.13%
Single SVR	115.529	55.741	0.24%

V. CONCLUSION

Forecasting sales of branches is a crucial aspect of the marketing and inventory management in IT chain store. In this paper, we used three different ICA algorithms including tICA, sICA and stICA for sales forecasting and compared the feature extraction performance of the three different ICA algorithms. Four sales forecasting methods including tICA-SVR, sICA-SVR, t-stICA-SVR and s-stICA-SVR were presented in this study. In the proposed sales forecasting methods, we first used three different ICA algorithms (i.e., tICA, sICA and stICA) on the predictor variables to estimate ICs. The ICs can be used to represent underlying/hidden information of the predictor variables. The ICs are then used as the input variables of the SVR for building the prediction model. A real weakly sales data of an IT chain store was used for evaluating the performance of the sales forecasting methods. Experimental results showed that the t-stICA-SVR and s-stICA-SVR models can produce the lowest prediction error. They outperformed the comparison methods used in this study. Thus, compared to tICA and sICA algorithms, stICA can estimate more effective ICs and improve sales forecasting performance for IT chain store. Moreover, we also found that, compared to spatial ICs, the temporal ICs are more suitable features for forecasting branch sales.

ACKNOWLEDGMENT

This work is partially supported by the National Science Council of the Republic of China, Grant no. NSC 102-2221-E-231-012-.

REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley & Sons, 2001.

[2] C. J. Lu, Y. E. Shao and B. S. Li, "Mixture control chart patterns recognition using independent component analysis and support vector machine," *Neurocomputing*, vol. 74, pp. 1908-1914, 2011.

[3] C. J. Lu, "An independent component analysis-based disturbance separation scheme for statistical process monitoring," *Journal of Intelligent Manufacturing*, vol. 23, pp. 561-573, 2012.

[4] C. J. Lu, T. S. Lee, C. C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," *Decision Support Systems*, vol. 47, pp. 115-125, 2009.

[5] L. J. Kao, C. C. Chiu, C. J. Lu and J. L. Yang, "Integration of nonlinear independent component analysis and support vector regression for stock price forecasting," *Neurocomputing*, vol. 99, pp. 534-542, 2013..

[6] W. Dai, J. Y. Wu and C. J. Lu, "Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes," *Expert Systems with Applications*, vol. 39, pp. 4444-4452, 2012.

[7] C. J. Lu, Y. E. Shao, "Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine," *Mathematical Problems in Engineering*, vol. 2012, article ID 831201, 15 pages, 2012.

[8] L. J. Kao, C. C. Chiu, C. J. Lu and C. H. Chang, "A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting," *Decision Support Systems*, vol 54, pp. 1228-1244, 2013.

[9] K. Kiviluoto, and E. Oja, "Independent component analysis for parallel financial time series", in *Proceedings of the Fifth International Conference on Neural Information*, Tokyo, Japan, 1998, pp. 895–898.

[10] C. J. Lu, J. Y. Wu, T. S. Lee, "Application of independent component analysis preprocessing and support vector regression in time series prediction," in *The 2009 International Symposium on Applied Computing and Computational Sciences (ACCS 2009)*, Sanya, Hainan Island, China, 2009, pp. 468-471.

[11] C. J. Lu and Y. W. Wang. "Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting," *International Journal of Production Economics*, vol. 128, pp. 603-613, 2010.

[12] V. D. Calhoun, T. Adali, G.D. Pearlson, and J. J. Pekar, 2001, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms." *Hum Brain Mapping*, vol. 13, pp. 43-53, 2001.

[13] J. V. Stone, *Independent Component Analysis: A Tutorial Introduction*, MA : MIT Press, 2004.

[14] J. V. Stone, J. Porrill, N. R. Porter, and I. D. Wilkinson, "Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions," *Neuroimage*, vol. 15, pp. 407-21, 2002.

[15] V. N. Vapnik, *The Nature of Statistical Learning Theory*, NY: Springer-Verlag, 2000.