Available online at www.sciencedirect.com
ScienceDirect
Procedia Social and Behavioral Sciences
Procedia - Social and Behavioral Sciences 129 (2014) 601 – 609

ICIMTR 2013

International Conference on Innovation, Management and Technology Research, Malaysia, 22 – 23 September, 2013

Gasoline Price Forecasting: An Application of LSSVM with Improved ABC

Zuriani Mustaffa[a*], Yuhanis Yusof,[b] Siti Sakira Kamaruddin[c]

[a,b,c]School of Computing, Universiti Utara Malaysia, 06010, UUM Sintok, Kedah, Malaysia

Abstract

Optimizing the hyper-parameters of Least Squares Support Vector Machines (LSSVM) is crucial as it will directly influence the predictive power of the algorithm. To tackle such issue, this study proposes an improved Artificial Bee Colony (IABC) algorithm which is based on conventional mutation. The IABC serves as an optimizer for LSSVM. Realized in gasoline price forecasting, the performance is guided based on Mean Absolute Percentage Error (MAPE) and Root Mean Square Percentage Error (RMSPE). The conducted simulation results show that, the proposed IABC-LSSVM outperforms the results produced by ABC-LSSVM and also the Back Propagation Neural Network.

Keywords: Least Squares Support Vector Machines; Artificial Bee Colony; Price forecasting.

1. Introduction

A novel machine learning technique, namely Least Squares Support Vector Machines (LSSVM) (Suykens, Van Gestel, De Brabanter, De Moor, & Vandewalle, 2002) is a variant of Vapnik's theory, viz. Support Vector Machines (Vapnik, 1995). As a modified version, LSSVM encompasses similar advantages as SVM, but LSSVM offers a linear system of equation rather than Quadratic Programming. This is achieved by applying equality constraints instead of inequality constraints and uses square errors

* Corresponding author.
E-mail address: zuriani.m@gmail.com

doi:10.1016/j.sbspro.2014.03.718

to replace the nonnegative errors in the problem formulation (Suykens, et al., 2002). Hence, such an approach simplifies the training process of a standard SVM to a great extent.

In LSSVM, there are two hyper-parameters of interest, namely regularization parameter, γ and kernel parameter, σ^2. Both play a significant role for the algorithm performance. An inappropriate selection of hyper-parameters value may cause the LSSVM prediction model vulnerable to over fitting and under fitting problem (Cheng, Qian &Guo, 2006). With respect to that matter, empirically derivation of the parameter values is inefficient since it would lead to selection of randomness. Hence the results produced may be unreliable (Dos Santos, Luvizotto, Mariani, & Dos Santos Coelho, 2012). Nevertheless, to date, the manual selection for parameters of interest still being utilized, such as presented in Zhai and Huang (2013) in solving marketing prediction.

From the literature, it is observed that, apart from manual selection approach, basically there are two common approaches in optimizing the LSSVM hyper-parameters. First is Cross Validation (CV) procedure in grid search and second is theoretical technique (Afshin, Sadeghian, & Raahemifar, 2007). Example of CV-LSSVM is such as presented in short term prediction of meteorological time series data (Mellit, Massi Pavan, & Benghanem, 2013). In the study, the comparison of proposed model was made against ANN based prediction technique and evaluated using five metrics: Root Mean Square Error (RMSE), Coefficient of determination (R^2), Mean Bias Error (MBE), Mean Absolute Percentage Error (MAPE) and Kolmogorov-Smirnov (KS) test. The finding of the study suggested that CV-LSSVM offers lower error rate as compared to the others. However, as CV requires an exhaustive search over the parameter space, in term of time–wise, it is inefficient (Zhang, Niu, Li, & Li, 2013). Besides, in term of error rate, the produced results tend to be unsatisfying (Yu, Chen, Wang, & Lai, 2009), such as examined in Xiang and Jiang (2009).

Meanwhile, the second approach which is based on theoretical, it includes meta-heuristic search algorithm such as Evolutionary Algorithm (EA) and Swarm Intelligence (SI) technique (Afshin, Sadeghian, & Raahemifar, 2007). Categorized as subset of EAs, Genetic Algorithm (GA) has been widely applied in optimizing LSSVM hyper-parameters and the application of GA-LSSVM can be seen in various fields (Fu, Liu, & Sun, 2010; Sulaiman, Mustafa, Aliman, Khalid, & Shareef, 2012; Yang, Gu, Liang, & Ling, 2010). Besides GA, the application of Particle Swarm Optimization (PSO) (Jiang & Zhao, 2013; Zhou & Shi, 2010) is also encouraging in the literature. In 2005, another optimization technique, namely Artificial Bee Colony (ABC) (Karaboga, 2005) has been introduced and its performance is proven to be competitive to the other existing optimization approaches. Even it is still considered as relatively new in the optimization community, the extensive application in various areas give an impressive feedback to its performance (Bolaji, Khader, Al-Betar, & Awadallah, 2013). Nonetheless, besides its excellent capability, there is always a room for improvement. From literature, it is observed that the ABC inclined to fall into local minimum (Gao & Liu, 2012).

Motivated from such situation, this study proposes an improved ABC (IABC) for optimizing LSSVM parameters. The improvement is put forward to ABC as it assists the LSSVM in achieving satisfactory generalization capability. The remainder of this paper is structured as follows: The next section provides a brief review on LSSVM while the details of ABC and its improvement is provided in Section 3. Section 4 describes the hybridization of IABC and LSSVM while Section 5 emphasizes the experiment task. The results are provided in Section 6 while Section 7 concludes the findings of the study.

2. Regression based on Least Squares SUpport Vector Machines

Given a training set of N points $\{x_i, y_i\}^N$ with the input values x_i and the output values y_i, for nonlinear regression, the objective is to estimate a model of the following form (Suykens, et al., 2002):

$$y(x) = w^T \varphi(x^i) + b + e_i \quad (1)$$

where w is the weight vector, b is the bias and e_i is the error between the actual and predicted output at the ith sample point. The input, x_i and output, y(x) are described in Section 5. The coefficient vector w and b can be obtained through the optimization problem which is formulated as follows (Suykens, et al., 2002):

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \quad (2)$$

Subject to the equality constraints

$$y_i = w^T \varphi(x_i) + b + e_i \text{ , } i = 1, 2 , \ldots, N$$

Applying the Lagrange multiplier to (2) yields:

$$L(w,b,e;\alpha) = J(w,e) - \sum_{i=1}^{N} \alpha_i \{w^T \varphi(x_i) + b + e_i - y_i\} \quad (3)$$

where α_i are Lagrange multipliers, γ is the regularization parameter which balances the complexity of the LSSVM model, i.e. $y(x)$, and the training error. Differentiating (3) with w, b, e_i *and* α_i, the Karush-Kuhn-Tucker (KKT) conditions for optimality of this problem can be obtained by setting all derivatives equal to zero, as express in the following:

$$\begin{aligned} &\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i \\ &\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ &\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \\ &\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0 \end{aligned} \qquad i = 1, 2, \ldots, \mathrm{N} \quad (4)$$

By elimination of w and e_i, the optimization problem can be transformed into the following linear equations:

$$\left[\frac{\alpha}{b}\right]\left[\frac{0}{1_v} \; \frac{1_v^T}{\Omega + I/\gamma}\right] = \left[\frac{0}{y}\right] \quad (5)$$

The resulting of LSSVM model for regression in (1) becomes:

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{6}$$

where α and b are the solutions of (5). In (6), there are several available kernel functions $K(x, x_i)$, namely Radial Basis Function (RBF) kernel, Multilayer Perceptron (MLP) kernel or quadratic kernel. In this study, the RBF kernel is used. It is expressed as:

$$K(x, x_i) = e^{\frac{\|x - x_i\|^2}{2\sigma^2}} \tag{7}$$

where σ^2 is a tuning parameter which is associated with RBF kernel. Another tuning parameter, which is regularization parameter, γ can be seen in (2).

3. Optimization using Artificial Bee Colony

The ABC algorithm (Karaboga, 2005) is motivated from intelligent behavior of honey bees. In the proposed model, the artificial colony consists of three groups of bee, namely Employed Bee (EB), Onlooker Bee (OB) and Scout Bee (SB). Each of them plays a significant role in achieving a common objective, which is maximizing the amount of nectar. The main steps of the algorithm are as follows:

A. Initialization Phase

Initial food sources are produced randomly within the range of the boundaries of the parameters (Karaboga, 2005) viz. γ, σ^2 (see Section 2). This is expressed as follows:

$$x_{ij} = x_j^{\min} + rand(0,1)(x_j^{\max} - x_j^{\min}) \tag{8}$$

where i = 1,…, SN, j = 1,…, D. SN is the number of food sources and D is the number of parameters of interest. Meanwhile, $x_j^{\max}$ and $x_j^{\min}$ are the upper and lower bound of parameters of interest respectively. After initialization of the population, the fitness of food source is calculated and is defined as follows (Karaboga, 2005):

$$fit_i = \frac{1}{(1 + obj.Fun_i)} \tag{9}$$

where *obj.Fun* is the objective function.

B. Employed Bee Phase

The size of EB and OB are both equal to the number of food sources, denoted by *SN*. For each food source's position, an EB is assigned to it. A new food source is obtained according to (10) (Karaboga, 2005):

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}) \tag{10}$$

Where i = 1,2,…, SN; j = 1,2,…, D; φ = a randomly generalized real number within the range [-1, 1], k = is a randomly selected index number in the colony. It must be noted that k has to be different from i. In (10), if the produced values of the parameters produced are exceeding their boundary, they are automatically switched onto the boundary values (Karaboga & Akay, 2009). After producing the new solution, v_{ij}, its fitness is calculated and compared to the previous solution, x_{ij}. If the new solution is better than previous one, the bee memorizes the new solution; otherwise she memorizes the previous solution.

C. Onlooker Bee Phase
The OB selects a food source to be exploited with the probability values related to the fitness values of the solution. This probability is calculated using the following equation (Karaboga, 2005):

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \tag{11}$$

where fit_i is the fitness of the solution *v*. *SN* is the number of food sources positions. Later, the OB searches a new solution in the selected food source site using (10), the similar way as EBs exploit. As the quality of the food increases, the number of OB visiting the food source increases too. It should be noted that even though both EB and OB phases use similar way for food source exploitation, however, in EB phase, every solution will be updated while in OB phase, only the selected one will go through that process.

D. Scout Bee Phase
In SB phase, if the fitness of a found food source by EB has not been improved for a given number of trial (denoted by limit), it is abandoned. The EB of that food source becomes a SB and makes a random search using (12).

$$x_{id} = x_d^{\min} + r\left(x_d^{\max} - x_d^{\min}\right) \tag{12}$$

where r = a random real number within the range [0-1] ; $x_d^{\min}$ and $x_d^{\max}$ = the lower and upper borders in the *d*th dimension of the problem space.

3.1 Improved ABC

The improvement introduced in this study is based on mutation approach (Haupt & Haupt, 1998). The main objective is to prevent the model from falling into local minimum. As highlighted before, in ABC algorithm, if the generated parameter value exceeds the pre-defined boundary, it is automatically switched onto the boundaries. However, in IABC, if such situation occurs, mutation approach is employed. With the boundaries are set to [1, 1000], it is expressed as follows:

$$new\_param = (ub - lb) * r \tag{13}$$

where *new_param* = new parameter; *ub* = upper bound ; *lb* = lower bound; *r* = random number between [0-1].

3.2 Parameters Optimization of LSSVM based on IABC

In optimizing the hyper-parameters of LSSVM utilizing IABC, each food source position represents a possible solution, viz. parameters combination. In this study, as RBF kernel is selected, the combination of parameters of interest are γ and σ^2. Meanwhile, the fitness function (i.e. generalization performance indicator) of each possible solution is evaluated using MAPE (see section 5.3). In the proposed study, LSSVM will be embedded in IABC algorithm as fitness function evaluation, so the optimized value of parameters of interest can be obtained after a maximum number of iteration has been achieved. The objective function to be minimized is MAPE, where the lower the MAPE, the better the prediction accuracy.

4. Experiments: Algorithm Implementation

4.1 Data Description

In this study, four energy fuels prices time series data were selected as empirical samples. The output variable was the daily spot price one month into the future (21 trading days) of gasoline (HU) prices. The samples covered is from December 1, 1997 to November 27, 2002 and available at *Barchart* (Barchart, 2012). From the data set, 70% is set for training, 15% for validation and the remaining 15% is reserved as testing. Table 1 shows the input and output variables utilized in the simulation. Besides the daily closing price, another three derivative features were also being fed to the forecasting model. The purpose is to assist the model in learning the underlying relationship that is constant over time (Malliaris & Malliaris, 2008).

Table 1: Input and Output Variables

Input	Variable	Output
Daily closing price of crude oil, heating oil, gasoline and propane	CL, HO, HU, PN	HU21owd
Percent change (%Chg) in daily closing spot prices from the previous day of CL, HO, HU and PN	CL%Chg, HO%Chg, HU%Chg, PN%Chg	
Standard deviation (sd) over the previous 5 days trading days of CL, HO, HU and PN	CLsd5, HOsd5, HUsd5, PNsd5	
Standard deviation (sd) over the previous 21 days trading days of CL, HO, HU and PN	CLsd21, HOsd21, HUsd21, PNsd21	

4.2 Data Normalization

Prior to training, data normalization was performed using Min Max Normalization (Al-Shalabi, Shaaban, & Kasasbeh, 2006) where each feature component is normalized into specific range. This ensure that larger input value do not overwhelm the smaller one.

4.3 Evaluation Criteria

To assess the forecasting performance, two commonly utilized evaluation criteria, viz. MAPE and Root Mean Square Percentage Error (RMSE) are employed in this work (Hyndman & Koehler, 2006). Both metrics interpret the generalization capability of experimented model in prediction.

5. Empirical Results

The hybridization of IABC-LSSVM is realized using LSSVMlab Toolbox (Pelkmans et al., 2002). The properties of IABC/ABC are defined as follows: *SN* = 20, *MCN* = 100 and *Limit* = *SN**D*. Table 2 summarizes the results obtained upon completing the experiment for HU price prediction. From the table, the researching results of parameters achieved by IABC are γ = 979.2209 and σ^2=112.8787. With such combination, the MAPE produced was 5.6199%, which is the lowest among all experimented techniques. Hence, better prediction accuracy is achieved. On the other hand, the MAPE recorded by ABC-LSSVM is 6.3779% while the BPNN produced 12.3283% of MAPE. As for ABC-LSSVM, too small value of σ^2 has made the model being exposed to over fitting which consequently affect the predictive capability. The visual result is illustrated in Figure 1. It can be seen that the prediction value produced by BPNN are strayed widely from the target value while closer prediction value is offered by IABC-LSSVM.

Table 2: HU Price Prediction

	IABC-LSSVM	ABC-LSSVM	BPNN
γ	979.2209	377.9901	-
σ^2	112.8787	1	-
MAPE (%)	5.6199	6.3799	12.3283
RMSPE (%)	0.0704	0.0821	0.1533

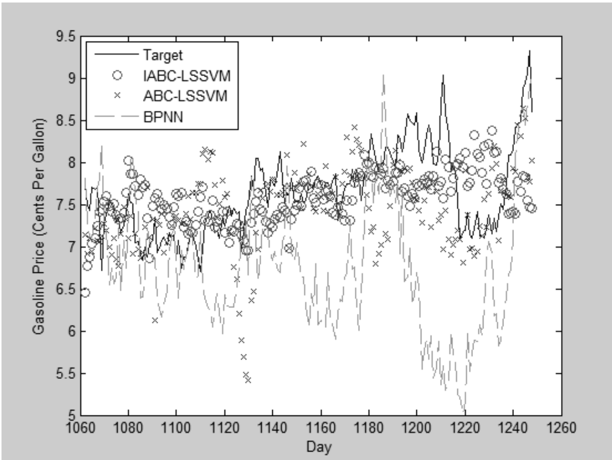

Figure 1: Hu Price Prediction

6. Conclusion

This paper presents a new parameter selection method for LSSVM, which is based on IABC algorithm. The improvement introduced in IABC addresses the local minimum issue found in standard ABC. Tested on gasoline (HU) price forecasting, the simulation results indicated that the IABC-LSSVM is superior than ABC-LSSVM and also BPNN. Thus, it is concluded that the proposed technique is an effective approach for the context study.

References

Afshin, M., Sadeghian, A., & Raahemifar, K. (2007). *on efficient tuning of ls-svm hyper-parameters in short-term load forecasting: a comparative study.* Procedings of the IEEE Power Engineering Society General Meeting

Al-Shalabi, L., Shaaban, Z., & Kasasbeh, B. (2006). Data mining: a preprocessing engine. *Journal of Computer Science, 2*(9), 735-739.

Bolaji, A. L. a., Khader, A. T., Al-Betar, M. A., & Awadallah, M. A. (2013). Artificial bee colony algorithm, its variants and applications: a survey. *Journal of Theoritical and Applied Information Technology, 47*(2), 434-459.

Cheng, J., Qian, J.-S., & Guo, Y.-N. (2006). Least squares support vector machines for gas concentration forecasting in coal mine. *International Journal of Computer Science and Network Security, 6*(6), 125-129.

Dos Santos, G. S., Luvizotto, L. G. J., Mariani, V. C., & Dos Santos Coelho, L. (2012). Least squares support vector machines with tuning based on differential evolution approach applied to the identification of a thermal process. *Expert Systems with Applications, 39*(2012), 4805-4812.

Fu, H., Liu, S., & Sun, F. (2010). *Ship motion prediction based on AGA-LSSVM.* Proceedings of the International Conference on Mechatronics and Automation (ICMA).

Gao, W., & Liu, S. (2012). A modified artificial bee colony. *Computers & Operations Research, 39*(2012), 687-697.

Haupt, R. L., & Haupt, S. E. (1998). *Practical genetic algorithms*. New York: Wiley Pub.

Hyndman, R. J., & Koehler, . B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecastin, 22*(2006), 679-688.

Jiang, B. T., & Zhao, F. Y. (2013). Particle swarm optimization-based least squares support vector regression for critical heat flux prediction. *Annals of Nuclear Energy 53*(2012), 69-81.

Karaboga, D. (2005). *An Idea based on honey bee swarm for numerical optimization* (Technical Report): Erciyes University.

Karaboga, D., & Akay, B. (2009). A comparative study of artificial bee colony. *Applied Mathematics and Computation, 214*(2009), 108-132.

Malliaris, M. E., & Malliaris, S. G. (2008). Forecasting inter-related energy product prices. *The European Journal of Finance, 14*(6), 453-468.

Mellit, A., Massi Pavan, A., & Benghanem, M. (2013). Least squares support vector machine for short-term prediction of meteorological time series. *Theory Application Climatol, 2013*(111), 297-307.

Pelkmans, K., Suykens, J. A. K., Gestel, T. V., Brabanter, J. D., Lukas, L., Hamer, B., et al. (2002). LS-SVMlab: A Matlab/c toolbox for least squares support vector machines. from www.esat.kuleuven.be/sista/lssvmlab/.

Sulaiman, M. H., Mustafa, M. W., Aliman, O., Khalid, S. N. A., & Shareef, H. (2012). Real and reactive power flow allocation in deregulated power system utilizing genetic-support vector machines technique. *International Review of Electrical Engineering, 5*(5), 2199-2208.

Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least squares* *support vector machines*. Leuven, Belgium: World Scientific Publishing Co. Pte. Ltd.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory 2nd ed.* New York: Springer-Verlag.

Yang, Z., Gu, X. S., Liang, X. Y., & Ling, L. C. (2010). Genetic algorithm-least squares support vector regression based predicting and optimizing model on carbon fiber composite integrated conductivity. *Materials and Design, 31*(2010), 1042-1049.

Yu, L., Chen, H., Wang, S., & Lai, K. K. (2009). Evolving least squares support vector machines for stock market trend mining. *IEEE Transactions on Evolutionary Computation, 13*(1), 87-102.

Zhang, W., Niu, P., Li, G., & Li, P. (2013). Forecasting of turbine heat rate with online least squares support vector machine based on gravitational search algorithm *Knowledge-Based Systems, 39*(2013), 34-44.

Zhou, B., & Shi, A. (2010). *LSSVM and Hybrid particle swarm optimization for ship motion prediction.* Proceedings of the International Conference on Intelligent Control and Information Processing (ICICIP).