

A Hybridization of Enhanced Artificial Bee Colony-Least Squares Support Vector Machines for Price Forecasting

Zuriani Mustaffa and Yuhanis Yusof
School of Computing, College of Arts and Sciences,
University Utara Malaysia, Kedah, Malaysia

Abstract: Problem statement: As the performance of Least Squares Support Vector Machines (LSSVM) is highly rely on its value of regularization parameter, γ and kernel parameter, σ^2 , man-made approach is clearly not an appropriate solution since it may lead to blindness in certain extent. In addition, this technique is time consuming and unsystematic, which consequently affect the generalization performance of LSSVM. **Approach:** This study presents an enhanced Artificial Bee Colony (ABC) to automatically optimize the hyper parameters of interest. The enhancement involved modifications that provide better exploitation activity by the bees during searching and prevent premature convergence. Later, the prediction process is accomplished by LSSVM. **Results and Conclusion:** Empirical results obtained indicated that feasibility of proposed technique showed a satisfactory performance by producing better prediction accuracy as compared to standard ABC-LSSVM and Back Propagation Neural Network.

Keywords: Artificial Bee Colony (ABC), energy fuel price time series, financial prediction, Least Squares Support Vector Machines (LSSVM)

INTRODUCTION

Time series prediction corresponds to uncertainty. It is vital an essential in various fields which leads people to study and present numerous of approaches in order to predict what will happen in the future. The tendency to choose a better prediction tools is inevitable for several reasons. Some of them are to allow proper plans to be executed about potential developments, reduce unwanted risk that may cause serious loss, no matter in form of non material or otherwise, reducing cost and in particular field, to provide better customer service (Makridakis and Hibon, 2000).

In the field of finance, there are many predictor tools that have been used since years ago, especially conventional statistical prediction tool. One of the most common applied is Box Jenkins method, which is technically known as Autoregressive Integrated Moving Average (ARIMA) (Palit and Popovic, 2005). The major shortcoming of ARIMA is the pre assumption on the linearity of the model which is clearly not suitable with commodity price of interest. Such dataset deals with high nonlinearity and complexity (Liu, 2009) where the mean and variance of the data sets can change over time (Kumar and Thenmozhi, 2007). Another common used approach is exponential

smoothing method (Palit and Popovic, 2005). Nonetheless, this approach also faced similar problem as ARIMA where it is difficult to deal with non linear feature in time series (Lai *et al.*, 2006).

The difficulties and demerits that own by those techniques have pave a way for many researchers to suggest a Computational Intelligence (CI) based predictor tool, which is Artificial Neural Network (ANN) to overcome the complexity in choosing an appropriate prediction approach. ANN is one of the well known CI techniques to be helpful prediction tool. By applying ANN, the requirement for analyst in specifying form of the model is not compulsory (White, 1994). This technique also has enjoyed considerable success in prediction where a good numbers of published studies about it can be seen and the reserach is continuously carry on (Jammazi and Aloui, 2012).

However, there are several limitations in developing the model. The adaptation of Empirical Risk Minimization (ERM) in ANN leads to numerical instabilities and poor generalization performance (Xiang and Jiang, 2009). The complexity in determining a great volume of control parameters worsens the situation and make it more complicated to be applied (Kumar and Thenmozhi, 2007). In addition, due to the employment of gradient descent based training method, ANN exposed to

over fitting problem (Ying and Hua, 2008). All of these have restricted its further application.

As compared to ANN, the emerging of Statistical Learning Theory based method, namely Support Vector Machines (SVM) has been developed as a robust tool for regression and classification. This technique was pioneered by Vapnik (1995) where it comes with the adaptation of Structural Risk Minimization (SRM). As opposed to ERM, this scheme seeks to minimize an upper bound of generalization error rather than only minimizing the training error.

Since SVM is based on Quadratic Programming (QP) which arises computational burden problem, the emerging of variant SVM, namely Least Squares Support Vector Machines (LSSVM) (Suykens *et al.*, 2002) which leads to solve a set of linear equations makes it becomes favorable among academia from different fields.

For the purpose of parameter tuning of LSSVM, there are many approaches have been presented regardless of the application domain. To date, the implementation of Evolutionary Computation (EC) algorithm, which includes Evolutionary Algorithm (EA) and Swarm Intelligence (SI), can be seen as the most promising. As compared to man-made selection approach in determining the value of parameters of interest, the hybridization of LSSVM with these optimization techniques seems much reliable and systematic. Before the emerging of SI, EA techniques first gained interest among academia where the most prominent technique under the group was Genetic Algorithm (GA) where the hybridization of GA and LSSVM have been extensively done by several researchers from different areas (Sun and Zhang, 2008; Yu *et al.*, 2009; Mustafa *et al.*, 2011). This makes GA more dominant than other EA techniques, such as Evolutionary Programming (EP) and Genetic Programming (GP) (Karaboga *et al.*, 2012).

On the other hand, there has been an avalanche of studies since the emerging of SI technique and the hybridization between SI with LSSVM has become increasingly popular during the last decade. The SI technique mimics the intelligent behavior of swarm of social insect, flocks of birds, or school of fish (Blum and Li, 2008). Among them, the application of Particle Swarm Optimization (PSO) has been researcher's favorable choice due to its non-complicated concept and ease of use (Park *et al.*, 2010; Ping and Jian, 2009). The hybridization of PSO-LSSVM for chaotic time series prediction has been proposed by Ping and Jian (2009). The proposed method which was tested on several important chaotic time series proved the

feasibility and effectiveness by defeating ANFIS and Fuzzy Cluster. In related work, PSO-LSSVM has been presented in solving water quality prediction by Xiang and Jiang (2009). By comparing with several single prediction models, the proposed technique proved to be superior by yielding lowest Mean Absolute Percentage Error (MAPE). In other field, mid long term load prediction has been presented by Niu *et al.* (2009) utilizing an improved PSO-LSSVM. A bit different with both studies, a modification in inertia weight leads the algorithm to faster searching.

Another SI approach, namely Artificial Fish Swarm Algorithm (AFSA) has been presented in Chen *et al.* (2008) for predicting electricity load in China. By making PSO-LSSVM as a competitor in the study, AFSA-LSSVM proved its capability to escape from local minima by producing better prediction capability. Meanwhile, a hybridization of Ant Colony Optimization (ACO) and LSSVM can be seen in Fang and Bai (2009) in predicting share price.

On the other hand, a relatively new optimization technique, namely Artificial Bee Colony (ABC) (Karaboga and Akay, 2005) has captured much attention from academia since its invention. Besides its simplicity and ease of use, it is also comes with modest requirement and less control parameters to be tuned (Gao *et al.*, 2012). Interestingly, this Karaboga's approach has been proven to be competitive to other SI technique. The remarkable performance of ABC can be seen in many studies regardless of area, including in the hybridization of ABC-LSSVM in solving gold price prediction (Yusof and Mustafa, 2011).

In this study, several modifications are introduced for enhancing the capability of ABC. Firstly, Lévy Probability Distribution (LPD) will be integrated in improving the exploitation process of the bees and mutation approach is later introduced in preventing premature convergence. The proposed prediction model is tested on correlated energy fuel price time series, namely propane.

A brief review on support vector machines and least squares support vector machines:

The nonlinear technique of Vapnik, which is known as Support Vector Machines (SVM) (Vapnik, 1995) was first introduced in 1995. It is developed based Statistical Learning Theory. The elegance of SVM is due to the employment of kernel, which is often called the kernel trick (Suykens *et al.*, 2002). It offers the capability to map nonlinear input data into a high dimensional feature space that are essentially linear (Fig. 1). In the view of ANN, this feature space is known as high dimensional hidden layer (Suykens *et al.*, 2002).

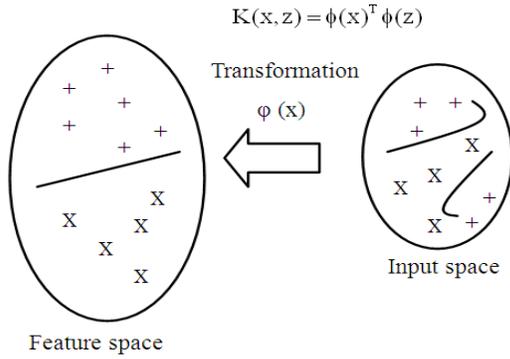


Fig. 1: Mapping of the input space to a high dimensional feature space (Suykens *et al.*, 2002)

Here, the optimization process can be duplicated as in the linear case (Sapankevych and Sankar, 2009). With the adaptation SRM, a main objective of Vapnik-Chervonenkis (VC) theory is minimize an upper bound of generalization error, which consequently leads to the superiority of SVM as compared to ANN (Kumar and Thenmozhi, 2007).

However, due to the employment of QP solvers in SVM, a computational burden of SVM is inevitable (Yu *et al.*, 2009). Due to that matter, a reformulation of SVM by Suykens *et al.* (2002), namely LSSVM comes to fill the gaps. This technique is reported to consume less computational effort in the huge-scale problem compared to standard SVM's. As a modified version of Vapnik's technique, LSSVM appears to outperforms SVM in many regression problems (Tarhouni *et al.*, 2011). By still maintaining the advantages of its original form, LSSVM appears with some user friendly properties regarding the implementation and computational time during training (Valyon and Horvath, 2007). Thus, LSSVM able to optimize more precise due to its short computation time (Wang and Hu, 2005). A brief introduction of LSSVM is presented here while a detailed description can be found in Suykens *et al.*, (2002).

The standard framework for LSSVM is based on the primal-dual formulation. Given the dataset $\{x_i, y_i\}_{i=1}^N$, the aim is to estimate a model of the form (Suykens *et al.*, 2002) Eq. 1:

$$y(x) = w^T \phi(x) + b + e_i \tag{1}$$

where, $x \in \mathbb{R}^n$, $y \in \mathbb{R}$ and $\phi(\cdot): \mathbb{R}^n \rightarrow \mathbb{R}^{n_h}$ is a mapping to a high dimensional feature space. The following optimization problem is formulated (Suykens *et al.*, 2002):

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

Subject to:

$$y_i = w^T \phi(x_i) + b + e_i, i = 1, 2, \dots, N$$

With the application of Mercer's theorem (Vapnik, 1995) for the kernel matrix Ω as $\Omega_{ij} = K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$, $i, j=1, \dots, N$ it is not required to compute explicitly the nonlinear mapping $\phi(\cdot)$ as this is done implicitly through the use of positive definite kernel functions K . From the Lagrangian function Eq. 3:

$$\zeta(w, b, e; \alpha) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i (w^T \phi(x_i) + b + e_i - y_i) \tag{3}$$

where, $\alpha_i \in \mathbb{R}$ are lagrange multipliers. Differentiating (3) with w, b, e_i and α_i , the conditions for optimality can be described as follow Eq. 4:

$$\begin{cases} \frac{\partial \zeta}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i \phi(x_i) \\ \frac{\partial \zeta}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial \zeta}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma_i e_i, i = 1, \dots, N \\ \frac{\partial \zeta}{\partial \alpha_i} = 0 \rightarrow y_i = w^T \phi(x_i) + b + e_i \end{cases} \tag{4}$$

By elimination of w and e_i , the following linear system is obtained Eq. 5:

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \gamma & \Omega + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{5}$$

With $y = [y_1, \dots, y_N]^T$, $\alpha = [\alpha_1, \dots, \alpha_N]^T$. The resulting LSSVM model in dual space becomes Eq. 6:

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

Usually, the training of the LSSVM model involves an optimal selection of regularization parameter γ and kernel parameter σ^2 . Several kernel functions, viz. Gaussian Radial Basis Function (RBF) Kernel, linear Kernel and quadratic Kernel are available. For this project, the RBF Kernel is used which is expressed as Eq. 7:

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

where, σ^2 is a tuning parameter which associated with RBF function. Another tuning parameter is γ which showed in Eq. 2.

Artificial bee colony: For the purpose of obtaining optimized hyper-parameters of LSSVM, a new SI population-based meta-heuristic approach, namely ABC is applied into LSSVM. ABC algorithm was first introduced by Karaboga (2005) for real parameter optimization. It was enlightened by simulating the foraging behavior of honey bee swarm. This technique has been proven to be uncomplicated structure and easy to be applied with modest requirement. In the standard ABC algorithm, the colony of artificial bees is classified into three groups of bees: employed bees which associated with specific food sources, onlooker bees which responsible to watch the dance of employed bees within the hive in order to choose a food source and scout bees which randomly search for food source. For onlookers and scouts, they are also known as unemployed bees (Karaboga *et al.*, 2012). Half of the colony consists of employed bees and the rest are of the onlooker bees. The number of food sources/nectar sources is equal with the employed bees, which means that one employed bee is responsible for a single nectar source. The objective of the whole colony is to maximize the amount of nectar. In ABC algorithm, the local and global exploration is combined, hence leads better balance in food exploitation and exploration of the bees (Dongli *et al.*, 2011).

The duty of employed bees is to search for food sources (solutions). Later, the amount of nectars (solutions' qualities/fitness value) is calculated. Then, the information obtained is shared with the onlooker bees which are waiting in the hive (dance area). The onlooker bees decide to exploit a nectar source depending on the information shared by the employed bees. For this purpose, the onlooker bees watch various dances before choosing a food source position where the duration of dance by employed bees is proportional to the nectar's content (fitness value) of food source currently being exploited by the employed bees. The onlooker bees also determine the source to be abandoned and allocate its employed bee as scout bees. For the scout bees, their task is to find the new valuable food sources. They search the space near the hive randomly.

In ABC algorithm, suppose the solution space of the problem is D-dimensional, where D is the number

of parameters to be optimized. The fitness value of the randomly chosen site is formulated as follows Eq. 8:

$$fit_i = \frac{1}{(1 + obj.Fun_i)} \quad (8)$$

The size of employed bees and onlooker bees are both SN, which is equal to the number of food sources. For each food source's position, one employed bee is assigned to it. For each employed bee whose total numbers are equal to the half of the food sources, a new source is obtained according to (9):

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (9)$$

Where:

$i = 1, 2, \dots, SN$

$j = 1, 2, \dots, D$

γ = a random generalized real number within the range[-1,1]

k = is a randomly selected index number in the colony

After producing the new solution, $v'_i = \{x'_{i1}, x'_{i2}, \dots, x'_{iD}\}$, it is compared to the original solution $v_i = \{x_{i1}, x_{i2}, \dots, x_{iD}\}$. If the new solution is better than previous one, the bee memorizes the new solution; otherwise she memorizes the former solution. The onlooker bee selects a food source to exploit with the probability Eq. 10:

$$P_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \quad (10)$$

where, fit_i is the fitness of the solution v . SN is the number of food sources portions. Later, the onlooker bee searches a new solution in the selected food source site by Eq. 9, the same way as exploited by employed bees. After all the employed bees exploit a new solution and the onlooker bees are allocated a food source, if a source is found that the fitness hasn't been improved for a given number (denoted by limit) steps, it is abandoned and the employed bee associated with it becomes a scout and makes a random search by Eq. 11:

$$x_{id} = x_d^{min} + r(x_d^{max} - x_d^{min}) \quad (11)$$

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 problems space

Basic steps of ABC algorithm are as follows:

- Initialize the food source positions (population)
- Each employed bees is assigned on their food sources
- Each onlooker bee select a source base on the quality of her solution produces a new food source in selected food source site and exploits the better source
- Decide the source to be cast aside and assign its employed bee as scout for discovering new food sources
- Memorize the best food source found so far
- If requirement are met, output the best solution, otherwise repeat steps 2-5 until the stopping criterion is met

MATERIALS AND METHODS

Materials and methods under study include research data and data preparation, correlation coefficient among time series employed, data proportion, data normalization, experiment setup, performance evaluation metric and finally enhanced of ABC

Research data and data preparation: In this study, data sets utilized consist of four energy fuels price time series which are highly correlated, namely Crude Oil (CL), Heating Oil (HO), gasoline (HU) and Propane (PN). All of time series are in daily basis, spanning from December 1997 to November 2002, which is similar as employed in Malliaris and Malliaris (2008).

Correlation:As to determine the correlation coefficient between the said time series, Pearson Product Moment Coefficient (Lomax, 2007) was utilized. In this study, it is defined as $r = 0.8-1.0$, which indicates as high correlation (Farid, 2010). Table 1 shows the correlation between the time series of interest.

Data proportion: Prior to conduct the experiment processes, the time series is split into three different sets, namely training, validation and testing set. The prediction model is build using the training set while to assess the performanceof the model, the validation set takes place. Finally, the testing set is used to assess the expected performance of the model (Williams, 2011). The proportion of each set is as tabulated in Table 2.

Table 1: Correlation between time series utilized

	CL	HO	HU	PN
CL	1.0000			
HO	0.9597	1.0000		
HU	0.9649	0.9262	1.0000	
PN	0.8422	0.8812	0.8473	1

Table 2: Number of samples for training, validation and testing processes

Data Types	Samples (Days)	Data Proportion (%)
Training	1-8740	70
Validation	875-1061	15
Testing	1062-1248	15

Table 3: Samples of original input

CL	HO	HU	PN
18.6300	0.5144	0.5338	0.3260
18.7000	0.5190	0.5316	0.3188
18.6000	0.5220	0.5328	0.3213
18.5900	0.5218	0.5272	0.3198
18.7000	0.5276	0.5262	0.3200

Table 4: Sample of normalized input

CL	HO	HU	PN
0.2989	0.2804	0.2970	0.1803
0.3015	0.2860	0.2943	0.1698
0.2977	0.2896	0.2957	0.1734
0.2973	0.2894	0.2889	0.1712
0.3015	0.2965	0.2877	0.1715

Table 5: Assigning input and output variables

Input	Variable	Output
Daily closing price of crude oil, heating oil, gasoline and propane	CL, HO, HU, PN	
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	PN21

Data normalization: The importance of data normalization has been emphasized in many studies (Shalabi *et al.*, 2006; Kotsiantis *et al.*, 2006) due to its capability in improving the prediction accuracy of the prediction model. In this study, Min Max Normalization is employed where the purpose is to independently normalize each feature component to specified range. This ensures that the larger input values do not overwhelm the smaller input values. The formula for min max normalization is as followed Eq. 12:

$$v' = \left((v - \min_a) / (\max_a - \min_a) \right) * (\text{new} - \max_a - \text{new} - \min_a) + \text{new} - \min_a \tag{12}$$

In min max normalization, it maps a value of v of A to v' in the range $[\text{new-max}_a, \text{new-min}_a]$ by solving the equation above. Data tabulated in Table 3 and 4 show the samples of original data used in this work and the normalized data.

Experiment setup: Data tabulated in Table 5 shows the variables assigned to features involved. The input arrangement employed in study includes four daily closing price of time series of interest and also another three derivation inputs, namely percent change in daily closing prices from the previous day, standard deviation over the previous 5 trading days and standard deviation over the previous 21 trading days. While for the output, the prediction begin from day 21 onwards of propane's price (PN21owd). The purpose of including the derivation variables is to helps the model to discover an underlying relationship that is constant over time (Malliaris and Malliaris, 2008).

Performance evaluation metric: In this study, two indexes were utilized in measuring the performance of the prediction model. The first metric is Mean Absolute Percentage Error (MAPE). MAPE is one of the most common evaluation metric applied in prediction (Klimberg *et al.*, 2010) while another metric is Prediction Accuracy (PA). The definitions of both metrics are as followed Eq. 13 and 14:

$$\text{MAPE} = \frac{1}{n} \left[\sum_{i,j=1}^n \left| \frac{y_i - p_i}{y_i} \right| \right] \quad (13)$$

Where:

$i, j = 1, 2, \dots, n$

y = Actual value

p = Prediction value

$$\text{PA} = \text{mean}(100\% - (\text{MAPE} \times 100)) \quad (14)$$

Enhanced artificial bee colony: In this study, the standard ABC-LSSVM is improved by introducing two modifications namely Levy Mutation and Simple Mutation.

Levy mutation to enhance bee's exploitation process: First modification that has been made is by applying Lévy Probability Distribution (LPD). The purpose is to enhance the exploitation process of the bees in standard ABC. This approach, termed as Lévy Mutation ABC (lvABC) is introduced in generating new food source, which involved employed and onlooker bee phase. The lvABC is based on LPD which is first introduced by (Levy, 1954):

$$lv_{\alpha,\theta}(y) = \frac{1}{\pi} \int_0^{\infty} e^{-\theta q} \cos(qy) dq \quad (15)$$

From (15), the distribution is symmetric with respect to $y = 0$ and has two parameters α and θ . α controls the shape of the distribution, requiring $0 < \alpha < 2$ while θ is the scaling factor satisfying $\theta > 0$. LPD possesses a power law in the tail region that is characteristic of the fractal structure of nature. Since the analytic form of Eq. 15 is unknown for general α , an algorithm to generate Levy random number is commonly utilized, such as suggested in Lee and Yao (2004) or algorithm by McCulloch (1996) from Ohio State University.

For θ , it can be set to $\theta = 1$ without loss of generality. To describe this, rescale y to $y' = by$ by with some constant b . Then, from (15), the following relation can be obtained:

$$L_{\alpha,\theta}(by) = \frac{1}{b} L_{\alpha,\theta}(y) \quad (16)$$

where, $\theta = \theta b^{-\alpha}$. In particular, by setting $\theta' = 1$, Eq. 16 and 17 becomes:

$$L_{\alpha,\theta}(y) = \gamma^{\frac{1}{\alpha}} L_{\alpha,1}(y) \quad (17)$$

Implying that θ is nothing but an overall scaling factor. Thus, with the distribution of $\theta = 1$, the distribution of any other θ can be obtained (Lee and Yao, 2004).

In standard ABC, the exploitation and exploration process carried out by employed and onlooker bees rely on the same equation, which is Eq. 9.

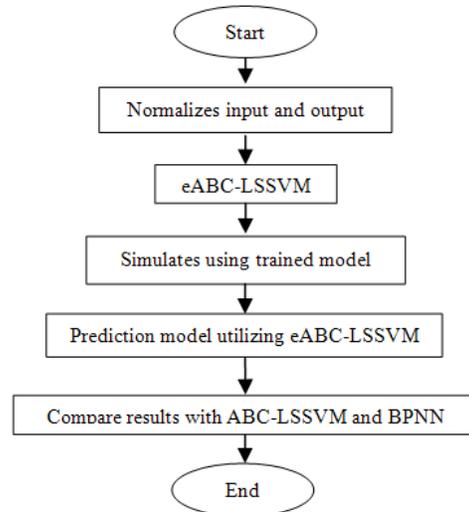


Fig. 2: Flow of flow of proposed correlated non renewable natural resources commodity price prediction model

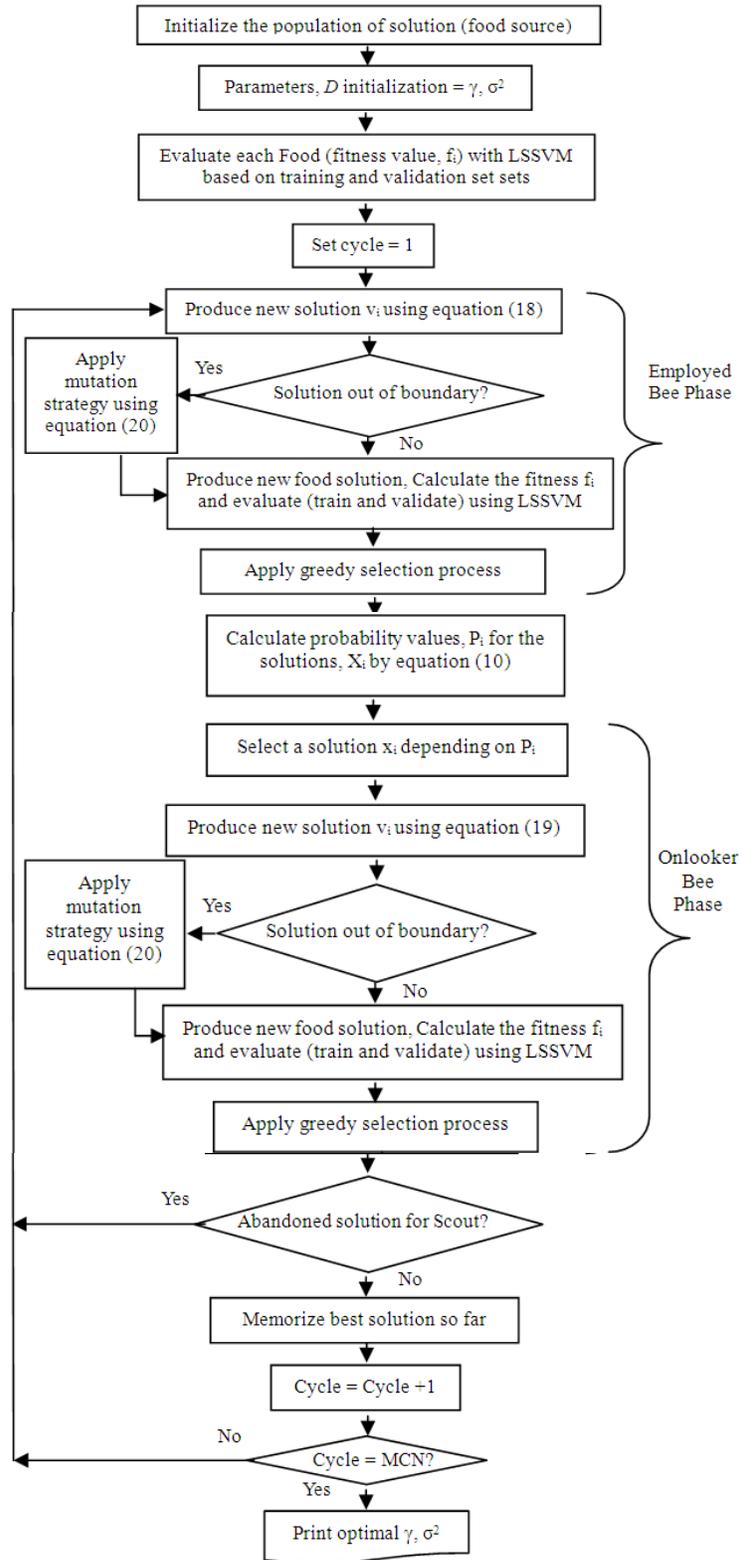


Fig. 3: Flow of eABC-LSSVM
1686

However, in the proposed approach, for employed bee phase, instead of applying Eq. 9, the following equation is introduced:

$$v_{ij} = x_{ij} + (x_{ij} - x_{kj-}) + lv \tag{18}$$

On the other hand, in onlooker bee phase, the following equation is used as to replace Eq. 19:

$$v_{ij} = x_{ij} + lv \tag{19}$$

Equations 18 and 19 are designed through experimental approach. The main objective is to emphasize the exploitation process at the strong (promising) solutions. In this study, all parameters involved are automatically tuned by eABC-LSSVM.

Simple Mutation to Prevent Premature Convergence: Secondly, as to prevent the premature convergence, another modification that has been made is by applying a mutation approach, termed as Simple Mutation ABC (SMABC). This modification involved in determining the value of parameters of interest. In ABC-LSSVM (Yusof and Mustafa, 2011), if the generated parameter value is out of boundaries, it is automatically shifted onto the boundaries. Nonetheless, in smABC-LSSVM, instead of forcing the parameter value to the boundary, when the unwanted situation aforementioned occurred, instead of forcing the parameter value to the boundary, mutation strategy is introduced. This operation is executed by multiplying the generated random number with the range of boundary that has been determined. In this study, the boundaries are set to the range of between [1, 1000] Eq. 20:

$$\text{new_param} = (\text{ub} - \text{lb}) * \text{rand_num} \tag{20}$$

Where:

- New-param= New parameter
- rand_num = Random number
- ub = Upper bound
- lb = Lower bound

The purpose of this modification is to prevent the model from trapping in local minima which may cause premature convergence. This might happen if the area explored by the model is not a global minima area. By applying mutation strategy, the model is induced to explore other areas in order to look for global minima rather than local minima (Haupt and Haupt, 1998).

Figure 2 shows the simplified form of the proposed prediction model while the flow of eABC algorithm with the modification that has been made is illustrated in Fig. 3. From Fig. 3, it can be seen that both modifications, namely smABC-LSSVM and lvABC-LSSVM involved in both employed and onlooker bee phases. For smABC-

LSSVM, the modification is done after producing new food solution while lvABC-LSSVM involved in the phases of producing new food solution.

RESULTS

The proposed method is elaborated by designing an appropriate eABC-LSSVM in Matlab environment utilizing LSSVMlab Toolbox which can be obtained in Pelkmans *et al.* (2000). Prior to that, the properties of ABC and competitor chose in the study, namely Back Propagation Neural Network (BPNN) was determined, as tabulated in Table 6 while the empirical results were showed in Table 7. The prediction accuracy yielded by ABC-LSSVM was 87.4138% and is achieved using $\gamma = 135.4851$ and $\sigma^2 = 995.0188$. An increment of 0.1325 and 0.2105% prediction accuracy was recorded by smABC-LSSVM and lvABC-LSSVM respectively.

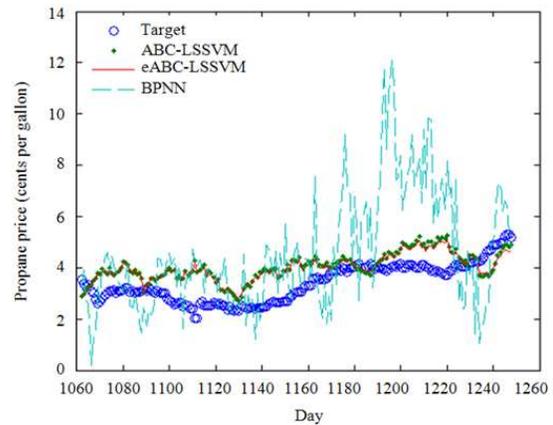


Fig. 4: Correlated energy fuel price prediction: propane

Table 6: Properties of control parameters for ABC and BPNN

Parameters	ABC	BPNN
Size of employed or onlooker bees (SN)	10	-
Limit	SN * D	-
Maximum Iteration	100	5000.0
Hidden Node	-	15.0
Learning rate	-	0.9
Momentum	-	0.7

Table 7: Correlated energy fuel price prediction: PN

	ABC -LSSVM	smABC -LSSVM	lvABC -LSSVM	eABC -LSSVM	BPNN
γ	135.4851	109.1151	82.5221	67.0728	-
σ^2	995.0188	984.0230	826.6000	927.0040	-
MAPE(%)	12.5862	12.4537	12.3757	12.0599	15.6629
PA(%)	87.4138	87.5463	87.6243	87.9401	84.3371

Table 8: Paired samples T-Test: PN

	Paired differences			95% Confidence Interval of the Difference		t	DF	Sig. (2-tailed)
	Mean	SD	Std. Error Mean	Lower	Upper			
	Pair 1ABC -LSSVM	0.0035	0.0075	0.0005	0.0024			
- eABC -LSSVM	465.0000	125.0000	494.0000	627.0000	303.0000	6.00	6	0.00

With the combination of both said approaches, the eABC-LSSVM able to produce higher prediction accuracy, which is 87.9401%, with the difference of 0.5263% from results obtained by standard ABC-LSSVM. For BPNN, the method was defeated with 3.0767% prediction accuracy by eABC-LSSVM. The visual result of experiment conducted is illustrated in Fig. 4 where the blue round mark represents the actual value, the red straight line represents eABC-LSSVM while the green dot and dashed line represents ABC-LSSVM and BPNN respectively.

As to test the significance level of modifications that have been made, T-Test was executed using SPSS (Hinton, 2004). As showed in Table 8, the significance level between pre modification (ABC-LSSVM) and post modification (eABC-LSSVM) was 0.000 (seelastcolumn). This indicates that there is significance difference in prediction accuracy when the eABC-LSSVM is utilized.

DISCUSSION

The performance of proposed technique was evaluated based on MAPE which reflect the PA obtained. From the empirical results, it shows that there is a positive increment of PA produced from standard ABC-LSSVM to eABC-LSSVM. Even increment in PA produced by smABC-LSSVM and lvABC-LSSVM are quite small, however when both techniques are combined, better PA can be obtained, as shown in Table 7. This shows that the combination of several mutation approach at different stages can lead to better performance of optimization techniques, which finally yield a good prediction result. By adopting mutation based on LPD, the search exploitation of the bees is enhanced due to a wider search region. This implies that large variation of mutation are quite likely and reduces the chance of revisiting the same food source. On the other hand, the integration of Simple Mutation in decision making process induced the bees to further the exploration instead of going to the boundary of the search space. This finally lead to better convergence as the maximum and minimum values of the boundary do not mean the optimal (Lendasse *et al.*, 2005).

CONCLUSION

This study reports the empirical results which examines the feasibility of proposed approach, eABC, which combines the modifications in lvABC and smABC when hybrid with LSSVM in predicting propane’s price. The highly dependence of the algorithm in the single equation for exploitation process has being considered by introducing lvABC. The advantage of standard ABC which comes with fast convergence feature also not taken for granted as it may increase the risk of facing with premature convergence. With respect to both matters, appropriate actions have been taken. With a modest requirement, the feasibility of proposed technique in predicting the time series of interest showed a satisfactory performance by producing better prediction accuracy than the competitors chosen under study.

REFERENCES

Blum, C. and X. Li, 2008. Swarm intelligence in optimization. *Natural Comput. Series*, 1: 43-85. DOI: 10.1007/978-3-540-74089-6_2

Chen, X., J. Wang, D. Sun and J. Liang, 2008. Time series forecasting based on novel support vector machine using artificial fish swarm algorithm. *Proceedings of the 4th International Conference on Natural Computation Oct. 18-20, IEEE Xplore Press, Jinan*, pp: 206-211. DOI: 10.1109/ICNC.2008.48

Dongli, Z., X. Guan, Y. Tang and Y. Tang, 2011. Modified artificial bee colony algorithms for numerical optimization. *Proceedings of the 3rd International Workshop on Intelligent Systems and Applications, May 28-29, IEEE Xplore Press, Wuhan*, pp: 1-4. DOI: 10.1109/ISA.2011.5873266

Fang, X.Y. and T. Bai, 2009. Share price prediction using wavelet transform and ant colony algorithm for parameters optimization in SVM. *Global Cong. Intell. Syst.*, 3: 288-292. DOI: 10.1109/GCIS.2009.85

Farid, J.A., 2010. *Risk Frameworks and Applications*. 2nd Edn., Alchemy Technologies, pp: 254.

- Gao, W., S. Liu and L. Huang, 2012. A global best artificial bee colony algorithm for global optimization. *J. Comput. Applied Math.*, 236: 2741-2753. DOI: 10.1016/j.cam.2012.01.013
- Haupt, R.L. and S.E. Haupt, 1998. *Practical Genetic Algorithms*. 1st Edn., Wiley, New York, simultaneously in Canada, ISBN-10: 0471188735, pp: 177.
- Hinton, P., 2004. *SPSS Explained*. 2nd Edn., Routledge, ISBN-10: 0415274095, pp: 400.
- Jammazi, R. and C. Aloui, 2012. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Econ.*, 34: 828-841. DOI: 10.1016/j.eneco.2011.07.018
- Karaboga, D. and B. Akay, 2005. *Artificial Bee Colony (ABC), harmony search and bees algorithm on numerical optimization*. Erciyes University.
- Karaboga, D., 2005. *An idea based on honey bee swarm for numerical optimization*. Erciyes University.
- Karaboga, D., B. Gorkemli, C. Ozturk and N. Karaboga, 2012. A comprehensive survey: Artificial Bee Colony (ABC) algorithm and applications. *Artifi. Intell. Rev.* DOI: 10.1007/s10462-012-9328-0
- Klimberg, R.K., G.P. Sillup, K.J. Boyle and V. Tavva, 2010. Forecasting performance measures – what are their practical meaning? *Adv. Bus. Manag. Forec.*, 7: 137-147. DOI: 10.1108/S1477-4070(2010)0000007012
- Kotsiantis, S.B., D. Kanellopoulos and P.E. Pintelas, 2006. Data preprocessing for supervised learning. *Int. J. Comput. Sci.*, 1: 1306-4428.
- Kumar, M. and M. Thenmozhi, 2007. Support vector machines approach to predict the S and P CNX NIFTY index returns. *J. Acad. Bus. Econ.*
- Lai, K.K., L. Yu, S. Wang and W. Huang, 2006. Hybridizing exponential smoothing and neural network for financial time series prediction. *Comput. Sci.*, 3994: 493-500. DOI: 10.1007/11758549_69
- Lee, C.Y. and X. Yao, 2004. Evolutionary programming using mutations based on the Levy probability distribution. *IEEE Trans. Evol. Comput.*, 8: 1-13. DOI: 10.1109/TEVC.2003.816583
- Lendasse, A., Y. Ji, N. Reyhani and M. Verleysen, 2005. LS-SVM hyperparameter selection with a nonparametric noise estimator. Helsinki University of Technology.
- Levy, P., 1954. *Theorie de l'Addition des Variables Aleatoires*. 2nd Edn., Paris, Gauthier-Villars, France, pp: 385.
- Liu, C., 2009. Price forecast for gold futures based on ga-bp neural network. *Proceedings of the International Conference on Management and Service Science*, Sept. 20-22, IEEE Xplore Press, Wuhan, pp: 1-4. DOI: 10.1109/ICMSS.2009.5302242
- Lomax, R.G., 2007. *An Introduction to Statistical Concepts*. 2nd Edn., Routledge, New Jersey, ISBN: 0805857397, pp: 488.
- Makridakis, S. and M. Hibon, 2000. *The M3-Competition: Results, conclusions and implications*. *Int. J. Forec.*, 16: 451-476.
- Malliaris, M.E. and S.G. Malliaris, 2008. Forecasting inter-related energy product prices. *Eur. J. Fina.*, 14: 453-468. DOI: 10.1080/13518470701705793
- McCulloch, J.H., 1996. *Adaptive learning and adaptive least squares*. The Ohio State University.
- Mustafa, M.W., M.H. Sulaiman, H. Shareef and S.N.A. Khalid, 2011. *Transmission loss allocation in deregulated power system using the hybrid genetic algorithm-support vector machine technique*. Cyber. J.
- Niu, D., H. Lv and Y. Zhang, 2009. Combination forecasting model for mid-long term load based on least squares support vector machines and a mended particle swarm optimization algorithm. *Proceedings of the International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*, Aug. 3-5, IEEE Xplore Press, Shanghai, pp: 525-528. DOI: 10.1109/IJCBS.2009.16
- Palit, A.K. and D. Popovic, 2005. *Computational Intelligence in Time Series Forecasting Theory and Engeneering Applications* 1st Edn., Springer London, ISBN-10: 1852339489, pp: 393.
- Park, J.B., Y.W. Jeong, J.R. SHin and K.Y. Lee, 2010. An improved particle swarm optimization for nonconvex economic dispatch problems. *IEEE Trans. Power Syst.*, 25: 156-166. DOI: 10.1109/TPWRS.2009.2030293
- Pelkmans, J., E. Vos and L.D. Mauro, 2000. Reforming Product Regulation in the EU: A Painstaking, Iterative Two-Level Game. In: *Regulatory Reform and Competitiveness in Europe*, Galli, G. and J. Pelkmans, (Eds.). Elgar, Cheltenham, pp. 238-291.
- Ping, L. and Y. Jian, 2009. Application of least square support vector machine based on particle swarm optimization to chaotic time series prediction. *Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems*, Nov. 20-22, IEEE Xplore Press, Shanghai, Beijing, China, pp: 458-462. DOI: 10.1109/ICICISYS.2009.5357656

- Sapankevych, N.I. and R. Sankar, 2009. Time Series Prediction Using Support Vector Machines: A Survey. *IEEE Comput. Intell. Mag.*, 4: 24-38. DOI:10.1109/MCI.2009.932254
- Shalabi, A.L., Z. Shaaban and B. Kasasbeh, 2006. Data Mining: A Preprocessing Engine. *J. Comput. Sci.*, 2: 735-739. DOI: 10.3844/jcssp.2006.735.739
- Sun, W. and J. Zhang, 2008. Forecasting day ahead spot electricity prices based on GASVM. *Proceedings of the International Conference on Internet Computing in Science and Engineering*, 28-29-2008, Harbin, pp: 73-78. DOI: 10.1109/ICICSE.2008.50
- Suykens, J.A.K., T.V. Gestel, J.D. Brabanter, B.D. Moor and J. Vandewalle, 2002. *Least Squares Support Vector Machines*. Scientific, Co. Pte. Ltd.
- Tarhouni, M., K. Laabidi, S. Zidi and M.K. Lahmari, 2011. A nonlinear MIMO system identification based on improved multi-kernel Least Squares Support Vector Machines (improved multi-kernel LS-SVM). *Proceedings of the 8th International Multi-Conference on Systems, Signals and Devices*, Mar. 22-25, IEEE Xplore Press, Sousse, pp: 1-7. DOI: 10.1109/SSD.2011.5767454
- Valyon, J. and G. Horvath, 2007. Selection methods for extended least squares support vector machines. *Int. J. Intell. Comput. Cyber.*, 1: 69-93. DOI: 10.1108/17563780810857130
- Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory* 2nd Edn., Springer-Verlag, New York, Berlin, ISBN 0387945598, pp: 188.
- Wang, H. and D. Hu, 2005. Comparison of SVM and LS-SVM for regression. *Proceedings of the International Conference on Neural Networks and Brain*, Oct. 13-15, IEEE Xplore Press, Beijing, pp: 279-283. DOI: 10.1109/ICNNB.2005.1614615
- White, S.C., 1994. *Forecasting Business Time Series with Radial Basis Function Networks*. Ph.D. Thesis, Texas A and M University.
- Williams, G., 2011. *Data Mining with Rattle and R the Art of Excavating Data for Knowledge Discovery*. 1st Edn. Springer, New York ISBN-10: 1441998896, pp: 394.
- Xiang, Y. and L. Jiang, 2009. Water quality prediction using Ls-SVM and particle swarm optimization. *Proceedings of the 2nd International Workshop on Knowledge Discovery and Data Mining*, Jan. 23-25, IEEE Xplore Press, Moscow, pp: 900-904. DOI: 10.1109/WKDD.2009.217
- Ying, X. and Z. Hua, 2008. Water supply forecasting based on developed LS-SVM. *Proceedings of the 3rd IEEE Conference on Industrial Electronics and Applications*, Jun. 3-5, IEEE Xplore Press, Singapore, pp: 2228-2223. DOI: 10.1109/ICIEA.2008.4582913
- Yu, L., H. Chen, S. Wang and K.K. Lai, 2009. Evolving least squares support vector machines for stock market trend mining. *IEEE Trans. Evolutionary Comput.*, 13: 87-102. DOI: 10.1109/TEVC.2008.928176
- Yusof, Y. and Z. Mustaffa, 2011. Optimizing LSSVM using ABC for non-volatile financial prediction. *Aus. J. Basic Applied Sci.*, 5: 549-556.