Lévy Mutation in Artificial Bee Colony Algorithm for Gasoline Price Prediction

Zuriani Mustaffa[1] and Yuhanis Yusof[2]

[1]*Universiti Utara Malaysia (UUM),Malaysia, zuriani.m@gmail.com*
[2]*Universiti Utara Malaysia (UUM), Malaysia, yuhanis@uum.edu.my*

ABSTRACT

In this paper, a mutation strategy that is based on Lévy Probabily Distribution is introduced in Artificial Bee Colony algorithm. The purpose is to better exploit promising solutions found by the bees. Such an approach is used to improve the performance of the original ABC in optimizing Least Squares Support Vector Machine hyper parameters. From the conducted experiment, the proposed *lv*ABC shows encouraging results in optimizing parameters of interest. The proposed *lv*ABC-LSSVM has outperformed existing prediction model, Backpropogation Neural Network(BPNN), in predicting gasoline price.

Keywords: Artificial Bee Colony, Least Squares Support Vector Machines, Mutation, Levy Probability Distribution, Correlated Time Series.

I INTRODUCTION

An application of Evolutionary Algorithm (EA) such as Genetic Algorithm (GA), Evolutionary Programming (EP) and many others have been revolutionized in Artificial Intelligence (AI) domain. Their outstanding performances have drawn much attention to the research community, which finally led to the extensive research publications in literature. The benefits of these methods have been exploited in optimizing machine learning control parameters, such as Artificial Neural Network (ANN), Support Vector Machines (SVM) and its variant, Least Squares Support Vector Machines (LSSVM).

The utilization of hybrid methods between EA, especially GA which has dominated other EA technique (Karaboga & Akay, 2009) with machine learning approach has crossed various domains, such as electrical field which has been presented by Mustafa, Sulaiman, Shareef and Khalid (2011). By proposing the application of GA and LSSVM, the hybrid method is utilized in solving reactive power tracing. In similar field, Wei and Jie (2008) presented a combination of GA and SVM to predict spot electricity prices while the combination of GA and LSSVM in finance has been proposed by Yu, Chen, Wang and Lai (2009).

Recently, the emergence of Swarm Intelligence techniques has boosted the field of scientific research. This methods, which is also known as foraging algorithms (El-Abd, 2012) have been applied in a large number of optimization problems, including in tuning parameters of machine learning technique. The Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995), Ant Colony Optimization (ACO) (Dorigo, Gambardella, Middendorf, & Stutzle, 2002), Artificial Fish Swarm Algorithm (AFSA) (Li, Shao, & Qian, 2002) and Artificial Bee Colony (ABC) (Karaboga, 2005) are examples of this group to name a few. As has been experienced by EA formerly, these methods also gain favorable attention among academia, especially in proposing the idea to hybrid with machine learning method. Among them are hybrid method of LSSVM with Particle Swarm Optimization (PSO) in predicting chaotic time series (Ping & Jian, 2009), various bio inspired optimization approaches with SVM (Rossi and De Carvalho, 2008) and financial prediction utilizing ABC and LSSVM (Yusof & Mustaffa, 2011). Besides LSSVM, ABC also been chosen in optimizing Multi Layer Perceptron (MLP) in predicting the earthquake magnitude by Shah and Ghazali (2011).

As a relatively new optimization technique, ABC has proved to be better than or at least comparable to the other existing optimization technique with fewer control parameters (El-Abd, 2012). With the main objective to improve the capability of standard ABC, this paper is devoted to develop an enhanced ABC algorithm by utilizing Lévy Probabily Distribution. The newly *lv*ABC is later utilized to optimize LSSVM's hyper parameters, namely regularization parameter, γ and kernel Radial Basis Function (RBF) parameter, σ^2. Later, the prediction of gasoline price is done by LSSVM. Gasoline which is primarily used for internal combustion engines is the main product produced from refining crude oil, where for each barrel of crude is refined, it produces about 3.05 billion barrels of finished motor gasoline (Dunsby, Eckstein, Gaspar, & Mulholland, 2008).

This paper is organized as follows: In Section 2, a brief review on LSSVM is described. In Section 3, the algorithm of ABC is presented, followed by the integration of mutation strategy that is based on Lévy Probabily Distribution in the ABC. The methodology implemented is elaborated in Section 5 while results and discussion is presented in Section 6. Finally, Section 7 presents the conclusion of the study.

II REGRESSION USING LEAST SQUARES SUPPORT VECTOR MACHINES

The standard framework for LSSVM is based on the primal-dual formulation. Given the dataset $\{x_i, y_i\}^N_{i=1}$, the aim is to estimate a model of the form (Suykens, Van Gestel, De Brabanter, De Moor, & Vandewalle, 2002):

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

where $x \in R^n$, $y \in R$, and ϕ (.): $R^n \rightarrow R^{n_h}$ 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…,N.

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

From the Lagrangian function:

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

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

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

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

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

with y =$[y_1,\ldots,y_N]^T$, $\alpha=[\alpha_1,\ldots,\alpha_N]^T$. The resulting LS-SVM model in dual space becomes:

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

Usually, the training of the LSSVM model involves an optimal selection of regularization parameter, γ and kernel parameter, σ^2. Several kernel functions, namely Gaussian Radial Basis Function (RBF) Kernel, MLP Kernel and quadratic Kernel are available. For this study, the RBF Kernel is used which 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 associated with RBF function. Another tuning parameter, which is regularization parameter, γ can be seen in Eq. (2).

III ARTIFICIAL BEE COLONY

Artificial Bee Colony (ABC) algorithm is a population-based algorithm which was first introduced by Karaboga in 2005 (Karaboga, 2005). It is inspired by the intelligent behavior of honey bees. Theoretically, the colony of artificial bees consisted of three groups of bees; employed, onlooker and scout bees. Half of the colony is composed of employed bees and the rest are of the onlooker bees. The number of food sources/nectar sources is equal with the employed bees. This 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, while employed and onlooker bees handle the exploitation process, the scout bees carry out the exploration process. The duty of employed bees is to search for food sources (solutions). Later, the amount of nectars (solutions' qualities/fitness value) is calculated. Sharing of information with onlooker bees is carried out by employed bees. In the hive (dance area), the onlooker bees decide to exploit a nectar source depending on the information shared by the employed bees. Onlooker bees watch different dances before doing the selection of food source position according to the probability proportional to the quality of that food source. The onlooker bees also determine the source to be abandoned and allocate the responsible employed bee as scout bees. For the scout bees, their task is to find the new valuable food sources at which point it once again becomes employed bee. 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. In this work, the parameters of interest are γ and σ^2.The fitness value of the randomly chosen site is formulated as follows:

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

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 Eq. (9):

$$v_{ij} = x_{ij} + \varphi_{ij}\left(x_{ij-}x_{kj}\right) \tag{9}$$

where;

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

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

φ = 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, $v'_1 = \{x'_{i1}, x'_{i2},\ldots,x'_{iD}\}$, it is compared to the original solution $v_1 = \{x_{i1}, x_{i2},\ldots x_{iD}\}$. If the new solution is better than previous one, the bee memorizes the new solution; otherwise she memorizes the previous solution. The onlooker bee selects a food source to exploit with the probability values related to the fitness values of the solution. This probability is calculated using the following equation:

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \tag{10}$$

where fit_i is the fitness of the solution *v*. *SN* is the number of food sources potions. 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 which becomes a scout and makes a random search by Eq. (11).

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

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:

1. Initialize the food source positions (population)
2. Each employed bees is assigned on their food sources.
3. 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.
4. Decide the source to be abandoned and assign its employed bee as scout for discovering new food sources.
5. Memorize the best food source (solution) found so far.

IV ARTIFICIAL BEE COLONY USING LÉVY PROBABILITY DISTRIBUTION MUTATION STRATEGY

As to enhance the exploitation process in standard ABC, Lévy Mutation ABC-LSSVM (*lv*ABC-LSSVM) is applied in generating new food sources, which involved employed and onlooker bee phase. The *lv*ABC-LSSVM is based on Lévy Probability Distribution which is first introduced by P. Levy in 1930 (Lee & Yao, 2004).

$$lv_{\alpha,\theta}(y) = \frac{1}{\pi}\int_0^\infty e^{-\theta q^\alpha}\cos(qy)dq \tag{12}$$

From (12), 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.

In employed bee phase, instead of applying Eq. (9), the following equation is introduced:

$$v_{ij} = x_{ij} + \left(x_{ij-}x_{kj-}\right) + lv \tag{13}$$

On the other hand, in onlooker bee phase, the following equation is used as to replace Eq. (9):

$$v_{ij} = x_{ij} + lv \tag{14}$$

Eqs. (13) and (14) are designed through experimental approach. The main objective is to emphasize exploitation process on strong (promising) solutions. In this study, all parameters involved are automatically tuned by *lv*ABC-LSSVM.

V METHODOLOGY

This section discusses methodology utilized in the study. The discussion covers on research data and data preparation, data normalization, experiment setup and performance evaluation metric utilized.

A. Research Data and Data Preparation

In this study, four correlated energy fuels price (time series data) were employed as input, namely crude oil (CL), heating oil (HO), gasoline (HU) and propane (PN). The time series data covered are from December 1997 to November 2002. This is similar to the one utilized by Malliaris and Malliaris (2008). The correlation between the four energy fuel is as tabulated in Table 1 while the proportion for training, validation and testing is as indicated in Table 2.

Table 2. Correlation between Energy Fuel

	CL	HO	HU	PN
CL	1			
HO	0.9597	1		
HU	0.9649	0.9262	1	
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	1-874	70
Validation	875-1061	15
Testing	1062-1248	15

A. Data Normalization

Prior to training process, all input and output were normalized using Decimal Scaling. The objective is to independently normalize each feature component to specified range where this ensures that the larger input values do not overwhelm the smaller input values. Using the technique of Decimal Scaling, the data is normalized by moving the decimal point of value of the attribute. The number of decimal point moved depends on the maximum obsolete value of A. The formula of Decimal Scaling is as shown in Eq. (15).

$$v' = \frac{v}{10^{j}} \qquad (15)$$

where j is the smallest integer such that **Max**(|v'|)<1.

Table 3 and 4 show the samples of time series utilized before and after normalization processes.

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. Samples of Input after Decimal Scaling

CL	HO	HU	PN
0.1863	0.0514	0.0534	0.3260
0.1870	0.0519	0.0532	0.3188
0.1860	0.0522	0.0533	0.3213
0.1859	0.0522	0.0527	0.3198
0.1870	0.0528	0.0526	0.3200

B. Experiment Setup

Table 5 indicates the variables assigned to features involved. The input arrangement employed is as suggested in Malliaris and Malliaris (2008). The output was the daily spot price of gasoline from day 21 onwards.

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	HU
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	

C. Performance Evaluation Metric

In this study, other than prediction accuracy, the Mean Absolute Percentage Error (MAPE) is also utilized as the indicator of the prediction model performance. In MAPE, the errors are measured relative to the data values. The definition of MAPE is as shown in Eq. (16).

$$MAPE = \frac{1}{n}\left[\sum_{i,j=1}^{n}\left|\frac{y_i - p_i}{y_i}\right|\right] \quad (16)$$

where $i, j = 1, 2, \ldots, n$

y= Actual value

p= Prediction value

VI RESULTS AND DISCUSSION

In this section, the performance of proposed *lv*ABC-LSSVM is discussed and compared against the ABC-LSSVM and also Back Propagation Neural Netwrok (BPNN). Referring to the data depicted in Table 6, it is noted that the *lv*ABC-LSSVM produced an encouraging result by generating significantly lower MAPE which is 7.5829% as compared to 12.3162% and 15.2389% yielded by ABC-LSSVM and BPNN respectively.

Table 6. Correlated Energy Fuel Price Prediction: HU

	γ	σ^2	Prediction Accuracy (%)	MAPE (%)
ABC-LSSVM	28.9195	973.5299	87.6838	12.3162
*lv*ABC-LSSVM	354.387	750.777	92.4171	7.5829
BPNN	-	-	84.761	15.2389

The visual performances of results are shown in Figure 1, which plot the actual and predicted value of proposed technique and the competitors. The blue solid line represents the actual price, the green circle mark show the ABC-LSSVM predicted price, while the red cross mark and turquoise dash line represent the predicted value obtained using *lv*ABC-LSSVM and BPNN respectively. From the figure, it can be seen that *lv*ABC-LSSVM able to predict the gasoline price better than ABC-LSSVM and BPNN.

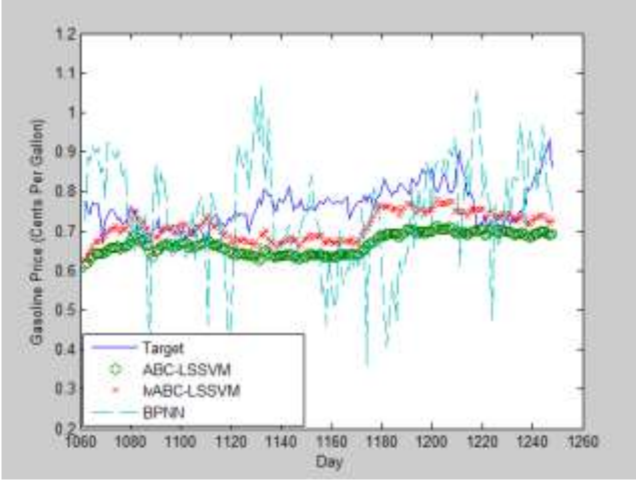

Figure 1. Gasoline Price Prediction

VII CONCLUSION

This paper proposed an extension of standard ABC algorithm with the aim of improving exploitation process of the bees in the search space. The encouraging initial results obtained indicate a positive opportunity to be explored in the future and may become a promising prediction model for the context of study.

REFERENCES

Chen, X., Wang, J., Sun, D., & Liang, J. (2008). *Time Series Forecasting Based on Novel Support Vector Machine Using Artificial Fish Swarm Algorithm.* Proceedings of the Fourth International Conference on Natural Computation (ICNC).

Dorigo, M., Gambardella, L. M., Middendorf, M., & Stutzle, T. (2002). Guest editorial: special section on ant colony optimization. *IEEE Transactions on Evolutionary Computation, 6*(4), 317-319.

Dunsby, A., Eckstein, J., Gaspar, J., & Mulholland, S. (2008). *Commodity Investing: Maximizing Returns Through Fundamental Analysis*. New Jersey: John Wiley & Sons.

El-Abd, M. (2012). Performance Assessment of Foraging Algorithms vs. Evolutionary Algorithms. *Information Sciences, 182*(2012), 243-263.

Karaboga, D. (2005). *An Idea Based on Honey Bee Swarm for Numerical Optimization*: Erciyes University, Engineering Faculty, Computer Engineering Department, .

Karaboga, D., & Akay, B. (2009). A comparative study of Artificial Bee Colony. *Applied Mathematics and Computation, 214*(2009), 108-132.

Kennedy, J., & Eberhart, R. (1995). *Particle Swarm Optimization.* Paper presented at the Proceedings of the International Conference of Neural Networks.

Lee, C.Y, & Yao, X. (2004). Evolutionary programming using mutations based on the Levy probability distribution. *IEEE Transactions on Evolutionary Computation, 8*(1), 1-13.

Li, L.X, Shao, Z.J, & Qian, J. X. (2002). An optimizing method based on autonomous animals: fish-swarm algorithm. *System Engineering Theory and Practice, 22*(11), 32-38.

Malliaris, M.E., & Malliaris, S.G. (2008). Forecasting Inter Related Energy Product Prices. The European Journal of Finance, 14(6), 453-468

Mustafa, M.W., Sulaiman, M.H., Shareef, H., & Abd. Khalid, S.N. (2011). Transmission Loss Allocation in Deregulated Power System Using Hybrid Genetic Algorithm-Support Vector Machine Technique. Cyber Journals: *Multidiscplinary Journals* *in Science and Technology, Journal of Selected Areas in Renewable and Sustainable Energy (JRSE) (January Edition)*.

Ping, L., & Jian, Y. (2009, 20-22 Nov. 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 (ICIS).

Rossi, A. L. D., & De Carvalho, A. C. P. L. F. (2008). *Bio-inspired Optimization Techniques for SVM Parameter Tuning.* Proceedings of the 10th Brazilian Symposium on Neural Networks.

Shah, H., & Ghazali, R. (2011, 6-8 Dec. 2011). *Prediction of Earthquake Magnitude by an Improved ABC-MLP*. Paper presented at the Proceedings of the Developments in E-systems Engineering (DeSE).

Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least Squares Support Vector Machines*: World Scientific Publishing Co. Pte. Ltd.

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

Wei, S., & Jie, Z. (2008). Forecasting Day Ahead Spot Electricity Prices Based on GASVM. Proceedings of the International Conference on Internet Computing in Science and Engineering (ICICSE)

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.

Yusof, Y., & Mustaffa, Z. (2011). Optimizing LSSVM using ABC for Non-Volatile Financial Prediction. *Australian Journal of Basic and Applied Sciences 5*(11), 549-556.