Data Normalization Techniques in Swarm-based Forecasting Models for Energy Commodity Spot Price

Yuhanis Yusof, Zuriani Mustaffa and Siti Sakira Kamaruddin

School of Computing, College of Arts and Sciences,
Universiti Utara Malaysia, Kedah, MALAYSIA
yuhanis@uum.edu.my

***Abstract*:** — Data mining is a fundamental technique in identifying patterns from large data sets. The extracted facts and patterns contribute in various domains such as marketing, forecasting, and medical. Prior to that, data are consolidated so that the resulting mining process may be more efficient. This study investigates the effect of different data normalization techniques, which are Min-max, Z-score and decimal scaling, on Swarm-based forecasting models. Recent swarm intelligence algorithms employed includes the Grey Wolf Optimizer (GWO) and Artificial Bee Colony (ABC). Forecasting models are later developed to predict the daily spot price of crude oil and gasoline. Results showed that GWO works better with Z-score normalization technique while ABC produces better accuracy with the Min-Max. Nevertheless, the GWO is more superior than ABC as its model generates the highest accuracy for both crude oil and gasoline price. Such a result indicates that GWO is a promising competitor in the family of swarm intelligence algorithms.

***Keywords*—** Artificial Bee Colony, data normalization, price forecasting, Grey Wolf Optimizer.

1. Introduction

Data mining extracts interesting non-trivial previously unknown and potentially useful information or patterns from data in large databases [1]. The discovered patterns can later be used to increase revenue, cuts costs, or both, in a business operation. It allows users to analyze data from different dimensions or angles, categorize it, and summarize the relationships identified. Data Preprocessing is a key step in building data mining models. In real-world business settings, a great proportion of time is spend on tasks such as data consolidation, data cleaning, data transformation, feature construction, feature selection, etc. More importantly, the quality of the resulting predictive models will largely depend on the ability to adequately preprocess the raw data and to create meaningful features from it. Realizing the importance of data preprocess in data mining, this study investigates the effect of different data normalization in swarm-based forecasting models for energy commodity. Three data normalization techniques involved are the Min Max, Decimal Scaling and Z-Score [1].

An efficient forecasting model for energy commodity price is essential to avoid unwanted risk, reducing loss and gaining high profit for a business. Furthermore, patterns obtained from such data mining process contribute to the strategic planning of the business. Classified as non renewable natural resources commodity, crude oil is very limited in production and irreplaceable in human time frame [2]. With the limitation in resources and continuously increasing demand, this situation leads to only one result; higher prices. As for investors, this means opportunity, however, for public people, this indicates inflation [3]. Due to that matter, the importance of price forecasting for such data has resulted to a large growing body of literature and research among the community is continuously carried out [4].

In literature, there are avalanche of studies which present various forecasting techniques for the said time series data. In [4], monthly crude oil price forecasting was implemented based on an improved Back Propagation Neural Network (BPNN). Realized in West Texas Intermediate (WTI) crude oil price, the BPNN model is compared against conventional BPNN. The finding of the study was in favour to the improved BPNN. Meanwhile, a hybridization of Genetic Algorithm and Feed Forward Neural Network (FFNN) with BP algorithm has also been demonstrated in crude oil price forecasting [5]. In the study, GA was employed to improve the learning algorithm and reduce the complexity in determining the control parameters of ANN. Later, the prediction process is continued by the FFNN. The experimental process involved two time series data of crude oil prices, viz. WTI and Iran crude oil prices and comparison was conducted against conventional Artificial Neural Network (ANN). Upon completing the experiment, it is indicated that the results produced by GA-FFNN are closer to actual data. In related work, the combination of Pattern Modelling and Recognition System (PMRS), Error Correction model (ECM) and Neural Networks (NN) has been presented to forecast the monthly WTI crude oil price [6]. The empirical results suggested that the presented model give good forecasting performance relative to the Mean Absolute Percentage Error (MAPE) and Root Mean Square Percentage Error (RMSPE). These methods, to a certain extent, all improve the accuracy of crude oil price forecasting.

In this study, two swarm-based forecasting models are presented to forecast daily crude oil and gasoline prices; Grey Wolf Optimizer (GWO) [7]and Artificial Bee Colony (ABC) [8]. This crude oil time series data is chosen due to its significant role not only in human life survival but also contributes to the global economic activities. In forecasting, the swarm algorithm (i.e GWO and ABC) is used to identify optimal values of the parameters in the prediction function, as applied in existing work [9] [10].

2. Related Work

Swarm Intelligence (SI) which is categorized as a subset of Evolutionary Computation (EC) [11], imitates the social behavior of animals or insects such as wolf, birds, ants, termites and bees.

2.1 GREY WOLF OPTIMIZER

GWO is considered as apex predators, which makes them placed as the top in food chain. In GWO, there are 4 hierarchies in grey wolf population, namely alpha, beta delta and omega. In the alpha level, it consists of male and female grey wolf and is responsible for decision making on hunting, sleeping place and others. Due to its dominant role, they are placed at the top of the hierarchy. The second level, beta, is responsible to help the alpha in decision making or any other activities of the pack. The beta can be male or female and will be the best candidate in replacing the alpha if one of the alpha passes away or become old. The beta acts as an advisor for the alpha in undertaking discipline of the pack. The group of wolves placed at the delta level is required to forward solutions to alpha and beta but they dominate the omega. This group consists of scouts, sentinels, elders, hunters and caretakers. Lastly, the omega, which is ranked last in the hierarchy, plays the role as scapegoat.

In GWO, the fittest solution is represented by alpha (α), followed by the second and third best solutions which are the beta (β) and delta (δ) respectively. Meanwhile, the balance of the candidate solutions is considered as omega (ω). The hunting (optimization) is guided by α, β and δ while the ω follows the three earlier groups. During hunting, the wolves tend to encircle their prey using the following equation:

$$\vec{D} = | \vec{C}.\vec{X}_p(t) - \vec{X}(t) | \tag{1}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}.\vec{D} \tag{2}$$

where t *=current iteration,* $\vec{A}$ *and* $\vec{C}$ *= coefficient vectors,* $\vec{X}_p$ *= position vector of the prey and* $\vec{X}$ *= position vector of the grey wolves.*

For vectors $\vec{A}$ and $\vec{C}$, it is calculated as follows:

$$\vec{A} = 2\vec{a}.\vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{C} = 2.\vec{r}_2 \tag{4}$$

where components of $\vec{a}$ *are linearly decreased from 2 to 0 over the curse of iterations. Meanwhile,* r_1 *and* r_2 *are random vectors in the range of [0,1].*

Commonly, hunting is guided by the alpha. However, both beta and delta may also involve in hunting, occasionally. In GWO, the alpha, i.e. the fittest candidate solution, beta and delta, are the experts on potential location of prey. Thus, the first three best solutions obtained are stored while the other agents (including omegas) are induced to update their positions based on the position of the best search agents. This defined by:

$$\vec{D}_\alpha = | \vec{C}_1.\vec{X}_\alpha - \vec{X} |, \vec{D}_\beta = | \vec{C}_2.\vec{X}_\beta - \vec{X} |, \vec{D}_\delta = | \vec{C}_3.\vec{X}_\delta - \vec{X} \tag{5}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1.(\vec{D}_\alpha), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2.(\vec{D}_\beta), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3.(\vec{D}_\delta) \tag{6}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{7}$$

Based on the GWO theory, its pseudo code is presented as in Figure 1.

```
Initialize the grey wolf population
Initialize a, A and C
Calculate the fitness of each agent
Xα = the best search agent
Xβ = the second best search agent
Xδ = the third best search agent
while (t < Max number of iterations)
   for each search agent
      Update the position of the current search agent
   end for
   update a, A and C
   Calculate the fitness of all search agents
   Update Xα, Xβ and Xδ
   t = t + 1
end while
return Xα
```

Figure 1: Pseudo code of GWO [7]

2.2 Artificial Bee Colony

In the bee based algorithm, it is observed that many bee swarm algorithms have been proposed in literature such as Virtual Bee Algorithm (VBA) [12], Bee System (BS) [13] and also Artificial Bee Colony (ABC) [14]. Among these algorithms, the ABC can be considered as the most broadly utilized algorithm in the literature [15]. The ABC algorithm is a simplified mathematical model of intelligent behavior of honey bees in food searching. Unlike any other meta heuristic algorithm which embodies a number of algorithmic parameters, in ABC, apart from population size and maximum iteration, it only has one additional parameter, namely limit. Interestingly, the limit parameter can be pre-defined based on population size and number of parameters of interest. Hence, ABC possess only two tuning parameters [16]. As compared to other optimization techniques such as Genetic Algorithm and Particle Swarm Optimization which suffer with more tuning parameters, this feature is an advantage in ABC algorithm [17]. Moreover, with the employment of basic mathematical operations, it results to a simple algorithm and easy to be implemented [18]. These two features are essential for the sake of easy adaptation to the problem in hand.

The ABC algorithm composes of three groups of bees, viz. Employed Bee (EB), Onlooker Bee (OB) and Scout Bee (SB). Each of them perform a simple task, yet results in complex behavior as a whole system. The distribution of colony members are equal, where half of the colony is comprised of EB and the other half filled by OB. The number of food sources is equal with the EB. This means a single EB corresponds to a single nectar source. The goal of the whole colony is to maximize the amount of nectar.

The work flow of artificial bee begins by discovering new food sources, which is carried out by the SB. Later, the EBs search for food around the food sources and the amount of nectars is calculated. Here, the amount of nectars is related to the fitness value of the solution under study. Upon completing the searching process, the EBs share the information collected with OBs which are waiting in the hive. This situation represents a multiple iteration in the intelligent swarm system. The OBs need to decide which nectar source will they exploit and the decision depends on the information shared by the EBs. For that matter, OBs watch various dances performed by the EBs before doing the selection of food source position according to the probability which is proportional to the quality of that food source. The OBs also decide the source to be casted away and allocate the responsibled EB as SB. For the SBs, their task is to find the new valuable food sources randomly and will once again becomes EB.

Supposed 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:

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

where obj.Fun is the objective function

The size of EBs and OBs are both equal to the number of food sources denoted by SN. For each food source's position, one EB is assigned to it. A new food source is produced according to [8]:

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

where $i = 1, 2, \ldots, SN$; $j = 1, 2, \ldots, D$; φ = a random generalized real number within the range [-1, 1], and k = is a random index number in the colony.

After producing the new solution, *v*, it is compared to the original solution, *x*, by applying greedy selection mechanism. If the new solution is better than previous one, the bee memorizes the new solution; otherwise she memorizes the previous solution. The OB selects a food source to exploit with the probability values related to the fitness values of the solution. This probability is calculated based on [8]:

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

where fit_i is the fitness of the solution v and SN is the number of food sources positions.

Later, the OB searches a new solution in the selected food source site, the similar way as exploited by EB. In SB phase, if the fitness of a found food source has not been improved for a given number of trial (denoted by limit), it is abandoned, and the EB of that food source becomes a SB and makes a random search using :

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

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 dth dimension of the problems space.

Basic steps of ABC algorithm [16] are as follows:

Initialize the food source positions (population)
Each EB is assigned on their food sources.
Each OB selects a source based on the quality of its solution, produces a new food source in the selected food source site and exploits for better source.
Decide the source to be cast aside and appoints its EB as SB for discovering new food sources.
Memorize the best food source position (solution) found so far.
If requirement is met, presents the best solution, otherwise repeat steps 2-5 until the maximum number of iteration is achieved.

Figure 2: Flow of ABC [16]

3. Experiments

In this study, real data of West Texas Intermediate (WTI) crude oil and gasoline prices are utilized in the experiments. Such datasets are included as they are the benchmark datasets in price forecasting [19]. The time series data covered in this experiment starts from December 1, 1997 to June 30, 1998 and is obtained from Barchart website [20]. Prior to dividing the dataset into training and testing sets, the data in-hand is pre-processed. This is done by transforming the commodity price via data normalization and deriving additional statistical attributes as undertaken in [21].

The variables assigned to features involved in predicting crude oil is as tabulated in Table 1. The undertaken experiment utilizes the daily spot price of crude oil for one month ahead (21 trading days) as the output.

Table 1. Input and Output Variables for Crude Oil

Input	Variable	Output
Daily closing price of crude oil	CL	Daily spot price of crude oil from day 21 onwards (CL21)
Percent change in crude oil daily closing spot price from the previous day	%Chg	
Standard deviation over the previous 5 trading days of crude oil price	Std5	
Standard deviation over the previous 21 trading days of crude oil price	Std21	

The utilized normalization techniques includes the Min-Max, Decimal Scaling and Z-Score [1]. The first technique, i.e Min-Max normalization, linearly transforms the actual data to a newly specified range via:

$$y' = \frac{actual\ data - min}{\max - min} * [max' - min'] + min' \quad (12)$$

On the other hand, the Decimal Scaling divides the value by 10 power n, where n is the number of digits of the maximum absolute value.

$$y' = \frac{y}{10^n} \quad (13)$$

The third technique is useful when the extreme value is unknown or outlier dominates the extreme values. Typically the scale will be [0 to 1].

$$y' = \frac{actual\ data - mean}{standard\ deviation} \quad (14)$$

4. Results

In evaluating time series prediction models, the utilization of appropriate evaluation metric is important. This is to justify the obtained results from the conducted experiments. In literature, various metrics have been employed and one of the popular metric is Mean Absolute Percentage Error (MAPE) [22]. As represented by its name, MAPE is a metric that is based on percentage error and it is a scale independent metric. The metric is measured by obtaining the sum of all absolute percentage errors and computing their average value. Another metric utilized is the prediction accuracy (PA).

For comparison purposes, the forecasting performance of GWO is compared against the results produced by ABC and Differential Equation. Results of MAPE and PA are depicted in Table 2 and 3. From the result tables, it is noted that GWO model utilizing Z-Score normalization technique produces similar accuracy (i.e above 95%) for both energy commodity. On the other hand, ABC model reacts differently; it achieves it highest prediction using different normalization for the crude oil and gasoline. And such a result can also be seen in the DE model.

Table 2. Results for Crude Oil Price Forecasting

	Min Max		Decimal Scaling		Z-Score	
	MAPE	PA	MAPE	PA	MAPE	PA
GWO	5.0514	94.9486	9.4768	90.5232	4.5393	**95.4607**
ABC	5.1708	94.8292	10.5595	89.4405	4.5481	**95.4519**
DE	5.3851	**94.6149**	10.8013	89.1987	5.6023	94.3977

Table 3. Results for Gasoline Price Forecasting

Gasoline

	Min Max		Decimal Scaling		Z-Score	
	MAPE	PA	MAPE	PA	MAPE	PA
GWO	5.2038	94.7962	6.8458	93.1542	4.9448	**95.5552**
ABC	4.6068	**95.3932**	8.6526	91.3474	5.0026	94.9974
DE	7.1572	92.8428	7.4499	92.5501	5.8893	**94.1107**

5. Conclusion

This study presents the utilization of recent swarm intelligence algorithms in the forecasting of energy commodity spot price. The success of the utilized forecasting model is partially contributed by the techniques used in the data pre-processing stage. Based on the undertaken study, it can be concluded that the Grey Wolf Optimzer works best while employing the Z-score normalization. Nevertheless, more work is required in determining the best normalization for the Artificial Bee Colony and Differential Equation models. The future work will also investigate the justification of such result.

References

[1] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, Third Edition ed. Waltham: Morgan Kaufmman Publishers, 2012.

[2] D. D. Kemp, *Exploring Environmental Issues: An Integrated Approach* London: Routledge, 2004.

[3] X. Zhang, Q. Wu, and J. Zhang, "Crude Oil Price Forecasting Using Fuzzy Time Series," in *Proceedings of the 3rd International Symposium on Knowledge Acquisition and Modeling*, 2010.

[4] R. Jammazi and C. Aloui, "Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling," *Energy Economics*, vol. 34, pp. 828-841, 2012.

[5] R. Tehrani and F. Khodayar, "A hybrid optimized Artificial Intelligent Model to Forecast Crude Oil using Genetic Algorithm," *African Journal of Bussiness Management*, vol. 5, pp. 13130-13135, 2011.

[6] D. Xu, Y. Zhang, C. Cheng, W. Xu, and L. Zhang, "A Neural Network-Based Ensemble Prediction Using PMRS and ECM," in *47th Hawaii International Conference on System Sciences*, 2014, pp. 1335-1343.

[7] S. Mirjalilia, S. M. Mirjalilib, and Andrew Lewisa, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46-61, March 2014.

[8] D. Karaboga, "An Idea Based on Honey Bee Swarm for Numerical Optimization," Erciyes University, Technical ReportOctober, 2005 2005.

[9] E. Hadavandi, A. Ghanbari, and S. Abbasian-Naghneh, "Developing a Time Series Model Based on Particle Swarm Optimization for Gold Price Forecasting," in *Proceedings of the Third International Conference on Business Intelligence and Financial Engineering (BIFE)*, 2010, pp. 337-340.

[10] Z. Mustaffa, Y. Yusof, and S. S. Kamaruddin, "Gold Price Prediction using Artificial Bee Colony Algorithm," in *Proceedings of the 2nd International Conference on Engineering and Applied Science (ICEAS)*, Tokyo, Japan, 2013.

[11] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: artificial bee colony (ABC) algorithm and applications," *Artificial Intelligence Review*, pp. 1-37, 2012.

[12] X. S. Yang, "Engineering Optimizations via Nature-Inspired Virtual Bee Algorithms," in *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach*. vol. 3562, J. Mira and J. R. Invarez, Eds., ed: Springer, 2005, pp. 317-323.

[13] P. Lucic and D. Teodorovic, "Bee System: Modeling Combinatorial Optimization Transportation Engineering Problems by Swarm Inteligence," in *Preprints of the Tristan IV Triennial Symposium on Transportation Analysis*, SaoMiguel, 2001, pp. 441-445.

[14] D. Karaboga and B. Akay, "Artificial Bee Colony (ABC), Harmony Search and Bees Algorithm on Numerical Optimization," 2005.

[15] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: artificial bee colony (ABC) algorithm and applications," in *Artificial Intelligence Review*, ed: Springer, 2012.

[16] D. Karaboga, B. Akay, and C. Ozturk, "Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Netwrok," in *Modeling Decisions for Artificial Intelligence*. vol. 4617, ed Berlin: Springer, 2007, pp. 318-329.

[17] P. Civicioglu and E. Besdok, "A Conceptual Comparison of the Cuckoo-Search, Particle Swarm Optimization, Differential Evolution and Artificial Bee Colony Algorithms," *Artificial Intelligence Review*, vol. 39, pp. 315-346, 2013.

[18] A. L. a. Bolaji, A. T. Khader, M. A. Al-Betar, and M. A. Awadallah, "Artificial Bee Colony Algorithm, its Variants and Applications: A Survey," *Journal of Theoritical and Applied Information Technology*, vol. 47, pp. 434-459, 2013.

[19] K. He, L. Yu, and K. K. Lai, "Crude Oil Price Analysis and Forecasting using Wavelet Decomposed Ensemble Model," *Energy Economics*, vol. 46, pp. 564-574, 2012.

[20] (2012, November, 2011). *Barchart*. Available: http://www.barchart.com/historicalquote.php?view=quote&sym=CLY00&txtDate=12%2F31%2F09&submit=Get+Data

[21] M. E. Malliaris and S. G. Malliaris, "Forecasting Inter-Related Energy Product Prices," *The European Journal of Finance*, vol. 14, pp. 453-468, 2008.

[22] R. J. Hyndman and A. B. Koehler, "Another Look at Measures of Forecast Accuracy," *International Journal of Forecasting*, vol. 22, pp. 679-688, 2006.