An Empirical Study of Double-Bridge Search Move on Subset Feature Selection Search of Bees Algorithm

Aras Ghazi Mohammed Al-dawoodi[1, 2], M. Mahmuddin[1]

[1]*School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia.*
[2]*College of Computer and Mathematical Science, Tikrit University, Tikrit, Iraq.*
ady@uum.edu.my

Abstract—**The application of Bees Algorithm in wrapper feature selection (BAFS) has been implemented but yet too far from perfect and has few weaknesses. The algorithm performs combination of exploitative neighbourhoods and random explorative search. This creates a heavy computational time, and in the same time could affect the overall accuracy subset selection. To rectify this issue, a double-bridge move proposed and benchmark dataset have been used to determine the performance of the proposed method. The obtained results from the experiment confirmed that the proposed extension of the search neighbourhood have provided better accuracy with suitable time than the original BAFS.**

Index Terms—**Wrapper Feature Selection; Local Neighbourhoods Search; Data Mining.**

I. INTRODUCTION

Generally, feature selection (FS) can be viewed in two main approaches known as 'Wrapper' and 'Filter'. Filter is faster but less accurate and it is computationally intensive than the Wrapper approach [9, 16]. A comparative analysis of these approaches is important to understand their accuracy and efficiency. A learning algorithm with the optimisation that uses the Wrapper approach incorporates an optimisation tool and evaluates a model, whereas the filters approach are similar to wrappers in the search approach, but instead of evaluating against a model, a simpler filter is evaluated. In the other words, inductive algorithms are used by wrapper methods as the evaluation function whereas filter methods are independent of the inductive algorithm [10]. In the context of FS, the Wrapper approach is one of the most widely used approaches due to its adequate results and efficiency in handling larger and more complex dataset as compared to the Filter approach [22]. However, it is an expensive technique as it involves a complex process of building a classifier with hundreds of items to evaluate one feature subset and dispensing huge numbers of features [6, 20]. Examples of the use of the Wrapper and Filter approaches are [16] and [2, 21] respectively. According to Schiezaro [19], the Wrapper approach achieves relatively superior results than Filter.

While working through these optimisation problem, one commonly used techniques is bio-inspired such as the Bees Algorithm (BA) due to their robustness, simplicity, and efficiency [3, 11]. It represents one of the fundamental search techniques and is based on the food foraging activities of bees. In order for the BA to compete successfully with other optimisation algorithms, it is important to analyse the issues and challenges faced by the same. One of the main issues of

the BA is the long computational time required during the process. The BA features a neighbourhood search where each bee is assigned to a number of random sites. For selecting the bees that can produce the best solution, BA calculates the fitness of each site. The selected bees are then assigned to the group of the most efficient bees known as "elite" bees. This further consumes more computational time.

An improvement of the BA is needed to avoid repetitive neighbourhood search as well as to reduce the total computational time and lack of accuracy. In order to overcome these limitations, the current study uses the Wrapper approach for optimisation. It also develops extension of swapping mechanism while combining extension of swapping mechanism for a faster and better solution [13, 18]. The purpose of extension approaches is not only to accelerate processing but also to maintain the abilities to overcome the drawbacks of the local optimal solution. The study aims to evaluate the proposed approach in terms of accuracy and time, and compare the results with original BAFS that was done by [2]. With data being available massively across the globe, some of it may be of inaccurate formatting or missing attribute [2]. Recent years have witnessed the accessibility to high-dimensional data on the Internet on a larger scale [13, 16]. These data contain large number of features which can be relevant, irrelevant, and redundant [5, 8]. In addition, the availability of massive data represents a challenge for analysing them [1, 15]. These data consume relatively more time and lack accuracy during the KDD process [8, 20, 22].

II. BEES ALGORITHM FEATURE SELECTION

The primary objective of FS is to improve performance including learning and accuracy. Moreover, it generates a lesser m number of features of the original data N where $m < N$, $m \neq 0$. As feature dimensionality is reduced, fewer features will be evaluated causing less computational time and thus faster learning.

A. Wrapper BAFS

The Wrapper FS local search uses the BA are based from the idea of mutation exist in genetic algorithm. BA itself does not have a special mechanism in findings best local search in identifying best or right combination of selected features. Due to that, and motivated from the genetic algorithm genetic operation, the same idea is adopted this work in the same way the genetic algorithm works within the chromosome. This is

also to ensure that dispersions of selection in identify best possible features combination.

The proposed method uses multilayer perceptron (MLP) networks as learning algorithms to guide FS in the form of feedback to the selection process as to how well a given set of features characterises patterns from various classes [2]. For the FS process, the method requires a dataset comprising of patterns, each with Nt features. All the patterns used in the training set are of known classes. Based on the original dataset, new dataset is constructed in which patterns only have a subset of the original features. In other words, a pattern in a new dataset will have Ns features *(1 ≤ Ns ≤ Nt)* selected from the original set of Nt features [2, 17].

The proposed algorithm will start with the parameter initialisation phase that includes the following parameters: the total number of Bees (n), the total number of "elite" Bees (e), the number of sites selected for neighbourhood search (m), the number of Bees around selected location (nsp), and the number of Bees around each "elite" locations (nep). Setting of these parameters is already discussed in the previous Section [2].

The initialisation process will generate subset features for each bee – feat1, feat2, … , featn. This phase generates random subset features and calculates the total MSE and fitness, based on Equation 1.

$$MSE = \frac{1}{N}\sum\nolimits_{i=1}^{N+1} x_i \qquad (1)$$

B. BAFS Local Search Mechanism

Feature in BAFS is organized into 2 main parts namely selected feature and fitness selected feature. Representation of a feature subset in the BAFS requires individual feature representation. As a result, each feature is represented in binary form. Bit '1' indicates the selection of a particular feature of the whole feature set, or else it will be '0'. For example, if the total number of the original features is 8, and only 7 of them are selected and a few representations of the new subset features are generated, it will use: a) '01111111' which means that the first feature from the total of the original features will not be selected and b) '1111110' which means that only the first seven features are selected. To extend this mechanism, a few bit search mechanism that have been introduced in [13] including simple swap, 2-Opt, and 3-Opt.

III. EXPERIMENTAL OF 4-OPT LOCAL SEARCH

A. 4-Opt Search for BAFS

The general structure of the proposed Wrapper FS method will be discussed in this section. Wrapper FS using the BA on dataset will describe in the rest of this section. As mentioned in the previous section, the proposed method uses MLP networks as learning algorithms to guide FS in the form of feedback to the selection process as to how well a given set of features characterises patterns from various classes [2]. For the FS process, the method requires a dataset comprising of patterns, each with Nt features. All the patterns used in the training set are of known classes. Based on the original dataset, new dataset is constructed in which patterns only have a subset of the original features. In other words, a pattern in a new dataset will have Ns features (1 ≤ Ns ≤ Nt) selected from the original set of Nt features [16].
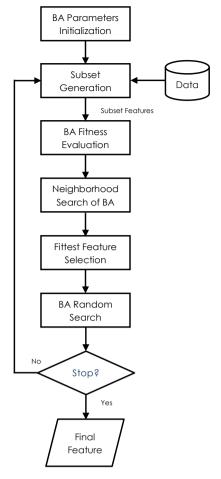

Figure 1: Simplified overall process flow of the BAFS

A bee represents a subset of Ns features that can be uniquely identified by a binary string (e.g. 010110111) where the total number of bits is Nt and the total number of non-zero bits is Ns. The position *i (1 ≤ i ≤ Ns)* of a bit along the string indicates a particular feature. If a feature is selected to form a new dataset, the corresponding bit is 1. Otherwise, it is zero.

FS starts with the random generation of a population of binary strings (or bees). For each string, a new dataset is constructed using the selected features specified in the string. The training data of the dataset are used in training an MLP whereas the remaining data or the test data are employed to evaluate the classification accuracy of the trained MLP. Figure 2 shows the FS method of the Bees Algorithm.

The 4-Opt search approach (in literature known as) is an extension to simple swap, 2-Opt, and 3-Opt operators. In the cases of the simple swap and 2-Opt operator, there is only one way to reconnect the tour fragments after deleting the two selected edges [2, 7]. The 3-Opt operator chooses the best triple edges that are not yet connected to the current tour [14]. In contrast, the 4-Opt is used as the perturbation technique, and a stochastic 2-Opt is used as the embedded local search heuristic. The 4-Opt move involves partitioning a permutation into 4 pieces (a, b, c, d) and putting it back together in a specific and jumbled ordering (a, d, c, b) in the TSP problem [4].
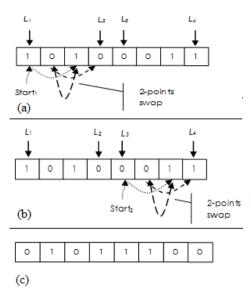
Figure 2: Simplified overall process flow of the BAFS

Therefore, this method can be termed as an extension from the previous methods proposed in [17]. For this, reference points are generated randomly at L_1, L_2, L_3 and L_4 where $L_1 > L_2 > L_3 > L_4$. Feature values at index L_1, L_2, L_3 and L_4 (F_{L1}, F_{L2}, F_{L3} and F_{L4}, respectively) use a 2-Opt-based operation as two sequential parts. The process is implemented by swapping the feature values at indices L_1 (F_{L1}) and L_2 (F_{L2}) as the first part and L_3 (F_{L3}) and L_4 (F_{L4}) as the second part. The swapping process continues with ($F_{L1} + 1 = F_{L2-1}$) and ($F_{L3} + 1 = F_{L4-1}$) until (L_1 and L_2) and (L_3 and L_4) have the same value. Figure 2 represents a simple example of this 4-Opt-based operation. Figure 3 shows the pseudo code of neighbourhood procedure of FS in the modified BA by 4-Opt operation and mutation.

```
For (itr=1 to maxRun){
  If (totalFeat > 2^totalFeat  -2 )
    stop process
  For (i=1 to m) {
    If (totalFeat > 2^totalFeat  -2)
      stop process
    For (j=1 to recuited bee) {
      If (totalFeat > 2^totalFeat  -2)
        stop process
      Do 4-Points-swap(F i )
      Do SimpleMutation(F i )
      If feat new is not exit,
        save in tabu list
      Evaluate and calculate feature,
F new
      If fitness F new <= Ft
        i save feature, new F
    }
  }
}
```

Figure 3: Pseudo code of the proposed integrated 4-Opt Local Search

B. Data Collection and Analysis

The datasets in this work comes are from the University of California at Irvine (UCI) repository [12] including from fields as varied as Pima Indians Diabetes, Wine, Hepatitis, SPECT Heart, Ionosphere, Soybean and Lung-cancer. Summary of the chosen Pima Indians datasets as describe in Table 1. A few motivations on why these datasets have been

chosen mainly due to total number of features of the datasets are more than 8 features. And the data themselves are in real or integer datatype.

Table 1
Summary of Datasets Used from UCI Dataset

No	Dataset Name	Number of Feature	Number of Instance	Data Type
1	Pima Indians Diabetes	8	768	Integer, real
2	Wine	13	178	Real
3	Hepatitis	19	155	Categorical, integer, real
4	SPECT Heart	22	267	Continuous
5	Ionosphere	34	351	Continuous
6	Soybean	35	47	Real
7	Lung-cancer	56	32	Integer

For this purpose, the fitness is calculated based on equation that has been proposed in [13] as showed in Equation 2.

$$Fitness = \frac{1}{k_1 \times MSE + k_2 \times \frac{N_s}{N_t}} \qquad (2)$$

C. Parameters Settings

As been mentioned earlier, MLP has been used in this work for evaluation. Standard setting by many been used for MLP is adopted. For this, only 1 hidden layer is used, with desired error 0.001, learning momentum 0.1, learning rate 0.3, number of epoch 500 and standard 10 accuracy cross-validation.

In the same time, the BA standard configuration setting is also been adopted from [13]. For this total number of scout bees n is 25, number of bees for best elite site nep=15, number of bees recruited of selected sites, nsp=20, number of bees for elite sites, e=2, number of sites for neighbourhood, m=5. Furthermore, the fitness is calculated based on Equation 2.

IV. RESULTS AND DISCUSSION
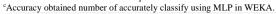
A. Speed Improvement

Table 2 shows that the extension had overcome its counterparts clearly in term of time than the original approach with the following datasets: Pima Indians Diabetes, Wine, Hepatitis, SPECT Heart, Ionosphere and Soybean. However, same result is obtained from Lung-cancer dataset. Figure 4 indicates that BAFS works well with the proposed extension and in the same time the accuracy.

Table 2
Summary of performance of each dataset

Datasets Name	# Selected Feature [a]	MSE	Fitness [b]	Time (sec.)	Acc. [c] (%)
Pima Indians Diabetes	2/8	0.0000	0.7999	50.3	67.0
	3/8	0.0000	0.8888	44.4	75.5
Wine	4/13	0.1645	0.6793	52.8	93.8
	4/13	0.1469	0.6874	52.6	91.6
Hepatitis	7/19	0.4303	0.5404	34.9	82.6
	6/19	0.4619	0.979	32.4	85.1
SPECT Heart	7/22	0.1941	0.6612	29.1	79.4
	7/22	0.1981	0.6595	27.2	83.8
Ionosphere	12/34	0.0000	0.7390	146.1	92.3
	13/34	0.0001	0.7390	130.6	91.5
Soybean	11/35	0.0034	0.7590	28.4	100
	10/35	0.0044	0.7745	25.5	100
Lung-cancer	24/56	0.0600	0.6882	17.0	46.9
	24/56	0.0003	0.7347	17.0	50.0

[a]Total number of selected feature. For example 4/12 means 4 out of 12 total features have been selected.
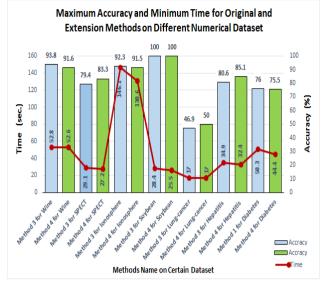
[b]Fitness values obtained based on Equation 2.
[c]Accuracy obtained number of accurately classify using MLP in WEKA.

Figure 4: Maximum Accuracy and Minimum Time for Original and Extension Methods on Different Numerical Dataset

B. Accuracy Improvement

It can also be noted from Figure 4 that classification accuracies comparable with those for the full-feature cases were achieved despite large reductions in the number of features. This confirms the ability of the proposed method to choose informative features. This study has empirically shown in Table 3, that the effect of applying proposed extension. The accuracy and time of the algorithm were computed with absolute differences of the best results. This table also shows that majority of the test dataset produce results that closer to 90%.

Table 3
Summary of performance of each dataset

Datasets Name	# Selected Feature [a]	MSE	Fitness [b]	Time (sec.)	Acc. [c] (%)
Wine	4/13	0.1641	0.6794	49.56	91.6
	2/13	0.1372	0.7745	43.45	78.65
SPECT Heart	6/22	0.2069	0.6758	30.54	93.75
	4/22	0.2036	0.7217	24.44	80.00
Ionosphere	12/34	0.0000	0.7390	136.28	92.30
	12/34	0.0001	0.7390	120.09	87.17
Soybean	10/35	0.0044	0.7745	25.53	100
	10/35	0.0401	0.7545	23.45	95.74
Lung-cancer	19/56	0.0116	0.7603	16.97	62.50
	24/56	0.0305	0.0250	14.59	53.12

[a]Total number of selected feature. For example 4/12 means 4 out of 12 total features have been selected.
[b]Fitness values obtained based on Eq. 2.
[c]Accuracy obtained number of accurately classify using MLP in WEKA.

V. CONCLUSION

An enhancement to neighbourhood local search and parameter numbers are represented in the BAFS. An extension operator 4-Opt and combination methods are introduced to reduce time consuming and increase the accuracy.

From the overall measurement results obtained, it can be noted that classification accuracies and time execution comparable with those for the full-feature cases were achieved despite large reductions in the number of features.

This confirms the ability of the proposed method to choose informative features. This study has empirically shown the effect of applying proposed extension and their combination of numerical dataset.

In this study, the validation of the result depends on two metrics: speed and accuracy. This determined by comparing the result was obtain between original methods from the original BAFS and the extension proposed of swapping mechanism of BAFS. Finally, the experimental results obtained confirmed that the proposed extension of the search neighbourhood that include 4-Opt and their combination approaches have provided better prediction accuracy with suitable time than the original BAFS.

REFERENCES

[1] Alelyani, S. 2013. On Feature Selection Stability: A Data Perspective. PhD, Arizona State University.
[2] Alsultanny, Y. A. 2011. Database Preprocessing and Comparison Between Data Mining Methods. *International Journal of New Computer Architectures and their Applications (IJNCAA)*, 1(1), 61-73.
[3] Biswas, A., Mishra, K. K., Tiwari, S., and Misra, A. K. 2013. Physics-inspired Optimization Algorithms: A Survey. *Journal of Optimization*. 1-16.
[4] Brownlee, J. 2012. Clever Algorithms: Nature-inspired Programming Recipes (1 ed.). Australia: lulu.com.
[5] Cheng, I. 2013. Hybrid Methods for Feature Selection. MSc., Western Kentucky University, Bowling Green, Kentucky. Retrieved from http://digitalcommons.wku.edu/theses/1244
[6] Dreyer, S. 2013. Evolutionary Feature Selection. MSc., Norwegian University of Science and Technology. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:no:ntnu:diva-24225
[7] Fosin, J., Davidović, D., and Carić, T. 2013. A GPU Implementation of Local Search Operators for Symmetric Ttravelling Salesman Problem. *PROMET-Traffic & Transportation*, 25(3): 225-234.
[8] Gantz, J. F. 2015. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, from http://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf
[9] Gupta, P., Jain, S., and Jain, A. 2014. A Review Of Fast Clustering-Based Feature Subset Selection Algorithm. *International Journal of Scientific & Technology Research*, 3(11): 20-26.
[10] Hossain, M. R., Oo, A. M. T., and Ali, A. B. M. 2013. The Effectiveness of Feature Selection Method in Solar Power Prediction. *Journal of Renewable Energy*, 9-9.
[11] Kumar, V., and Minz, S. 2014. Feature Selection. *SmartCR*, 4(3): 211-229.
[12] Lichman, M. 2013. UCI Machine Learning Repository. from University of California, School of Information and Computer Science http://archive.ics.uci.edu/ml.
[13] Mahmuddin, M. 2009. Optimisation Using Bees Algorithm on Unlabelled Data Problem. PhD. Retrieved from http://eserv.uum.edu.my/docview/1033191206?accountid=42599.
[14] Mahmuddin, M., and Yusof, Y. 2009, 2009. A Hybrid Simplex Search and Bio-Inspired Algorithm for Faster Convergence. *Paper presented at the International Conference on Machine Learning and Computing*, Australia.
[15] Mandli, I., and Panchal, M. 2014. Selection of Most Relevant Features from High Dimensional Data using IG-GA Hybrid Approach.
[16] Pham, D. T., Mahmuddin, M., Otri, S., and Al-Jabbouli, H. 2007. Application of the Bees Algorithm to the Selection Features for Manufacturing Data. *Paper presented at the International Virtual Conference on Intelligent Production Machines and Systems (IPROMS 2007)*, Whittles, Dunbeath, Scotland.
[17] Pham, D. T., Otri, S., Afify, A., Mahmuddin, M., and Al-Jabbouli, H. 2007. Data Clustering Using the Bees Algorithm. *Paper presented at*

the 40th CIRP International Manufacturing Systems Seminar 2007 Liverpool, UK.

[18] Rufai, K. I., Muniyandi, R. C., and Othman, Z. A. 2014. Improving Bee Algorithm Based Feature Selection in Intrusion Detection System Using Membrane Computing. *Journal of Networks*, 9(3): 523-529.

[19] Schiezaro, M., and Pedrini, H. 2013. Data Feature Selection Based on Artificial Bee Colony Algorithm. *EURASIP Journal on Image and Video Processing*, 2013(1), 47-47. doi: 10.1186/1687-5281-2013-47

[20] Song, Q., Ni, J., and Wang, G. 2013. A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data. *IEEE Transactions on Knowledge and Data Engineering*, 25(1): 1-14.

[21] Tang, W. Y. 2008. Feature Selection Algorithms for Very High Dimensional Data and Mixed Data. Nanyang Technological University. Retrieved from http://dr.ntu.edu.sg/handle/10356/41404.

[22] Varghese, S. M., and Sushmitha, M. N. 2014. Efficient Feature Subset Selection Techniques for High Dimensional Data. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(3).