

A New Discretization Approach of Bat and K-Means

Rozlini Mohamed¹, Noor Azah Samsudin²

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia (UTHM)
Parit Raja, Johor, Malaysia

Abstract—Bat algorithm is one of the optimization techniques that mimic the behavior of bat. Bat algorithm is a powerful algorithm in finding the optimum feature data collection. Classification is one of the data mining tasks that useful in knowledge representation. But, the high dimensional data become the issue in the classification that interrupt classification accuracy. From the literature, feature selection and discretization able to overcome the problem. Therefore, this study aims to show Bat algorithm is potential as a discretization approach and as a feature selection to improve classification accuracy. In this paper, a new hybrid Bat-K-Mean algorithm refer as hBA is proposed to convert continuous data into discrete data called as optimize discrete dataset. Then, Bat is used as feature selection to select the optimum feature from the optimized discrete dataset in order to reduce the dimension of data. The experiment is conducted by using k-Nearest Neighbor to evaluate the effectiveness of discretization and feature selection in classification by comparing with continuous dataset without feature selection, discrete dataset without feature selection, and continuous dataset without discretization and feature selection. Also, to show Bat is potential as a discretization approach and feature selection method. . The experiments were carried out using a number of benchmark datasets from the UCI machine learning repository. The results show the classification accuracy is improved with the Bat-K-Means optimized discretization and Bat optimized feature selection.

Keywords—Classification; discretization; feature selection; optimization algorithm; bat algorithm

I. INTRODUCTION

Bat Algorithm (BA) is a metaheuristic optimization algorithm that is found on echolocation characteristics of microbats where the emission and loudness has a varying pulse rates [1]. BA has been employed to resolve various optimization problems in many application. BA could possibly perform better than Particle Swarm Optimization (PSO) [2]. Further, unlike PSO, the parameter of BA can be varied. Moreover, the BA is capable to zoom the solving space, automatically which eventually allows local exploitation.

In [3] BA was used for classification in medical data. Moreover, this research [4] objective is to improve the classification performance in large data. This objective is achieved by employing BA as feature selection to select important feature in real-life dataset from UCI repository. In [5] hybridization of BA with Weighted Extreme Learning Machine (WELM) known as WELM-BAT was proposed to adjust the parameter for WELM. WELM-BAT was tested on real world medical diagnosis datasets and the result shows the performance of classification is improved.

The classification performance is improved using BA in the large data [6]. In [7] a hybrid Cross-Entropy method (BBACE) with Binary Bat Algorithm is used for feature selection in Big Data. To evaluate the effectiveness of the proposed method, the classifier such as Support Vector Machine and Naïve Bayes is used as a comparison regarding to time processing and accuracy. The results show that BA produced a better result compare to PSO.

Classification aims to assign data into predefined topic correctly. High dimensional, data redundancy and large amount of data are the factors that influence the classification performance. In data mining, classification is one of the popular problems that gets attention from researchers.

In real world datasets, the value of datasets are mixed data (continuous and discrete data), continuous data and discrete data. But some classification approaches require discrete data only. Thus, a process to convert from continuous data into discrete data is required. The process is known as discretization. The discretization task is to convert the continuous values into a small number of range based on the break points and then mapping each continuous value within the range, where each range is assigning to a discrete value. Discretization is needed before performing the feature selection task.

In classification problem, it is challenging to deal with irrelevant features caused by huge size of data [8]. Feature selection aims to select the relevant features and remove the irrelevant features. On the other hand, discretization aims to discretize the continuous value. Data representation in discrete value is easier to understand. Let consider this example, a person's weight has various readings, and it is difficult and ambiguous to interpret whether the reading is light or heavy. Discretization will place the weighted reading within a certain interval and the weight can be easily interpreted as to whether it is light or heavy.

This paper has proposed implementation of optimizing algorithm in discretization and feature selection steps. To implement the steps, BA is hybrid with the K-Means as a discretization approach. Then BA is then used as a feature selection to determine optimum features.

This paper is organized as follows: related works is presented in Section 2. Section 3 discusses the proposed approach based on K-Means and Bat algorithm and the description of the data sets. Section 4 discusses the experiment results and Section 5 presents the conclusion.

II. RELATED WORKS

This section briefly discuss on basic concept of Bat algorithm and reviewed the previous work on classification, discretization, and feature selection.

A. Bat Algorithm

BA is an intelligent optimization algorithm to solve complex engineering problems. The BA is motivated by the emitting sound pulse of bats known as echolocation. The echolocation of bats it used to search prey. In Fig. 1 shows bat will emit the sound to prey or obstacle refers as object. Then the object gives the reflection wave to bat by returning the echo wave to bat. From these echo wave, the bat can differentiate type of object; prey or obstacle event though in the darkness.

In this paper, the rules of bat algorithm are employs as follows:

1) *Rule 1:* The echolocation of bats used to sense the distance and they can recognize object and differentiate the kind object; obstacle or prey.

2) *Rule 2:* Bat randomly flies with varying wavelength, λ and loudness, A_0 , with velocity, v_i at position x_i with a fixed frequency, f_{min} to search for prey. The wavelength (or frequency) is automatically adjusted of their emitted pulses and adjust the rate of pulse emission $r \in [0,1]$, according to distance of target prey.

3) *Rule 3:* The loudness value are varied and can assume a large positive value, A_{max} to a small constant value, A_{min} . Thus based on the above rules, the location x_i^t and velocities v_i^t can be improve regarding to equation as follows;

$$f = f_{min} + (f_{max} - f_{min})\lambda, \quad (1)$$

$$v_i^t = v_i^{t-1} + (v_i^{t-1} - v_*)f_i, \quad (2)$$

$$x_i^t = x_i^{t-1} + v_i^t, \quad (3)$$

where $\lambda \in [0,1]$ is a random vector drawn from a uniform distribution.

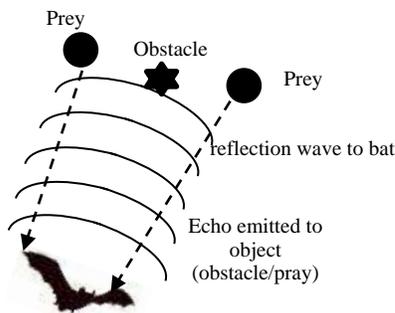


Fig. 1. Echolocation behavior by Bat.

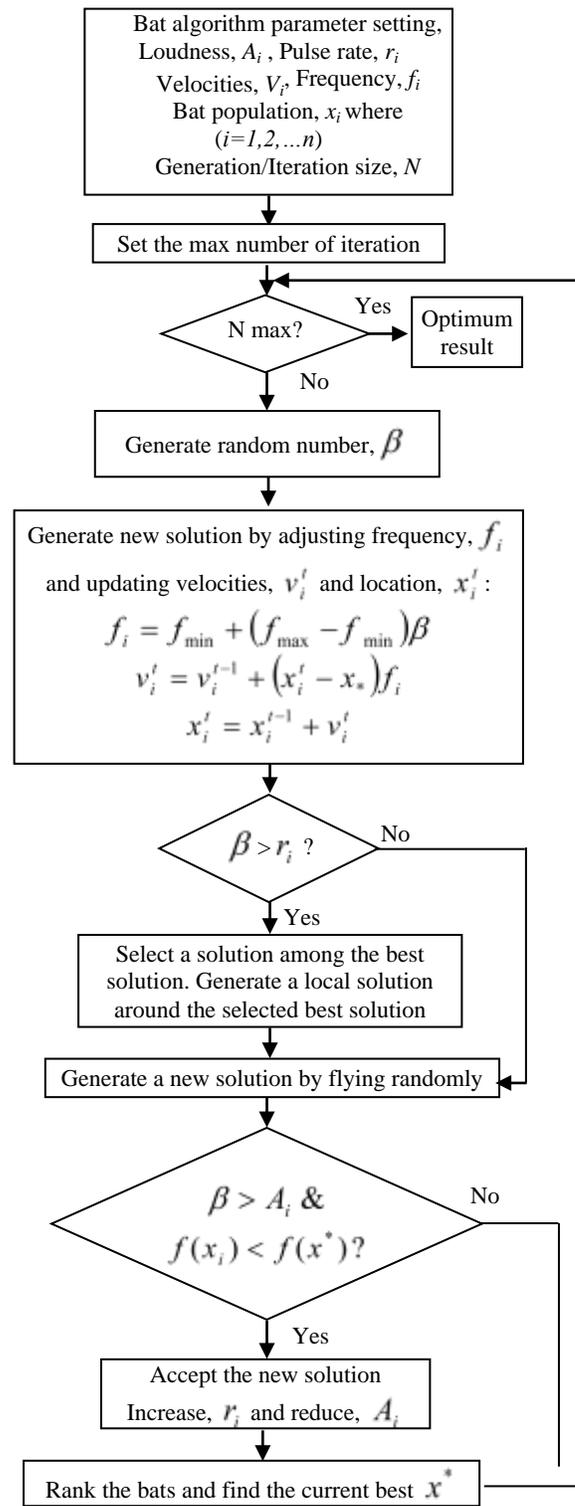


Fig. 2. Bat Algorithm [1].

Fig. 2 shows the flowchart for the Bat algorithm. The algorithm starts with initializing the value of loudness, pulse rate, velocity, frequency, number of bat population and the maximum number of iteration. While the number maximum iteration is not exceeded, the random number is generated to produce a new solution. In this new solution involve adjusting the previous frequency, updating the previous velocity as well

as updating the previous location. The equation 1, 2 and 3 are used to calculate these new frequencies, velocity and location, respectively. The next process is to proceed if the random number is higher than the pulse rate, the local best solution is determined among the solution if not continue to generate the new solution.

Later, if the value of the random number is higher than loudness and frequency in the previous location is lower than the frequency in a new location, accept the new solution otherwise continue to rank the best bat with their frequency. The process is repeated until reach the maximum iteration and when the iteration reaches the algorithm will come out the optimum results.

B. Classification

Various algorithms were designed for supervised data classification which are based on totally different approaches, for example, neural network [9] and text clustering [10]. Cano in 2019 [11] and Shafiq in 2020 [12] presented good review of data classification methods, including their performance measure and comparison.

Ayyagari [13] proposed a new classification algorithm that integrate the CART, k-NN and SVM. The proposed algorithm was used to handle the problem on imbalanced datasets and class overlap that influence the performance of classification algorithm.

One of the approaches can be used for supervised data classification is an optimization algorithm based on swarm. The most influential important factors on the classification performance are discretization and feature selection that occur in data preparation phase. For example, Zhou in their research, employed discretization in feature selection to find relevant feature [14]. In another research, Ucar employed P-Score in feature selection algorithms to improve classification performance [15].

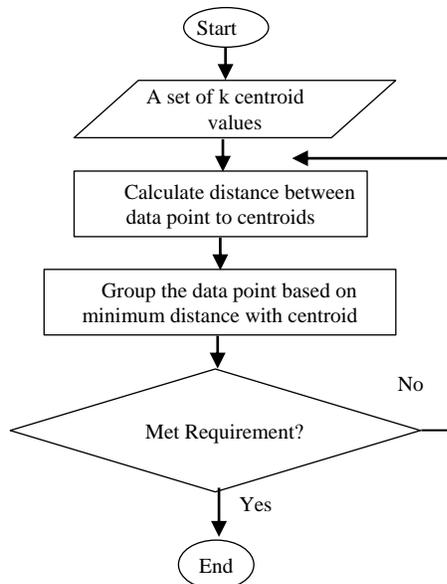


Fig. 3. K-Means Flowchart.

C. Discretization

Discretization can be defined as transforming continuous values into numerical values. In classification, discretization process is employed in data processing step before performing the feature selection process. Data preparation is one of the critical steps in classification problems [16]. The results from preparation is directly incorporated into the model and final results will be obtained. This means a good data source from data processing can increase the accuracy of classification performance.

A discretization approach was proposed to handle a mix of categorical and numeric safety data using fuzzy c means algorithm to discretize values [17]. The research in [18] has employed fuzzy discretization with Random Forest (RF) classifier to enhance labeling accuracy. In [19], Boltzmann machine is used in discretization process.

K-Means was describe by J. MacQueen in 1967 is an iterative algorithms [20]. The K-Means is used as the discretization algorithm [21]. The standard K-Means algorithm as shows in Fig. 3. From Fig. 3, the number k centroid is determined randomly. k is the number of class. Then the distance between data and each k centroid is calculated. The data is grouped into a pre-defined class according to the minimum distance between data and centroid. The process is repeated until the cluster centroid value stop changing.

This study has employed K-Means as pre-treatment method where data is discretized by K-Means before classification process. In [22] K-Means was employed to discretize the mixed data, in which the datasets present both, numerical and categorical attributes. In [23], K-Means is combined with discretization technique and Naïve Bayes to solve the problem in anomaly based intrusion detection for network intrusion detection system.

D. Feature Selection

In classification problem, feature selection is an effective method for dimensional reduction and helps in improving classification performance. Rachman and Rustam [24] in their proposed work researcher have use Fisher's Ratio algorithm to select cancer data. Their experiment reveals that feature selection gives a better result on classification accuracy.

Cheruku and Edla [25] highlight that the problems in medical data classification are redundant and high-dimensional. These problems were solved with improvements in data preparation before classification step. The research has proposed the Branch and Bound algorithm in feature selection step too.

Mithy and KrishnaPriya [26] has employed various feature selection methods in prediction of anemia for pregnant women. The feature selection methods employed rough set base fuzzy threshold in prediction and give promising results.

Reem Kadry and Osama Ismael [27] has improved the K-Nearest Neighbour performance in multidimensional data classification problem. In their work Particle Swarm Optimization (PSO) algorithm is employed to find out the optimum feature on Soybean datasets.

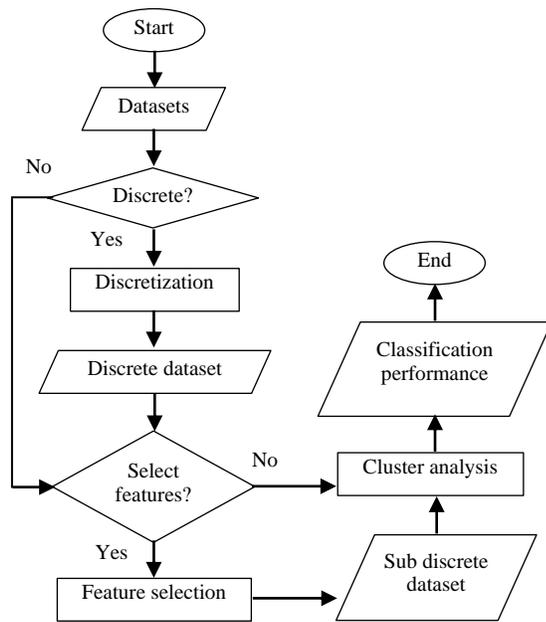


Fig. 4. Research Framework.

III. METHODOLOGIES

The experiments are carried out on nine benchmark datasets from UCI machine learning repository were used as input data and the detail about datasets are shown in Table I. The nine datasets refer as contS dataset and become the initial datasets. Assume the contS consist of the datasets as shown in Table I where,

$$\text{contS} = \{\text{DS1, DS2, DS3, DS4, DS5, DS6, DS7, DS8, DS9}\}$$

In this research, BA is hybrid with K-Means discretization refers as hBA and the hBA is enhance by employ BA as feature selection and refer as hfBA. Fig. 4 shows the framework in this research that involve two main process; 1) discretization process and 2) feature selection process. The implement of hBa and hfBA are explain in the following section.

A. hBA Approach

From Fig. 4, the approach start with the continuous datasets (contS) is convert into discrete data by using Bat-K-Means algorithm (hBA). From Fig. 2, the number k centroid is randomly define, but in hBA dataset, k centroid is according to optimum result from BA as shown in Fig. 3 and generate a new datasets which is optimize discrete dataset refers as hbaS dataset. In this approach, does not used feature selection so the algorithm will continue to the next process which is cluster analysis by using k-Nearest Neighbours. Finally, classification accuracy is measured.

B. hfBA Approach

From Fig. 4, the hfBA approach start in feature selection process. Where, BA algorithm as shown in Fig. 3 is applied to select the optimum feature after discretization with hBA approach. Then a new dataset refer as sub discrete dataset, hfBaS dataset is generated.

In hfBA approach, the size of bat represent by N length, where N is the total number of features in a dataset. The information regarding the instances is given by $I = \{i_1, i_2, \dots, i_n\}$ where n is the number of instance. Each instance is $I_i = \{a_{i1}, a_{i2}, \dots, a_{ik}\}$, where k is the number of attributes for the I_i -th instance in the DS dataset.

For example, it is assumed that DS dataset, has 10 attributes, 25 instances and 50 generations or repetitions. After 50 repetitions, the 10 -th bat or instance number 10^{th} represented by I_{10} is considered the best in the population. The optimum value for 10^{th} instance is $I_{10} = \{a_{10,1}, a_{10,2}, \dots, a_{10,10}\}$. Thus the optimum value for attributes is given by $att = \{a_{10,1}, a_{10,2}, \dots, a_{10,10}\}$. Based on BA, the minimum distance is assume as optimum value, so, in this study only value of $att = 0$ are selected. For example, $att = \{1,0,1,0,1,0,0,0,1,1,1,1,0,1,0\}$, so the attributes 1,3,5,9,10,11,12, and 14 are irrelevant features and will removes from dataset.

This research also employ K-Means discretization to contS datasets refers as discrete datasets (dkS). Besides, BA applied to select optimum feature from dkS and select feature from contS. The approach is follows the algorithm as shown in Fig. 4. The approach and the generated datasets are simplify in Table II.

TABLE I. DATASET INFORMATION

Dataset	No of Instances	No of Attributes
Credit Approval, DS1	690	15
Hill Valley, DS2	606	100
Image Segmentation, DS3	210	19
Libras Movement, DS4	360	90
Plant Species, DS5	1600	64
Steel Plates Faults, DS6	1941	27
Urban Land, DS7	507	147
Automobile, DS8	159	25
Yeast, DS9	1484	8

TABLE II. APPROACH AND DERIVED DATASET

Approach	Experiment	Datasets
	Continuous dataset or original datasets	contS={DS1, ..., DS9}
cfBA	To select feature from contS datasets using BA	cfbaS={cfbaS ₁ , ..., cfbaS ₉ }
dK	To convert contS into discrete dataset using K-Means discretization	dkS={dkS ₁ , ..., dkS ₉ }
dfBA	To select feature from dkS datasets using BA	dfbaS={dfbaS ₁ , ..., dfbaS ₉ }
hBA	To convert contS into discrete dataset using hybrid K-Means classifier with BA	hbaS={hbaS ₁ , ..., hbaS ₉ }
hfBA	To select feature from hBS datasets using BA	hfbaS={hfbaS ₁ , ..., hfbaS ₉ }

IV. EXPERIMENTS AND RESULTS

In this section, the experiments are set up to show Bat algorithm is potential as discretization approach and as feature selection to improve classification accuracy. Four experiments were conducted to show the BA in the discretization process and one experiment to show BA in discretization and feature selection process.

The comparison of hbaS dataset with contS, cfbaS, dkS and df of hfbaS datasets are conducted. The comparison results show in Table III that analyze by k-Nearest Neighbours for all datasets including initial datasets and new datasets. For easier understanding, the results are presented in a graphical form according to the following experiment where involving comparisons between two data sets only.

1) 1st experiment: the experiment is conducted to compare hbaS datasets with contS dataset and the result is shown in Fig. 5.

2) 2nd experiment: the experiment is conducted to compare hbaS datasets with cfbaS dataset and the result is shown in Fig. 6.

3) 3rd experiment: the experiment is conducted to compare hbaS datasets with dkS dataset and the result is shown in Fig. 7.

4) 4th experiment: the experiment is conducted to compare hbaS datasets with dfbaS and the result is shown in Fig. 8.

5) 5th experiment: the experiment is conducted to compare hbaS datasets with hfbaS and the result is shown in Fig. 9.

Fig. 5 presents the comparison between hbaS dataset with contS dataset. The purpose of this experiment is to identify which datasets give the better classification accuracy, either datasets with optimize discretization or continuous (original) datasets. From the results, classification accuracy of 3 datasets are improved by hBA which are DS1, DS5 and DS7. This experiment shows, by using hBA approach on the contS dataset can increase the classification performance results of dataset by 33%.

Fig. 6, presents the comparison between hbaS dataset with cfbaS dataset. The purpose of this experiment is to identify which datasets give the better classification accuracy, either datasets with optimize discretization or continuous dataset with optimize feature selection. The results shows hBA achieve the better performance in 3 datasets; DS3, DS5 and DS7. This experiment shows, hBA approach perform only in 33% from the entire datasets.

In the third experiment, the comparison results between hBA approach with dK approach shows in Fig. 7. The purpose of this experiment is to identify which datasets give the better classification accuracy, either datasets with discretization using optimize discretization or datasets with optimization using original K-Means discretization. Both of the approaches discretize the contS dataset. The results shows hBA achieve the better performance in 3 datasets; DS5, DS7 and DS9. This experiment shows, discretization with dK is better than discretization with hBA. hBA approach perform only in 33% from entire datasets.

In the third experiment, the comparison results between hBA approach with dK approach shows in Fig. 7. The purpose of this experiment is to identify which datasets give the better classification accuracy, either datasets with discretization using optimize discretization or datasets with optimization using original K-Means discretization. Both of the approaches discretize the contS dataset. The results shows hBA achieve the better performance in three datasets; DS5, DS7 and DS9. This experiment shows, discretization with dK is better than discretization with hBA. hBA approach perform only in 33% from entire datasets.

TABLE III. THE EXPERIMENTS RESULTS

Dataset	Accuracy					
	hbaS	contS	cfbaS	dkS	dfbaS	hfbaS
DS1	0.810	0.807	0.816	0.828	0.722	0.842
DS2	0.484	0.584	0.594	0.584	0.498	0.600
DS3	0.864	0.869	0.860	0.895	0.514	0.897
DS4	0.837	0.864	0.866	0.861	0.333	0.880
DS5	0.829	0.754	0.753	0.739	0.262	0.830
DS6	0.683	0.718	0.689	0.718	0.478	0.850
DS7	0.809	0.785	0.768	0.804	0.269	0.817
DS8	0.700	0.788	0.797	0.836	0.745	0.846
DS9	0.551	0.524	0.521	0.522	0.476	0.558

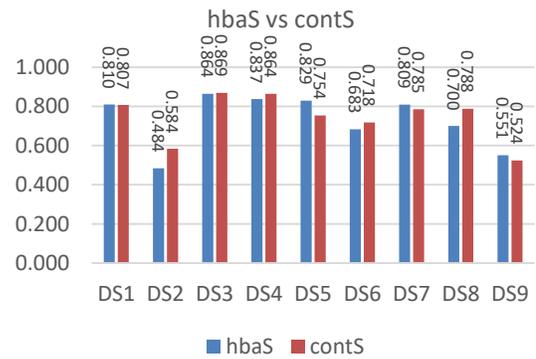


Fig. 5. The Classification Accuracy Comparison for Dataset hbaS with Dataset ContS.

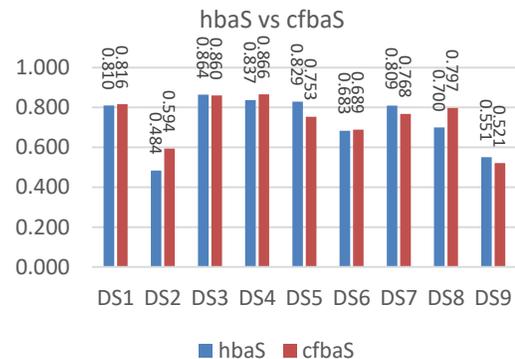


Fig. 6. The Classification Accuracy Comparison for Dataset hbaS with Dataset cfbaS.

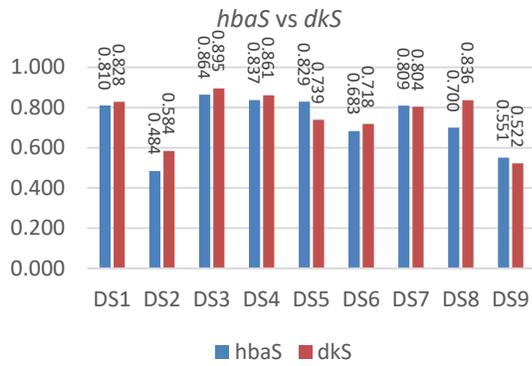


Fig. 7. The Classification Accuracy Comparison for Dataset hbaS with Dataset dkS.

In the fourth experiment, the comparison results between hBA approach with dfBA approach shows in Fig. 8. The purpose of this experiment is to identify which datasets give the better classification accuracy, either datasets with optimize discretization or dataset with K-Means discretization and optimize feature selection. The results shows hBA achieve the better performance in seven datasets; DS1, DS3, DS4, DS5, DS6, DS7 and DS9. This experiment shows, hBA approach perform in 78% from entire datasets.

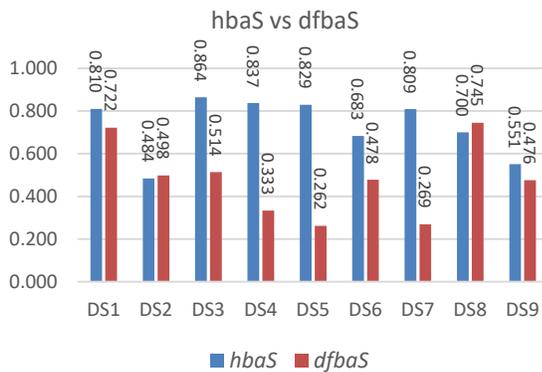


Fig. 8. The Classification Accuracy Comparison for Dataset hbaS with Dataset dfbaS.

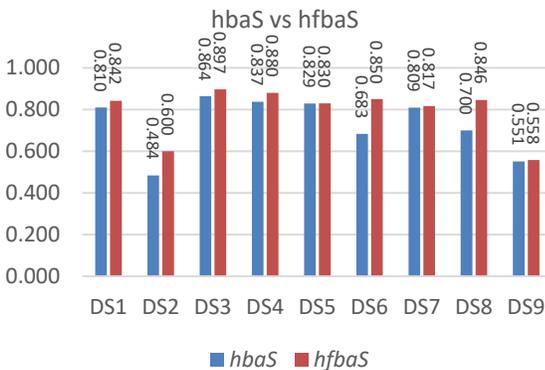


Fig. 9. The Classification Accuracy Comparison for Dataset hbaS with Dataset hfbaS.

Fig. 9 presents the comparison between hBA approach and hfBA. The purpose of this experiment is to identify which one is better in improving classification accuracy either optimize discretization or optimize discretization with optimize feature selection. From the results, hfBA able to improve the classification accuracy in all datasets.

V. CONCLUSION

The Bat algorithm is a powerful algorithm in finding the optimum feature data collection. The algorithm mimic the behavior of bat. Meanwhile, the high dimensional data in classification is interrupt classification accuracy. From the literature, feature selection and discretization is needed to solve. For that reason, this study show the Bat algorithm is potential as a discretization approach and as a feature selection to improve classification accuracy.

In this paper, a new hybrid Bat-K-Mean algorithm is proposed refer as hBA approach that used to convert continuous data into discrete data called as optimize discrete dataset (hbaS). The Bat algorithm is used to select the optimum feature in hbaS dataset for data dimension reduction and the new dataset refer as sub optimize discrete dataset (hfbaS) is created. The experiment was conducted to show the potential BA as discretization approach and feature selection approach to improve classification accuracy. k-Nearest Neighbor is used to evaluate the effectiveness of BA as discretization and feature selection in classification by comparing with dataset without feature selection, discrete dataset without feature selection, and dataset without discretization and feature selection. The experiments were carried out using a number of benchmark datasets from the UCI machine learning repository. The results show the classification accuracy is improved with the Bat-K-Means optimized discretization and Bat optimized feature selection.

The experiment results were present in Table III and Fig. 5 to Fig. 9. From the observation, conventional K-Means is not strong enough to improve classification performance (Fig. 7). If only the optimize discretization used in a continuous dataset is not able to improve the classification accuracy (Fig. 8). Optimize feature selection is required to select the optimum feature in the discrete dataset (Fig. 9). The combination of optimize discretization and optimize feature selection archive the better classification performance compare to continuous (original) in entire datasets.

In conclusion, BA is a powerful optimization technique and potentially to be used in discretization and feature selection. In the future, the research will extend for discretizing the mixed dataset value instead of continuous value only.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Higher Education, Malaysia for supporting this research under Fundamental Research Grant Scheme Vot K213 (FRGS/1/2019/ICT02/UTHM/02/2) and Universiti Tun Hussein Onn Malaysia for Multidisciplinary Research, Vot H511.

REFERENCES

- [1] X.S. Yang, "A new metaheuristic bat-inspired algorithm," *Studies in Computational Intelligence*, vol. 284, pp.65–74, 2010.
- [2] W. Kongkaew, "Bat algorithm in discrete optimization: a review of recent applications," *Songklanakarin J. Sci. Technol.*, vol. 39, pp. 641–650, 2017.
- [3] Y. Aboubi and H. Drias, N. Kamel, "BAT-CLARA: BAT-inspired algorithm for clustering LARge Applications," *IFAC-PapersOnLine*, vol. 49(12), pp. 243–248, 2016.
- [4] F. Yonis and O. Qasim, "Classification of diseases using a hybrid binary bat algorithm with mutual information technique," *Clinical Oncology and Cancer Biology*, vol. 1(1), pp. 1–5. 2019.
- [5] S. Priya and R. Manavalan, "Tuning the parameters of weighted elm using iwo and bat algorithm to improve the classification performance," in *Proceedings of the 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) 2018, 2018*, pp. 547-552.
- [6] D. Gupta, J. Arora, U. Agrawal, A. Khanna, and V. H. C. de Albuquerque, "Optimized Binary Bat Algorithm for classification of White Blood Cells," *Measurement*, vol. 143, pp. 180-190, 2019.
- [7] G. Li and C. Le, "Hybrid binary bat algorithm with cross-entropy method for feature selection," in *Proceedings of the 4th International Conference on Control and Robotics Engineering (ICCRE) 2019, 2019*, pp. 165–169.
- [8] D. Renuka Devi and S. Sasikala, S, "Online feature selection (ofs) with accelerated bat algorithm (aba) and ensemble incremental deep multiple layer perceptron (EIDMLP) for big data streams," *Journal of Big Data*, vol. 6(103), pp. 1-20, 2019.
- [9] Silva J, Borré JR, Piñeres Castillo AP, Castro L, Varela N. Integration of Data Mining Classification Techniques and Ensemble Learning for Predicting the Export Potential of a Company. *Procedia Computer Science*. 2019; 151(2019): 1194–1200.
- [10] A. O. Adeleke, N. A. Samsudin, A. Mustapha and N. Nawi, "Comparative analysis of text classification algorithms for automated labelling of quranic verses," *Int. J. Adv. Sci. Eng. Inf. Technology*, vol. 7(4), pp. 1419, 2017.
- [11] J. R. Cano, P. A. Gutiérrez, B. Krawczyk, M. Woźniak and S. García, "Monotonic classification: an overview on algorithms, performance measures and data sets," *Neurocomputing*, vol. 341(2019), pp. 168–182, 2019.
- [12] M. Shafiq, Z. Tian, A. K. Bashir, A. Jolfaei and X. Yu, "Data mining and machine learning methods for sustainable smart cities traffic classification: a survey," *Sustainable Cities and Society*, vol. 60(2020), pp. 102177, 2020.
- [13] M. R. Ayyagari, "Classification of imbalanced datasets using one-class svm, k-nearest neighbors and CART algorithm," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 11(11), pp. 1-5, 2020.
- [14] Y. Zhou, J. Kang, S. Kwong, X. Wang and Q. Zhang, "An evolutionary multi-objective optimization framework of discretization-based feature selection for classification," *Swarm and Evolutionary Computation*, vol. 60(2021), pp. 1-16, 2021
- [15] M. K. Ucar, "Classification performance-based feature selection algorithm for machine learning: P-Score," *Innovation and Research in BioMedical engineering*, vol. 41(4), pp. 229-239, 2020.
- [16] R. Zhao, Y. Qu, A. Deng and R. Zwiggelaar, "Inconsistency measure associated discretization methods to network-based intrusion detection," in *Proceedings of the 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2018, pp. 1–6.
- [17] K. Dhalmahapatra, R. Shingade and J. Maiti, "An innovative integrated modelling of safety data using multiple correspondence analysis and fuzzy discretization techniques," *Safety Science*, vol. 130, pp. 104828, 2020.
- [18] M. Fikri, "The impact of fuzzy discretization's output on classification accuracy of random forest classifier," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp.3950–3956, 2020.
- [19] N. Hranisavljevic, A. Maier and O. Niggemann, "Discretization of hybrid CPPS data into timed automaton using restricted Boltzmann machines," *Engineering Applications of Artificial Intelligence*, vol. 95, pp. 103826, 2020.
- [20] J. MacQueen. "Some methods for classification and analysis of multivariate observations." in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 1967, pp. 281–297.
- [21] R. Zhao, Y. Qu, A. Deng and R. Zwiggelaar, "Inconsistency measure associated discretization methods to network-based intrusion detection," in *Proceeding of the 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2018, pp. 1–6.
- [22] D. Maryono, P. Hatta and R. Ariyuana, "Implementation of numerical attribute discretization for outlier detection on mixed attribute dataset," in *Proceeding of the 2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 715–718.
- [23] H. M. Tahir, A. M. Said, N. H Osman, N.H Zakaria, P. N. . M Sabri and N. Katuk, " Oving K-Means Clustering using discretization technique in Network Intrusion Detection System," in *Proceeding of the 2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, 2016, pp. 248–252.
- [24] A. A. Rachman and Z. Rustam, "Cancer classification using Fuzzy C-Means with feature selection," in *Proceedings of the 12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA) 2016, 2016*, pp. 31–34.
- [25] R. Cheruku and D. Edla, "Bin-BB: Binning with Branch & Bound feature selection for improved diabetes classification," in *Proceedings of the 14th IEEE India Council International Conference (INDICON) 2017, 2017*, pp. 1–4.
- [26] M. D. Mithy and V. KrishnaPriya, "Anemiascreening in pregnant women by using feature selection with data classification algorithms," *Materials Today: Proceedings*, In Press, Corrected Proof, Available online 31 October 2020.
- [27] R. Kadry and O. Ismael, "A new hybrid kNN classification approach based on particle swarm optimization," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11(11), pp. 291-296, 2020.