

Research Article

Bat Algorithm Based Hybrid Filter-Wrapper Approach

Ahmed Majid Taha,¹ Soong-Der Chen,² and Aida Mustapha¹

¹Soft Computing and Data Mining Center, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

²College of Information Technology, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia

Correspondence should be addressed to Ahmed Majid Taha; ahmed.majid.taha@gmail.com

Received 11 April 2015; Revised 27 August 2015; Accepted 20 September 2015

Academic Editor: Viliam Makis

Copyright © 2015 Ahmed Majid Taha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a new hybrid of Bat Algorithm (BA) based on Mutual Information (MI) and Naive Bayes called BAMI. In BAMI, MI was used to identify promising features which could potentially accelerate the process of finding the best known solution. The promising features were then used to replace several of the randomly selected features during the search initialization. BAMI was tested over twelve datasets and compared against the standard Bat Algorithm guided by Naive Bayes (BANV). The results showed that BAMI outperformed BANV in all datasets in terms of computational time. The statistical test indicated that BAMI has significantly lower computational time than BANV in six out of twelve datasets, while maintaining the effectiveness. The results also showed that BAMI performance was not affected by the number of features or samples in the dataset. Finally, BAMI was able to find the best known solutions with limited number of iterations.

1. Introduction

A number of studies have illustrated hybrid approach that combines the good characteristics of both filter and wrapper techniques. They are more efficient than wrapper methods while at the same time providing comparable accuracy [1–4]. Lemma and Hashim [5] proposed a hybrid approach by using boosting technique and integrated some of the features in wrapper methods into a fast filter method. The results show that the proposed method is competitive with wrapper methods while selecting feature subsets much faster. Reference [6] developed a hybrid method based on Markov Blanket filter for high-dimensional genomic microarray data. The experimental results stated that the proposed method led to feature subsets outperforming those of regularization methods.

Hu et al. [7] investigated filter and wrapper methods for biomarker discovery from microarray gene expression data for cancer classification. They proposed a hybrid approach where Fisher's ratio was employed as the filtering method. The proposed method was tested extensively on real datasets, and the results demonstrate that the hybrid approach is computationally more efficient than the simple wrapper method. Furthermore, the results showed that the hybrid approach significantly outperformed the simple filter method.

Huda et al. [8] presented a flexible hybrid feature selection method based on floating search methods to increase flexibility in dealing with the quality of result against computational time trade-off. The performance of the hybrid method was tested using real-world datasets. The authors stated that the proposed method significantly reduced the search time while achieving comparable accuracies with those of the wrapper methods.

The measures of relevance are of fundamental importance in a number of applications. MI is widely accepted in quantifying linear or nonlinear relevance degree between random variables [9–11]. MI of two random variables is a quantity that measures the mutual dependence of the two variables. The earliest studies to use MI for selecting features in building models were by Lewis [12] and Lashin et al. [13]. Huang et al. [14] developed an MI technique that utilized feature similarity for redundancy reduction in an unsupervised feature selection.

Tay and Shen [15] developed a method that targeted an efficient estimation of MI in high-dimensional datasets. From their observation, a feature is relevant to the classes if it embodies important information about the classes; otherwise, the feature is irrelevant or redundant. This method was based on both information theory and statistical tests,

whereby selected feature is conditional and the information given by this feature must allow for statistical reduction of overlapping class. Using both synthetic and real-world datasets, authors stated that the hybrid method was able to eliminate irrelevant and redundant features, even in very large feature spaces, and performed more efficiently than the pure wrapper methods.

Tomar and Agarwal [16] proposed a hybrid Genetic Algorithm (GA) with MI for finding a subset of features that are most relevant to the classification task. They optimized the MI between the predictive labels in a trained classifier and the true class labels instead of optimizing the classification error rate. The method was validated using real-world datasets and the results indicated that the hybrid method outperformed the accuracy performance of filter methods. They also concluded that their hybrid method was more efficient than the wrapper methods.

BA has shown superior performance in handling feature selection problem [17–21] as well as other problems [22–24]. However, the aim of this paper is to develop an algorithm that extracts and combines the best characteristics of filter-based and wrapper-based approach into one algorithm. The filter model is based on the Mutual Information (MI) and wrapper model is based on Naive Bayes classifier. The proposed algorithm BAMI will be tested over twelve benchmarking datasets compared against BANV which has been proposed previously by Goyal and Patterh [25]. In this section, hybrid models in feature selection problem have been presented. The rest of this paper is organized as follows. Section 2 presents the proposed algorithm BAMI with the relevance and redundancy of features based on MI. Section 3 compares the proposed hybrid BAMI with the BANV and reports the computational efficiency and iterations. Finally, Section 4 discusses the results and Section 5 concludes the paper.

2. Proposed Bat Algorithm with Mutual Information

Our main motivation to build a hybrid feature selection model is to strike a good balance between the computational efficiency of a filter model and the accuracy performance of a wrapper model. In the conventional wrapper model such as the BANV, all bats are initialized with randomly selected features. In this hybrid model, we propose a small fraction of the bats to be initialized with “good” features ranked by the Maximum Relevance Minimum Redundancy (mRMR) method. The injection of the good features aims to strategize part of the search in effort to speed up the swarm convergence towards the best known solution.

2.1. Maximum Relevance Minimum Redundancy. A “good” feature is defined as the one that has the best trade-off between minimum redundancy within the features and maximum relevance to target variable. Chen and Cheng [26] proposed a method called the Maximum Relevance Minimum Redundancy (mRMR) to gauge the “goodness” of a feature. The mRMR method is a sequential forward selection algorithm that evaluates the importance of different features.

This algorithm uses MI to select M features that best fulfill the minimal redundancy and maximal relevance criterion. It is found to be very powerful for feature selection. The relevance and redundancy are measured by the MI as defined in

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (1)$$

where x and y are two random variables, $p(x, y)$ is their joint probability density, and $p(x)$ and $p(y)$ are their marginal probability densities, respectively. Let F represent the entire feature set, while F_s denotes the already-selected feature set which contains m features, and F_t denotes the yet-to-be-screened feature set which contains n features. Relevance D of the feature f in F_t with the target c can be calculated by

$$D = I(f, c). \quad (2)$$

The redundancy R of the feature f in F_t with all the features in F_s can be calculated by (3):

$$R = \frac{1}{m} \sum_{\bar{f} \in F_s} I(f, \bar{f}), \quad (3)$$

$$\max_{f \in F_t} \left[I(f, c) - \frac{1}{m} \sum_{\bar{f} \in F} I(f, \bar{f}) \right]. \quad (4)$$

To obtain the feature F_t with maximum relevance and minimum redundancy, (3) and (4) are combined with the mRMR function. For a feature set with M features, the feature evaluation will continue M rounds. After these evaluations, we will get a feature set s by the mRMR method as illustrated in (5). The feature index h indicates the importance of the respective feature. Better features will be extracted earlier with a smaller index h :

$$s = \{f_1, f_2, f_3, \dots, f_h, \dots, f_M\}. \quad (5)$$

2.2. Algorithm Procedure. The main steps in the proposed algorithm have been cleared in Figure 1; the shaded region refers to the main difference from previous proposed BANV. In our proposed BAMI algorithm, Maximum Relevance Minimum Redundancy (mRMR) method is used to analyze the “goodness” of each feature. A particular number of the top-ranked features (K) will be used to initialize one bat in the swarm. As shown in (5) and (6), K are dynamic parameters that give flexibility to the proposed algorithm by making changes according to the swarm size and the number of features in the dataset:

$$K = b * \text{Number of Features}. \quad (6)$$

Next, all the bats will be evaluated by a Naive Bayes classifier. Accordingly, if the initialized bats really contain informative features, one of these bats will become the global best solution and speed up the swarm convergence to a promising area within the search space. Otherwise, the proposed BAMI would proceed with ordinary BANV procedure; hence, the solution quality will not be affected.

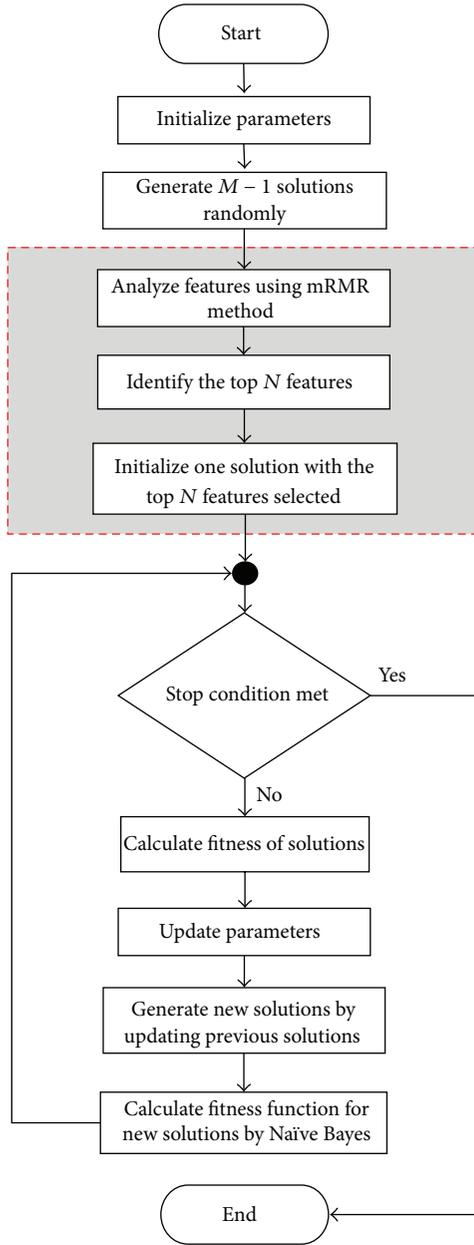


FIGURE 1: Flowchart of Bat Algorithm with Mutual Information.

3. Experiments and Results

The objective of the experiments is to evaluate the performance of the proposed algorithm BAMI versus a traditional Bat Algorithm with Naïve Bayes classifier (BANV). Note that the searching efficiency in this study is evaluated based on the speed to converge to best known solution. The speed is measured in terms of number of iterations and execution time; therefore, once the algorithm obtains the best known solution, we will record the time and the number of iterations. Twelve benchmark datasets were used for the evaluation. The datasets have been selected from various domains; furthermore, each dataset has different number of features and samples as shown in Table 1. Both algorithms BANV

TABLE 1: Characteristics of datasets.

Datasets	Number of features	Number of samples
Lung	56	32
WQ	38	521
Derm2	34	358
Derm	34	366
LED	24	2000
Mushroom	22	8124
Credit	20	1000
Vote	16	300
Heart	13	294
Exactly2	13	1000
Exactly	13	1000
M-of-N	13	1000

TABLE 2: Average iterations and runtimes to obtain the best known solution.

Datasets	ATM	ATB	AIM	AIB
M-of-N	11.3	15.70	7.76	10.73
Exactly	4.25	4.99	2.43	3.3
Exactly2	2.16	2.9	3.85	4.88
Heart	3.92	6.32	8.1	13.6
Vote	3	4.36	5.26	8.16
Credit	9.49	15.52	5.63	10.06
Mushroom	630.44	639.12	46.33	51.23
LED	4.41	46.19	1	15.13
Derm	16.71	28.52	25.33	44.93
Derm2	14.83	26.15	23.16	44.53
WQ	63.86	82.14	70.8	89.03
Lung	4.81	4.94	38.03	39.53

and BAMI were run for 30 times with the maximum of 250 iterations; the population size is set to 10 for both algorithms, and the value b equals 0.5.

Table 2 shows the average time and number of iterations (over 30 runs) when algorithms obtain the best known solution. In this table, ATB refers to average time of standard BA, ATM refers to average time for BAMI, AIB refers to average iteration for BA, and AIM refers to average iteration for BAMI.

Next, we investigated the significance of enhancement in terms of number of iterations required to get the best known solution. In achieving this, a set of statistical tests were carried out. The results were verified by Kolmogorov-Smirnov and Levene tests, whereby the outcome showed that only some of the data met the assumptions of normality distribution and equality of variance, while the remaining data did not. Because of this, t -test was used for normal data and the Wilcoxon Test was used for nonnormal data. Table 3 presents the results from statistical tests for both BAMI and BANV. Between the brackets is the algorithm that outperformed the other. Figure 2 illustrates the deference between average numbers of iterations for both algorithms across all datasets.

TABLE 3: Significant test.

Datasets	Significance test	Test type
M-of-N	.013 (BAMI)	Wilcoxon Test
Exactly	.010 (BAMI)	Wilcoxon Test
Exactly2	.010 (BAMI)	Wilcoxon Test
Heart	.035 (BAMI)	Wilcoxon Test
Vote	.000 (BAMI)	<i>t</i> -test
Credit	.000 (BAMI)	<i>t</i> -test
Mushroom	.572	Wilcoxon Test
LED	.000 (BAMI)	Wilcoxon Test
Derm	.004 (BAMI)	Wilcoxon Test
Derm2	.042 (BAMI)	Wilcoxon Test
WQ	.178	Wilcoxon Test
Lung	.480	<i>t</i> -test

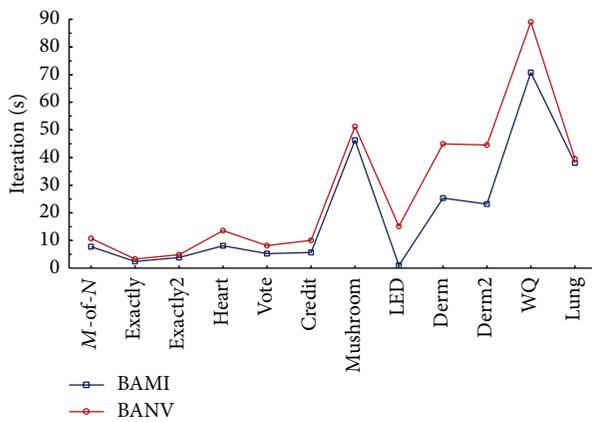


FIGURE 2: Average numbers of iterations across all datasets.

4. Discussion

The results showed that the performance of the proposed algorithm BAMI is superior and more efficient than BANV. As shown in Figure 2, BAMI recorded less time consumption and lower number of iterations to obtain the good solutions in all datasets. Statistically, it can be seen in Table 3 that BAMI performed significantly better than a standard BANV in nine out of 12 datasets. Next, we will discuss the results according to the time saving from biggest to smallest by observing the numbers of features and samples to see whether these factors are related with the time saving or the algorithm performance.

In the *LED* dataset, the proposed method achieved best known solution in the first iteration within 4.41 seconds, with an average of 90.45% time saving across all dataset. With more features but smaller samples in the *Derm* and *Derm2* datasets, the result for BAMI was very close to BANV, whereby the time saving is 41.40% in *Derm* and 43.28% in *Derm2*. In the *credit* dataset, with a decreased number of features and increased number of samples as compared to the previous dataset, the time saving is 38.85%.

The results also showed that in *Heart* dataset the time saving is 37.97%, which is very close to the *Credit* dataset in

spite of variation in number of features and samples, while in *Vote* dataset, with comparable number of samples and slightly higher number of features as compared to the *Heart* dataset, the decrease in time saving is up to 31.19%. It can also be seen that *M-of-N* dataset which has the same number of features as *Heart* dataset only has time saving of 28%.

Although both *Exactly* and *Exactly2* datasets have the same number of features and samples, their time saving is different. For *Exactly2*, the time saving is 25.51% while *Exactly* has time saving of 14.82%. In *WQ* dataset, the time saving is 22.25%, which is lower by 50% as compared to *Derm2* dataset in spite of the fact that both datasets have approximately the same number of features. In *Lung* and *Mushroom* datasets, the time saving is small, which is only 2.63% and 1.35%, respectively. From the results, it can be seen that the performance of BAMI is not affected by the number of features and samples in the dataset. The subsets obtained by both algorithms are the same, which implies that the proposed algorithm BAMI is able to maintain the same effectiveness while having higher efficiency.

5. Conclusion

In this study, a hybrid filter-wrapper approach named BAMI is presented. BAMI approach structurally integrates the MI model within BA using the Naive Bayes classifier. BAMI aims to bring together the efficiency of filter approach with higher accuracy from wrapper approach. In BAMI, MI was used to identify promising features which could potentially accelerate the process of finding the best known solution. The promising features were then used to replace several of the randomly selected features during the search initialization. BAMI was compared to BANV using the twelve datasets. The results showed that BAMI outperformed BANV in most datasets in terms of the computational time. The statistical test indicated that BAMI has significantly lower computation time than BANV in six out of twelve datasets. More importantly, this research presented a new feature selection technique that provided a good start point for further investigation and enhancement.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] M. H. Khooban and T. Niknam, "A new intelligent online fuzzy tuning approach for multi-area load frequency control: self Adaptive Modified Bat Algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 71, pp. 254–261, 2015.
- [2] X.-S. He, W.-J. Ding, and X.-S. Yang, "Bat algorithm based on simulated annealing and Gaussian perturbations," *Neural Computing & Applications*, vol. 25, pp. 459–468, 2014.
- [3] A. Alihodzic and M. Tuba, "Improved bat algorithm applied to multilevel image thresholding," *The Scientific World Journal*, vol. 2014, Article ID 176718, 16 pages, 2014.

- [4] K. Premkumar and B. V. Manikandan, "Speed control of Brushless DC motor using bat algorithm optimized Adaptive Neuro-Fuzzy Inference System," *Applied Soft Computing*, vol. 32, pp. 403–419, 2015.
- [5] T. A. Lemma and F. B. M. Hashim, "Use of fuzzy systems and bat algorithm for exergy modeling in a gas turbine generator," in *Proceeding of the IEEE Colloquium on Humanities, Science and Engineering (CHUSER '11)*, pp. 305–310, Penang, Malaysia, December 2011.
- [6] E. P. Xing, M. I. Jordan, and R. M. Karp, "Feature selection for high-dimensional genomic microarray data," in *Proceedings of the 18th International Conference on Machine Learning (ICML '01)*, pp. 601–608, San Francisco, Calif, USA, 2001.
- [7] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, "Hybrid filter–wrapper feature selection for short-term load forecasting," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 17–27, 2015.
- [8] S. Huda, M. Abdollahian, M. Mammadov, J. Yearwood, S. Ahmed, and I. Sultan, "A hybrid wrapper–filter approach to detect the source(s) of out-of-control signals in multivariate manufacturing process," *European Journal of Operational Research*, vol. 237, no. 3, pp. 857–870, 2014.
- [9] M. R. Sathya and M. Mohamed Thameem Ansari, "Load frequency control using Bat inspired algorithm based dual mode gain scheduling of PI controllers for interconnected power system," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 365–374, 2015.
- [10] T.-T. Nguyen, C.-S. Shieh, M.-F. Horng, T.-G. Ngo, T.-K. Dao, and T.-T. Nguyen, "Unequal clustering formation based on bat algorithm for wireless sensor networks," in *Knowledge and Systems Engineering*, V.-H. Nguyen, A.-C. Le, and V.-N. Huynh, Eds., vol. 326 of *Advances in Intelligent Systems and Computing*, pp. 667–678, Springer, 2015.
- [11] Z. W. Ye, M. W. Wang, W. Liu, and S. B. Chen, "Fuzzy entropy based optimal thresholding using bat algorithm," *Applied Soft Computing*, vol. 31, pp. 381–395, 2015.
- [12] R. Lewis, "Neuroscience rough set approach for credit analysis of branchless banking," in *Foundations of Intelligent Systems*, T. Andreasen, H. Christiansen, J.-C. Cubero, and Z. Raś, Eds., vol. 8502 of *Lecture Notes in Computer Science*, pp. 536–541, Springer, 2014.
- [13] E. F. Lashin, A. M. Kozae, A. A. Abo Khadra, and T. Medhat, "Rough set theory for topological spaces," *International Journal of Approximate Reasoning*, vol. 40, no. 1-2, pp. 35–43, 2005.
- [14] G. Huang, W. Zhao, and Q. Lu, "Bat algorithm with global convergence for solving large-scale optimization problem," *Application Research of Computers*, vol. 30, no. 5, pp. 1–10, 2013.
- [15] F. E. H. Tay and L. Shen, "Fault diagnosis based on Rough Set Theory," *Engineering Applications of Artificial Intelligence*, vol. 16, no. 1, pp. 39–43, 2003.
- [16] D. Tomar and S. Agarwal, "Hybrid feature selection based weighted least squares twin support vector machine approach for diagnosing breast cancer, hepatitis, and diabetes," *Advances in Artificial Neural Systems*, vol. 2015, Article ID 265637, 10 pages, 2015.
- [17] W. Wang, Y. Wang, and X. Wang, "Bat algorithm with recollection," in *Intelligent Computing Theories and Technology*, D.-S. Huang, K.-H. Jo, Y.-Q. Zhou, and K. Han, Eds., vol. 7996 of *Lecture Notes in Computer Science*, pp. 207–215, Springer, Berlin, Germany, 2013.
- [18] A. M. Taha, S.-D. Chen, and A. Mustapha, "Natural extensions: bat algorithm with memory," *Journal of Theoretical and Applied Information Technology*, vol. 79, no. 1, 2015.
- [19] A. M. Taha, S.-D. Chen, and A. Mustapha, "Multi-swarm bat algorithm," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 10, no. 12, pp. 1389–1395, 2015.
- [20] A. M. Taha, A. Mustapha, and S.-D. Chen, "Naive Bayes-guided bat algorithm for feature selection," *The Scientific World Journal*, vol. 2013, Article ID 325973, 9 pages, 2013.
- [21] A. M. Taha and A. Y. C. Tang, "Bat algorithm for rough set attribute reduction," *Journal of Theoretical and Applied Information Technology*, vol. 51, no. 1, pp. 1–8, 2013.
- [22] T.-S. Pan, T.-K. Dao, T.-T. Nguyen, and S.-C. Chu, "Hybrid particle swarm optimization with bat algorithm," in *Genetic and Evolutionary Computing*, H. Sun, C.-Y. Yang, C.-W. Lin, J.-S. Pan, V. Snasel, and A. Abraham, Eds., vol. 329 of *Advances in Intelligent Systems and Computing*, pp. 37–47, Springer, 2015.
- [23] P. Dash, L. C. Saikia, and N. Sinha, "Automatic generation control of multi area thermal system using Bat algorithm optimized PD–PID cascade controller," *International Journal of Electrical Power & Energy Systems*, vol. 68, pp. 364–372, 2015.
- [24] J. Sadeghi, S. M. Mousavi, S. T. A. Niaki, and S. Sadeghi, "Optimizing a bi-objective inventory model of a three-echelon supply chain using a tuned hybrid bat algorithm," *Transportation Research E: Logistics and Transportation Review*, vol. 70, no. 1, pp. 274–292, 2014.
- [25] S. Goyal and M. S. Patterh, "Performance of BAT algorithm on localization of wireless sensor network," *International Journal of Computers & Technology*, vol. 6, no. 3, 2013.
- [26] Y.-S. Chen and C.-H. Cheng, "Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry," *Knowledge-Based Systems*, vol. 39, pp. 224–239, 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

