Journal of Scientific Research and Development 2 (14): 158-165, 2015
Available online at www.jsrad.org
ISSN 1115-7569

Bat with Dempster theory for feature selection: A framework

Rozlini M. [1,*], Munirah M.Y. [1], Nawi N.M. [1], Wahid N. [2]

[1]University Tun Hussein Onn Malaysia (UTHM), Department of Software Engineering, Faculty of Computer Science and Computer Technology, 86400, Parit Raja, Johor, Malaysia
[2]University Tun Hussein Onn Malaysia (UTHM), Department of Multimedia, Faculty of Computer Science and Computer Technology, 86400, Parit Raja, Johor, Malaysia

Abstract: The important of good classifier and feature selection process need to be considering in achieving high classification performance. However most of the feature in datasets that did not contribute to end result is unknown. Some of the features need to be diminished to obtain better classification result. Therefore this research intends to overcome this issue. The objective of the research is to propose a feature selection framework based on enhancement of Bat algorithm (BA) with Dempster-Shafer. The research outcome is proposed framework that integrates BA with Dempster-Shafer had significantly improve the classification performance. Ongoing research will develop feature selection simulation based on our proposed framework.

Key words: Feature selection; Bat algorithm; Dempster-Shafer theory; Classification

1. Introduction

Probabilistic identification of pattern in data is required in most problems in data mining. Data mining also known as knowledge discovery is the process of data analyzing for pattern discovery in huge dataset. Data mining allow for the acquirement of meaningful information from large-scale data using a computer program. Many researchers interested to employ various data mining techniques either classical techniques or next generation techniques in order to solve the data mining problems such as classification, clustering, anomaly detection and feature learning.

Classification is one of the prominent problem exist in data mining. In classification, high classification performance is depending on accuracy of classifier to classify the data into predefined topic. To achieve this aim there two aspects need to be consider, the classifier and feature selection. Most of the feature in datasets that did not contribute to end result is unknown. Some unimportant or irrelevant feature need to be diminished in order to reduce the classification complexity and time processing. This matter is related with the data reduction where feature selection is needed in order to dimensionality reduction and performs the better performance in classification.

The most problem in classification is related with the number of features, while the problem in feature selection in finding the optimum features. Therefore, the problem in finding subset feature selection has a connection with optimization problem. Feature selection technique try to find the minimum number or attribute or in other word reduce attribute with maximum performance.

Some researcher tend to employ statistical method such as global information gain for instance (Shang et al., 2013) employ global information gain (GIG) to solve the redundancy ignores for text. While mutual information method were applied by a researcher to solve several problems, for instance (Lee and Kim, 2015a) multiple labels of feature selection tasks as well as (Doquire and Verleysen, 2013) applied together with greedy search algorithm for multi-label problem and in (Zhao et al., 2015) to select the most informative features. In the other research, heuristic method such as genetic algorithm (Lee et al, 2015) ACO (Kashef, and Nezamabadi, 2015), (Wang et al., 2015) and (Zhang et al., 2015) in mimetic feature selection, noisy data, spam email, binary variables; respectively.

Furthermore, Dempster (Baraldi et al., 2014) and rough set theory (Jensen and Parthalain, 2015) also employ to reduce the number of feature. Today, rough set theory gets more interest from the researcher. However, these two techniques are strongly connected (Skowron and Grzymala-Busse, 1994); (Wu et al., 2010); (Yao and Lingras, 1998). It demonstrated from the various rough set approximations space are associated with various belief. Also, the different dual pairs of lower and upper approximation operators from rough approximation space is corresponding to dual pairs of belief and plausibility functions from belief structures. Dempster-Shafer theory is a mathematical theory as general framework for uncertainty and incomplete information.

Bat algorithm (BA) or bat-inspired algorithm was developed by Xin-She Yang in 2010 (Yang, 2010). BA

* Corresponding Author.

is met heuristic optimization algorithm that based on echolocation characteristics of micro bats with varying pulse rates of emission and loudness. BA becomes the powerful technique for solving engineering optimization tasks (Yang and Hossein Gandomi, 2012). Also, be one of the possible solutions to solve data mining problems such as classification and feature selection. Combinatorial growths of possible solution through searching exist in classification and feature selection seen as an optimization problems. Classification aims to assign new data into the most suitable predefined topic, while feature selection aims to identify the most useful information from a given set of features. There are many potential issues have been highlighted on BA by previous researchers. Those issues involve the enhancement of BA to improve classification problems (Wang and Guo, 2013; Jaddi et Al., 2015), feature selection (Rodrigues et al., 2014; Rezghi and Obulkasim, 2014), and also exploration and exploitation of BA in global and local search (Yılmaza and Kücüksille, 2015).

Therefore, this research aims to propose the framework for enhancement of the BA with Dempster-Shafer theory to find the optimum feature or subset feature to improve the classification performance.

The rest of this paper is organized to provide a brief explanation of Feature Selection, Bat Algorithm and Dempster-Shafer in Literature Review section. Then, the following section will be discussed about previous work from the other researcher. While, in the next section will come out with proposed framework that discuss the details of each stage in framework. Discussion and conclusion become our final section.

2. Literature review

2.1. Feature selection

Feature selection has been an active and fruitful field of research area in pattern recognition, machine learning, statistics and data mining communities (Han et al., 2011). It is a dimensionally reduction technique that main goal is to reduce irrelevant data and finding a features that increase classification accuracy. The main objective of feature selection is to choose a subset of input variables by eliminating features, which are irrelevant or of no predictive information. It has been proven in both theory and practice to be effective in enhancing learning efficiency, increasing predictive accuracy and reducing complexity of learned results (Almuallim and Dietterich, 1994), (Koller and Sahami, 1996).

2.1.1. Feature selection framework algorithm

There are four basic steps in a typical feature selection process as shown in Fig. 1 (Hall and Smith, 1997).

The process of feature selection is as below;

- The generation procedure to generate the next candidate subset from original feature set
- The evaluation function to evaluate the subset to determine the relevancy towards the classification task using measure for instances distance, dependency, information and consistency
- Stopping criteria to decide when to stop. This is where it determine the relevant subset or optimal feature subset
- Validation procedure is to check whether the selected feature subset is valid

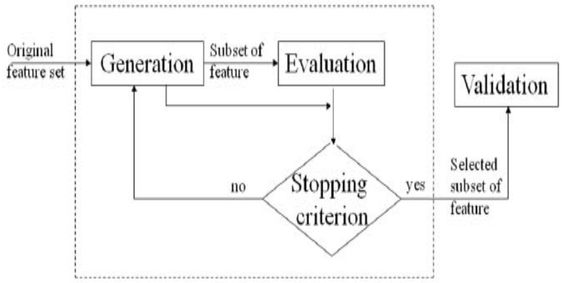

Fig. 1: Feature selection process (Hall and Smith, 1997).

2.2. Bat algorithm

Bat algorithm (BA) or bat-inspired algorithm was developing by Xin-She Yang in 2010 (Yang, 2010). BA is meta heuristic optimization algorithm that based on echolocation characteristics of micro bats with varying pulse rates of emission and loudness. The development of BA can be done in many ways but for simplicity BA use the following approximate or idealized rules:

- All bats use echolocation to sense distance, and they also 'know' the difference between food/prey and background barriers in some magical way;
- Bats fly randomly with velocity v_i at position x_i with a fixed frequency f_{min}, varying wavelength λ and loudness A_0 to search for prey. They can automatically adjust the wavelength (or frequency) of their emitted pulses and adjust the rate of pulse emission $r \in [0,1]$, depending on the proximity of their target;
- Although the loudness can vary in many ways, we assume that the loudness varies from a large (positive) A_0 to a minimum constant value A_{min}.

Therefore, from the three rules above the location x_i^t and velocities v_i^t can be update according to following equations;

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

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

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

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

These simplified assumptions also use the following approximations, for simplicity. In general the frequency f in a range $[f_{min}, f_{max}]$

corresponds to a range of wavelengths $[\lambda_{\min}, \lambda_{\max}]$. For example a frequency range of $[50\ khz, 500\ khz]$ corresponds to a range of wavelengths from 0.5mm to 50mm.

Bat algorithm
Objective function f(x), x=(x1,....xd)^T
Initialize the bat population x_i (i=1,2,...,n) and v_i
Define pulse frequency f_i at x_i
Initialize pulse rates r_i and the loudness A_i
***while** (t<Max number of iterations)*
Generate new solutions by adjusting frequency, and updating velocities and locations/solutions [equations 1 to 3]
***if** (rand > r_i)*
Select a solution among the best solutions
Generate a local solution around the selected best solution
end if
generate a new solution by flying randomly
***if** (rand < A_i & $f(x_i) < f(x_*)$)*
Accept the new solutions
Increase r_i and reduce A_i
end if
Rank the bats and find the current best x_*
end while
Post process results and visualization

Fig. 2: Pseudo code of the bat algorithm (BA)

Any value of wavelength can be used in problem for the ease implementation. In actual implement, the value of range can be adjust with adjusting the wavelengths (or frequencies), and the detectable range (or the largest wavelength) should be chosen such that it is comparable to the size of the domain of interest, and then toning down to smaller ranges. Moreover, differ value of frequency can be used while fixing the wavelength λ. In fact, λ and f are associated because of λf is constant.

For simplicity, let $f \in [0, f_{\max}]$. The higher frequencies have short wavelengths and travel a shorter distance. For bats, the typical ranges are a few meters. The rate of pulse can simply be in the range of [0, 1] where 0 means no pulses at all, and 1 means the maximum rate of pulse emission. Based on these approximations and idealization, the basic steps of the Bat Algorithm (BA) can be summarized as the pseudo code shown in Fig 2.

2.3. Dempster Shafer theory

Dempster-Shafer theory is an approach to combining evidence that was originated by Dempster's (Dempster, 1967) in 1967 and extended by Shafer (Shafer 1976) as a theory. Dempster-Shafer theory is a mathematical theory as general framework for uncertainty and incomplete information. The significant of this theory that allocate the probability mass of set or information, belief and plausibility function. The Dempster-Shafer theory requires satisfying the following hypotheses:

- Each hypothesis in a hypothesis set must be mutually exclusive.
- There exists only one hypothesis that we look for in the hypothesis set In this paper, A_i denotes the domain of attributes in Ω. $\{a_1, a_2, \ldots a_n\}$ denote the set of hypotheses. The number of subsets obtained from the hypothesis set is $2^n - 1$. The subsets can be denoted as follows:

$$\Omega = \{A_1, A_2, \ldots A_{(2^n - 1)}\} \quad (4)$$

Each of subsets is evaluated with basic probabilities $m(A_i)$ assigned to subset A_i. m is also called mass function. A mass function satisfies the following equations:

$$m(\phi) = 0 \quad (5)$$

$$\sum_{A_i \subseteq \Omega} m(A_i) = 1 \quad (6)$$

where ϕ denotes empty set.

The basic probability assigned on Ω can be interpreted as the lack of knowledge and information. It is because this basic probability $m(\Omega)$ cannot be assigned to a more concrete set of hypotheses.

Based on the Dempster-Shafer theory we can evaluate the upper and lower probabilities of each subset A_i using the basic probabilities. These upper and lower probabilities are called plausibility and belief functions denoted by $Pl(A_i)$ and $Bel(A_i)$, respectively.

$$Bel(A_i) = \sum_{A_j \subseteq A_i} m(A_j) \quad (7)$$

$$Pl(A_i) = 1 - Bel(\overline{A}_i) = \sum_{A_j \cap A_i \neq 0} m(A_j) \quad (8)$$

$$Bel(A_i) \leq Pl(A_i) \quad (9)$$

The Dempster rule of combination enables us to combine evidence from different sources without any knowledge of their distributions. The Dempster combination rule evaluates the basic probability $m(A_k)$ when independent information of 1 and 2 obtained at the same time and these basic probabilities $m_1(A_{1i})$ and $m_2(A_{2j})$ are given for each of information 1 and 2, respectively. The combination rule is written in the following:

$$m(A_k) = \frac{\sum_{A_{1i} \cap A_{2j} = A_k} m_1(A_{1i}) \cdot m_2(A_{2j})}{1 - \sum_{A_{1i} \cap A_{2j} = \phi} m_1(A_{1i}) \cdot m_2(A_{2j})}$$

$$m(\phi) = 0 : \forall A_k \neq \phi \quad (10)$$

In other words, the new separation of the hypothesis set is given by the combination between the information 1 and 2. Each intersection $A_{1i} \cap A_{2j}$ will be a new subset but these are given the same name when these subsets indicate the same set, and a null set will be evaluated automatically as $m(\phi) = 0$ where $A_{1i} \cap A_{2j} = \phi$ even if $m_1(A_{1i})$ and $m_2(A_{2j})$ are positive.

3. Related works

3.1. Feature selection

Most of the feature in datasets that did not contribute to end result is unknown. Some feature is not important or irrelevant need to diminish the number of feature in order to reduce the complexity and time processing. This matter is related with the data reduction where feature selection is needed in order to dimensionality reduction and performs the better performance in classification. There are several Meta heuristic algorithms are used in classification, but the well-known such as Genetic Algorithm (GA), Cuckoo Search (CS), Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO) and Bat Algorithm (BA). Some feature selection method that used by researcher as literatures are shown in Table 1.

3.2. Bat algorithm

BA is one of the popular meta heuristic techniques. The problem or issue in BA and techniques are used together either to enhanced or embedded with BA are summarize in Table 2. From the literature researcher likely to employ BA together with other techniques in order to solve the problem related with optimization in classification, feature selection and other.

Some scholar likely to solve the problem in feature selection, or instance in (Nakamura et al., 2012) was proposed Bat algorithm for feature selection that seen as optimization problem. Since that the problem the feature have two condition selected or not and search space given by a Boolean hypercube, a sigmoid function for new bat position. However, this research did not consider the feature weight to show degree of important feature in dataset. Other scholar, (Taha et al., 2013) also proposed bat algorithm guided by Naïve Bayes algorithm as a feature selection. This research has put forward on frequency, velocity, position adjustment and other part in BA algorithm in order to improve the classification performance. The proposed algorithm did not cover the very high-dimensional dataset and in gene expression. Also did not consider the feature weight.

Table 1: Feature selection techniques

Author(s)	Problems	Technique(s)
(Shang et al., 2013)	only evaluate features individually but completely ignore the redundancy between them	global information gain (GIG)
(Lee and Kim, 2015a).	multiple labels feature selection task	Mutual Information, interaction information
(Zhang et al., 2015).	feature selection problem with binary variables	Bare Bones Particle Swarm Optimization (BPSO)
(Zhao et al., 2015).	Features in huge dataset are irrelevant or redundant, which typically deteriorates the performance of machine learning algorithms	conditional mutual information-based criterion, feature vector graph
(Doquire and Verleysen, 2013).	Multi-label	Greedy search algorithm, mutual information, pruned problem transformation
(Lee and Kim, 2015b).	Premature convergence	Genetic algorithm (GA)
(Wang et al., 2015).	Spam	frequency based feature selection method (ODFFS), optimal term frequency based feature selection (OTFFS), global best harmony oriented harmony search (GBHS),
(Kashef and Nezamabadi-pour, 2015).	Noisy, irrelevant and redundant data.	ACO , for feature selection, binary ACO (BACO)
(Lee et al., 2015).	Data consists of vast volume and noises from the human body	Fuzzy–Rough Feature Selection, Multi tree Genetic Programming.

A hybrid meta heuristic method was proposed in (Wang and Guo, 2013) that to improve the bat algorithm by combining with harmony search algorithm. This research found that BA cannot perform global search well due to trapping in local optima. The improvement that has been made is used fixed frequency and loudness value instead of variety of value.

From the literature, some scholar put the focus on search space such as use the fixed value of parameter and reformulated the formula in BA. However, the feature weights are not considered.

3.3. Dempster-Shafer theory

Dempster-Shafer theory is a probability theory that able to solve data mining problems. The importance of information or data in determine with real number in range between 0 and 1. Other techniques required to compute probabilistic on data such as fuzzy and rough set (Pawlak, 1982). Furthermore, rough set and Dempster theory are strongly connected. It demonstrated from the various rough set approximations space are associated with various belief. Also, the different dual pairs of lower and upper approximation operators from rough approximation space is corresponding to dual pairs of belief and plausibility functions from belief structures (Skowron and Grzymala-Busse, 1994); (Wu et al., 2010); (Yao and Lingras, 1998). Because of the strong connection with these two techniques, the researcher try to employ Dempster-Shafer theory instead of rough set theory in a few problems as shown in Table 3.

Table 2: Related work on feature selection with BA

Author	Problem	Other method(s)
(Rodrigues et al., 2014)	optimizing classifier predictive performance and addressing the curse of the dimensionality	Optimum-Path Forest (OPF)
(Nakamura et al., 2012)	To find the most informative feature in a search space that given in boolean hypercube in high dimensionality in object description.	Optimum-Path Forest (OPF)
(Taha et al., 2013)	Increasing the amount of data and information	Naïve bayes

Table 3: Demspter for attribute reduction

Author(s)	Problem	Techniques	Domain/ Application
(Baraldi et al., 2014)	Uncertainty	Dempster-Shafer theory of evidence, possibility theory and fuzzy random variable	Nuclear power plant
(Liu et al., 2014)	Incomplete	Rough set	Medical diagnosis
(Dai and Xu, 2012)	Incomplete and uncertainty	Rough set and information theory approach	Nine real-life data sets available from the UCI Repository of Machine Learning Database at University of California

Consequently, it is very motivating to discover Dempster theory in decision making. For instance in (Tang, 2015) have discover Dempster theory with fuzzy to solve problem in medical diagnosis. Dempster was applying in this research to combine the number of parameter or evidence with degree into collective evidence. Then, the solution form these several of evidence or parameter will be derive using Dempster without knowing much prior information.

In (Troiano et al., 2015) was employed Demspter theory to discover user preferences from item characteristic. This research is to solve the problem of filtering the interesting product for customer from an e-shop catalogue. The item preferred by user according to measurement of Belief and Plausibility functions.

In another research (Yue et al., 2015), Dempster theory was apply to multi-software reliability allocation in multimedia system. In this research, Dempster was able to identify relative reliability weights.

Therefore, the decision making can be considered one of the issues can be relate with optimization problem. A decision has to decide either feature is relevant or not to predefine group and need to dismiss or not. When making a decision, it is supposed to have a good decision that may prompt action or in other word the optimum solution. There are many features in dataset, but the best features or the degree of important for each feature are unknown.

From (Wang and Guo, 2013) BA cannot perform global search well due to trapping in local optima. This research will overcome the above problem with apply Dempster-Shafer theory to generate a new solution to select the feature according to feature weight. From the literature above, Dempster-Shafer is one of the powerful probability techniques that based on belief and plausibility function.

4. The proposed framework

In this research, the features of BA and Dempster are find out. The outstanding research from some scholar and the issue in features selection and classification are also discovered. Therefore, it is very meaningful to explore the Dempster-Shafer theory with BA to solve the feature selection problem in classification.

As a result, the research framework is proposed in this research that consists of three main phases as shown in Fig. 3. The phases involved in the research are:

- Phase 1: Dataset acquisition
- Phase 2: Feature selection
- Phase 3: Performance evaluation

4.1. Phase 1: Dataset acquisition

In this research, the experiment will be set up to test the proposed framework on the number of datasets.

The dataset from UCI Machine Learning repository from different domain will use for this experiment. Then, selected datasets will be group according to number of attribute as shown in Table 5. The dataset group will be divided into three scales depending to the number of attributes which is small-scale for attribute less than 20, medium-scale in from 20 to less than 50 attribute and large-scale for attributes in range from 50 to 100 according to (Kudo and Sklansky 2000).

4.2. Phase 2: Feature selection

During this stage, all the parameters in original BA need to assign initial position, x , velocity, v, and frequency, f for each bat. Also the initial value for loudness, A and pulse rate r. All these value are according the researcher or in other word it gives by random within range that already specify.

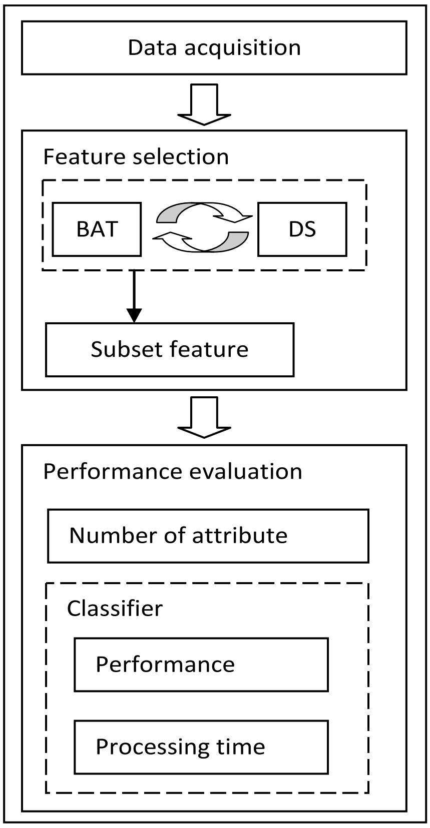

Fig. 3: The proposed research framework

Table 5: Dataset for research experiment

Scale	Datasets
Small scale ($attribute < 20$)	Heart Abalone Credit approval Zoo Yeast
Medium scale ($20 \le attribute < 50$)	Lung cancer Dermatology Mushroom Soybean Trans
Large scale ($50 \le attribute \le 10$	Audiology Libras movement

Each bat moves in the search space and will update their new velocity and location. For each movement the local solution or local best will be generated around the best solution according to loudness and pulse rate. However, in case of feature selection, the search space is can be represent by dataset in domain. Since that a number of attribute or feature in dataset, BA need to find out the optimum feature according to the best solution that generated for each bat in each iteration. As mention before, all the initial value for all the parameter in BA not follow any rules and the search are completely depend on random walk.

For instance, the number of feature and generation or iteration is determined. Let determine dataset with 8 attribute and 5 generation represent 8 features and 5 Bats, respectively. While, the Bat position is corresponding to the selected feature. However not all the feature or attribute are selected. For each generation or iteration the value of local best and global best is computed. Then, a sub features with highest global best is selected. Feature selected according to the maximum value of global best among bat is selected as the best.

The importance of feature is different. The feature that contribute or important to end result is unknown. Might be the same feature but have the different degree of important to dataset as well as classifier. The optimum feature can be selected according to high degree of important. In this research, during the bat walk in search space to find the features, Dempster-Shafer will apply to control the bat movement. Here, without consider prior knowledge, Dempster will compute the mass movement for attribute or feature in datasets. At the end of the process, the selected feature or subset feature will be determined. Feature for the best global best is choosing as selected feature or sub features. The subset feature will evaluate the effectiveness through the performance of classifier.

4.3. Phase 3: Performance evaluation

Sub feature from the phase 2 will be evaluated. Also, other feature selection method such as gain ratio and DPSORS (Chung and Wahid, 2012) will employ for dataset in Table 5. Then, the number of attribute will compare with proposed techniques, gain ratio and DPSORS (Chung and Wahid, 2012). Then, in order to proof that proposed techniques is outperform, the experiment will be done with several benchmark classifiers such as k-Nearest Neighbor (kNN) and Naive Bayes to observe the effectiveness of the new technique. Lastly, processing time will be compute and compare with original Bat algorithm.

5. Discussion and conclusion

This research was discovered the advantages of BA and Dempster-Shafer theory as well as feature selection issue or problem in classification. High classifier accuracy indicates the high classification performance. However, most feature or attribute in dataset did not contribute to end results is unknown. In this research, irrelevant feature for classification process is does know. But, feature and classification performance are related.

The decision making are considered as one of the issues can be relate with optimization problem. There are many features in dataset, but the best features or the degree of important for each feature are unknown. BA is one of the powerful techniques to solve the optimization problem but, from (Wang and Guo, 2013) BA cannot perform global search well due to trapping in local optima.

Therefore, this research will overcome the above problem with apply Dempster-Shafer theory to generate a new solution to select the feature according to feature weight. From the literature above, Dempster-Shafer is one of the powerful probability techniques that based on belief and plausibility function. In this research, Dempster-Shafer will apply to control the bat movement. Here, without consider prior knowledge, Dempster will compute the mass movement for attribute or feature in datasets. At the end of the process, the selected feature or subset feature will be determined.

Consequently, this research proposed a framework for feature selection technique that utilizes the benefits from bat algorithm and Dempster-Shafer theory to improve classification performance. This research is ongoing research; the proposed framework will apply in the stage of research. To prove that the proposed technique is outperformed, in future experiments the comparison will be done with benchmarks feature selection techniques and classifier techniques. The number of attribute from proposed technique will compare with gain ratio and DPSORS (Chung and Wahid (2012)). Furthermore, another two issues performance and processing time, the comparison will be done with k-NN and Baiyes Naives for classifier techniques.

References

A. Skowron, and J. Gryzymala-Busse (1994). From rought set theory to evidence theory, in:R.R. Yager, M. Fredrizzi, J. Kacprzyk (Eds), Advance in the Dempster-Shafer Theory of Evidence, Wiley, New York, pp. 193-236.

A.M. Taha, A. Mustapha, and S.D. Chen (2013). Naive Bayes-Guided Bat Algorithm for Feature Selection, The Scientific World Journal, vol. 2013, Article ID 325973, 9 pages, 2013. doi:10.1155/2013/325973.

A.P. Dempster (1967). Upper and lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics, vol. 38 (2), pp. 325-339, April 1967.

C. Shang, M. Li, S. Feng, Q. Jiang, and J. Fan. (2013). Feature selection via maximizing global Information Gain for text classification, Knowledge-Based Systems, vol. 54, pp. 298-309, December 2013.

D. Koller, and M. Sahami (1996). Toward optimal feature selection. In Proceedings of the Thirteenth International Conference on Machine Learning (ICML'96), pp. 284–292.

D. Liu, T. Li, and J. Zhang (2014). A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems, International Journal of Approximate Reasoning, vol. 55(8), pp. 1764-1786, November 2014.

D. Rodrigues, L.A.M. Pereira, R.Y.M. Nakamura, K.A.P. Costa, X.S. Yang, A.N. Souza, and J.P. Papa (2014). A wrapper approach for feature selection based on bat algorithm and optimum-path forest. Expert Systems with Applications, vol. 41(5), pp. 2250–2258, April 2014.

F. Yue, G. Zhang, Z. Su, Y. Lu, and T. Zhang (2015). Multi-software reliability allocation in multimedia systems with budget constraints using Dempster-Shafer theory and improved differential evolution, Neurocomputing, vol. 169, pp. 13-22, May 2015.

G. Doquire, and M. Verleysen (2013). Mutual Information-based feature selection classification, Neurocomputing, vol. 122, pp. 148-155, December 2013.

G. Shafer (1976). A mathematical theory of evidence, Princeton University Press, Pricenton, 1976.

G. Wang, and L. Guo, (2013). Novel hybrid bat algorithm with harmony search for global numerical optimization. Journal of Applied Mathematics, vol. 2013. Article ID 696491, doi.org/10.1155/2013/696491.

G. Zhao, Y. Wu, F. Chen, J. Zhang, and J. Bai (2015). Effective feature selection using feature vector graph for classification, Neurocomputing, vol. 151(Part 1), pp. 376-389, March 2015

H. Almuallim, and T.G. Dietterich (1994). Learning boolean concepts in the presence of many irrelevant features. Artificial Intelligence, vol. 69 (1-2) , pp. 279–305, September 1994.

H. Tang (2015). A novel fuzzy soft set approach in decision making based on greyrelational analysis and Dempster–Shafer theory of evidence, Applied Soft Computing, vol. 31(C), pp. 317-325, June 2015.

J. Dai, and Q. Xu (2012). Approximations and uncertainty measures in incomplete information systems. Information Sciences: an International Journal, vol. 198, pp. 62-80, September 2012.

J. H. Lee, J.R. Anaraki, C.W. Ahn, and J. An. (2015). Eficient Classification System based on Fuzzy-Rough Feature Selection and Multitree Genetic Programming for intension Pattern Recognition using Brain Signal, Expert Systems with Applications, vol. 42(3), 1644-1651, February 2015.

J. Han, M. Kamber, and J. Pei (2011). Data Mining: Concepts and Techniques. 2nd edition. Morgan Kaufmann, San Francisco, CA. June 2011.

J. Lee, and D. W. Kim (2015a). Memetic feature selection algorithm for multi-label classification, Information Sciences, vol. 293, pp. 80-96, February 2015.

J. Lee, and D.W. Kim (2015b). Mutual information-based multi-label feature selection using interaction information, Expert Systems with Applications, vol. 42(4) , 2013-2025, March 2015.

L. Troiano, L. J. Rodriquez-Muniz and I. Diaz (2015). Discovering user preferences using Dempster-Shafer theory, Fuzzy Sets and Systems, vol.278, pp. 98-117, June 2015.

M. Dorigo, V. Maniezzo, and A. Colorni (1996). The ant system: optimization by a colony of cooperating agents, IEEE Trans. On Systems, Man, and Cybernatics-Part B, vol. 26(1), pp. 29-41, February 1996.

M. Kudo and J. Sklansky (2000). Comparison of algorithms that select features for pattern classifiers, Pattern Recognition, vol. 33 (1), pp. 25-41, January 2000.

M. Rezghi, and A. Obulkasim (2014). Noise-free principal component analysis: an efficient dimension reduction technique for high dimensional molecular data. Expert System with Application, vol. 41(17), pp. 7797-7804, December 2014.

M.A. Hall, and L.A. Smith (1997). Feature subset selection: a correlation based filter approach, In Proc. Of the 1997 International Conference on Neural Information Processing and Intelligent Information Systems, pp. 855-858, Springer Heilderberg.

N. S. Jaddi, S. Abdullah, and A. R. Hamdan (2015). Multi-population cooperative bat algorithm-based optimazation of artificial neural network model. Information Sciences, vol. 294, pp. 628-644, February 2015.

P. Baraldi, M. Compare, and E. Zio (2014). Uncertainty treatment in expert information system for maintenance policy assessment, Applied Soft Computing, vol. 22, pp. 297-310, September 2014.

R. Jensen and N. Mac Parthalain (2015). Towards scalable Fuzzy-Rough feature selection, Information Sciences, vol.323 , pp. 1-15, December 2015.

R.Y.M Nakamura, L.A. Pereire, K.A. Costa, D. Rodigues, and J.P. Papa (2012), BBA: a binary bat algorithm for feature selection, 2012 XXV SIBGRAPI Conference on Graphics, Patterns and Images, Aug 22-25, 2012, pp. 291-297.

S. Kashef and H. Nezamabadi-pour (2015). An advanced aco algorithm for feature subset selection, Neurocomputing vol. 147, pp. 271-279, January 2015.

S. Yılmaza, and E.U. Kücüksille (2014). A new modification approach on Bat Algorithm for solving optimization problems. Applied Soft Computing,vol. 28(c), pp. 259–275, March 2015.

W.Z. Wu, Y. Leung, and W.X. Zhang. (2010). Connections between rough set theory and Dempster-Shafer theory of evidence, International Journal of General System, vol. 31(4), pp. 405-430, September 2010

X.S. Yang (2010). A new metaheuristic Bat-inspired algorithm, in: J.R. Gonzalez, et al.(Eds.), Nature Inspired Cooperative Strategies for Optimization (NISCO 2010). Studies in Computational Intelligence, Springer Berlin, 284, Springer, Berlin,2010, pp. 65–74, August 2010.

X.S. Yang, and A. Hossein Gandomi (2012). Bat algorithm: a novel approach for global engineering optimization. International Journal for Computer-Aided Engineering and Software. vol. 29(5). pp. 464-483.

Y. Wang, Y. Liu, L. Feng, and X. Zhu (2015). Novel feature selection method based on harmony search for email classification, Knowledge-Based Systems, vol. 73, pp. 311-323, January 2015.

Y. Zhang, D. Gong, Y. Hu, and W. Zhang (2015). Feature Selection Algorithm based on Bare Bones Particle Swarm Optimazation, Neurocomputing, vol. 148, pp.150-157, January 2015.

Y.Y. Chung, and N. Wahid (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO), Applied Soft Computing, vol. 12(9), pp. 3014–3022, September 2012.

Y.Y. Yao, and P.J. Lingras (1998). Interpretations of belief functions in the theory of rough sets, Information Sciences, vol. 104(1-2), pp. 81-106, January 1998.

Z. Pawlak (1982). Rough sets, International Journal of Computer and Information Sciences, vol. 11(5), pp. 341-356, October 1982.