International Journal of Mathematics and
Computer Science, **16**(2021), no. 1, 213–228

Arabic Text Classification Using Hybrid Feature Selection Method Using Chi-Square Binary Artificial Bee Colony Algorithm

Musab Hijazi[1,3], **Akram Zeki**[2], **Amelia Ismail**[3]

[1]Department of Computer Science and Software Engineering
College of Engineering
Al Ain University
Al Ain,UAE

[2]Department of Information Technology
Kulliyyah of Information and Communication Technology
International Islamic University Malaysia
Kuala Lumpur, Malaysia

[3]Department of Computer Science
Kulliyyah of Information and Communication Technology
International Islamic University Malaysia
Kuala Lumpur, Malaysia

email: musab.hijazi@aau.ac.ae, akramzeki@iium.edu.my,
amelia@iium.edu.my

(Received October 21, 2019, Revised March 2, 2020,
Accepted May 13, 2020)

Abstract

Text classification is a popular method in data mining. It is utilized to get valuable information from the vast quantity of data. Feature selection is a crucial step in Text classification. It is a vital preprocessing technique for powerful data analysis, where only a subset from the original data features is chosen by removing noisy, irrelevant, or

Key words and phrases: Arabic Text Classification, Artificial Bee Colony, Feature Selection, Text Mining.
AMS (MOS) Subject Classifications: 68T20.
ISSN 1814-0432, 2021, http://ijmcs.future-in-tech.net

redundant features. In this paper, a feature selection method utilizing the combination of chi-square and Artificial Bee Colony (ABC) is proposed. Chi-square, a filter method that is computationally fast, simple and has the ability to deal with a large dimensional feature, is used as the first level of the feature selection process. After that, the wrapper method, Artificial Bee Colony algorithm, is used as the second level where Naive Base is used as a fitness function. The results showed that a reduced number of features outperformed classification accuracy to that using the original features set. Furthermore, the proposed method had a better performance compared with the chi-square method and the ABC algorithm as a feature selection method.

1 Introduction

Documentation is the best method to illustrate knowledge, which implies that the substantial repositories of information are documents [1]. Due to the rapid expansion of the internet, there is a massive growth in the number of electronic documents, which complicate the process of information retrieval especially documents that dont have keywords or they have but unfortunately, the keywords dont reflect the content of documents. Text classification and text clustering are vital and effective ways that will help to ease the process of information retrieval [2].

The process of grouping or classifying documents into a pre-defined set of classes based on a set of criteria that defined in advance is called text classification [3]. Text classification has been exploited in various applications such as documents organization, automated indexing of documents, spams filtering, text filtering, word sense disambiguation [4]. The main phases of text classification are: preprocessing, data division, feature extraction, feature selection, build the classifier model based on selected features, and finally evaluate the model.

Preprocessing is an attempt to improve text classification by removing of worthless information. It may include the removal of punctuation, numbers, and stop words which are words that can be found in any text like prepositions and pronouns to improve the text classification process [5].

After preprocessing, the data is divided into two parts, training data, and testing data. The classification algorithm will be trained using training data to produce a classification model. The performance of the classification model will be assessed using testing data. Since there is no ideal split of data to the training part and testing part, different scenarios have been used for

Arabic text classification research in a range from 25 percent for training and 75 percent for testing up to 80 percent for training and 20 percent for testing [1]. The k-fold cross-validation is usually used with an imbalanced dataset to overcome the overfitting problem where a set of partitions for training and testing are used to produce k-classification models.

Feature extraction is the process of extraction a list of features and their corresponding frequency in the training dataset. The output of the feature extraction is a long list of features, its range from thousand to hundreds of thousands. Not all of these features are useful for classification for several reasons: firstly, the performance of some classification algorithms is negatively affected by a large number of features due to which called the curse of dimensionality. Secondly, an over-fitting problem may occur when the classification algorithm is trained in all features. Thirdly, a large number of these features occur rarely, only once or twice, in the training data. Finally, some features are common in all or most of the classes which imply waste-time in the data analysis process [1]. So, having an accurate and efficient low-dimensional data from high-dimensional one is needed.

Analysis of enormously high-dimensional data, by removing unnecessary one, is a substantial process in data mining which called dimension reduction. Feature selection (FS) is one of the dimension reduction process [6]. Feature selection has two methods: wrapper and filter method. In the wrapper method, features are picked out or filtered according to the classification accuracy while in the filter method, features are filtered based on scoring metric [4] [6].

Usually, the data is represented as a matrix with n rows and m columns such that the rows correspond to the texts in the training data, and the columns correspond to the selected feature. The value of each cell in the matrix represents the weight of the feature in the text. The classification algorithm is trained using the training matrix that contains selected features to build a classification model. The performance of the classification model will be evaluated using different methods such as accuracy, precision and recall, and F-measure [4].

In this paper, a Feature selection method for Arabic text classification (ATC) using the chi-square method and Artificial Bee Colony (ABC) algorithm is proposed. Our work proposes two levels of feature selection. In the first level, the chi-square method that is computationally fast, simple, and can deal with a large dimensional feature is used as a filter method to diminish the number of features. In the second level, the binary Artificial Bee Colony (ABC) algorithm is applied to find out a subset of features that has

the highest classification accuracy to be used in building the classification model. The experimental results show that a reduced number of features can outperform the classification accuracy to that using the original features set and the proposed method has a better performance compared with using the chi-square only or the ABC algorithm as a feature selection method.

The paper is organized as follows: first of all, background knowledge about the Artificial Bee Colony (ABC) algorithm and related work are described. After that, the proposed feature selection method is presented. Experimental results for applying the proposed feature selection method on the BBC dataset are discussed. Finally, the conclusion of the research is presented

2 ABC Algo.-Background and related work:

Feature selection (FS) is the process of electing features as a subset of the whole features set, which can be considered as exploring into a state space. The whole space is explored either in a full or heuristic search strategy. The full search strategy is an impracticable strategy when we have a large number of features. In heuristic search, features that are not selected yet will be considered at each step for evaluation. A random subset will be generated using a random search approach. The heuristic search strategy is used in many bio-inspired and genetic algorithms. Moreover, The search can be divided into three approaches based on the behavior and initialization through the search step: a forward approach, the process will be started with empty feature subsets, through the selection process, features will be inserted in the subset; backward approach: the process will be started with a set of all features, through the feature selection process, features will be removed; a bidirectional approach: features either will be coming in or eliminated through the selection process. Furthermore, feature selection methods are divided into two approaches: wrapper and filter methods. In the filter method, a process of filtering will be started before the process of classification. For each feature, a score value will be calculated. Features with higher values are used to represent or outline the original dataset. In the wrapper method, a set of candidate or nominee features will be created by inserting and eliminating features based on the accuracy of the classifier. Usually, superior classification results will be achieved by wrapper methods compared to filter methods [7].

Evolutionary algorithms have been utilized as a feature selection method

by many researchers such as genetic and swarm algorithms[7]-[18]. One of swarm intelligent algorithms used in many research areas to deal with optimization problems is the Artificial Bee Colony (ABC) algorithm proposed by Karaboga [19] based on the attitude of bees while collecting honey. The process begins when bees fly to find out food sources (nectar). The nectar is kept in the bees stomach when finding it. The nectar is unloaded and a waddle dance is performed to partake information regarding food source (quantity of nectar, direction, and distance) when the bees come back to the beehive. New bees will recruit to scout most wealthy food sources [20]. The minimum model of Artificial Bee Colony represents the collective intelligence of bees swarm made up of three parts: food sources, employed, and unemployed bees [21]. Food sources which represent potential solutions. Employed bees that search for a food source have information regarding the quality of food source and partake this information with other bees. The number of employed bees and food sources should be the same. Unemployed bees can be onlooker bees or scout bees. Information regarding the goodness of food sources is shared with onlooker bees who select food sources with superior quality to scout about their neighborhood. Onlooker bees become employed bees when they select a food source to explore. When the food source is exhausted, employed bees become scout bees and trying to find out a new food source. A pseudo-code for the ABC optimization approach is shown in Figure 1 [20].

Algorithm 1 ABC optimization approach
1: Initialization Phase
2: **Repeat**
3: Employed Bee Phase
4: Onlooker Bee Phase
5: Scout Bee Phase
6: Memorize the best solution achieved so far
7: **until** (Cycle = Maximum Cycle Number of a Maximum CPU time)

Figure 1: A general pseudo-code for the ABC optimization approach

2.1 Initialization phase

An arbitrary creation of food sources was proposed in the original algorithm [22], such that each one of the food sources looks like a potential solution to

the problem.

$$x_{ij} = x_j^{min} rand\,(0,1)\left(x_j^{max} - \ x_j^{min}\right) \qquad (1)$$

where i = *1,...,* N, such that N represents the number of food sources and j = *1,....,* M, such that M is the number of optimization parameters.

2.2 Employed bee phase

The neighborhood of each food source x_i will be explored by an employed bee to find out a new food source v_i having more nectar and this done by changing one parameter of x_i. The exploration of the neighborhood is represented by the following formula.

$$v_{ij} = x_{ij} + \ \phi_{ij}\,(\ x_{ij} - \ x_{kj}\ ) \qquad (2)$$

A food source v_i for every food source x_i is specified based on the amendment of an optimization parameter j, which means the modification of x_{ij}. j and k are variables with a random value. The value of j is in the range *1,....,* M and the value of k is in the range *1,...,* N and should differ from *i.* ϕ_{ij} is a real number range from -1 to 1. When v_i is produced, the fitness or quality of the food source is acquired by

$$fitness_i = \begin{cases} \frac{1}{1+f_i}, & f_i \geq 0 \\ 1 + abs(f_i), & f_i < 0 \end{cases} \qquad (3)$$

Where f_i is a cost function. f_i function can be considered as a fitness value for maximization problems. The information regarding the goodness of the food source will be shared with the onlooker bees after the search has been conducted by employed bees. The probability of choosing a food source to be explored by the onlooker bee is based on its fitness value as follow:

$$p_i = \frac{fitness_i}{\sum_{n=1}^{F} fitness_i} \qquad (4)$$

Based on probability value, the onlooker bees determine which food sources to be explored.

2.3 Onlooker bee phase

The onlooker bees will explore the food sources with better probability which means that they will become employed bees. The neighborhood or vicinity of the chosen food sources is explored as shown in the phase of the Employed bee.

2.4 Scout bee phase

The algorithm verifies if there is an exhausted source to be deserted. To do that, the *LIMIT* variable is used which will be updated during the search. If its value reaches *MAX END*, then the food source is considered exhausted and it will be deserted. A newly discovered food source by scout bee will replace any food source deserted by its bee. A newly randomly created food source will be associated with the scout bee.

2.5 Related Works

In [14], a hybrid method that combines the ABC technique with a differential evolution algorithm was proposed as a feature selection. In [23], the ABC algorithm was utilized for feature selection. In [16], a hybrid method that combined Ant Colony and ABC algorithms were proposed as a feature selection method. The feature selection method for removing duplicated information using the ABC algorithm was proposed in [24]. In [25], the FS method using a random walk and ABC was proposed. In [26], an improved feature selection method using the chi-square filter method for Arabic Text Classification (ATC) was proposed. In [27], an efficient feature selection method using a firefly algorithm for ATC was proposed. Gharib et al. proposed an improved filter feature selection method for ATC [28]. Aisha et al. proposed a feature selection method using statistics of the compound word for ATC which has superior results compared with the set of filter FS methods [29].

3 Proposed Method: Chi-Square with ABC Algorithm as Feature Selection

In optimization problems, vectors of real values represent the expected problem solutions while in the feature selection problem, bit vectors represent the possible solutions. A bit vector of size N represents each food source, where N represents the number of features. In the vector, each position corresponds with or represents a feature to be assessed. If the value at the corresponding position is 1, it means that the feature will be included in the subset to be evaluated. Otherwise, the feature will not be included in the subset to be evaluated.

Furthermore, there is the quality (fitness) associated with every food source which is estimated by the accuracy of the classifier based on the subset

of features represented by the bit vector. Steps of the proposed FS method are represented in Figure 2 which are described as the following:

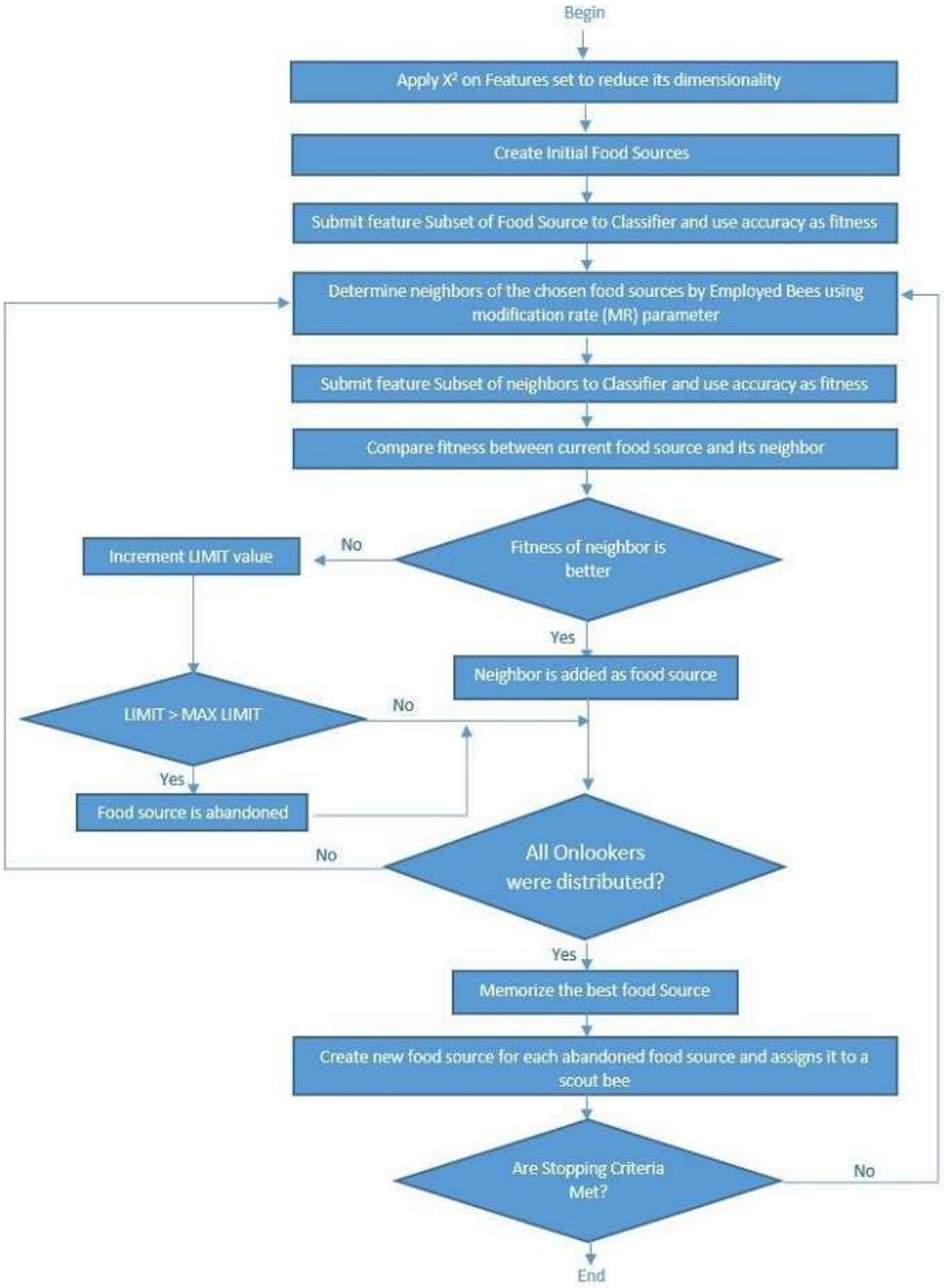

Figure 2: The main steps of the proposed FS method

1. The chi-square method is applied first to reduce the number of features. The method determines the level of interdependency between t term or word (feature) and a predefined document class c.

$$x^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{5}$$

The reason behind using chi-square as a first level of feature selection is that it is computationally fast, simple and can deal with a large dimensional feature. Meanwhile, the ABC algorithm is a wrapper that is relatively slow to deal with the massive number of features. So we tried to utilize the chi-square method first to reduce the number of features after that ABC algorithm is used to find out the best subset of features with higher accuracy as shown below.

2. Applying the ABC algorithm as the second level of the feature selection process to come up with a subset of features (words) that has a higher classification accuracy.

   (a) Initial food sources will be created: for feature selection, it is preferable to find out the best accuracy based on the minimum number of features. Therefore, the forward search strategy is used in the proposed method. Each food source represents a bit vector with either a value 1 for a feature to be considered in the subset to be evaluated or a value 0 for the feature that will not be considered. In the original ABC algorithm, for each position in each food source, a random number R_i is generated in the range between 0 and 1. If the value of R_i is less than MR value, the feature is considered as a part of the subset (food source), which means that the value of the position for that feature will be 1. Otherwise, the feature will not be considered and the value of the position for that feature will be 0. In the proposed method, one more control variable *nFeatures* which is a random number in range 7 to 50 is added to control the number of features that will be considered in each food source. Equation 6 represents how the position will be initialized:

$$x_i = \begin{cases} 1, & R_i < MR \\ x_i, & R_i \geq MR \end{cases}, \; and \; 7 \leq nFeatures \leq 50 \tag{6}$$

(b) The subset of features for every food source is evaluated by the classifier, and classification accuracy will be used as the fitness value of food source.

(c) Based on the value of the (MR) parameter, the neighbors of food sources will be determined by employed bees using Equation 6. The range 7-50 features to be modified in every step is used to make sure that we don't have a big jump in the number of selected features in each step.

(d) The subset of features for neighbors is evaluated by the classifier, and the accuracy of the classifier will be used as the fitness of neighbors.

(e) If the newly created food source has better quality than the food source under consideration, then it will be considered as a new food source and its quality information will have partaken with other bees. Otherwise, the variable $LIMIT$ for the food source is incremented. If its value reached the $MAX\ END$, then the food source is deserted which means exhausted food source. For each exhausted source, a scout bee is created randomly to search for a new food source.

(f) Quality Information about food sources visited by employed bees is collected by the onlookers. Based on that information, a food source with a better probability of exploration or better fitness will be chosen by onlooker bees which become employed bees and execute step C

(g) The best food source will be memorized after the distribution of all onlooker bees.

(h) For each deserted food source, newly food source will be created and assigned to scout bee which will be employed and perform step C.

4 Experimental results

4.1 Arabic Dataset

Several datasets are used in Arabic text classification. One of the most widely used datasets is the BBC Arabic dataset. It is free, public, and contains an appropriate number of documents. It comprises of seven classes with 4222

documents. Table 1 shows the number of documents in each category of the BBC Arabic dataset.

Table 1: Number of documents in each category of the BBC Arabic dataset.

Category	Number of documents
Middle East News	2145
World News	1348
Business and Economics	280
Sport	147
Magazine	41
Science and Technology	197
Collection (Art and Culture)	64
Total	4222

4.2 Text Preprocessing

In the preprocessing phase, non-meaningful words will be eliminated such as stopwords, punctuation marks, and numbers so all dataset documents have been preprocessed as follow:

1. Conversion to UTF-8 encoding.
2. Remove hyphens, punctuation marks, numbers, digits, non-Arabic letters, and diacritics.
3. Remove stopwords.

4.3 Data Division

In this research, the k-folds cross-validation method is used which means that the dataset is split into k different subsets or folds such that *k-1* subsets will be used to train the proposed model and the last subset will be used for validation. In this research k=10.

4.4 Data representation

Boolean representation is used as data representation for feature vector (bag of Words) which means that if the value is 1, the word will be considered as a part of the feature set. Otherwise, the word will not be considered as a part of the feature set. Finally, no stemming algorithm is applied.

4.5 Feature Selection

Feature selection aims to improve classification accuracy and computational efficiency of the classification technique by removing irrelevant and redundant features and selecting feature which contains sufficient information. Base on that, the main contribution in this research is to speed up the performance of the ABC algorithm by using the chi-square method as a first-level feature selection method which will reduce the search space for ABC algorithm, and improving the classification accuracy by selecting the best subset of features that contains sufficient information needed to build the classification model.

4.6 Parameters of ABC Algorithm

In the ABC algorithm, the following values are used:

1. The forward search strategy is used in the feature selection phase. Due to, the massive number of features that we have, it is impractical and computation cost to have n food sources and each of them initialized with one feature so the swam size or the number of food sources is set to be 200 swarms.

2. $MAXLIMIT$=10

3. MR=0.1

4. Number of iterations = 200

In the proposed method, the Naive Bayesian classifier which is one of the most efficient classifiers from the computation perspective is used as a fitness function [6].

All experiments have been implemented using java programming language through the Netbeans 8.2 software to handle the feature selection process using the ABC algorithm with the support of Weka machine-learning software for the preprocessing, and classification phases. Naive Bayes classifier, J48 which is weka implementation of the C4.5 decision tree, and Support Vector Machine SVM classifiers are used to evaluate the selected subset of features. 10-fold cross-validation is used in this research where the dataset is randomly partitioned into 10 equal size subsets. a single subset is used as the validation data for testing the model while the remaining 9 subsets are used as training data. Weighted F1-measure is used as a classification evaluation metric. F1-measure is the harmonic mean of precision (p) and recall (r) which provides

a more realistic measure of a tests performance. Weighted-F1 weights the F1-score of each class C_i by the number of documents N_i from that class.

$$F1 - Measure = 2 \; . \; \frac{p \; . \; r}{p + r} \quad (7)$$

$$Weighted \; F1 - Measure = \frac{\sum_{i=0}^{n} F1 - measure \, (C_i) * N_i}{D} \quad (8)$$

such that D is the total number of documents.

4.7 Discussion

Table 2 shows that the proposed method find out the minimum number of features that will be used as feature set compared with the original dataset and other methods. Table 3 shows the weighted F1-measure using selected features. Notably, the proposed model has superior results compared with the others. This is because chi-square is used first as a filter method to reduce the dimensionality of feature space. After that, the ABC algorithm tries to find out the subset of features that has a lower number of features and higher the weighted F1-measure value, by modifying food sources in the range from seven to fifty features in each step and find out the weighted F1-measure after the modification until we have a subset with a possible minimum number of features and higher weighted F1-measure. Furthermore, the combination of two methods reduces the computation time for feature selection and increases the weighted F1-measure value compared with the ABC algorithm.

Table 2: Number of Features using the proposed method and other methods.

Method	Number of Features
Original Dataset	82103
CHI	2444
ABC	2559
CHI-ABC	1812

Table 3: Weighted F1-measure for the proposed method and other methods.

Classifier	CHI FS	ABC FS	Proposed FS
NB	76.45	77.41	77.72
SVM	77.60	76.10	77.60
J48	72.63	72.44	73.05

5 Conclusion

A feature selection method utilized a combination of Chi-square with ABC algorithm is proposed by this research. The results show that a higher weighted F1-measure value can be achieved using a reduced number of features. The proposed method has superior results compared with other methods. As future work, the Arabic text classification based on semantic relations will be studied. Arabic WordNet will the source of knowledge that will be used.

References

[1] M. S. Khorsheed, A. O. Al-Thubaity, Comparative evaluation of text classification techniques using a large diverse Arabic dataset, Lang. Resour. Eval., **47,** no. 2, (2013), 513–538.

[2] S. A. Yousif, V. W. Samawi, I. Elkabani, Arabic text classification: The effect of the awn relations weighting scheme, in Proceedings of the World Congress on Engineering, **2,** (2017).

[3] R. Elhassan, M. Ahmed, Arabic Text Classification review, Int. J. Comput. Sci. Softw. Eng., **12** , no. 1, (2015).

[4] S. A. Yousif, V. W. Samawi, I. Elkabani, R. Zantout, The effect of combining different semantic relations on Arabic text classification, World Comput. Sci. Inform. Tech. J., **5**, no. 1, (2015), 112–118.

[5] H. K. H. Chantar, New techniques for Arabic document classification, Ph. D Thesis, Heriot-Watt Univ., 2013.

[6] M. Hijazi, A. Zeki, A. Ismail, Arabic Text Classification: Review Study, J. Eng. Appl. Sci., **11**, no. 3, (2016), 528–536.

[7] M. Schiezaro, H. Pedrini, Data feature selection based on Artificial Bee Colony algorithm, EURASIP J. Image Video Process., **2013,** no. 1, 47.

[8] H. K. Chantar, D. W. Corne, Feature subset selection for Arabic document categorization using BPSO-KNN, in 2011 Third World Congress on Nature and Biologically Inspired Computing, (2011),546–551.

[9] T. Prasartvit, B. Kaewkamnerdpong, T. Achalakul, Dimensional reduction based on artificial bee colony for classification problems, in International Conference on Intelligent Computing, (2011), 168–175.

[10] M. J. Meena, K. R. Chandran, A. Karthik, A. V. Samuel, n enhanced ACO algorithm to select features for text categorization and its parallelization, Expert Syst. Appl., **39**, no. 5, (2012), 5861–5871.

[11] D. Jia, X. Duan, M. K. Khan, Binary Artificial Bee Colony optimization using bitwise operation, Comput. Ind. Eng., **76**, (2014), 360–365.

[12] D. Karaboga, B. Gorkemli, C. Ozturk, N. Karaboga, A comprehensive survey: artificial bee colony (ABC) algorithm and applications, Artif. Intell. Rev., **42**, no. 1, (2014), 21–57.

[13] M. H. Aghdam, S. Heidari, Feature selection using particle swarm optimization in text categorization, J. Artif. Intell. Soft Comput. Res., **5**, no. 4, (2015), 231–238.

[14] E. Zorarpac?, S. A. zel, A hybrid approach of differential evolution and artificial bee colony for feature selection, Expert Syst. Appl., **62**, (2016), 91–103.

[15] A. S. Ghareb, A. A. Bakar, A. R. Hamdan, Hybrid feature selection based on enhanced genetic algorithm for text categorization, Expert Syst. Appl., **49**, (2016), 31–47.

[16] P. Shunmugapriya, S. Kanmani, A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid), Swarm Evol. Comput., **36**, (2017), 27–36.

[17] H. Naji, W. Ashour, M. Al Hanjouri, Text Classification for Arabic Words Using BPSO/REP-Tree, Int. J. Comput. Linguist. Res., **9**, no. 1, (2018).

[18] M. A. Basir, Y. Yusof, M. Saifullah, Optimization of Attribute Selection Model Using Bio-Inspired Algorithms, J. ICT, **18**, no. 1, (2019), 35–55.

[19] D. Karaboga, An idea based on honey bee swarm for numerical optimization, Technical report-tr06, Erciyes university, engineering faculty, computer ?, 2005.

[20] D. Karaboga, B. Akay, A survey: algorithms simulating bee swarm intelligence, Artif. Intell. Rev., **31**, nos. 1-4, (2009), 61–85.

[21] L. Bao, J. Zeng, Comparison and analysis of the selection mechanism in the artificial bee colony algorithm, in 2009 ninth international conference on hybrid intelligent systems, **1**, (2009), 411–416.

[22] B. Akay, D. Karaboga, A modified artificial bee colony algorithm for real-parameter optimization, Inf. Sci., **192**, (2012), 120–142.

[23] M. S. Uzer, N. Yilmaz, O. Inan, Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification, Sci. World J., (2013).

[24] Y. Wang, L. Feng, J. Zhu, Novel artificial bee colony based feature selection method for filtering redundant information, Appl. Intell., **48**, no. 4, (2018), 868–885.

[25] L. Feng, Y. Wang, W. Zuo, Novel feature selection method based on random walk and artificial bee colony, J. Intell. Fuzzy Syst., **32**, no. 1, (2017), 115–126.

[26] S. Bahassine, A. Madani, M. Al-Sarem, M. Kissi, Feature selection using an improved Chi-square for Arabic text classification, J. King Saud Univ. Inf. Sci., 2018.

[27] S. L. Marie-Sainte, N. Alalyani, Firefly algorithm based feature selection for Arabic text classification, J. King Saud Univ. Inf. Sci., 2018.

[28] A. S. Ghareb, A. A. Bakara, Q. A. Al-Radaideh, A. R. Hamdan, Enhanced filter feature selection methods for Arabic text categorization, Int. J. Inf. Retr. Res., **8**, no. 2, (2018), 1–24.

[29] A. Adel, N. Omar, M. Albared, A. Al-Shabi, Feature selection method based on statistics of compound words for Arabic text classification, Int. Arab J. Inf. Technol., **16**, no. 2, (2019), 178–185.

[30] S. A. Yousif, V. W. Samawi, I. Elkabani, R. Zantout, ?Enhancement of Arabic Text Classification Using Semantic Relations with Part of Speech Tagger, W. Trans. Adv. Electr. Comput. Eng., (2015), 195–201.

[31] University of Waikato Machine Learning Group, Weka Software. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/index.html.