Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 30(1), 25-31 ss., Haziran 2015
Çukurova University Journal of the Faculty of Engineering and Architecture, 30(1), pp. 25-31 , June 2015

An Artificial Bee Colony Based Algorithm for Feature Selection

Ezgi ZORARPACI[*1], Selma Ayşe ÖZEL[1], Süleyman GÜNGÖR[2]

[1]*Ç.Ü., Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Adana*
[2]*Ç.Ü., Fen-Edebiyat Fakültesi, Fizik Bölümü, Adana*

Geliş tarihi: 09.01.2015 *Kabul tarihi: 24.06.2015*

Abstract

The aim of the feature selection is to reduce the number of features to be used during the classification process to improve run-time performance and efficiency of the classifier. In this study, Artificial Bee Colony (ABC) Optimization Technique, which is a recent successful swarm intelligence algorithm, based feature selection method is proposed for classification tasks. The algorithm was experimented on fifteen datasets from the UCI Repository which are commonly used in classification problems. The experimental results of this study showed that the proposed ABC based algorithm is able to select good features for classification tasks.

Keywords: Feature selection, Artificial bee colony, Classification.

Nitelik Seçimi için Yapay Arı Kolonisi Tabanlı Bir Algoritma

Özet

Nitelik seçiminin amacı çalışma zamanını ve sınıflandırıcının verimliliğini iyileştirmek için sınıflandırma işlemi sırasında kullanılacak olan özellik sayısını azaltmaktır. Bu çalışmada, sınıflandırma işlemleri için yakın zamanda geliştirilmiş başarılı bir sürü zekası algoritması olan Yapay Arı Kolonisi (YAK) Optimizasyon Tekniğine dayalı bir nitelik seçim yöntemi geliştirilmiştir. Geliştirilen yöntem UCI veritabanından elde edilen ve sınıflandırma problemlerinde sıklıkla kullanılan 15 veri kümesi üzerinde sınanmıştır. Deney sonuçları önerilen YAK tabanlı algoritmanın sınıflandıma çalışmaları için iyi nitelikleri seçebildiğini göstermiştir.

Anahtar Kelimeler: Nitelik seçimi, Yapay arı kolonisi, Sınıflandırma.

[*] Yazışmaların yapılacağı yazar: Ezgi ZORARPACI, *Ç.Ü., Mühendislik-Mimarlık Fakültesi, Bilgisayar Bölümü, İstanbul. ezorarpaci @cu.edu.tr*

1. INTRODUCTION

Feature selection, also known as attribute selection or dimension reduction, is the method of selecting an optimum subset of relevant features which represents original feature set with the least error for the classification model. Thanks to feature selection techniques, we have some benefits such as improved model interpretability, shorter training times, enhanced generalization by reducing overfitting during the learning process [1]. Theoretically, a feature selection method must search through the subsets of features, and find the best one among the all candidate subsets according to a certain evaluation criterion. However, this procedure cannot be possible in a reasonable amount of time as it tries to find the best subset. Also it is too costly and restrictive in general. So, instead of the best subset, a (sub)optimum feature subset not reducing or least reducing classification accuracy may be accepted. Heuristic and random search methods can be used to find these optimum subsets. Various metaheuristic search methods have been used including Tabu Search (TS), Simulated Annealing (SA), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Differential Evolution (DE) and Artificial Bee Colony (ABC) to search solution space for feature selection problem [2]. Artificial Bee Colony Algorithm (ABC) [3], defined by Derviş Karaboğa, is an optimization algorithm that mimics the intelligent foraging behaviour of honey bee swarm. ABC algorithm has good properties such as easy to implement, strong robustness, high flexibility and fewer control parameters [4]. In this study, a modified ABC algorithm was proposed to find the best features to improve the classification time and accuracy of the classifiers. In our implementation, feature subset solutions were represented in binary form. Thus, a new producing of neighborhood of food source operator of ABC was applied to feature subsets. This paper is organized as follows: Initially, previous works related to feature selection based on ABC algorithm are described. The proposed ABC method for feature selection is then presented in detail. At the end of the paper, experimental results obtained through the proposed method on datasets are given and discussed.

2. RELATED WORKS

Referring to literature, search methods based on ranking of features such as Information Gain, ChiSquare etc. have been used for selecting attributes. A joint disadvantage of these filter methods is that they neglect the mutual effect with the classifier model, and this may lead to decrease classification accuracy when compared to other types of feature selection methods. While *filter methods* are able to find a good feature subset without depending on a classifier model, *wrapper methods* accommodate a classifier model into their search strategy for good feature subset. In this manner, it is possible that a search procedure evaluates a specific subset of features by using a classification model and, then it can obtain better classification performance. However, for n features we have 2^n feature subsets so, the computational time for the search strategy which finds the best feature subset among all feature subsets is exponential in wrapper methods. Therefore meta-heuristic search methods, which will discover a (sub)optimum feature subset, can be used with a classifier model to construct a wrapper method [5]. In literature several meta-heuristic search methods including ABC algorithm for feature selection problem have been proposed. Palanisamy and Kanmani (2012) have performed a wrapper based feature selection approach for classification problem. According to this approach, ABC is used as a feature selector and generates the feature subsets, and a classifier (J48) is used to evaluate each feature subset generated by ABC algorithm. This study has been implemented and tested using 10 datasets taken from UCI Repository. Hence, they have showed that algorithm has resulted in reduced feature size of the feature subset, increased classification accuracies, low computational complexity [6].

Prasartvit et al. (2013) have proposed a novel method of ABC for data dimension reduction in classification problems. The proposed method applies ABC wrapping with a k-Nearest Neighbor (kNN) classifier kNN has been used for evaluation criteria to compute the fitness value of

the new feature subsets generated by ABC. In this method, employed bees and the onlooker bees have generated new candidate food sources, which are the subsets of selected features, and kNN has been performed to evaluate the classification accuracy (i.e., objective function) of the new candidate food sources. The proposed method has been validated in two distinct application domains: Gene Expression Analysis, and Autistic Behaviors. For Autistic Behaviors classification dataset, they have obtained 85% accuracy value with 25% of the features selected from the original dataset. For Gene Expression Analysis; the rates of genes (features) selected for Colon_Cancer, Acute_Leukemia, Hepatocellular_Carcinoma, High-grade_Glioma, and Prostate_Cancer datasets have been reduced to 3.15%, 3.39%, 4.38%, 3.61%, and 3.59% respectively, with accuracy values changing between 89.5% and 100% [7].

Schiezaro and Pedrini (2013) have implemented a feature selection method using ABC for classification of different datasets. Heart-c, Hepatitis, Lung Cancer, Image Segmentation, Iris, Winsconsin, Labor, and Diabetes datasets from the UCI repository have been used to demonstrate the effectiveness of their method. With the proposed method, they have obtained accuracy values changing from 71.48% to 98.46% by using SVM classifier [8].

Uzer et al. (2013) have offered a wrapper approach that uses ABC for feature selection and SVM for classification. The purpose of this study was to examine the effect of elimination of unimportant and obsolete features from the datasets on the success of the classification performance. To test their approach, Hepatitis, Liver-Disorders and Diabetes datasets from the UCI Repository have been used, and the proposed system has reached classification accuracies of 94.92%, 74.81%, and 79.29% respectively [9].

In our study, a different neighborhood producing operator was employed for ABC algorithm to produce new neighbours (i.e., feature subsets). To generate a new neighbour (i.e., new feature subset) for a current food source (i.e., current feature subset), firstly we find the difference component (i.e., feature) between two randomly chosen food sources (i.e., feature subsets). So, the difference component is a random different feature which is found in only one of the randomly chosen two feature subsets (i.e., food sources). Then, we combine this random feature with current feature subset, if this feature does not exist in the current feature subset and the following combination condition holds. To compute the combination condition, we generate two random numbers between 0 and 1. If the first generated random number is greater than the second random number, the condition holds. By combining the difference component with the current feature subset we obtain new feature subset (i.e., new neighbour food source). The random values are used to simulate the coefficient of neighborhood producing operator in standart ABC algorithm that is the random value which is between -1 and 1. Through these random values, we preserve the low convergence speed of the ABC algorithm.

3. METHOD

3.1. Artificial Bee Colony

Artificial Bee Colony (ABC) [3] is a swarm based meta-heuristic algorithm that was introduced by Karaboğa in 2005 for optimizing numerical problems. It simulates the intelligent foraging behavior of bees. Foraging model of honey bees includes some important constituent such as food sources, employed foragers, and unemployed foragers [3]. The stages of ABC algorithm can be described as follows:

Step 1. *Initialization:* First, ABC generates a randomly distributed *SN* food source positions, where *SN* denotes the size of employed bees or food sources. Each food source, defined as X_i, where $i=1,2,..,SN$ is a vector with dimension D which represents the number of parameters for the optimization problem. Generally, the beginning food source positions are randomly produced by using equation 1.

$$X_i^j = X_{min}^j + rand(0,1).\left(X_{max}^j - X_{min}^j\right) \quad (1)$$

where, $j=1,2,\ldots,D$; X_{max}^j and X_{min}^j are the upper and lower values of j th paremeter of the problem; $rand(0,1)$ is a random value between 0 and 1.

Step 2. *Nectar amount (fitness value) evaluations of the food sources:* In this step, the nectar amount is calculated for each food source.

Step 3. *Employed bee process:* After initialization, each employed bee goes to a food source and searches for a new food source having more nectar amount (i.e., quality) of own food source within its neighborhood. For an employed bee X_i, neighboring food source position is V_i, and is produced by equation 2.

$$V_i^{jrand} = X_i^{jrand} + rand.\left(X_i^{jrand} - X_k^{jrand}\right) \quad (2)$$

where X_k is a randomly selected food source, $k \in \{1,2,\ldots,SN\}$ is randomly determined and has to be different from *i*, $jrand \in \{1,2,\ldots,D\}$ is a random integer number, $rand$ is a random value between -1 and 1, *D* is the number of parameters of the problem at hand.

$$p_i = \frac{fitness_i}{\sum_{n=1}^{SN} fitness_n} \quad (3)$$

where $fitness_i$ is the quality (i.e., fitness value) of the food source (i.e., solution) *i* evaluated by its employed bee. After calculating the probability, onlooker bee X_i finds the neighboring food source according to the equation 2 and evaluates the nectar amount of new candidate food source. If the nectar is higher than that of the previous, the bee memorizes the new source position (i.e., solution) and forgets the old.

Step 6. *Memorizing the best food source:* In this step, the best food source which has the highest nectar amount (i.e., fitness) is stored.

Step 7. *Scout bee process*: In scout bee process, a new food source is determined by a scout bee and replaced with the abandoned one. For this process, a counter is used for each bee in the swarm. If there is a bee that her counter value exceeds maximum limit, she abandones the food source (i.e., solution)

and searches for new food source. To search a new food source for a scout bee, equation 1 is used.

3.2. The Proposed Artificial Bee Colony for Feature Selection

The main steps of the proposed artificial bee colony algorithm are described as the following:

Step 1. *Specifying initial food sources*: The initial food sources which are binary-coded solution vectors are created using binary random values 0 and 1. In the original ABC, the initial food sources are real-valued vectors.

Step 2. *Nectar amount (i.e., fitness value) evaluations of the food sources:* In this step, the nectar amount (i.e., fitness value) is calculated for each food source. The fitness value of a food source is computed by using Weka J48 classifier with 3-folds crossvalidation; and weighted F-measure value returned from Weka is taken as the fitness value for the food source.

Step 3. *Employed bee process:* Each employed bee goes to a food source and searches for a new food source having more nectar amount (i.e., quality) of own food source within its neighborhood. For a food source X_i, two random food sources X_{r1} and X_{r2} are chosen in the swarm. These food sources are different from each other and the main food source ($r1 \neq r2 \neq i$). After this selection process, a random component (*jrand*) for the food source is specified and only for this component, (*jrand*), the difference of one of the chosen food sources $X_{r1}$to the other X_{r2} is found, which is called difference component. Equation 4 expresses how to find the difference component. After finding the difference component, a random value r1(0,1) between 0 and 1 which simulates the coefficient for neighborhood solution production of the ABC, is generated to decide whether "OR" logic operator will be applied to the component of

food source using difference component. Equation 5 explains how to construct the neighbor of the food source. The fitness value of the neighbor of the food source is calculated using Weka J48 classifier with 3-folds crossvalidation; and weighted F-measure value returned from Weka is assigned as the fitness value of it. If the fitness value of neighbor is greater than that of current food source X_i, the neighbor food source replaces the current food source. Step 3 is repeated for each employed bee in the swarm.

Step 4. *Calculating fitness value probabilities*: Fitness value probability is calculated for each food source using equation 3 as described in section 3.1.

Step 5. *Onlooker bee process*: In this step, a random value between 0 and 1 is generated and compared to the probability value of a food source p_i for an onlooker bee. If p_i is greater than this random value, this food source is chosen by an onlooker bee and new neighboring food source is searched with equation 4 and 5, respectively. If the fitness value (i.e., nectar) of this new food source is better than that of current food source X_i, the bee memorizes the new food source and forgets the current food source.

Step 6. *Memorizing the best food source*: In this step, the best food source which has the highest fitness value is memorized.

Step 7. *Scout bee process*: If the counter value of a food source is the maximum among those of food sources and exceeds limit, a new food source (i.e., feature subset solution) is created by a scout bee using binary random values 0 and 1.

The steps from 3 to 7 are repeated until a predetermined termination criterion is satisfied. The best feature subset solution found so far is taken as the optimum solution for our problem.

$$ifference\ component = \begin{cases} 0 & , \ if\left({X_{r1}}^{jrand} = {X_{r2}}^{jrand}\right) \\ {X_{r1}}^{jrand} & , \ otherwise \end{cases} \quad (4)$$

$$neighbor = X_i$$

$$neighbor^{jrand} = \begin{cases} 1 & , \ if\left(difference\ component = 1\ and\ r1(0,1) > r2(0,1)\right) \\ {X_i}^{jrand} & , \ otherwise \end{cases} \quad (5)$$

4. EXPERIMENTAL RESULTS

All implementations were performed with Java programming language under NetbeansIDE 7.2 platform. Our computer had Windows 7 Home Premium operating system, 4 GB of RAM, and Intel Core i5-2430 M 2.4 GHz processor. The proposed method was tested on fifteen UCI machine learning repository datasets (http://archive.ics.uci.edu/ml/). Randomly chosen 75% of the instances in each dataset were specified as training instances, and the rest of the instances were used in the testing phase. The distributions of classes and instances in the training and test data for each dataset are shown in Table 1. For the proposed method, number of iterations was set to 300. The number of bees in the swarm was set to 50.

In 2010, Saraç and Özel compared the classification performances of some classifiers namely J48, NaiveBayes, RBF Networks, Voted Perceptron, Threshold Selector, Voting Feature Intervals (VFI) for URL based Web page classification problem [10]. According to this study, J48 classifier had the highest classification F-measure. Therefore J48 classifier was chosen to evaluate the feature subsets of the proposed algorithms. The best and average weighted F-measure values with the number of selected features through the proposed method at the end of 10 runs in the test phase, and also F-measure values with all attributes (i.e., without making any feature selection) are shown in Table 2.

According to Table 2, the proposed ABC algorithm shows quite good classification performance on all datasets in terms of the best F-measure values. When we analyze the classification results of the proposed algorithm in terms of average F-measure values, the proposed algorithm is successful on eight datasets among fifteen datasets.

To evaluate the classification performance of the algorithm completely, we should look at the numbers of selected features which are given in Table 3. As shown in Table 3, the proposed feature selection algorithm reduced the number of features more than 50% without decreasing the classification accuracy in majority of the datasets. When we evaluate the classification performances with the numbers of selected features by the proposed method, the classification accuracies decrease with the proportion of 4.26%, 1.7%, 2.64%, 3.11%, 3.65%, 7.1% and 1.91% respectively in terms of average F-measure values while the numbers of average selected features reduce with the proportion of 60%, 58.8%, 73.6%, 50%, 45%, 75% and 52.9%, respectively for Autos, Dermatology, Hepatitis, Lymph, Credit-g, Sonar and Zoo datasets. Taking account of this, we can say that our proposed method is successful for these seven datasets since it reduces the numbers of selected features without decreasing the average F-measure values too much.

5. CONCLUSION

This study was designed to solve feature selection problem on classification tasks. ABC is a recent successful swarm intelligence algorithm. In this work, ABC was used to choose the optimum feature subsets by using the training datasets. Test datasets were employed with selected features to evaluate classification performance. The experimental results showed that the proposed ABC algorithm was able to select good features for classification tasks by not reducing or least reducing classification accuracies.

6. REFERENCES

1. He, X., Zhang, Q., Sun, N., Dong, Y. 2009. Feature Selection with Discrete Binary Differential Evolution. In Artificial Intelligence and Computational Intelligence, AICI'09 International Conference, 4, p. 327 330.

2. Frohlich, H., Chapelle, O., Scholkopf, B., 2003. Feature Selection for Support Vector Machines by means of Genetic Algorithm. In Tools with Artificial Intelligence, 15th IEEE International Conference, p. 142-148.

3. Karaboğa, D., 2005. An Idea based on Honey Bee Swarm for Numerical Optimization. Erciyes University, Engineering Faculty, Computer Engineering Department, Technical report.

4. Bolaji, A. L. A., Khader, A. T., Al-Betar, M. A., Awadallah, M. A. 2013. Artificial Bee Colony Algorithm, Its Variants and Applications: A Survey. Journal of Theoretical and Applied Information Technology, 47(2), p. 434-459.

Table 3. Number of selected features for the best/average f-measure values at the end of 10 runs

Dataset	Best/Average # of features	Total # of features
Autos	10/10	25
Breast-w	3/5	9
Car	5/5	7
Glass	5/6	9
Heart-c	3/7	13
Dermatology	13/14	34
Hepatitis	2/5	19
T.Surgery	1/7	16
Lymph	11/9	18
Credit-g	13/11	20
Sonar	24/15	60
Ionosphere	10/8	34
Liver-Disorders	5/5	6
Vote	2/4	16
Zoo	8/8	17

5. Saeys, Y., Inza, I., Larrañaga, P., 2007. A Review of Feature Selection Techniques in Bioinformatics. Bioinformatics, 23(19), p. 2507-2517.
6. Palanisamy, S., Kanmani, S., 2012. Artificial Bee Colony Approach for Optimizing Feature Selection. International Journal of Computer Science Issues, 9(3), p. 432-438.
7. Prasartvit, T., Banharnsakun, A., Kaewkamnerdpong, B. and Achalakul, T., 2013. Reducing Bioinformatics Data Dimension with ABC-kNN. Neurocomputing, 116, p. 367-381.
8. Schiezaro, M. and Pedrini, H., 2013. Data Feature Selection based on Artificial Bee Colony Algorithm. EURASIP Journal on Image and Video Processing, 1, p. 1-8.
9. Uzer, M. S., Yilmaz, N., Inan, O., 2013. Feature Selection Method based on Artificial Bee Colony Algorithm and Support Vector Machines for Medical Datasets Classification. The Scientific World Journal, p. 1-10.
10. Saraç, E., Özel, S. A., 2010. URL-Based Web Page Classification. ASYU Symposium, p. 13-17.