Feature Selection for Image Classification Based on Bacterial Colony Optimization

Hong Wang[1], Zhuo Zhou[1], Yixin Wang[1], and Xiaohui Yan[2](✉)

[1] College of Management, Shenzhen University, Shenzhen 518060, China
[2] School of Mechanical Engineering, Dongguan University of Techonlogy, Dongguan 523808, China

Abstract. Image classification is an important issue in pattern recognition, the high dimension features is a challenging task since only a few number of them are effective in classification. To improve the classification efficiency, it is necessary to reduce the dimensionality of image features before classification. This study provides a novel image classification application based on Bacterial Colony Optimization, which can decreases the computation burden and improves the classification's efficiency. Specifically, the elimination strategy in original algorithm is removed, and the communication, chemotaxis, migration, and reproduction strategies are kept. Additionally, the communication and chemotaxis step size of the Bacterial Colony Optimization are modified for feature selection in image classification. Several comparision experiments on two public image datasets are conducted to verify the effectiveness of the method. Experimental results prove that the method can greatly improve the classification accuracy and efficiency.

Keywords: Feature selection · Bacterial colony optimization · Image classification

1 Introduction

In recent years, with the deepening of image recognition research, the research objectives have become more complex, and the image feature dimension has become higher. The feature space of many high-dimensional data objects contains redundant features and noise features. These features may reduce the accuracy of classification or clustering on the one hand and increase the time and space complexity of learning and training on the other hand. Therefore, when classifying or clustering high-dimensional data, it is usually necessary to reduce the dimensionality of features to reduce machine learning time and space complexity.

Feature selection is an important feature dimensionality reduction method, and a large number of feature selection algorithms have been proposed. Feature selection uses certain evaluation criteria to select feature subsets from the original feature space. The purpose is to screen out invalid, incomplete and redundant features as much as possible. Besides, a good feature subset can use fewer features to describe most of the original data's attributes and better retain the information that the original data can convey.

Y. Tan and Y. Shi (Eds.): ICSI 2021, LNCS 12690, pp. 430–439, 2021.
https://doi.org/10.1007/978-3-030-78811-7_40

Finding the optimal feature combination is a difficult task. It is almost impossible to consider all the feature combinations to find the optimal feature combination.

Due to the advantages of heuristic algorithms, such as fewer parameters to be adjusted and irrelevant to the optimization target's gradient, more research focuses on using these heuristic algorithms to deal with feature selection problems. At present, many scholars have applied different evolutionary algorithms to the feature selection of image recognition problems. Zhang, X et al. applied the Particle Swarm Optimization [PSO] to the feature selection of hyperspectral image classification [1]. Naeini et al. used PSO for target-based very high spatial resolution satellite image feature selection [2]. Chen, L et al. used Ant Colony Optimization (ACO) or an improved ACO for image feature selection [3, 4]. Rutuparna Panda used the improved Bacterial Foraging Optimization (BFO) for the feature selection of face images [5], Lei L. et al. used Genetic Algorithms (GA) for the feature selection of insect images [6], Uroš Mlakar et al. used differential evolution algorithm for humans Feature selection of facial expression images [7].

By simulating the life cycle behavior of bacterial colonies, a novel heuristic algorithm is proposed in 2012, namely Bacterial Colony Optimization (BCO) [8]. Considering the application of the algorithm, this paper proposes an image classification method based on Bacterial Colony Optimization Feature Selection (BCOFS). In the experiment, two public databases were used to test the effectiveness of the classification method, and a variety of other feature selection methods were used for comparison experiments. Experimental results show that the classification method has shown good performance on both databases.

The rest of the paper is organized as follows: Sect. 2 introduces the image features used in the experiment of this paper. Section 3 provides a brief description of BCO and BCOFS. In Sect. 4, the experiment results and analyses of the experiment are shown and discussed. Finally, the conclusion of this paper is provided in Sect. 5.

2 Image Feature Extraction

The feature is the key to image classification, which directly determines the final classification result as a basis for classification. At present, the three essential features of images are shape feature, texture feature and color feature [9]. The above features describe the image target from different angles to extract effective information to identify the image. Note that different features are often selected for different classification tasks. When selecting features, It is necessary to select features based on specific issues.This paper uses four image features: Average curvature, Histogram of Oriented Gradient (HOG), Grayscale, and Gabor. The following is a detailed description of them.

Average curvature is an external bending measurement standard in differential geometry, which describes the curvature of a curved surface embedded in the surrounding space. The average curvature is the average value of any two orthogonal curvatures at a point on a space surface.

The essence of the HOG feature is the gradient's statistical information, and the gradient mainly exists at the edge. It constructs features by calculating and counting the histogram of the gradient direction of the local area. First, the image needs to be divided into small connected regions, called cells. Then collect the histogram of each direction's gradient or edge direction and each pixel in the cell. Finally, these histograms are combined to form a feature descriptor.

The texture is formed by the repeated occurrence of grayscale distribution in space. A certain grayscale relationship between two pixels is separated by a certain distance in the image space. This relationship is called the spatial correlation of grayscale. The gray-level co-occurrence matrix describes the texture by studying the spatial correlation of gray levels, describing the relationship between adjacent pixels in the image, and reflecting different aspects (uniformity, uniformity, etc.). It gives adjacent information about the intensity conversion between pixels. For a grayscale image I with a size of N $\times$ N, the grayscale co-occurrence matrix P can be defined as:

$$P(i,j) = \sum\nolimits_{x=1}^{\infty}\sum\nolimits_{y=1}^{\infty}\{1, \text{if} I(x,y) = i \text{ and } I(x+\Delta x, y+\Delta y) = j\} \tag{1}$$

Among them, $(\Delta x, \Delta y)$ represents the positional relationship between two pixels.

Haralick et al. proposed 14 types of statistics calculated based on the gray-level co-occurrence matrix [10]. Following is only the part used in this paper: Correlation, Uniformity, Contrast, Energy, Entropy.

The Gabor feature is generally obtained by convolving the image and the Gabor filter, which is defined as the Gaussian function's product and the sine function. The Gabor filter's frequency and direction representation are close to the frequency and direction representation of the human visual system, so it is often used for texture representation and description. Gabor filter is a complex exponential signal constructed by Gaussian function. A two-dimensional Gabor filter using Cartesian coordinates in the space domain and polar coordinates in the frequency domain is defined as:

$$g_{x_0,y_0,f_0,\theta_0} = \mathrm{e}^{i[2\pi f_0(x_0\cos\theta_0 + y_0\sin\theta_0)+\varphi]} \cdot gauss(x_0, y_0) \tag{2}$$

$$gauss(x_0, y_0) = K \cdot \mathrm{e}^{-\pi\left[a^2(x_0\cos\theta_0 + y_0\sin\theta_0)^2 + b^2(x_0\sin\theta_0 - y_0\cos\theta_0)^2\right]} \tag{3}$$

a and b describe the shape of the feature, θ_0 indicates the direction, φ indicates phase shift.x_0, y_0, f_0,and θ_0 indicate the position in the space domain and frequency domain respectively.

3 Bacterial Colony Optimization for Feature Selection

Feature selection is a complex combination problem with discrete high-dimensional parameters. Considering the adaptability of image feature selection, the Bacterial Colony Optimization Feature selection, abbreviated as BCOFS, has made some changes to its structure and strategies on the basis of BCO, the following is a detailed description of BCO and BCOFS.

3.1 Bacterial Colony Optimization (BCO)

The BCO [8] is proposed based on simulating the bacterial colony's life cycle behavior. The evolutionary process of bacterial growth model and the growth process of colonies will show the best solution to the problem.

The basic behavior of bacteria in the lifecycle can be simply divided into five parts: chemotaxis, communication, migration, reproduction and elimination. Under the action of these 5 parts, the entire colony system finally obtains the best source of nutrition. The following is a detailed introduction of these 5 parts. For more detailed information, please refer to [8].

3.1.1 Communication and Chemotaxis

Throughout the optimization process, chemotaxis is accompanied by communication. Bacteria will swim and tumble in the environment. However, they must also provide their information to the colony in order to exchange overall information, thereby guiding their direction and methods of action. In the bacterial individual's update process, when the objective function value of the bacterial individual's spatial position is better than the previous position's objective function value, the bacterial individual will adopt a forward movement mode. When the objective function value of the individual's spatial position is When the objective function value is not superior last time, it is simulated that the bacteria's current area environment is not as good as its parent's location. The individual bacteria will not move in this direction, but will tumble around in place, that is, search for nearby spatial locations.

The process of bacterial swim can be formulated as (4), The process of bacterial tumble can be formulated as (5), Their chemotaxis step size is shown in (6).

$$X_{k+1} = X_k + C_k[R_1 \cdot (G_{best} - X_k) + R_2 \cdot (P_{best} - X_k)] \tag{4}$$

$$X_{k+1} = X_k + C_k[R_1 \cdot (G_{best} - X_k) + R_2 \cdot (P_{best} - X_k) + Tu_k] \tag{5}$$

$$C_k = C_{min} + \left(\frac{iter_{max} - iter_k}{iter_{max}}\right)^n \cdot (C_{max} - C_{min}) \tag{6}$$

In the formula, X_{k+1} indicates where the bacteria will move to in the next step, X_k indicates the current location of the bacteria, G_{best} is the global best location. P_{best} is the best location for individual bacteria. R_1 and R_2 are randomly generated constant, the specified value range belongs to [0,1], $R_1 + R_2 = 1$. C_{k} is an adaptive chemotaxis step size, $iter_{max}$ indicates the highest number of iterations, $iter_k$ indicates the current number of iterations. Tu_k is tumble direction variance of the bacterium.

3.1.2 Migration

Individual bacteria will migrate according to certain conditions. The purpose is to find a better source of nutrients and reduce the possibility of falling into the best local source of nutrients. The migration strategy in this paper is expressed as random. If the conditions are met, the bacteria will migrate to a random area according to a function, it can be described as:

$$X_{k+1} = rand \cdot (B_u - B_l) + B_l \tag{7}$$

The value range of $rand$ is [0, 1], B_l and B_u are the lower and upper boundary.

3.1.3 Reproduction and Elimination

Similar to natural selection in reality, bacteria in locations with low nutrient content will be eliminated. In contrast, bacteria in places with high nutrient content will get more energy for survival and reproduction. In this algorithm, the bacteria with the higher fitness value will split into two identical bacteria, which is advantageous for the subsequent search.

3.2 Bacterial Colony Optimization Feature Selection (BCOFS)

The original BCO is a good method to solve continuous optimization problems, but in this paper, the research goal is to apply it to the feature selection of the image, which is a discrete combination problem with high dimensions. In order to adapt BCO to feature selection, BCOFS has made some changes on the basis of BCO. BCOFS removes the elimination strategy while retaining the communication, chemotaxis, migration and reproduction strategies. Some improvements are made in communication and chemotaxis strategies. It can be formulated as:

$$X_{k+1} = X_k + R_1 \cdot (G_{best} - X_k) + (1 - R_1) \cdot (P_{best} - X_k) \tag{8}$$

$$X_{k+1} = X_k - C_{\mathrm{k}} \cdot Tu_k \tag{9}$$

$$C_k = (\frac{iter_{max} - iter_k}{iter_{max}}) \cdot R \tag{10}$$

$$Tu_k = \Delta_k / \sqrt{\Delta_k^T \cdot \Delta_k} \tag{11}$$

where X_{k+1} and X_k are the location of the bacteria, G_{best} and P_{best} are the global best position and the individual best position. $R_1 \in [0, 1]$. C_{k} is tumble step size, $iter_{max}$ indicates the highest number of iterations, $iter_k$ indicates the current number of iterations. R is a constant coefficient depending on the total number of iterations. The randomness parameter Tu_k is calculated from Δ_k, $\Delta_k = (2 \cdot rand - 1) \cdot rand$.

The main idea of the algorithm is as follows: The bacteria at the beginning is placed in the solution space and some mechanisms are stipulated. Bacteria update their position through communication and chemotaxis, and the migration mechanism, and move to a place with high nutrient concentration. If the number of bacteria movement reaches a certain set and it still does not reach a higher concentration position, such bacteria will be died, and then the bacteria in the high nutrient concentration will start to reproduce to replace the died bacteria, so as to maintain the population. It uses the number of iterations or accuracy as the end condition like other swarm intelligences.

Figure 1 shows the flow chart of BCOFS for image feature. After preprocessing the image database (multi-feature extraction), it enters the process of feature selection and classification.The whole process can be simply divided into five steps:

Step 1: parameters of population are initialized, the classification accuracy is used as an evaluation function to calculate the fitness of each bacteria, the individual best fitness value and the global best fitness value is recorded.

Step 2: communication and chemotaxis operations are used to update the location of the bacteria, the fitness value of the bacteria is calculated again, the individual best fitness value and the global best fitness value are updated.
Step 3: if the migration conditions are fulfilled, the bacterial individuals will migrate to a random location.
Step 4: after the bacteria positions are updated, the reproduction operation is used to replicate individuals with good positions instead of individuals with poor positions.
Step 5: if the algorithm reaches the specified number of iterations, the best feature combination and best accuracy will be output. If not, repeat the above steps.

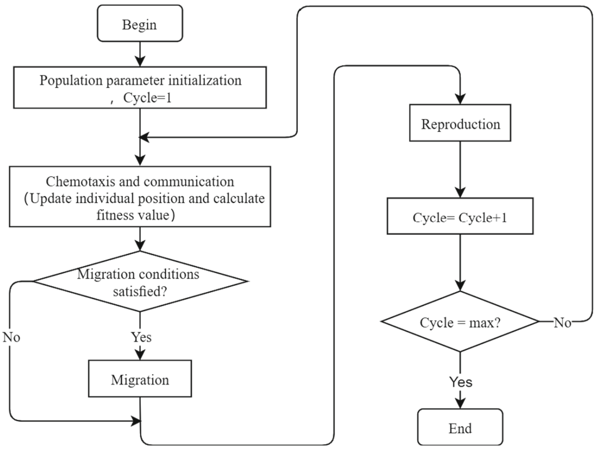


Fig. 1. Flow chart of BCOFS for image feature

4 Experiments

4.1 Experimental Setting

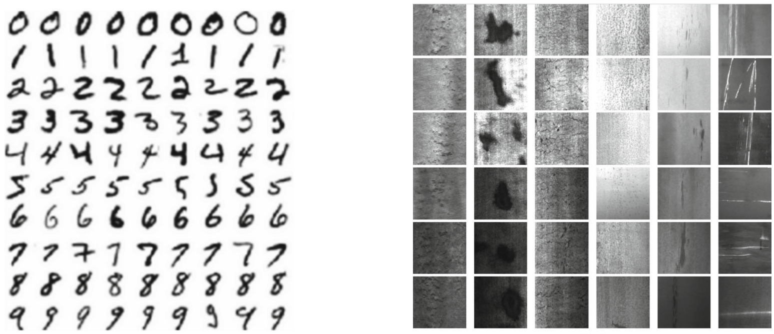
This paper conducts experiments on the MNIST handwritten digits dataset and the NEU surface defect dataset. In the comparison experiment, four other feature selection algorithms were used for comparison, namely the classic Bacterial Foraging Optimization (BFO) [11], the traditional feature selection algorithm Correlation-based Feature Selection (CFS) [12] and the widely used swarm intelligence algorithm Genetic Algorithm (GA) and Binary Particle Swarm Optimization Algorithm (BPSO) [13]. After many experiments, it is found that all algorithms will converge within 50 iterations, so the number of iterations is set to 50, and the number of each algorithm population is set according to the specific data set (The dimension of features extracted from the MNIST database is higher, so the population set to 50, and the dimension of features extracted from the NEU database is lower, so set to 26). KNN algorithm selects $K = 1$, SVM uses linear kernel function. In addition, all experiments were repeated 30 times.

4.2 Dataset Introduction and Multi-feature Extraction

The MNIST database of handwritten digits comes from the National Institute of Standards and Technology (NIST). There are 70,000 processed two-dimensional grayscale images, including 60,000 training set images and one Ten thousand test set images. Handwritten numbers do the image from 250 different people, 50% of them are high school students, and 50% are staff from the Census Bureau. Due to the excessive number of images in the original data set, this experiment uses 1000 pictures as the data set.

In the Northeastern University (NEU) surface defect database [14], six kinds of typical surface defects of the hot-rolled steel strip are collected, i.e., Rolled-in Scale (RS), Patches (Pa), Crazing (Cr), Pitted surface (PS), Inclusion (In) and Scratches (Sc). The database includes 1,800 grayscale images: 300 samples each of six different kinds of typical surface defects.

Figure 2 shows some examples in the two data sets. Figure 2(a) is the MNIST database of handwritten digits and Fig. 2.(b) is NEU surface defect database.



(a) (b)

Fig. 2. Examples of the two databases

For the MNIST database of handwritten digits, four types of features, namely grayscale, Gabor, average curvature, and HOG, are extracted and integrated together to obtain 247 features. For the NEU surface defect database, two types of features, grayscale and Gabor are extracted and integrated, and the total feature dimension is 69.

4.3 Experiment Results and Analyses

Table 1 shows the classification results of each algorithm. Shown in brackets is the average number of features selected by the algorithm. Only BFO and GA can compare the classification accuracy based on feature dimensions, so Fig. 3 and Fig. 4 only show the relationship between the classification accuracy of these methods and the number of selected features.

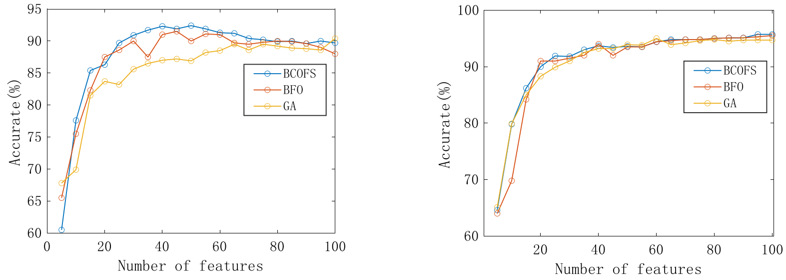
The data shows the accuracy of BCOFS is better than the accuracy of no feature selection process. It indicates that the algorithm removes redundant or useless features. Compared with other algorithms, BCOFS has the highest accuracy rate on the two experimental data sets. This proves the effectiveness of BCOFS in image feature selection.

In the actual application process of the classification method, the classification has two phases: training and use. The use phase of the classification method is longer than the training phase, the efficiency of the use phase accounts for a larger weight. The efficiency of the use stage is determined by the number of features selected by the algorithm, so the efficiency of the classification method should be analyzed according to the number of features selected by each method. In the training phase, it is the fastest not to use the feature selection algorithm. However, if the feature selection is not performed in the training phase, it will reduce the efficiency of the use phase, A large number of features will reduce classification efficiency. Therefore, in the long run, the efficiency of not using the feature selection algorithm is the worst. Other classification methods that use feature selection algorithms have correspondingly reduced the number of features. Although the time of the training phase is different, they all show a certain feature selection ability. From the average number of selected features, the efficiency of the classification method based on CFS is the highest. BCOFS does not have a big advantage, only superior to BFO and PSO, and GA shows a big difference between the two databases.

From the perspective of comprehensive classification accuracy and classification efficiency, BCOFS has the highest classification accuracy and the second highest classification efficiency.

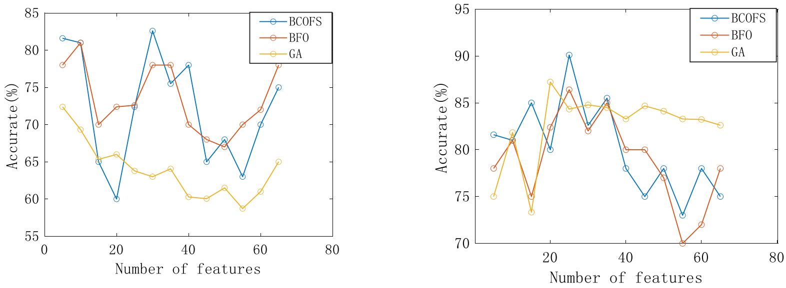
Table 1. Comparison of accuracy of each algorithm

Algorithm		Database			
		MNIST database		NEU database	
		Mean(%)	Best(%)	Mean(%)	Best(%)
KNN	ALL	89.21	89.50	59.6	60.2
	BCOFS	92.17(51.67)	92.50	82.6(30)	83
	BFO	91.5(55)	92	80.1(35)	81.5
	CFS	89.60(25)	90.60	76.11(7)	76.6
	GA	90.4(100)	91.2	72.39(5)	72.5
	BPSO	91.6(96)	92	71.06(33)	72.25
SVM	ALL	93.53	93.80	80.8	81.2
	BCOFS	95.72(88.33)	96.2	90.1(25)	90.5
	BFO	95.5(90)	96	86.4(40)	87.0
	CFS	90.70(25)	91.6	86.33(10)	86.5
	GA	95(95)	95.4	87.22(20)	88
	BPSO	94.7(102)	95.2	86.56(35)	87.1



(a) Use KNN classifier (b) Use SVM classifier

Fig. 3. Classification results of some methods in the MNIST database



(a) Use KNN classifier (b) Use SVM classifier

Fig. 4. Classification results of some methods in the NEU database

5 Conclusions

For the feature selection and classification of images, this paper proposes a classification method based on BCOFS. This method aims to improve the classification accuracy as much as possible while selecting as few features as possible. Experiments on the MNIST database of handwritten digits and the NEU surface defect database have proved BCOFS's effectiveness in image feature selection. The experimental results show that the classification method based on BCOFS is effective. There is a lot of application space in the field of image recognition. However, the generalization of this classification method in the field of image recognition requires further research.

Acknowledgement. This work is partially supported by The National Natural Science Foundation of China (Grants Nos. 71901152, 61703102), Natural Science Foundation of Guangdong Province (2020A1515010752), Guangdong Basic and Applied Basic Research Foundation (Project No. 2019A1515011392), and Natural Science Foundation of Shenzhen University (85303/00000155).

References

1. Zhang, X., Wang, W., Li, Y., Jiao, L.C.: PSO-based automatic relevance determination and feature selection system for hyperspectral image classification. Electron. Lett. **48**, 1263–1265 (2012)
2. Alizadeh Naeini, A., Babadi, M., Mirzadeh, S.M.J., Amini, S.: Particle Swarm optimization for object-based feature selection of VHSR satellite images. IEEE Geosci. Remote Sens. Lett. **15**, 379–383 (2018)
3. Chen, L., Chen, B., Chen, Y.: Image feature selection based on ant colonny optimization. Aust. Joint Conf. Artif. Intell. **7106**, 580–589 (2011)
4. Chen, B., Chen, L., Chen, Y.: Efficient ant colony optimization for image feature selection. Signal Process. **93**, 1566–1576 (2013)
5. Panda, R., Kumar, M., Panigrahi, B.K.: Face recognition using bacterial foraging strategy. Swarm Evol. Comput. **1**, 138–146 (2011)
6. Lei, L., Peng, J., Yang, B.: Image feature selection based on genetic algorithm. In: Zhong, Z. (ed.) Proceedings of the International Conference on Information Engineering and Applications (IEA) 2012. LNEE, vol. 219, pp. 825–831. Springer, London (2013). https://doi.org/10.1007/978-1-4471-4853-1_101
7. Mlakar, U., Fister, I., Brest, J., Potočnik, B.: Multi-objective differential evolution for feature selection in facial expression recognition systems. Expert Syst. Appl. **89**, 129–137 (2017)
8. Niu, B., Wang, H.: Bacterial colony optimization. Discret. Dyn. Nat. Soc. **2012**, 1–28 (2012)
9. Wang, Y., Li, S., Mao, J.: Computer image processing and recognition technology. J. China High Education Press, pp. 31–105 (2001)
10. Haralick, R.M., Shanmugam, K.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **SMC-3**, 610–621 (1973)
11. Passino, K.M.: Biomimicry of bacterial foraging for distributed optimization and control. IEEE Control Syst. Mag. **22**, 52–67 (2002)
12. Hall, M.A.: Correlation-based feature selection of discrete and numeric class machine learning (2000)
13. Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. In: IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, vol. 5, pp. 4104–4108 (1997)
14. He, Y., Song, K., Meng, Q., Yan, Y.: An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Trans. Instrum. Meas. **69**, 1493–1504 (2020)