Feature Selection for Cotton Foreign Fiber Objects Based on PSO Algorithm

Hengbin Li, Jinxing Wang, Wenzhu Yang, Shuangxi Liu, Zhenbo Li, Daoliang Li

Feature Selection for Cotton Foreign Fiber Objects Based on PSO Algorithm

Hengbin Li[1], Jinxing Wang[1], Wenzhu Yang[2], Shuangxi liu[1], Zhenbo Li[2] , Daoliang Li[2*]

[1]Mechanical and Electronic Engineering College, Shandong Agricultural University, Taian 271018, China
[2]College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
{ytu-lhb, jingxingw, wzyservers,shuangxl,zhboli,daoliangli}@163.com

Abstract. Due to large amount of calculation and slow speed of the feature selection for cotton fiber, a fast feature selection algorithm based on PSO was developed. It is searched by particle swarm optimization algorithm. Though search features by using PSO, it is reduced the number of classifier training and reduced the computational complexity. Experimental results indicate that, in the case of no loss of the classification performances, the method accelerates feature selection.

Key Words: Cotton, Foreign fiber, Feature selection， Particle swarm optimization

Introduction

The foreign fibers in cotton is non-cotton fiber and non-nature color cotton fiber, such as chemical fiber, hair, hemp rope and so on, which are missing with the cotton and have serious influence on the quality of cotton and its product. Foreign fiber has serious impact on textile products even the quantity is very small. Once the fiber is mixed into cotton fiber, it will not only affect the textile spinning capacity, but also form color spots in dyed fabric which may seriously affect the appearance of the fabric and make a great damage to the cotton textile industry. In China, a number of cotton processing enterprises mainly rely on large numbers of workers employed to pick the foreign fiber. It takes a great deal of manpower and material resources. It’s very necessary for improving efficiency and reducing costs to online monitoring technology in cotton foreign fiber based on machine vision.

Feature extraction of cotton foreign fiber can obtain the initial characteristics set, which can describe the feature. By the premise of he classification accuracy is not reducing, select the optimal feature subset with powerful ability of classification from the initial feature set. It can minimize complexity of the classifier design and increase classification speed, which is the premise and guarantee to real-time classification.

Particle swarm optimization (abbreviation: PSO) algorithm is a new global optimization algorithm. It searches optimal solution by the group of each particle kept moving and the motion of each particle is determined by current local optimal solution and global optimal solution. PSO as a heuristic search algorithm is concise, fast convergence, little parameters, etc., have been used in feature selection, classification and many other aspects. It can help the feature optimization and be able to extract optimal feature set of image by using the PSO in characteristics of cotton foreign fiber. The optimal feature subset can shorten the time of classification; even can be improve the classification accuracy in a certain extent.

1 Cotton Foreign Fiber Objects

It has many different foreign fibers in cotton. It is almost impossible to accurately classify in using certain features for the preprocessing foreign cotton fiber's target. Many different types of foreign fibers in cotton, the fiber pretreated to be the target of the opposite sex alone is almost impossible to use certain features on the accurate classification of foreign fibers of cotton, combined with the color, shape and texture features, you can improve the classification accuracy of the opposite sex fibers. So combing color, shape and texture features can improve the accuracy of foreign fibers' classification.

The characteristic of cotton foreign fibers can be expressed by 75-dimensional feature vectors, and the 75 characters include 24 color features, 43 texture features and 8 shape features.

It has 234 sample images, containing foreign fiber with 4000*500 24-bit true color images, in this sample image. These images were collected from real-time images of 4000election.

The sample image was processing segment and the small target was divided to six types, which are plastic, cloth, hemp rope, hair, polypropylene fiber silk and feather. Selected 135 small targets for each type. And obtained 810 samples of 75-dimensional feature data set according feature extracting to small target. The data set of sample characteristics randomly equal divided into 10 parts, each time taking one of them as a test set and the other 9 parts being merged into the training set.

2 Feature selection based on PSO algorithm

2.1 feature selection

Feature selection can be described like that, selecting a feature subset which can make the categorizer have the best classification performance in the d-dimensional feature set. There are two problems which need to be solved, one is the criteria of selection, namely the evaluation criterion which is used to evaluate the properties of the different feature subsets; the other one is the searching strategy, if d has been defined, there are C_D^d kinds of possible combinations. If all of the possible combinations are calculated in one pass, the amount of calculation is very large. Thus a effective searching method which can searched out a group of characteristics in the allowed time is needed.

The method of feature selection can be divided into two kinds: one is Filter method, the other is Wrapper method. Filter method is one kind of high efficient method which can eliminate the big kinds noise, it generally choose information entropy, correlation and class separation distance as the evaluation criterions. When the correlation between the categorizer and features is large, the method does not necessarily selectee a good feature subset. Although the sped of Wrapper method is slower than Filter method, the method is closely associated with the corresponding study arithmetic, then avoids the situation of selecting the wrong feature subsets. Nowadays,more and more researchers choose Wrapper method to selecte features.

2.2 The binary basic principle of particle swarm optimization algorithm

2.2.1 The basic principle of particle swarm optimization algorithm

PSO algorithm is the earliest be by Eberhart and Kennedy put forward 1995. Like other swarm algorithm, pso searches optimal solution by the continual moving of each individual. The motion direction of each particle is determined by local optimal solution Pbest and global optimal solution Gbest.

Each particle represents a D-dimensional solution space of a point, and its next position is determined by its current position and speed.

The location of i-th particle in the k-th iteration can be expreessed by

$\mathrm{X}_{\mathrm{id}}(k) = (x_{i1}(1), x_{i2}(2), ..., x_{iD}(k))$.

And the speed $\mathbf{V}_{\mathrm{i}}\ (k) = (v_{i1}(1), v_{i2}(2), ..., v_{iD}(k))$ is a vector in D-dimensional space.

Let $\mathbf{pBest}_{\mathrm{i}}\ (k) = (\mathrm{pBest}_{\mathrm{i1}}(1), \mathrm{pBest}_{\mathrm{i2}}(2), ..., \mathrm{pBest}_{\mathrm{iD}}(\mathrm{k}))$ be local optimal solution of $\mathbf{X}_{\mathrm{i}}$ and $\mathbf{gBest}_{\mathrm{i}}\ (k) = (\mathrm{gBest}_{\mathrm{i1}}(1), \mathrm{gBest}_{\mathrm{i2}}(2), ..., \mathrm{gBest}_{\mathrm{iD}}(\mathrm{k}))$ be global optimal solution. So in the iteration of K + 1 times, the entire particle positions and speeds is updated as following formula:

$$v_{id}(k+1) = wv_{id}(k) + c_1 r_1(k)[pBest_{id}(k) - x_{id}(k)] + c_2 r_2(k)[gBest_d(k) - x_{id}(k)] \qquad (1)$$

$$x_{id}(k+1) = x_{id}(k) + \beta v_{id}(k+1) \qquad (2)$$

In the formula, c_1, c_2 is learning factor, usually the value of its in the span 0 to 2. The r_1 and r_2 are random number , its value in the span 0 to 1. The w is weighted coefficients，and the β is constraint factor. d = 1, 2… D.

It can be used to deal with continuous space search problem by PSO, but can not be used to solve optimal combination problem. Therefore, Binary Particle Swarm Optimization (abbreviation: BPSO) is be by Eberhart and Kennedy put forward 1997 to solve discrete problem.

In BPSO, the value of $\mathbf{X}_{\mathrm{id}}$ is 0 or 1, and $\mathbf{V}_{\mathrm{id}}$ is the probability of $\mathbf{X}_{\mathrm{id}}$ taking 1, limited by the transfer function to the range of 0 to 1. Usually Sigmond function is be used, which is

$$\mathrm{sig}(\mathrm{v}_{\mathrm{id}}(\mathrm{k})) = 1/[1+\exp(-\mathrm{v}_{\mathrm{id}}(\mathrm{k}))] \qquad (3)$$

In BPSO, the formula of particle position update as follow:

$$\mathrm{x}_{\mathrm{id}}(\mathrm{k}+1) = \begin{cases} 1, & \mathrm{r}_{\mathrm{id}} \le \mathrm{sig}(\mathrm{v}_{\mathrm{id}}(\mathrm{k}+1)) \\ 0, & \text{otherwise} \end{cases} \qquad (4)$$

It is an important research direction to improve the speed of Wrapper by heuristic search method in order to overcome the shortcomings of slow speed of feature selection with Wrapper. As a heuristic search method, PSO has the characteristics of simple, few parameters, and fast convergence and so on. Therefore, BPSO was be used to be a search algorithm of feature selection in this article.

2.2.2 Fitness function

Define particle's evaluation standard of evaluation standard performance as follows:

$$F(i) = p(i) / (1 + \lambda n(i)) \tag{5}$$

In the formula, $F(i)$ is fitness value of generating solution of particle i. And $p(i)$ is the mean rate of correct classification using this subset. $n(i)$ is the number of selected features,

$$n(i) = x_{i1} + x_{i2} + ... + x_{id} \tag{6}$$

λ is weight parameter of the number of feature, and λ generally is 0.01. The larger the value of F, the better indicate that the selected subset has better performance, obtaining a higher classification accuracy with fewer features.

2.3 classification and recognition based on support vector machine

A new optimum design criterion to linear classifier was put forward by Vapnik and others on the basis of year's research of statistical learning theory. The principle of support vector machine (abbreviation: SVM) is based on linearly separable, extended to linearly inseparable problem, even extended to using nonlinear function.

In this system, take optimal feature subset of cotton foreign fiber objects selected by BPSO algorithm as input vector of svm. And use radial basis function as kernel factions, which can obtain relatively superior classification effect. Classification learning of dataset, verify the quality of feature selection by the SVM classifier.

2.4 Algorithm steps

The process of feature selection for cotton foreign fiber objects based on PSO algorithm:

1) Initializing the group of particle, obtain SVM classifier s according training samples set x, taking i=0;

2) For each particle $p_i = x_i$, calculate the characteristics of that subset of the new sample $\mathrm{x}_{\mathrm{j}}(k)\mathrm{Y}$;

3) If the new moderate value is greater than moderate local optimal solution, the

particle x_i local optimal solution update value: $\mathrm{pBest}_i = x_i$;

4) If the new moderate value is greater than global optimal solution, the global optimal solution update value: $\mathbf{gBest}_i = x_i$;

5) Calculating the particle's new speed:

$$v_{id}(k+1) = wv_{id}(k) + c_1 r_1(k)[pBest_{id}(k) - x_{id}(k)] + c_2 r_2(k)[gBest_d(k) - x_{id}(k)] \quad (7)$$

6) Calculating the particle's new location：

$$x_{id}(k+1) = x_{id}(k) + \beta v_{id}(k+1)\mathrm{c}_2 \quad (8)$$

7) Self-increasing i, if I is less than the maximum the number of iterations, then turn step 2), otherwise the algorithm terminates.

3 Results of the experiment and discussion

The algorithm used java into programming, using K-fold cross-validation in this experiment, in which K is 10. The data set of sample characteristics randomly equal divided into 10 parts, each time taking one of them as a test set and the other 9 parts being merged into the training set. In this study, some parameters were set as: $c_1 = c_2 = 2$, PN (particle number) = 20, N (the maximum iteration number) = 1000.

In this experiment, performance of feature selection was compared between pso algorithm and ant colony algorithm [5] (abbreviation: IACA). Each algorithm run 10 times and takes the average. The iteration number of PSO is 718, and running time is 3120s； The iteration number of IATA is 1472, and running time is 5152s. The iteration number of PSO is less 52% than IACA and running time is less 40%, which can prove PSO has higher searching efficiency.

Types	The original number of features	The number of features after feature select of IACA	The number of features after feature select of PSO
Color characters	24	10	11

Texture characters	43	11	9
Shape characters	8	1	0
All characters	75	22	20

Table 1. Comparison of the number of original feature and selected feature set selection

To show the accuracy of PSO algorithm for improving the effectiveness of classification, compare the sets chose by two algorithms (show in Table 1) and its classification performance (show in Table 2). As can be seen from Table 1, optimal feature subset is smaller and more conducive to classifier design after feature selected by PSO. Table 2 shows that PSO can improve the classification accuracy of classifier after comparing. The above results indicate that PSO for feature selection methods to some extent improve the classification accuracy is valid.

Types	Accuracy of original	Accuracy of selected by IACA	Accuracy of selected by PSO
Plastic	92	99	99
Cloth	82	94	95
Hemp rope	78	88	89
Hair	82	89	89
Polypropylene fiber silk	94	100	100
Feather	74	85	87

Table 2. Comparison of classification performance of original feature set and selected future set

4 conclusions

A fast feature selection for cotton foreign fiber objects based on PSO algorithm is be proposed in this paper. The time of feature selection is greatly reduced in the case of lossless of classification accuracy in this method. But the further research needs to improve the algorithm and its classification accuracy.

Acknowledgements

The authors would like to thank the Project supported by the National Natural Science Foundation of China (Grant No. 30971693) and supported by Program for New Century Excellent Talents in University (NCET-09-0731）.

References

1. Zhao,N.,Fu,C.X.:real-time vision-based detecation of foreign fibers in cotton.Journal of Automation an Instrumentation 3,24(2008)
2. Liu,Sh., Zheng,W.X., Li,H.B,Wang, J.X.:Research of Dynamic Identification Techonlogy on Cotton Foreign Fibers.Computer and Computing Technologies in Angriculture IV, 389(2010)
3. Zhu,Z.G., Ahi, D.J.: Digit limage processing. Publishing House of Elecronics Industry,beijing (2003)
4. Wen, Z.Q., Cai, Z.X.:Conwergence analysis of Mean Shift algortithm. Journal of Software 2(18),206-211(2007)
5. Zhao, X.H., Li, D.L., Yang, W.ZH. Feature Selection for Cotton Foreign Fiber Objects Based on Improved Ant Colony Algorithm. Transactions of the chinese Society for Agricultural Machinery,2011,42(4):168-173.
6. Yan, W.ZH., Li, D.L., Wei, X.H. et al. Feature selectiong for cotton foreign fiber objects based on improved genetic algorithm[J].Transactions of the chinese Society for Agricultural Machinery, 2010,41(4): 173 ~ 178.
7. Lin, S.W., Ying, K.C., Chen, S.C.,et al. Particle swarm optimization for parametermianation and feature selection of support vector machines[j]. Expert Systems with Applications, 2008 , 35(4): 1817-1824.
8. Shima K, Todoriki M, Suzuki A. SVM-based feature selection of latent semantic feature.Pattern Recognition Letters, 2004, 25(9):1051 – 1057.
9. Lin, J.Y., Ke, H.R., Chien, B.C., et al. Classifier design with feature selection and feature extraction using layered genetic programming. Expert Systems with Applications, 2008, 34(2), 1384~1393．
10. Uncu Özge, Türkşen I.B．A novel feature selection approach: Combining feature wrappers and filters. Information Sciences, 2007, 177(2): 449~466．
11. Horng Ming-Huwi．Multi-class support vector machine for classification of

the ultrasonic images of supraspinatus. Expert Systems with Applications, 2009, 36 (4): 8124~8133.

12. Sergio Bermejo, Brais Monegal, Joan Cabestany. Fish age categorization from otolith images using multi-class support vector machines. Fisheries Research, 2007, 84 (2): 247~253.