Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(6):928-932

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Application of ant colony optimization in flue-cured tobacco auto-grouping identification

Zhang Ying and Wang Yi*

Department of Computer Science, College of Science, HuaZhong Agricultural University, China

ABSTRACT

Flue-cured tobacco grouping is the basis for automatic grade recognition of tobacco acquisition quality. This paper presents an ant colony optimization that can be used in flue-cured tobacco grouping identification. Compared with other algorithms, it can be used to identify tobacco grouping more quickly. Firstly, collect tobacco from three different locations and of three different colors, and there are 20 samples respectively in each part location and each color. And then extract the image and spectral characteristics of each sample, altogether 16 parameters. Randomly classify the 140 sets of samples. Finally, achieve location and color automatic grouping of the samples by ant colony optimization. The results revealed that, compared with the original grouping, the location and color classification rate based on ant colony optimization is as high as 91.7%. This suggests that the use of ant colony optimization is feasible for classifying unknown tobacco groups.

Key words: Flue-cured Tobacco Leaves Grouping; Ant Colony Algorithm; Clustering Analysis

INTRODUCTION

As a giant country of tobacco production and consumption in the world, tobacco occupies a very important position in China's economy, as well as in state revenues. Currently the flue-cured tobacco acquisition process is mainly dependent on artificial sensory, ranking by experience, which is not only time-consuming but also difficult to ensure accuracy and objective fairness. This study will quantify tobacco quality features and develop intelligent tobacco inspection methods with simulated human vision and wisdom. Thusly, we can achieve the transition from the qualitative inspection of artificial vision to quantitative inspection of machine vision, which is of high economic value and broad application prospects for the automatic assessment of tobacco levels.

Nowadays, many scholars group and grade identify flue-cured tobacco by computer vision technology and spectroscopic technology. The research team, led by Han Liqun from College of Information Engineering, Beijing Technology and Business University, completed the software and hardware system development of flue-cured tobacco quality characteristic extraction and established a standard database management system of flue-cured tobacco. Meanwhile, they also employed artificial neural network technology in the automatic grading of tobacco maturity. Liu Jianjun et al. combined the support vector machine with infrared spectroscopy technique for the analysis of tobacco grading.

All of the above methods only explore the external qualities such as tobacco color, texture, and shape etc. It is difficult to associate its internal quality information with grouping and grading. Besides, the acquired spectral characteristic parameters are not relatively few. Therefore, in this paper, the author proposes to employ the ant colony optimization analysis technique in the identification of different groups of flue-cured tobacco, so as to get an effective method to identify tobacco grouping quickly.

EXPERIMENTAL SECTION

EXPERIMENTAL MATERIALS
Select the tobacco provided by the Yunnan tobacco company that has been artificially graded by experts. In order to ensure that the test results are representative, randomly select 20 samples, grade 1-4, from each group (upper lemon BL, upper orange BF, upper reddish brown BR, Central lemon CL, Central Orange CF, lower lemon XL, lower Orange XF) according to the tobacco leaf, and there are 140 samples altogether in these seven groups of tobacco. In the location and color classification modeling experiments, each sample data are used twice.

Before the test, press the tobacco leaf with two large pieces of glass in order to get satisfactory smoothness. In the scan measurement, take the main vein of each tobacco leaf as axial symmetry; divide two large regions A and B. In each large area, divide four small regions from leaf apex to petiole, namely A1, A2, A3, A4 and B1, B2, B3, B4, altogether eight. Measure these eight areas, and then take the average [1].

INSTRUMENT AND EQUIPMENT
The hyper spectral image data in this experiment is collected by the hyper spectral images system shown in Figure 1. The system mainly includes an imaging spectrometer (VNIR Concentric, Headwall photonics, USA) and high spectral camera (F1.9 35mm compact lens, Schneider Optics, USA), slit, mobile platforms and controllers, 150W visual matching illuminants, image processing machines, and other computer software packages like spectral imaging systems. In these instruments, the imaging spectrometer includes a set of linear array electron multiplication charge-coupled device with effective pixels of 1004*1002. The wavelength is 400 ~ 1000nm, and altogether 753 wavebands. The instantaneous field view of the imaging spectrometer is a thin line, which is dependent on the slit width installed on the imaging spectrograph[2].。

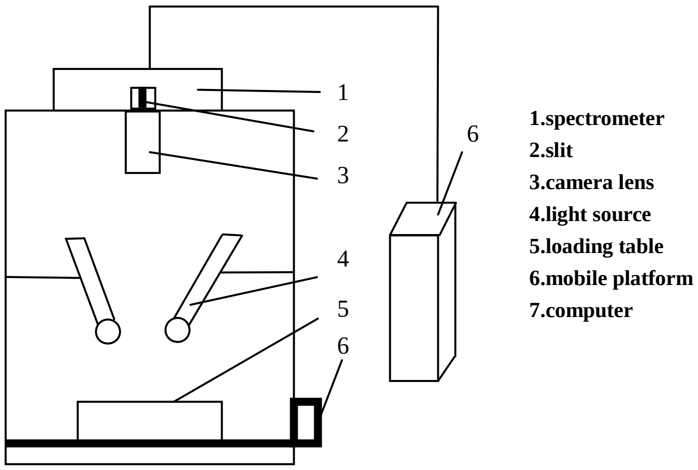

Fig.1 Hyperspectral imaging system

DATA ANALYSIS AND MODELING
DATA COLLECTION OF TOBACCO IMAGES AND SPECTRA
Collect data of tobacco images and spectra as standards of classification, including power of textural features, inertia moments, entropy, correlation and partial smoothness, altogether five parameters and absorption of depth and absorption area, two parameters. Parts of the statistics are listed in Table 1.

DESIGN OF ANT COLONY CLUSTER ALGORITHM
Ant Colony Optimization is a recently proposed search algorithm based on colony optimization[3]. This algorithm is founded on the basis of the collective behavior of real ants in nature. By means of the information exchange between each member ant, search for the optimizing features of the shortest path from the ant nest to food and figure out the optimization issue in the discrete system. Ant colony optimization has already been successfully employed in solving the traveling salesman problem and now expanded into more aspects such as multi-objective optimization, data classification, data clustering and pattern recognition [4]. So far it has already got satisfying experiment results.
Cluster analysis, as a very important method of data mining and knowledge discovery, it's different from classification rules. The categories required by cluster are unknown, so it is typical non-guidance learning algorithm. Cluster analysis is to assemble and divide the physical or abstract objects, forming a multiple-category analysis process combined by similar targets. The input is a set of unclassified objects and beforehand it is unknown how to classify. Nor is it known to classify into how many types. The output is to reasonably divide data sets based on a

certain standard (normally some distance) by analyzing statistics. Confirm the category of each object and assemble the similar objects in to one category.

Table 1 Parts of data

Samples & Serial number	Energy/ J	Moment of inertia/ (m · kg · s^2)	Entropy/ (J · mol^{-1} · K^{-1})	…	Depth/ mm	Area/ mm^2
BL-1	9.312	1.064	11.247	…	0.459	164.58
…	…	…	…	…	…	…
BL-20	9.209	1.125	12.020	…	0.421	166.38
BF-1	9.189	1.352	10.259	…	0.501	167.05
…	…	…	…	…	…	…
BF-20	9.090	1.193	10.302	…	0.489	165.52
BR-1	9.112	1.264	11.627	…	0.465	166.72
…	…	…	…	…	…	…
BR-20	9.315	1.016	11.241	…	0.521	167.91
CL-1	9.209	1.174	10.689	…	0.482	166.36
…	…	…	…	…	…	…
CL-20	9.388	1.201	11.516	…	0.510	166.41
CF-1	9.290	1.329	11.075	…	0.507	167.29
…	…	…	…	…	…	…

Constructing the objective function is a very important step of ant colony analysis. If there are N samples in the known sample set $\{X\}$, M models are classified and each sample is possessed with p characteristic parameters. Wherein, in the Initialize clustering center K-means algorithm is adopt to achieve the fast convergence effect. For the characteristic parameters of the n samples, describe the p characteristic values of the $i(i \in N)$ object as $x_{i1}, x_{i2},..., x_{ip}$, and describe the p characteristic values of the $j(j \in N)$ object as $x_{j1}, x_{j2},..., x_{jp}$. Therefore, we could use distance $d(i, j)$ to represent the differentiation of random samples. In the experiment, the contribution rate of each attribute characteristic parameters can be referred to, plus one weight, thus it can applied the weighted Minkowski distance to define.

$$d(i, j) = (w_1|x_{i1} - x_{j1}|^q + w_2|x_{i2} - x_{j2}|^q + ... + w_p|x_{ip} - x_{jp}|^q)^{1/p} \tag{1}$$

Wherein, $w_k, k \in p$ represents the weight of the kth attribute.

Nevertheless, as the measurement of each feature parameter directly affects the result of cluster analysis. Therefore, in order to avoid the dependence of feature parameters on measurement selection, we are obligated to standardize the data. For the measurement value of a set feature f, take standardized treatment z-score.

$$z_{if} = \frac{x_{ij} - m_f}{s_f} \tag{2}$$

The s_f in the formula is the sum of mean absolute deviation.

$$s_f = |x_{1f} - m_f| + |x_{2f} - m_f| + ... + |x_{Nf} - m_f| \tag{3}$$

$x_{1f}, x_{2f},..., x_{Nf}$ are the N characteristic parameters of f, m_f is the average value of x_{if} . After standardization treatment of z-score, the new characteristic parameter can be achieved.

Therefore, based on the above method, take the distance sum of each sample to the cluster center reaching the minimum value as the final objective function. The mathematical model is:

$$\min J(w,c) = (\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{p} w_{ij}\|x_{ik} - c_{jk}\|^q)^{1/q} \tag{4}$$

In this model, x_{ik} is the kth characteristic of the i sample. c_{jk} is the kth characteristic of the j cluster center.

TOBACCO GROUPING BASED ON ANT COLONY CLUSTER ANALYSIS

The grouping identification based on ant colony cluster analysis is to take the extracted color and texture of tobacco as the sample's characteristic parameter. The formula (4) is served as objective function, through iteration optimization, search the optimal solution. The specific steps are listed as follows:

1) Initialize the parameters of the ant colony, which include the antNum, the centerNum, parameter Q of the conversion rule, parameter Q of pheromone evaporation, local search threshold L and etc.
2) Initialize clustering center and the pheromone matrix.
3) Build the solution set for all the ants according to the pheromone matrix.
4) Calculate all the centers. Calculate the objective function when solving each ant, and sequence all the ants in ascending order based on the value of objective function.
5) Take the first H ants in the sequenced ant solution set as the exchanging ant samples, and conduct the local search on these ants.
6) Update the pheromone value, according to

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{s=1}^{H} \Delta\tau_{ij}^{s}(t) \tag{5}$$

Wherein, $\rho(0 < \rho < 1)$ is the pheromone evaporation parameter, $\Delta\tau_{ij}(t)$ is the pheromone concentration of the i sample and j sample at time t. If the sample i in ant s belongs to category j, then $\Delta\tau = Q / J_s$ otherwise $\Delta\tau_{ij}^{s} = 0$; wherein, J_s is the value of the object function of ant s, Q is the constant parameters.

7) If it doesn't reach at the maximum iteration, turn to the step (3) or output the optimal clustering solution set.

MODEL BUILDING AND IDENTIFICATION RESULT

This experiment takes the color and spectral signature as well as the data of tobacco as the research object. Using MATLAB software to programs, it takes the number of ants as 300, the maximum iterations as 400, classification amount as 3, transform rules as 0.5, and information element evaporation index as 0.1. And then it conducts local search on the former H=5 ant solution set with the local search threshold value as L=0.05 and gets the group results in Table 2. The MATLAB grouping effect picture is left out. Its total discriminant rate is 91.7%.

Table 2 Discriminating results

Group	Sample total	Discriminating results		
		Lemon	Orange	Red brown
Lemon	60	58	2	0
Orange	60	1	57	2
Red brown	20	1	1	18
Group	Sample total	Discriminating results		
		Upper	Central	Lower
Upper	60	60	0	0
Central	40	0	39	1
Lower	40	0	1	39

CONCLUSION

In this paper, the research makes use of the energy, moment of inertia, entropy, dependency, local stationarity, RGB, HIS, Lab, absorption depth, and absorption area of tobacco as the characteristic parameters for tobacco, and realize the automatic classification of tobacco's color and parts based on the ant colony clustering analytical method. Compared to the original group, the recognition rate can reach to 91.7% with a better grouping effect.

Acknowledgments

This work was financially supported by the Fundamental Research Funds for the Central Universities (Program No. 2013QC027).

REFERENCES

[1]Zhang Ying, He Liyuan. *Transactions of the CSAE*, **2011**,27(4),350-354.
[2]Zhou Zhu, Li Xiaoyu, Gao Hailong, et al. *Transactions of the CSAM*. **2012**,43(2),128-133,185.
[3]DORIGO M, GAMBARDELLA L M. Ant colony system: *a cooperative learning approach to the traveling*

salesman problem. IEEE Trans-actions on Evolutionary Computation, **1997**,1(1),53-66.
[4]Duan Haibin. *Principle and application of ant colony algorithm*.Beijing: Science Press, **2006**,35-36
[5]FELDMAN R． *Journal of Intelligent Information Systems*,**1998**,10(3),281-300.