Infrared Spectrum Analysis to Extract Information of Physical and Chemical Characteristics of Fruit Tree Leaf

Xiaoli WANG[1,2], Xiaoli ZHANG[1]* and Guomin ZHOU[2]

[1]Beijing Forestry University, BeiJing, 100083, China
[2]Institue of agricultural information, Chinese academy of agricultural sciences, BeiJing, 10008, China

***Abstract* — In this paper we propose infrared spectrum analysis using mixed differential particle swarm to extract information of the physical and chemical characteristics of fruit tree leaf, as follows: i) we study differential evolution algorithm and particle swarm algorithm, ii) for differential evolution algorithm, 4 different individuals are selected from the swarm randomly to generate a difference vector and mutation operation scheme is made for the individuals, iii) for particle swarm, a simplified particle swarm position transformation formula which does not contain speed item is selected and then cross fusion is made for differential evolution sub-swarm and particle evolution sub-swarm to promote optimization algorithm performance, iv) Fushi fruit tree leaf is selected as an experimental subject and the proposed optimization algorithm is used for characteristic wavelength optimization selection of infrared spectrum, v) we discuss the experimental results to verify the validity of the proposed method.**

Keywords - differential evolution; particle swarm algorithm; infrared spectrum; fruit tree leaf; physical and chemical characters

I. INTRODUCTION

Nondestructive detection for internal quality of fruit tree leaf is a hot research direction in agricultural engineering and physical and chemical characters are important indexes to evaluate fruit tree leaf. Traditional detection method uses sampling for chemical detection generally. But the analysis process is complicated. It takes long time and expensive costs and the technical condition is complex and it can not realize prompt monitoring and the samples need to be destroyed.

In recent years, as a nondestructive detection method, near infrared spectrum technology is widely used in information extraction for physical and chemical characters of fruit tree leaf. Literature [1] uses image information and near infrared spectrum data fusion technology to promote classification accuracy for physical and chemical characters of fruit tree leaf. Literature [2] uses ant colony optimization to optimize optimal variable of near-infrared wave length and it reduces use of wave length effectively and reduces complexity of model and also promotes prediction accuracy of the model. Literature [3] uses combined moving window and genetic algorithm-partial least squares algorithm and selects information variable to establish physical and chemical characters prediction model of fruit tree leaf. Literature [4] transfers hybrid linear analysis into near infrared spectrum detection of an impure constituent content index, namely physical and chemical characters of fruit tree leaf and model accuracy is superior to that of traditional model. At present, physical and chemical characters detection of fruit tree leaf conducted with both traditional detection method and near infrared spectrum technology, only aims at fruit tree leaf and fruit. There is no early-stage dynamic monitoring for final fruit quality of fruit tree leaf.

The research takes physical and chemical characters of fruit tree leaf as object of study and uses spectrum characters of fruit tree leaf and to track physical and chemical characters information of fruit tree leaf and fruit in whole process, and to explore contribution proportion for physical and chemical character accumulation of fruit so as to provide whole-process, convenient and nondestructive detection method for physical and chemical character detection of fruit tree leaf and fruit. With an information extraction for physical and chemical characters of fruit tree leaf based on mixed differential particle swarm infrared spectrum analysis and mixed differential particle swarm optimization algorithm, it provides theoretical foundation on the basis of fruit tree leaf and tree growth monitoring.

II. NHDEPSO ALGORITHM

A. Basic DE Algorithm

DE algorithm has features of individual optimal solution memory and population information sharing, namely it can optimize solution of problems through cooperation and competition among individuals of the population. It is a kind of greed genetic algorithm with real number encoding and excellence preservation thought essentially.

The algorithm needs to get a group of population which is initialized randomly in searching space:

$$X^0 = \left[x_1^0; x_2^0; \cdots; x_{N_p}^0 \right] \tag{1}$$

Np is scale of population. After a series of specified operation, the *t*th generation of individual evolution is:

$$x_i^t = \left[x_{i,1}^t, x_{i,2}^t, \cdots, x_{i,D}^t \right] \tag{2}$$

Where, *D* is number of dimensions for problems need to be optimized.

Variation can prevent evolution from falling into local extreme value and its methods are various. Two kinds of basic variation method are listed here: DE/rand/1 and DE/best/2:

$$x_m = x_{s3}^t + F * \left(x_{s1}^t - x_{s2}^t\right) \tag{3}$$

$$x_m = x_{gbest}^t + F * \left[\left(x_{s1}^t - x_{s2}^t\right) + \left(x_{s3}^t - x_{s4}^t\right)\right] \tag{4}$$

In Equation (1) and (2), $x_{s1}^t, x_{s2}^t, x_{s3}^t, x_{s4}^t$ are different random individuals; x_{gbest}^t is the individual with the best fitness in current population; $F \in [0,2]$ is zoom factor.

Cross strategy is: assume individual x_i^t and x_m in the population conduct cross operation and they generate test individual x_T. In order to ensure individual evolution, at least one x_T shall be contributed by x_m through random selection and others shall use cross probability factor CR. Cross operation equation is:

$$x_{Tj} = \begin{cases} x_{mj},\ rand \le CR \\ x_{ij},\ rand > CR \end{cases} \quad j = 1, 2, \cdots, D \tag{5}$$

Selection operation uses "greed" search strategy. That with high adaptive value shall be selected as filial generation:

$$x_i^{t+1} = \begin{cases} x_T,\ f\left(x_T\right) < f\left(x_i^t\right) \\ x_i^t,\ f\left(x_T\right) \ge f\left(x_i^t\right) \end{cases} \tag{6}$$

In above algorithm, differential vector of the difference between two different random individuals of parent generation shall be added to the 3rd individual selected randomly to form a variation individual. Then cross operation shall be conducted between parent generation individual and variation individual according to certain probability to generate a test individual. Then selection operation shall be conducted between parent generation individual and test individual according to fitness function value. Individual with better fitness shall be selected as filial generation and it can guarantee optimal direction of evolution [2].

B. Particle Swarm Algorithm

For standard PSO algorithm, the swarm consisting of *m* particles flies in *n*-dimensional searching space with certain speed. When particles search, they consider their best point in history and best points of other particles in the swarm. They make position change on this basis and they stop and output optimal solution when it meets the condition:

$$\begin{aligned} \underbrace{p_j^i(t+1)}_{\text{Position of next moment}} &= \underbrace{p_j^i(t)}_{\text{Current position}} + \underbrace{\omega \cdot v_j^i(t)}_{\text{Current speed reference}} + \\ &\underbrace{c_1 \cdot r_1 \cdot \left(pb_j^i(t) - p_j^i(t)\right)}_{\text{Individual optimum reference in history} Gb_j^i} + \\ &\underbrace{c_2 \cdot r_2 \cdot \left(pg_j^i(t) - p_j^i(t)\right)}_{\text{Swarm optimum reference} Gg_j^i} \end{aligned} \tag{7}$$

In the equation, superscript *i* corresponds to the *i*th individual ($i = 1, \ldots, N$, N is the scale of population); subscript j corresponds to the j-dimension ($j = 1, \ldots, n$); *t* represents current times of iteration; $P^i = \left[p_1^i \ldots p_n^i\right]^{\mathrm{T}}$ and $v^i = \left[v_1^i \ldots v_n^i\right]^{\mathrm{T}}$ represent position and speed of the ith particle and $\left|v_j^i\right| \le v_{\max}$; ω is inertia factor; c_1 and c_2 are accelerated factor for individual optimum and swarm optimum in history respectively; r_1 and r_2 are random number in [0,1].

C. Algorithm Improvement

Many scholars have researched improvement of differential evolution algorithm and particle swarm algorithm. Two simple improvement methods are adopted in the Thesis to promote performance of the algorithm. A new differential mutation method is proposed in the Thesis. Such mutation method selects 4 different individuals from the population randomly to generate differential vector and mutation operation scheme is made for the individual. This scheme can both promote convergence rate of the algorithm and keep high diversity of the population in certain degree. Individual mutation operation equation is:

$$x_i^{t+1} = x_i^t + F\left(\left(x_{r1}^t - x_i^t\right) + \left(x_{r2}^t - x_{r3}^t\right)\right) \tag{8}$$

In above mutation equation, for the *i*th individual x_i^t of the *t*th generation of population, base vector is still x_i^t, representing mutation on the basis of x_i^t. This helps to keep diversity of the initial population and also considers evolution direction of the population. The latter item $x_{r2}^t - x_{r3}^t$ is introduced as disturbing item. This can make individual keep as large difference as possible after mutation so as to keep diversity of the population.

Literature [13] thinks particle swarm algorithm in the past is based on two concepts: "position" and "speed" and improved algorithm also add operator such as hybridization and mutation on the basis of these two parameters. It makes algorithm description becomes more and more complex. After studying and analyzing related literature and material, the author proposes a simplified particle swarm position transformation formula which does not contain speed item:

$$p_j^i(t+1) = \omega p_j^i(t) + c_1 r_1\left(pb_j^i(t) - p_j^i(t)\right) + c_2 r_2\left(pg_j^i(t) - p_j^i(t)\right) \tag{9}$$

In above equation, the 1st item on the right is "history" part, representing influence of the past on present and it adjust influence degree through ω; the 2nd item is "cognition" part, representing particle's thought of itself; the 3rd item is "social" part, representing comparison and simulation to the neighboring particle and it realizes information sharing and cooperation among particles.

D. NHDEPSO Algorithm Steps

Step 1: set parameters such as N_p, D, F, CR, $x\min$, $x\max$, c_1 and c_2; set evolution algebra counter $t=0$; set maximum value of variable space times T.

Step 2: initialization; generate N_p individuals randomly in searching space $[x\min, x\max]$ as initial population $P(0)$.

Step 3: individual evaluation; calculate evaluation function value $f(i)$ for each individual in the population $P(t)$ and assign value to $pb_j^i(t)$ and $pg_j^i(t)$ according to the evaluation value.

Step 4: conduct mutation operation according to Equation (7) and generate population $P_1(t)$; conduct mutation operation according to Equation (8) and generate population $P_2(t)$.

Step 5: $P_1(t)$ and $P_2(t)$ conduct cross operation according to Equation (5) to generate next generation of population $P'(t)$.

Step 6: conduct selection operation for population $P_1(t)$ and $P'(t)$ according to Equation (6) so as to generate filial generation population $P(t+1)$; evaluate population $P(t+1)$ individual

Step 7: if it meets accuracy, evolution shall be stopped and optimum individual shall be output; if not, turn to Step 3.

III. EXPERIMENT PART

A. Experimental Data Collection

Selected Fushi fruit tree leaf comes from main producing areas of fruit tree leaf of China: Shandong, Shaanxi and special production area Xinjiang. 207 fruit tree leaves without defect, damage or pollutant are selected and 138 leaves (40 from Shaanxi, 48 from Shandong and 50 from Xinjiang) are selected randomly as calibration set and the remain 69 leaves (20 from Shaanxi, 24 from Shandong and 25 from Xinjiang) are deemed as prediction set. They will be numbered respectively and then placed in 4℃ ice tank for storage. Fruit tree leaf shall be taken from the ice tank and shall be placed in the laboratory for 12h before experiment, making comprehensive temperature of fruit tree leaf sample consistent with environment temperature. Temperature and humidity of the laboratory shall be unchanged basically during experiment.

Near infrared spectrum data acquisition equipment of Fushi fruit tree leaf is AntarisTM II type Fourier transform near-infrared spectrograph, which adopts InGaAs detector and Result Integration 3.0 acquisition software. Spectrum acquisition scope is 3800-14000cm. Instrument built-in background is reference and integrating sphere diffuse reflection acquisition method is adopted. Scan times are 32; resolution ratio is 4cm; sampling interval is 1.928cm-1. Each spectrum can gain 5291 variables. Spectrum of every fruit tree leaf sample shall be acquired for 3 times and acquisition point is located in the place with 120o interval away from equator position of fruit tree leaf. Mean value of 3 times of spectrum shall be deemed as original spectrum of the sample.

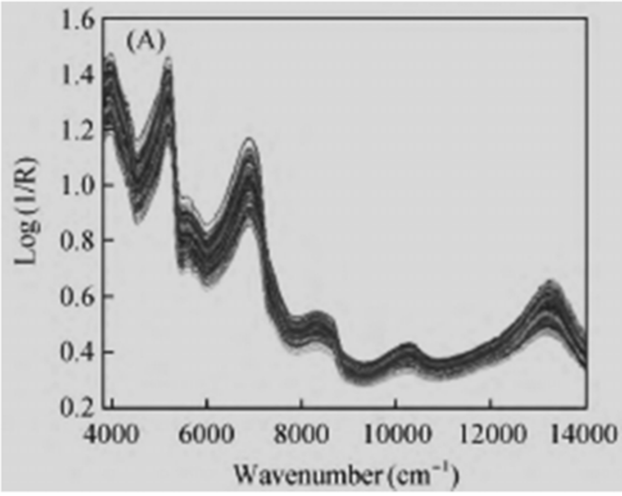
After fruit tree leaf spectrum is acquired, soluble solid content value shall be measured in the spectrum acquisition position. Measurement for soluble solid content of fruit tree leaf shall refer to GB 10651-2008. Arias 500 type semi-automatic Abbe refractometer shall be adopted and automatic temperature compensation is 20oC. 1-2 drops of juice shall be dripped on the center part of prism from fruit sample and then auxiliary prism shall be closed quickly. Readings where light and shade boundary coincides with cross hair point shall be the soluble solid percentage of fruit tree leaf under 20℃ condition. Mean value of soluble solid contents measured in 3 spectrum acquisition parts of the fruit tree leaf shall be deemed as reference value of the fruit tree leaf sample. Table 1 lists statistics for measured value of soluble solid of fruit tree leaf such as change scope, mean value, standard deviation and variable coefficient.

TABLE 1. STATISTICAL RESULT FOR MEASURED VALUE OF SOLUBLE SOLID CONTENT OF FRUIT TREE LEAF

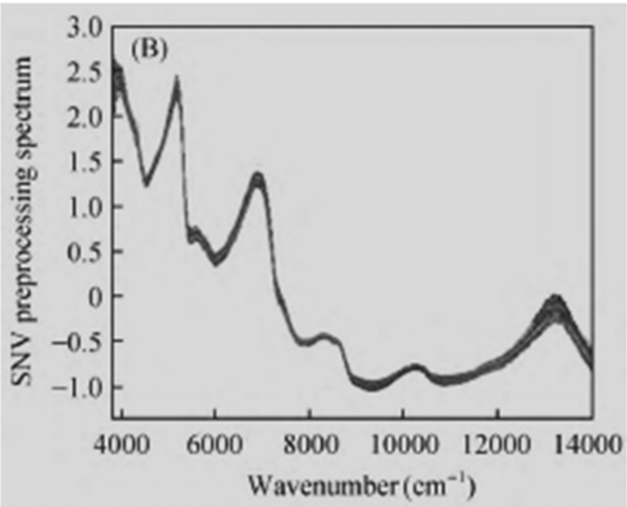
	Samples	range	mean	SD	RSD
Celebrative set	138	10.68-22.12	16.68	2.48	16.08
Predictive set	70	10.87-22.18	16.48	2.51	16.12

B. Spectrum Data Pre-processing

Near infrared spectrum is the absorption spectrum for frequency combination and frequency doubling which contain hydrogen groups. Its spectral line is smooth with small absorbance change. In the original near infrared spectrum acquired in 3800~14000 cm-1 scope, 3 obvious absorption peaks are characteristic peaks of moisture. Corresponding spectral absorption difference of soluble solid component to be measured is no huge and characteristic wavelength can not be determined directly. During near infrared spectrum acquisition, there are many noise information involved such as high-frequency random noise, baseline shift, surface towering and light scattering and this will influence correlation relationship between near infrared spectrum and soluble solid content and will influence reliability and stability of the model established directly. In order to eliminate noise influence, standard normal variable transformation (SNV) is adopted for near infrared spectrum preprocessing. Original near infrared spectrum and spectrum after SNV preprocessing are shown in Fig. 1.



(a) Original infrared spectrum



(b) Infrared spectrum after preprocessing

Figure 1. Infrared spectrum comparison before and after preprocessing

C. Characteristic Wavelength Optimization Selection

All 5291 wave points in near infrared spectrum are deemed as selection object. Control parameters of optimization algorithm are verified in several times of experiments. Setting of the research is: initial population size is 80; maximum number of iterations is 50; maximum number of circulation is 20; threshold value of variable selection probability is 0.3. Experimental results are shown in Fig. 2.

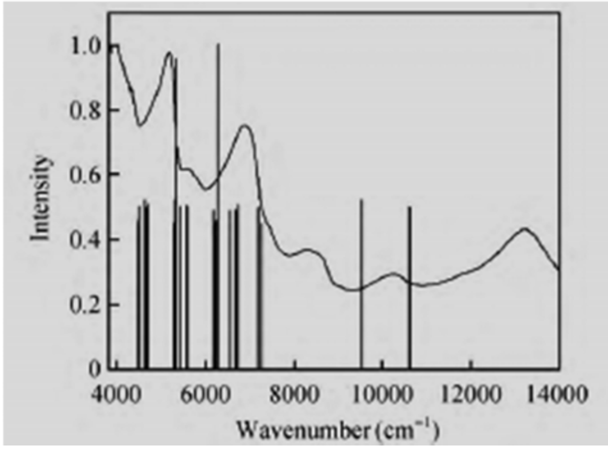


Figure 2. Characteristic wavelength of near infrared spectrum selected by optimization algorithm

Fig. 2. shows selective probability of wave point for once operation of ant colony optimization. There are 22 wave points with high selective probability such as 6275.2, 5341.9, 9538.2 and 4632.2cm. It can be deemed that these wave length points are closely related to soluble solid components. Selected wave points are not in the position of moisture absorption peak, which indicates that selection result of ant colony optimization has strong anti-interference capacity.

IV. CONCLUSIONS

The Thesis proposes an information extraction method for physical and chemical characters of fruit tree leaf based on infrared spectrum analysis of mixed differential particle swarm. For algorithm mutation basis, it conducts cross fusion for differential evolution sub-swarm and particle evolution sub-swarm to promote optimization algorithm performance. Then it conducts characteristic wavelength optimization selection for infrared spectrum of physical and chemical characters of fruit tree leaf with the proposed optimization algorithm. Experiment results verify validity of the method proposed.

ACKNOWLEDGMENTS

National High-tech R&D Program (863 Program): Orchard and canola crop production Digital Management Platform based on the model, No.2013AA102405; SFA Public Welfare Forestry Scientific Research Project: Study on Key Technologies for Monitoring and Warning of Major Forest Plague of Insects, No.201404401.

REFERENCES

[1] Wang Y M, Shu-Ling M A, Feng L Q. The Influence of Oil Heat Treatment on Wood Decay Resistance by Fourier Infrared Spectrum Analysis[J]. Guang pu xue yu guang pu fen xi = Guang pu, 2014, 34(3):660-663.

[2] Sebben D, Pendleton P. Infrared spectrum analysis of the dissociated states of simple amino acids[J]. Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy, 2014, 132(132C):706-712.

[3] Li L L, Chai X S, Tian Y X, et al. A new sample pretreatment method of infrared spectrum analysis for barium sulfate identification in

paper[J]. Chung-kuo Tsao Chih/China Pulp and Paper, 2015, 34(4):39-42.

[4] Chen B, Wu H Y, Han C, et al. [Study on Trace Water Influence on Electrical Properties of Insulating Oil Based on the Mid-Infrared Spectrum Analysis].[J]. Guang pu xue yu guang pu fen xi = Guang pu, 2015, 35(10):2740-2745.

[5] Yang A, Hua Q, Liu X, et al. Application of Infrared Spectrum Analysis on Heat Treatment Process of the Silver Detonating Cord[J]. Explosive Materials, 2015.

[6] Wang Z Y, Cao J J, Luo S Y, et al. X-Ray Diffraction and Infrared Spectrum Analysis of Fault Gouge in Wenchuan Seismic Belt[J]. Spectroscopy & Spectral Analysis, 2014, 34(34):1416-1420(5).

[7] Zhang N, Fan N, Yan H E. Application of near-infrared spectrum analysis to measure sugar liquid concentration[J]. Optical Instruments, 2014.

[8] Luo H L, Liu X X, Gong S J, et al. [Infrared Spectrum Analysis of Propolis and Tree Gum Collected from Different Areas].[J]. Guang pu xue yu guang pu fen xi = Guang pu, 2015, 35(11):3058-3062.

[9] Jinyu Hu and Zhiwei Gao. Distinction immune genes of hepatitis-induced heptatocellular carcinoma[J]. Bioinformatics, 2012, 28(24): 3191-3194.

[10] Jiang, D., Ying, X., Han, Y., & Lv, Z. (2016). Collaborative multi-hop routing in cognitive wireless networks. Wireless personal communications, 86(2), 901-923.

[11] Lv, Z., Tek, A., Da Silva, F., Empereur-Mot, C., Chavent, M., & Baaden, M. (2013). Game on, science-how video game technology may help biologists tackle visualization challenges. PloS one, 8(3), e57990.

[12] Jiang, D., Xu, Z., & Lv, Z. (2015). A multicast delivery approach with minimum energy consumption for wireless multi-hop networks. Telecommunication systems, 1-12.

[13] Lv, Z., Chirivella, J., & Gagliardo, P. (2016). Bigdata Oriented Multimedia Mobile Health Applications. Journal of medical systems, 40(5), 1-10.

[14] Chen B, Han C, Liu G. Detection on Particulate PollutantinTransformer oil Based on the Mid-Infrared Spectrum[J]. Acta Photonica Sinica, 2016.

[15] Reru Y, Wibowo N A, Rondonuwu F S. Construction of calibration model for sweetness degree on fructose using the Fourier-transform near-infrared spectroscopy and singular value decomposition analysis[J]. AIP Conference Proceedings, 2016, 1746(1):431-438.

[16] Jing L, Xiaoxuan G, Yuzhao J, et al. A Micro-Analysis of Buton Rock Asphalt[J]. Journal of Computational and Theoretical Nanoscience, vol. 12, issue 9, pp. 2751-2756, 2015, 12.

[17] Malik M, Wysokiński R, Zierkiewicz W, et al. Raman and Infrared Spectroscopy, DFT Calculations, and Vibrational Assignment of the Anticancer Agent Picoplatin: Performance of Long-Range Corrected/Hybrid Functionals for a Platinum(II) Complex[J]. Journal of Physical Chemistry A, 2014, 118(34):6922-34.

.