NEAR INFRARED SPECTRUM DETECTION OF SOYBEAN FATTY ACIDS BASED ON GA AND NEURAL NETWORK

Changli Zhang [1], Kezhu Tan [2], Yuhua Chai [2], Junlong Fang [1], Shuqiang Liu [3]

[1] *Engineering College, Northeast Agricultural University, Harbin, China, 150030*
[2] *Cheng Dong College, Northeast Agricultural University, Harbin, China, 150030*
[3] *Hei longjiang Engineering College, Harbin, China, 150050*

Abstract: This paper represented a way to build mathematical model on genetic multilevel forward neural network. Building the relationship between chemistry measurement values and near infrared spectrum datum. The near infrared spectrum data was input in this network, five kinds of content of fatty acids, which measured by chemistry method, were output. Training the weight of multilevel forward neural network by genetic algorithms, building the soybean fatty acids neural network detection model, and exploring the network model which can realize near infrared spectrum detection exactly and efficiently. The authors designed a multilevel forward neural network trained by genetic algorithms. Test showed that relative coefficient in five fatty acids of soybean can be round about 0.9, and can satisfy init detection of soybean breeding.

Keywords: near infrared, multilevel forward neural network, genetic algorithms, soybean, fatty acids

1. INTRODUCTION

With the development of soybean seed breeding and soybean production, breeding experts in breeding has attached more and more importance to the quality of breeding soybean. The quality of oil depends much on the proportion of fatty acid and fatty acid is composed of Kitool acid, Stearic

acid, Oleic acid, Suboleic acid, Flax acid. While the process of breeding detects soybean fatty acid more frequently and the traditional methods of detection exist some disadvantages, such as time-consuming detection, high cost and a few detecting samples. To these difficulties, near infrared spectrum can solve those disadvantages efficiently.

Methods applying to near infrared spectrum include MLR, SMR, PCR and PLS. Only some spectrum information can be calculated in the methods of MLR and SMR when the samples are analyzed, losing a big surplus of information, the instructed model will generate much adaptability easily. Although PCR eliminate the question of linear relationship, but when decomposing spectrums matrix, inner relationship between the matrix and content in density can't be considered. The model was constructed based on the method that GA and neural network in this paper and forecasting soybean fatty acid by this constructed model.

At present, the method of GA combining neural network has been applied in every field.GeHong.et. Combined neural network fuzzy controller, training method of variation textures with modified GA. This method not only realized self-study of membership grade but also realized regular self organization, it settled a difficult question in fuzzy control, and it improved primary BP arithmetic which had the disadvantages in getting into local part optimality, realized the control goal of object.(LuChun.etc, 2001)raised a new arithmetic: BP-GA which combined GA with neural network. Because there is a strong complementation between GA and BP, GA-BP's convergence probability and convergences pace are good. (Wen Shaochun.etc, 2001) discussed the main appliance of the method which based on GA and neural network and got a lot of experiment data, the experiment indicated that GA has the function with studying net weighting fast. (Luo Jian, Liu, Junxiang, 2001) raised a new operator-BP operator.

To summarize, with the rapid development of computer technology and artificial intelligence, the method that GA combined with neural network developed rapidly. They beat out a new path for agriculture production.

2.   MATERIALS AND METHODS

Haizhou opencast colliery which is the largest opencast colliery in Asia is located in the north temperature zone and is characterized by mountainous topography. This natural barrier has a strong influence on the meteorological conditions determining the air pollution situation. This site has been studied for approximately 5 years by researchers at China Agricultural University.

2.1   Instruments and devices

The soybean fatty acid contents detector of the experiment uses which is the perten8620 near-infrared instrument produced by the Sweden Bo Tong company. This instrument has 20 light filters, separately corresponds 20 wavelengths, and has the NIR software procedure.

The instrument to menstruate soybean fatty acid contents by Chemistry method uses GC-9A gas chromatography which produced by DaoJin in Japan.

2.2   Experimental materials

The experiment selects 25 different varieties soybeans which produced by Heilongjiang Province as the examination sample, and uses chemistry method to menstruate 5 varieties fatty acids content, concrete like table 1 show:

Table1. Examples of Experiment

Sample name	Kitool acid %	Stearic acid %	Oleic acid%	Suboleic acid %	Flax acid %
DongNong01-1357	10.82	4.116	23.7	50.69	10.27
KenNong19	10.33	3.774	25.81	51.09	8.671
HeiNong41	10.83	4.237	23.78	52.79	7.835
HeFeng42	10.55	3.193	29.48	49.45	7.056
Hei7-50	12.05	3.755	21.82	53.22	8.641
HeiNong44	10.57	4.345	25.61	50.49	8.335
KenFeng9	11.64	4.261	22.06	52.27	9.302
HeFu93155-16	12.29	4.1	21.05	53.95	7.985
SuiNong10	11.38	3.651	18.66	56.95	8.849
HeiHe99	9.768	4.403	24.99	52.43	7.969
HeiNong37	11.94	3.624	17.44	57.33	9.187
HeiHe97-1225	12.18	3.612	23.78	51.27	8.627
BaoFeng9	11.23	4.356	24.56	50.83	8.364
HeFeng41	9.907	4.28	36.62	41.41	7.178
Sui98-579	10.02	3.033	25.57	52.17	8.679
JiYu47	10.6	3.518	23.16	51.45	10.65
DongNong300	10.58	3.519	23.19	51.53	10.66
JiGuang441	11.81	3.599	25.44	50.92	7.671
BeiFeng16	11.66	4.229	21.38	54.06	8.123
NongDa5236	11.98	3.8	25.95	48.33	9.396
SuiHuaWuXing soybean	12.63	3.415	27.02	45.92	10.25
BeiJiang171	12.06	3.971	26.94	47.32	9.141
DongNong1620	10.28	3.95	24.83	52.57	7.789
ShanHe711	11.13	3.692	23.16	53.96	7.368
HeFeng43	11.33	3.867	22.2	5.308	8.142

3.   EXPERIMENTAL METHODS

When the near-infrared spectroanalysis technology carries out quantitative analysis. Firstly, choose having representative and can coverage the inspected parameter sample collection, according to the chemistry

determining value (as table 1 show) which being allotted for picket age assembles and rectify-forecast assembles. The chemistry determines value coming from corn of Department of Agriculture and products quality monitoring checking testing centre (Harbin) detecting result ,the method among them base on the national standard gas appearance chromatography

。In putting soybean powder sample inside the near-infrared Perten 8620 standard forms admeasurements instrument (the Sweden Bo tong instrument company produced) receiving sample system, collecting the light absorption value of every sample in each wavelength, and changes into (actual light absorption of log (1/ R) amounts) , finally stores up the belt ready-made a computer document , forming a spectrum strip

The spectrum data uses chemistry metrology method to build the quantify mathematic model between determine value and near-infrared chemistry spectrum data with picket age collection, the spectrum data uses intersection to confirm law later with the collection rectifying a forecast , checks what be built a model mass coming the relevance modulus by the fact that forecast standard deviation and the chemistry determining value and the near-infrared spectrum forecasting the value room. If built the model precision reached request, been available, it carries out quantitative analysis on the sample. This method characteristic is to analyze speed quickly, the sample is not to assumes any pretreatment, less dosages, no contaminating.

The main body of a paper is adopt to inherit the quantify mathematic model between determine value and near-infrared multilayer front make a present of neural networks (Zitzler et al, 1999) building-up chemistries spectrum data, the entering data looking on near-infrared spectrum data log (1/ R) as a network, takes using chemistry to follow the five kinds fatty acid contents determining as network output, making use of inheritance algorithm to train multilayer front make a present of neural networks right value again, the neural networks building soybean fatty acid is check a model , final, carry out a checkout again with the model that the collection correct rectifying a forecast gets.

4.   DATA PROCESSING AND RESULT ANALANCE

4.1   The principle of neural network combined with GA

Artificial neural network has strong ability of non-linear mapping, it uses the method like black box to remember and find the relationship between input and output. But, the designer of neural network still depends on experiential experts. In the actual study, we found that train the network

using the traditional BP algorithm and then the actual calculation , the effect is not very satisfactory, there is a major problem as follows : (1) Convergence speed is very slow, sometimes shock; (2) Smallest error is local minimum ; (3) Unable to accurately determine the optimal network structure . The appearance of GA brought neural network training with a new look, The objective function is neither to be consecutive, nor differentiable, only that the problem computable , And it always search throughout the entire solution space, easy access to the global optimal solution .Therefore, combine genetic algorithm with BP network, the mixed neural network training is a feasible way.

4.2   The principle of constructing model

The model was constructed based on the method that GA and neural network in this paper and forecasting soybean fatty acid by this constructed model. GA optimizes the weighting of neural network, BP arithmetic realize local part precise search.
Input of the network is X=[x1, x2, …x10] T, output of the network is Y=[y1,y2,…y25]T,Fig.1 is the NN model of soybean fatty acid:
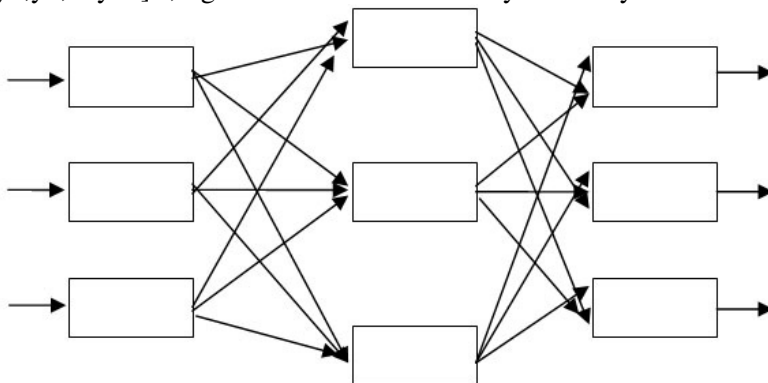

Fig.1 The NN model of soybean fatty acid

4.3   Calculation process of the model

Network training is a key step in artificial neural network running , BP network training is based on weight modify principles of error gradient reduce , Convergence in local minimum points is its own ineluctable

shortcoming. Genetic algorithm is based on the theory of biological evolution mathematical algorithm system, It is an effective global search tool, but not in the precise local search. The combination of artificial neural network and genetic algorithm achieved the complementary of strengths between them. For fixed structure neural network, evolutionary process of connection weights can be divided into the following steps:

(1) Determine weights coding schemes, generate initial population；
(2) decode each individual in the population, construct the corresponding neural network;
(3) compute network fitness according to certain performance evaluation criteria；
(4) Determine reproduce probability of each individual According to fitness size, then complete seed selection；
(5) Impose genetic operator on the selected population according to certain probability, in order to get a new generation of population；
(6) return to step 2, until the performance meets requirements

4.4   Realization of BP neural network and Genetic Algorithm program

Initialize, simulate and train the network using MATLAB; observe dynamic training process of changes through graphics. Specific training steps are as follow：

(1) Initialize weights w and thresholds b of each layer using a small random number, to ensure that the network was not saturated by large weighted inputs, at the same time the parameters are set or initialized.
(2) Generating network weights according to the generation method of initial population, while computing network input vector and network errors.
(3) The expression of chromosome and the formation of initial population.
(4) Genetic Algorithm operating steps that are selection, crossover and mutation。

Run self designed MATLAB program, Choose 20 samples training the network，Five Outputs are the content of palmitic acid, hard fatty acids, oleic acid, linoleic acid and linolenic acid，Specific steps are shown table 2 ：

Table 2 the Output Value of Neural Network

Sample name	Kitool acid %	Stearic acid %	Oleic acid%	Suboleic acid %	Flax acid %
DongNong01-1357	10.8265	4.1191	23.7055	50.6953	10.2773
KenNong19	10.3310	3.7741	25.8107	51.0906	8.6719
HeiNong41	10.8255	4.2212	23.7773	52.7862	7.8312
HeFeng42	10.5460	3.1964	29.4757	49.4461	7.0509
Hei7-50	12.0401	3.7390	21.8080	53.2071	8.6254
HeiNong44	10.5784	3.3444	25.6183	50.4973	8.3483
KenFeng9	11.6533	4.2640	22.0697	52.2788	9.3123
HeFu93155-16	12.3085	4.1113	21.0668	53.9660	8.0066
SuiNong10	11.3742	3.6677	18.6534	56.9456	8.8419
HeiHe99	9.7861	4.3923	25.0053	52.4430	7.9842
HeiNong37	11.9262	3.6378	17.4288	57.3207	9.1798
HeiHe97-1225	12.1858	3.6072	23.7862	51.2755	8.6316
BaoFeng9	11.2029	4.3638	24.5359	50.8088	8.3384
HeFeng41	9.8998	4.2939	36.6132	41.44049	7.1733
Sui98-579	10.0190	3.0303	25.5684	52.1683	8.6769
JiYu47	10.6032	3.5136	23.1628	51.4514	10.6530
DongNong300	10.5733	3.5205	23.1864	51.5265	10.6549
JiGuang441	11.8237	3.5941	25.4525	50.9310	7.6856
BeiFeng16	11.6543	4.2915	21.3755	54.0553	8.1141
NongDa5236	11.9803	3.7917	25.9495	48.3291	9.3939

5.   CONCLUSIONS

Having analyzed characteristics of artificial neural networks and genetic algorithms, we put the combination of them into the modeling of soybean fatty acid content detection. The established model has the following characteristics:

(1) High contrast，BP neural network simulates the way of human thinking, it judges on the basis of intuitionist ratiocination according to the essence of things, rather than some kind of pre-established pattern.

(2) Strong manipulation, providing unknown spectral data samples the trained network will output results of the analyzing, so it is simple and easy.

REFERENCES

Mroczyk W B, Michalski K M. Analyzed elementary compositions of beet tops by NIR [J]. Computers Chem. 1995, 19(3):299-301.

Bochereau L, Beurgine Petal. Classical date analysis and neural network to NIR to predict quality of the apple [J]. Journal of Agricultural Engineering Research. 1992, 51:207-216.

Fogel D B. An introduction to simulated evolu- tionary optimization [J]. IEEE Trans on Neural Networks, 1994, 5(1):3-14.

Srinivas M, Patnail LM. Adaptive Probabilities of Crossover and Mutations in Gas [J]. IEEE Trans on SMC, 1994, 24(4):656-667.

Zitzler E, Thiele L. Multi-Objective Evolutionary Algorithms: A Comparative Case Study And the Strength Pareto Approach [J]. IEEE Transactions of Evolutionary Computation, 1999, 3(4):257- 271.

Williams P C,Norris K,Gehrke CW and Bernstein K.1983. Comparison of near-infrared methods for measuring protein and moisture in wheat. Cereal Foods Worle .150:149-152

Williams P C,Stevenson S,Starkey P M and Hawtin G.1978.The application of near-infrared reflectance spectroscopy to protein-testing in pulse breeding- programmes. Journal of Science Food Agriculture.29: 285-292.

Morgan J E and Williams P C.1995.Starch damage in wheat flours: A comparison of enzymic, iodometric and near-infrared reflectance techniques. Cereal Chemistry.72 (2): 209-212.

Murray L and Williams P C.1987.Chemical principles of near-infrared technology. In Williams P C and Norris K ed. Near-infrared Technology. AACC St. Paul, MN.

Norris K H,Barnes R F,Moore J E and Shenk J S.1976.Predicting forage quality by infrared reflectance spectroscopy. Journal of Animal Science. 43: 889-897

Norris K H.1984.Multivariate analysis of raw material. In: Schmilt L W, ed. Chemistry and world food supplies. Manila: Reihold Publisher, 155-164.