Self-Organizing Map Analysis on Peanut Yield and Agronomy Characteristics

Yujian Yang, Mingchuan Ji

Self-Organizing Map Analysis on peanut yield and agronomy characteristics

Yujian Yang[1],Mingchuan Ji[2]*
1 S & T Information Engineering Technology Center of Shandong Academy of Agricultural Science, Information center of agronomy College of Shandong University
2 The institute of sustainable development of Shandong Agricultural Science Academy
Corresponding author, Address: The institute of sustainable development of Shandong Agricultural Science Academy, Number 202 Gongye North Road, Licheng District of Jinan, 250100, Shandong Province, P. R. China, Tel: +86-531-83179076, Fax:+86-531-83179821, Email:yyjtshkh@gmail.com

Abstract. The model system between peanut yield and agronomy characteristics which is nonlinear, irreversible and dissipative. The objective in the study was the peanut cultivated in the different ecological regions in Shandong province, aimed to establish the new non-nonlinear model based on Self-Organizing Maps (SOM) to improve the cultivation information of peanut growth process. In the article, applying SOM network achieved the cluster between peanut yield and agronomy characteristics about 4 variables, involved in plant height, branches, full pods and peanut yield ratio. MATLAB 7 software is used to classify 60 samplings of peanut yield and agronomy characteristics. It is concluded that the SOM network can respond the complicated information classification among each peanut yield, during the analysis, the results also showed SOM method is the most suitable for peanut yield and characteristics classification, especially analysis of clusters on basis of peanut agronomy parameters，so the study can be applied on agronomy characteristics and peanut yield of the different ecological regions in Shandong province.

Keywords: peanut ; Self-Organizing Mapping ; yield ; agronomy characteristics

1 Introduction

As the important oil and economic plants in China, peanut is widely planted in Shandong province, where is the core area of production and export of the superior peanut cultivars (Wan Shubo, et al., 2007, 2009)[1,2]. An important subject between peanut yield and all kinds of agronomy characteristics has been reported in recent years. The quantitative relationship focused in yield and its correlated traits (Zhang Haiyan, et al., 2006)[3]. The important study was mainly developed by the some statistical methods, such as correlation analysis and regression analysis, which carried on the variation of genetic factors and environmental conditions of peanuts (Wu Zhengfeng, 2008)[4].The capability of Self-Organizing Maps (SOM) to visualize high-dimensional data is well known, an important SOM subject between peanut yield and agronomy characteristics has been rarely reported at home and abroad. In recent years, the SOM is widely applied in many fields, such as model establishment, imaging application, text clustering. There is the rarely reports of clustering results considering peanut yield and agronomy characteristics with the support of SOM, so the main objective in the article is achievement of effective clusters about peanut yield. Correspondingly, the organization of the paper is as follows: a first section describes the basic SOM algorithm with a deeper insight of its capabilities. An approach, based on SOM visualization for peanut yield and agronomy characteristics is suggested in the next section for 60 samples. The last section discusses the obtained results, followed by the conclusion.

2 Material and Methodology

2.1 Materials and experimental designing

Materials and data were collected from multiple sites in 9 different ecological regions about 7 peanut varieties in Shandong province in 2007. Experimental sites with ecological regions includes Laixi city, Jiaozhou city, Changyi city, Qixia city, Haiyang city, Donggang District, Lanshan District, Dongchangfu District and Ningyang country in Shandong province. 7 peanut varieties were selected in the paper with obvious representative, they are respectively: “LuHua 11”,“LuHua 12”,“LuHua 14”,“Huayu 20”,“Weihua8”, especially,“Huayu22” ,“Huayu23”,“Huayu22” for the traditional big export-oriented peanut, “Huayu23” for the traditional small export-oriented peanut, the others is high yield peanuts. In the experimental designing, the local plant mode was selected for spring peanuts or the summer peanuts. Each site is the same planting specifications, for spring peanut, ridge is 80-85cm, wide is 50cm, the seeding approach was hill-drop, row spacing is 30cm, the hill spacing is 17cm, 9,800 plants were planted per 666.7 m2, interplanting cultivation mode of peanut was selected to be naked before 20d, 10,000 holes were dug per 666.7 m^2[5]. In the process of data measurement, 4 traits variables of peanut(plant height, branches, full pods and peanut yield ratio) was acquired involved in 60 samples. In order to achieve the SOM results, we determined the cluster numbers of peanut yield according to SOM algorithm, so we developed the preprocessing classification, there are 3 kinds of productions including 3000-4000 Kg/per ha, 4000-5000Kg/per ha and >5000 Kg/per ha. Use the SOM for clustering data correlated with the agronomy characteristics of input data (plant height, branches, full pods and peanut yield ratio), which provides a topology preserving mapping from the high dimensional space to map units. Map units, or neurons, usually form a two-dimensional lattice and thus the mapping is a mapping from high dimensional space onto a plane. In the literature [6,7], the effective characteristics of SOM method was determined by the sample distribution characteristics.

2.2 Use of the Self-Organizing Map algorithm

2.2.1 Self-Organizing Map algorithm

The Self-Organizing Map has been proven useful in many applications, it belongs to the category of competitive learning networks, so it is also called SOM, which is put forward by T. Kohonen(1981)[6]. It is a neural network that maps signals from a high-dimensional space to a one- or two-dimensional discrete lattice of neuron units, or is named after cluster algorithm on basis of neural network. SOM embodied the similarities in samples, achieved the transformation from high-dimensional data to low-dimensional data. In many perspectives, spatial structure and function of neurons of SOM is very important, the traditional model of neural network didn't consider the spatial structure characteristics, but SOM not only considering the structure, but also achievement of the similar input data responded the best matching units without supervision and prior knowledge[8,9,10].

SOM has the better status of non-linear clustering, it was also used to find correlations between the data by labeling the neurons of the SOM using the training set and finding the best-matching-units for every example. In theory, each neuron stores a weight. The map preserves topological relationships between inputs in a way that neighboring inputs in the input space are mapped to neighboring neurons in the map space. And by grouping units that respond to similar stimuli together. Nerve cells, neurons, in the cortex of the brain seem to be clustered by their function. During the computation, by defining the Hebbian learning rule is to determine the learning rate and update the relationship for best-matching units on the map. In essence the learning rule of the SOM defines the model as a collection of competitive units that are related through the neighborhood function. In practice, the units are placed on a regular low dimensional grid and the neighborhood is defined as a monotonically decreasing function on the distance of the units on the map lattice, thus creating a latent space, which has the dimension of the map grid and flexibility determined by the neighborhood function, and embedding when the dimension of the map grid matches the dimension of the input data manifold. In general, gaussian function is selected in the neighborhood function[11].

SOM, based on unsupervised learning, or high-dimensional observations projected to the two-dimensional coordinate system, which means that no human intervention is needed during the learning and that little needs to be known about the characteristics of the input data. In a word, Self-Organizing Maps is a categorization method, a neural network technique and the unsupervised characteristics.

The main computing steps of SOM, including construction of data sets, data preprocessing, initialization and training of input data and visualization and analysis of the correlated results. In the elaborate SOM computing process, important parameters should be also considered as the significant contents, such as Unified distance matrix (U-matrix), importance degree in the trained self-organizing map(D). Especially, U-matrix, the parameter representation of the Self-Organizing Map visualizes the distances between the neurons, the distance between the adjacent neuons is calculated and presented with different colorings between the adjacent nodes, for each node in the map, compute the average of the distances between its weight vector and those of its immediate neighbors. In the results, the different colors between the neurons presents the different distance significance and responds the cluster differences, average distance is a measure of a node's similarity between it and its neighbors, this can be a helpful presentation when one tries to find clusters in the input data without having any a priori information about the clusters. For the parameter D, it responds the important degree of the agronomy characteristics impact on peanut yield in the article[12,13].

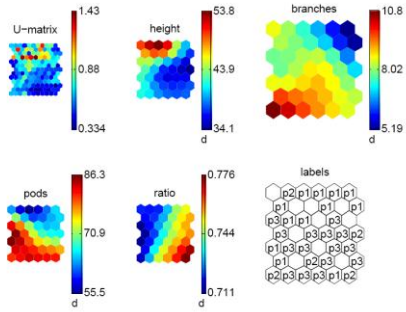
2.2.2 Data processing and data analysis

First, the data has to be brought into Matlab using construction of data, second, we used the function som_normalize for data preprocessing data to perform a linear scaling to unit variance. The function som_make is the basic function to use when creating and training a SOM, it is a convenient tool that combines the data of creating, initialization and training a SOM. After the data set is ready, the data set is loaded into Matlab7 and normalized, the variance normalization is used, a SOM is trained. Since the data set had labels, the map is also labeled using som_autolabel. There are a variety of methods to visualize the SOM, the basic tool is the function som_show, it can be used to show the U-matrix and the component planes of the SOM. After this, the SOM is visualized using som_show[14,15].

Here is the usage of the Matlab Toolbox to make and visualize a SOM of a data set about peanut yield and agronomy characteristics. This data set consists of four measurements from 60 samples: the first classification of peanut yield has 18 samplings, the second classification of peanut yield has 13 samplings and the third classification of peanut yield has 29 samplings. The data is in an ASCII file, the first line contains the names of the variables, the corresponding sequences, including plant height characteristics, branches characteristics, full pods characteristics, peanut yield ratio characteristics and peanut yield classification. Each of the following lines gives one data sample beginning with numerical variables and followed by labels[16].

3 Results and Analysis

We developed the cluster analysis based Matlab7 SOM Tool. the PCA-projection of both data and the map grid was required by the program computerization, in general, the projection also presented the four variables and the subspecies information from the SOM. Projection of the sample data set to the subspace spanned by its three eigenvectors with greatest eigenvalues. The SOM grid has been projected to the same subspace, and visualizes all four variables of the SOM plus the subspecies information using coordinates, neighboring map units are connected with lines, labels associated with map units are also shown.



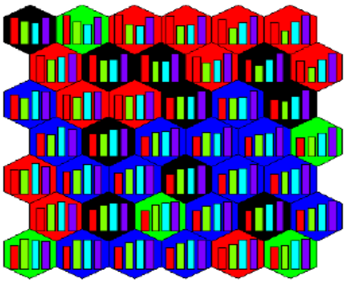


Figure1 Visualization of the SOM of peanut yield and agronomy parameters data
Figure 2 SOM ordination diagram of characteristics

U-Matrix visualization provides a simple way to visualize cluster boundaries on the map, the U-matrix visualizes distances between neighboring map units, and thus shows the cluster structure of the map, high values of the U-matrix indicate a cluster border, uniform areas of low values indicate clusters themselves. For each node in the map, compute the average of the distances between its weight vector and those of its immediate neighbors, average distance is a measure of a node's similarity between it and its neighbors. Each component plane shows the values of one variable in each map unit. On top of these visualizations, additional information can be shown: labels, data histograms and trajectories.

As illustrated in Figure 1, Gradient distribution of peanut characteristics in the trained SOM ordination diagram. U-matrix on top left, then component planes(including plant height characteristics; branches characteristics; full pods characteristics; peanut yield ratio characteristics), and map unit labels on bottom right. The six figures are linked by positions, in each figure, the hexagon in a certain position corresponds to the same map unit, additional hexagons exist between all pairs of neighboring map units. The map unit in top right corner has low values for plant height, branches and full pods, and relatively high value for peanut yield ratio. The label associated with the map unit is "p1" and from U-matrix it can be seen that the unit is very close to its neighbors. From the U-matrix it is easy to see that the top two rows of the SOM form a very clear cluster. By looking at the labels, it is immediately seen that this corresponds to the peanut yield (p1) subspecies, the two other subspecies peanut yield (p2) and peanut yield (p3) form the other cluster. The U-matrix shows no clear separation between them, but from the labels it seems that they correspond to two different parts of the cluster. From the component planes it can be seen that the branches characteristics and full pods characteristics are very closely related to each other in the certain degree.

Figure 2 visualizes all four variables of the SOM plus the subspecies information, 4 variables barcharts (including plant height characteristics ; branches characteristics; full pods characteristics; peanut yield ratio characteristics) in the topography maps. The distribution of 3 types peanut yield in the topography maps. In every topography map, the weight of each variable is shown in the figure. The four variables shown with the different barcharts in each map unit and in the background color indicates the subspecies.

4 Conclusions and Discussion

In this paper, SOM is used to classify 60 samplings of peanut data of the different ecological regions in Shandong province, the kind of yield classification method promotes the comparative analysis for the different ecological regions in Shandong province for the yield prospective. The results showed that SOM is an excellent tool in the visualization of high dimensional data about peanut yield and agronomy characteristics and as such it is most suitable for peanut traits and yield classification, especially analysis of clusters on basis of peanut agronomy parameters.

It is concluded that the SOM network can respond the complicated information among each peanut yield. The effect of classification is good and SOM considered the complicated characteristics of peanut agronomy parameters. And it can be applied on peanut characteristics and yield. Although SOM cluster applied in peanut yield and agronomy characteristics is a preliminary trail in the study, the SOM has the obvious characteristics from the high dimensional data to low dimensional data for peanut growth process, with the development of spatial analysis technology, combining with GIS data, the spatial evaluation on peanut agronomy characteristics and yield should be explored on the precision scale to clarify the spatial variability and spatio-temporal characteristics of complicated parameters under soil-peanut system in the next step.

References

1. Wan Shubo. Peanut Quality[M]. Beijing:China Agricultural Scientific Press, 2007
2. Wan Shubo, et al. The high Quality of synergism Cultivation Theory and Technology of Peanut[M]. Beijing: China Agricultural Scientific Press, 2009
3. Zhang Haiyan, Jiao Bichan, Li Guiquan. Analysis on the Relationship between Yield and Correlated Quantitative Character of Soybean. Journal of Shanxi ricultural Science, 2006, 34(2): 27-29(in Chinese)
4. Wu Zhengfeng, Wang Caibin, Zheng Yaping,et al. Analysis of characteristics and stability of peanut yield in different ecological regions of Shandong Province Chinese Journal of Eco-Agriculture, 2008, 16(6):1439-1443(in Chinese)
5. Yang Yujian, Wan Shubo.The grey correlation degree analysis on yield of peanut and traits in warm temperature regions. 2010, computer and computing technologies in agriculture Ⅱ,volume 24, Daoliang li and Simon X.Yang pp161-167
6. Kohonen, T. “The self-organizing map,” Proc. IEEE, (1990)78 (9),1480-1497
7. Juha Vesanto and Esa Alhoniemi. Clustering of the Self-Organizing Map, In IEEE Transactions on Neural Networks, Volume 11, Number 3, pp. 586−600, 2000.
8. Zhang Qindi, etc. Self-organizing feature map classification and ordination of Larix principisrupprechtii forest in Pangquangou Nature Reserve
9. Zhang Yancheng, etc. Soil Classification Based on Self□Organizing Feature Mapping Neural Networks
10. D. Moshou, I. Gravalos , D. Kateris, N. Sawalhi3 and S. Loutridis. Condition monitoring in centrifugal irrigation pumps with self-organizing feature visualization. EFITA/WCCA ,2011,116-124
11. Kohonen T., Self-Organizing Maps, volume 30 of Springer Series in Information Sciences. Springer, Berlin, Heidelberg, 1995.

12. Juha Vesanto, John Himberg, Esa Alhoniemi and Juha Parhankangas. SOM Toolbox for Matlab5. 2000
13. SURI Guga, ZHANG Jin-Tun1, TIAN Shi-Guang, ZHANG Qin-Di, ZHANG Bin, CHENG Jia-Jia, and LIU Su-Jun. Application of self-organizing map to quantitative analysis of mountain meadow in the Songshan Nature Reserve of Beijing, China. Chinese Journal of Plant Ecology, 2010, 34 (7): 811–818
14. Juha Vesanto,Johan Himberg, Esa Alhoniemi and Juha Parhankangas. Self-organizing map in Matlab: the SOM Toolbox,
15. Vesanto J., Alhoniemi E., Himberg J., Kiviluoto K., Parviainen J. Self-Organizing Map for Data Mining in MATLAB: the SOM Toolbox. Simulation News Europe 1999;25:54.
16. Vesanto J., Himberg J., Alhoniemi E., Parhankangas J. Self-organizing map in Matlab: the SOM Toolbox. In Proceedings of the Matlab DSP Conference 1999, pages 35-40, Espoo, Finland, November 1999