

# LAND-USE ANALYSIS PERFORMANCE IN PRECISION AGRICULTURE: EVOLUTION ALGORITHM OF DATA MINING

Deng Yuejin<sup>a</sup> Cai Jingjing<sup>a</sup> Yan Shijiang<sup>b</sup>

<sup>a</sup> State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, P.R.C., 430079-dengyuejin@geostar.com.cn, jinnycai@gmail.com

<sup>b</sup> Key laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, Jiangsu, P.R.C., 210046- anew2@hotmail.com

**KEY WORDS:** Image Interpretation, Percentage Vegetation Cover, Image Understanding, Data Mining, Precision Agriculture

## ABSTRACT:

With the rapid development of economy, Chinese rural land-use is experiencing tremendous change, and the constant retreat of the land-use area causes a huge reaction recently. In this paper, we take the change of crops area and the reasons why changes triggered as the subject of our study. We use the technology of Remote Sensing (RS) to gain data of crops area of Ningbo in the period from 1986 to 2005, when China was conducting the reform and opening up and reconstruction of agriculture, which caused a lot of change in crops area. We adopt Data Mining (DM) to gain knowledge and analyze this phenomenon, as well as predicate the tendency of the crops area in the future. In this paper, the evolution algorithm of data mining is used to model the changes and the tendency of crops coverage as time goes by. We take into account four factors: State Management (SM), Local Management (LM), Market (MK) and Weather (WH). In our research, we found firstly the factor of “SM” is a dominant factor but its influence will decline in the future, secondly the factor of “MK” is a rising and important factor, thirdly the factor of “LM” is in the rising stage but it wouldn’t have more share than the factor of “MK”, and finally “WH” factor which accounts for a large proportion in the past and now is declining to the lightest one. According to our experiment, it is predicated that Chinese crops area downward trend will decelerate, on the contrary it will have a continuous rising until it gets to the optimization conditions and then realize the harmonious use of the land.

## 1. INTRODUCTION

China is a large country in agriculture. Agriculture is of benefit to society, ecology, but with low economic returns. It is also a high-risky and weak industry which can be easily affected by the natural environment. Since the reform and opening up, especially from 1990, with the rapid development of economy and society and the strengthening of ecological construction, the reduction of the area of arable land in China accelerated over the past 20 years and is becoming a prominent problem. Not only does it affect seriously the food security, but also it endangers our social stability and harmony economic development. Ningbo, which is in ZheJiang Province and is famous with its beautiful, green and healthy scenery, is experiencing great changes of the arable land from 1986 to 2005. In order to find the discipline of these changes and make suggestions for the local government, we conducted a study on precision agriculture which is experiencing a vigorous development over the world. One of the purpose of precision agriculture is to make scientific management of the farming work. According to the data we got from the website of Ningbo Municipal Statistics Bureau Survey Office of the National Bureau of Statistics in Ningbo, we use evolution algorithms of data mining (DM) to analysis the relationship and interoperability between variable factors which have impacts on the change, to discover the movement tendency of the land area, to make a forecast report, and to provide a optimization guide for policy makers and economic agents. The research of precision agriculture needs the technology of remote sensing to gain extensive historical data, and the method of data mining.

The approach we used in this paper is that: firstly, we determine the interested area of food and other economic crops and get data from the website of Ningbo Municipal Statistics Bureau Survey Office of the National Bureau of Statistics in Ningbo.

Most of these data is obtained by remote sensing data acquisition technology and have verified accuracy. Secondly, we need to clean the data using the basic method of DM. Thirdly, we introduce Evolution Model, e.g. logistic model and an improved one. The logistic model is for one isolated species, this continuous version of this model was originally introduced by Verhulst [Ricardo Lopez-Ruiz, Daniele Fournier-Prunaret] in the nineteenth century as a counterpart to the Malthusian theories of human overpopulation. And the improved one is a discrete logistic model for the competition between two species, which displays coexistence, oscillations, chaos and fractal behavior of basins. In this paper, we will make experiments using both models.

## 2. BACKGROUND

After 20 years of reform and opening up, Chinese agriculture has entered into a new stage, the integrative agricultural production capacity has basically met the needs of food at this stage. At the same time, the distribution of agricultural products is not satisfied, e.g. the yield of grain, cotton is little because more and more lands has been used to grow other economic agricultural products, for example fruit, vegetable, which results in some economic agricultural products is overmuch and some non-economic products lack. The products can not entirely adapt to the change of market, so the farmers’ income doesn’t increase correspondingly than expected. After joining in the WTO, the impact of economic globalization on Chinese agriculture is becoming more obvious, and the solution of new problems lies in the readjustment of agricultural structure, so a new performance of agricultural restructuring launches in China, including our study area of Ningbo.

From the late 1990s, with the loss of arable land in China the adjustment of agricultural structure starts. According to the land-use change survey data of Ministry of Land and Resources, the adjustment of agricultural structure results in that the quantity of arable land experienced three stages : "an increase - a drastic reduction - slow to reduce" from 1997 and 2005. In our research and investigation, we consider four factors of the significant reduction of crops areas: market factors, state management factors, local management factors, and natural ecological factors (e.g. weather). In the following context we will explore the way how these factors impact on the reduction of the percentage of food crops areas and the increase of the percentage of vegetables and other economic crops areas.

Data mining technologies and algorithms are wide used in developed country, but is not much mature in China. Studies show that these technologies and algorithms are helpful for the management of farmland and forecast the output of crops. Some algorithms or models are briefly listed here:

- decision tree – CART (T.Waheed,R.B.Bonnell\*, et al.),
- the direct use of spatial data mining tools-SRSDM (Sheng-sheng Wang.,et al.),
- Bayesian network operator in the evolution (Kyung-Joong Kim.,et al),
- a classification based on the knowledge of genetics for XCS,
- one-dimensional Logistic model (Ricardo Lopez-Ruiz, Daniele Fournier-Prunaret),
- serial model of mining algorithms,
- artificial neural networks.

All these methods have their own advantages and disadvantages. In our study in Ningbo, we choose the Logistic model and an improved one to explore the principle in the complex of Chinese economy. Now agriculture, which is as the basic industry, is breaking away from the traditional agricultural model with the support from modern science and technology, is facing globalization and the market economy after the reform and opening up in China, and is facing with the global warming. We try to answer the following questions: what kind of changes will happen to agricultural structure, what are the main factors that cause the changes, what kind of impact of the reduction of land area will have on the change of Chinese economic construction, national food security and income of the people.

### 3. MATERIALS AND METHODS

#### 3.1 Experiment data

The experiment data we used are taken from the website of Ningbo municipal Statistics Bureau Survey Office. We selected the "Sown Areas and Yield of Major Farm Crops over the Years" in the agriculture statistical annual on the website, which contains the area of grain, cotton, oil plants and vegetables, which are classified into two species, the grain, cotton and oil plants are classified into food crops, the vegetables are classified into economic crops. All these data are from 1986 to 2005, when China is experiencing the period of "7th Five-Year Plan" to "10th Five-Year Plan". We figure out the percentage of two species area which is expressed in fig1. We will analysis the data to find rules in these data, and make a forecast report of it.

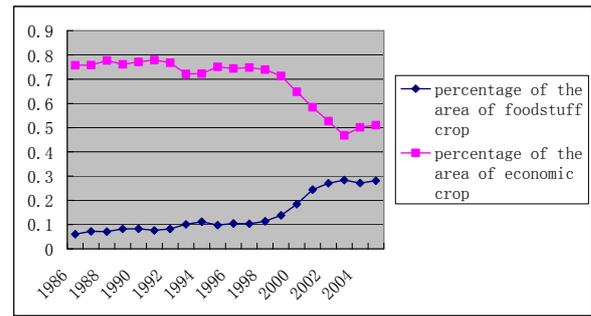


Figure 1. The tendency of the percentage of the crops area

#### 3.1.1 Factors

The factors are taken from many papers and websites such as <http://www.agri.gov.cn/> and <http://www.nongye.ningbo.gov.cn/>, and we classified these factors into four parts: state management (SM), local management (LM), whether (WH) and market (MK). These factors have their own specific characteristics which are described here:

1. SM: China is of large population, scarcity of arable land reserve resources and the deep-rooted "three rural issues", which request the central government to constitute different agriculture structural adjustment policies focus on different points in different periods. The SM factor plays a fundamental role in the change of arable land area.
2. LM: Local government management is mainly based on national policies, which develop the local policies to adjust to the local conditions and to accomplish the assigned task. It plays a direct role in the change of arable land area.
3. WH: Under the co-effect of the topography conditions together with the other natural or man-made conditions, the quality of the arable land is poor. While the land's quality of Ningbo is relatively good, but the main climate disasters due to typhoon and flood brought by the coastal climate will cause large-scale plant diseases and insect pests leads massive reduction of output will influence the positivity of farmers, and have impact on the change of arable land area indirectly.
4. MK: For a long time, Chinese planting industry is mainly on grain production which is a kind of land-intensive industry and has difficulty in getting rapid growth in economic benefits industry. Farmers have the profit-driven trend on the land management, however, under the situation of market economy, the usage of arable land resources is similar to other economic activities, the factors of production such as the capital, labor and land will flow to the high-yield industry as long as there is any interest difference. The MK factor is a very important factor that has impact on the change of arable land area.

We list the discrete values of the four factors in some years in table 1, where the value "M" represents for Middle, "H" represents for High, "L" represents for Low.

	1985	1990	1995	2000	2005
SM	M	M	M	H	H
LM	L	L	M	H	H
WT	M	L	M	H	L
MK	L	L	M	M	H

Table 1. Discrete values of our factors

### 3.1.2 Tendency of the experiment data

From 1986 to 1998, the change of the percentage of both two kinds of crops area is little, because the agricultural restructuring has not begun to work in this region yet, the most important factor is weather; from 1999 to 2003, the agricultural restructuring works, the percentage of both crops has undergone enormous changes, the trends of the two percentages are going in the opposite direction; And in the period from 2003 to 2005, they emerged into a stable state again, the declining trend of food crops in 2003 rebound after it met the minimum, while the rising trend of the economic crops area also slowed, it re-declined when it reached its highest point. We give the expression of the tendency of each crop as following:

1. From 1986 to 1998, the tendency of the percentage of food crops area experienced an age that the traditional Chinese cultivated method is moved to the modern scientific farming method, which helped us keeping in the status of self-sufficiency. From 1999 to 2003, under the reconstruction of agriculture, the tendency of the percentage of food crops area is basically declined as a straight line reaches to 39.218%, the minimum in 2003. In this period, the food supply is reducing but still can keep self-sufficiency because of the successful using of modern farming methods. And in 2004 and 2005, the percentage of food crops area is keeping stability, and the yield of the food crops is maintained in an increasing situation, the integrated capacity of food crops production is close to the highest level in history. In China, the large population faces with the reality that the quick declining of arable land area is a very cruel issue. So far, the arable land area together with the percentage of food crops area is down to the minimum value which is never happened in history.
2. Since 1986, the economic crops area kept a stable situation until 1997. From 1998 to 2003, the quick increase of the percentage of area of vegetables was happened when the percentage of food crops area decreased. In this period, the country carried out the agriculture reconstruction. In the next few years the value maintained a stable condition again.

After a complete analysis of the tendency of the percentage of the two crops area, we give the quantifiable criteria of the four factors. The discrete values of the attributes are high, middle and low. We scored each factor in each year according to the following criterion in a window width of 7 from 2 to 8, in order to facilitate calculation.

$$\begin{cases} L : 2 \leq l \leq 4; \\ M : 5 \leq l \leq 6; \\ H : 7 \leq l \leq 8. \end{cases}$$

### 3.2 Methods

Here we proposed an appropriate optimization approach to study the change tendency of the percentage of the crops area. Firstly we use Logistic to model the tendency, which is designed as a multi-target function, whose factors are classified into four as described above.

The logistic model for one isolated species can be represented by formula (1), where  $X_n$  represents for the population of an isolated species after  $n$  generations. Here we use the value of the percentage of one crop as  $X_n$  and this variable is supposed in the range  $0 < x_n < 1$ .

$$X_{n+1} = \mu X_n(1 - X_n) \tag{1}$$

where  $X_n$  = the value of the percentage of one crop  
 $\mu$  = growth rate

In order to assure  $0 < x_n < 1$ ,  $\mu$  should in the range  $0 < \mu < 4$ . The continuous version of this model was originally introduced by Verhulst in the nineteenth century as a counterpart to the Malthusian theories of human over population.

Secondly we introduce the multi-target logistic competition model. Suppose two species ( $x_n, y_n$ ) evolved according to a logistic-type dynamics. If they share the same ecosystem and they interact by each other, we can approach their evolutions by the coupled system by formula (2):

$$\begin{aligned} x_{n+1} &= \mu_x(y_n)x_n(1 - x_n), \\ y_{n+1} &= \mu_y(x_n)y_n(1 - y_n). \end{aligned} \tag{2}$$

where  $x_n$  = value of the percentage of economic crop area  
 $y_n$  = value of the percentage of food crop area  
 $\mu_x$  = the growth rate of the percentage of economic crop area  
 $\mu_y$  = the growth rate of the percentage of food crop area

The interaction between both species causes the growth rate  $\mu(z)$  varies with time. Thus,  $\mu(z)$  depends on the population size of the others. The simplest choice for this growth rate can be a linear increasing  $\mu_1$  or decreasing  $\mu_2$  function that expands the parameter interval in which the logistic map shows some activity, which is  $1 \leq \mu \leq 4$ . Hence,

$$\begin{aligned} \mu_1(z) &= \lambda_1(3z + 1), \\ \mu_2(z) &= \lambda_2(-3z + 4). \end{aligned} \tag{3}$$

These systems could be considered for further studies as the necessary bricks to build up more complex networks among interactive species. In particular, the model for the competition is

$$\begin{aligned} x_{n+1} &= \lambda(-3y_n + 4)x_n(1 - x_n), \\ y_{n+1} &= \lambda(-3x_n + 4)y_n(1 - y_n), \end{aligned} \tag{4}$$

where  $\lambda$  = strength of the mutual competitive interaction

When one of the species is null the typical logistic behavior is found for the other species. We divided these all data into 3 parts. The first part are from 86 to 99, and the second part is from 99-03, and the last part of the data is from 03 to 05. The main topic of this paper is to find whether we can find a year when the percentage of the crops area becomes stable. In the following, we will give the performance of each model used in these three phases.

1. The first phase is from 86~98. As it shown in fig 1, the percentage of crops is in a stable situation. We get  $\lambda$  of these two species shown in fig 2 according to the model (4). We can expect that if  $\lambda$  is invariable the tendency of the species will be also invariable, and vice versa. On the other hand, the value of  $\lambda$  is different when the initial values of the two variables are different.

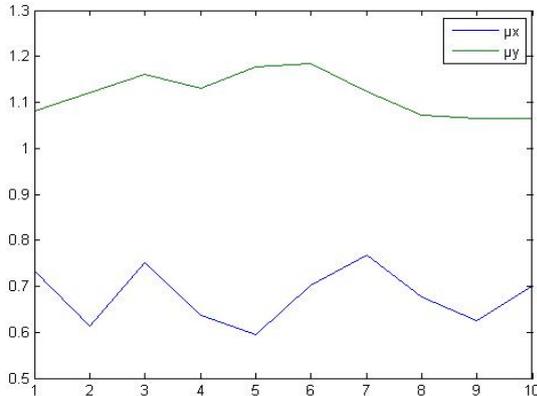


Figure 2. The value of  $\lambda$  of two species calculate according to model (4) ( $\mu_x$  means the  $\lambda$  of economic crops, the  $\mu_y$  means the  $\lambda$  of food crops)

From the analysis above, only when the value of  $\lambda$  of the two species is the same that the data can be modeled as an competitive condition. In this phase, we can only use model (1), and get the result that, when the growth rate is modified, the dynamical behavior of the logistic model (1) is following:

- ①  $3.93 < \lambda < 4.47$ , when the initial value of the percentage of food crops area is around 75%;
- ②  $1.00 < \lambda < 1.30$ , when the initial value of the percentage of economic crops area is around 10%.

2. In the period from 1999 to 2002, for the tendency of the two species, we experiment according to model (2), and the result is shown in fig 3. And the result approves our assumption, i.e. the value of  $\lambda$  of the two species approximate each other.

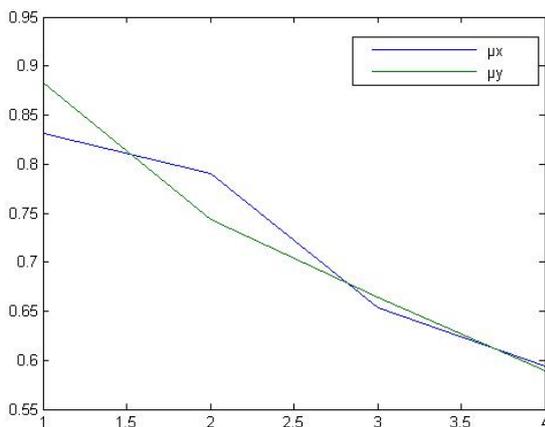


Figure 3. The value of  $\lambda$  of two species from 1999 to 2002

According to the result of the experiment, we got the value of  $\lambda$  of several years, then we founded equation-group bases on function (5). We have four equations and four variables  $l_1, l_2, l_3$  and  $l_4$ , and then we can calculate the final results of the four parameters,  $\alpha, \beta, \phi$  and  $\delta$ , which is shown in (6).

$$\mu(l_1, l_2, l_3, l_4) = \alpha \cdot l_1 + \beta \cdot l_2 + \phi \cdot l_3 + \delta \cdot l_4 \quad (5)$$

where  $\mu$  = the function to calculate the value of  $\lambda$   
 $l$  = value of four variables

$$\begin{cases} \alpha = 0.014; \\ \beta = 0.096; \\ \phi = 0.052; \\ \delta = -0.057. \end{cases} \quad (6)$$

When we get the value of parameters of the function, we can get  $\lambda$  of each year. In this period we can give the detail report of the importance of each factor according to the absolute value of these four parameters which have impact on four factors. The most important factor is LM, next is MK and third is WH, and the least is SM, which is also the same as our assumption. And the factor MK is anti-related with  $\lambda$ , which means the bigger the value of market factor is, the smaller the  $\lambda$  will be.

3. Then we can make the prediction. From the result of the experiment in the second phase, we can get  $\lambda$  of each year according to model (2). And in this period, the value of  $\lambda$  is different again, and the initial value of the two species is also different, then we can experiment according to model (1) to get the exactly year when the two species can have a stable condition, whose the results is shown in fig. 4.

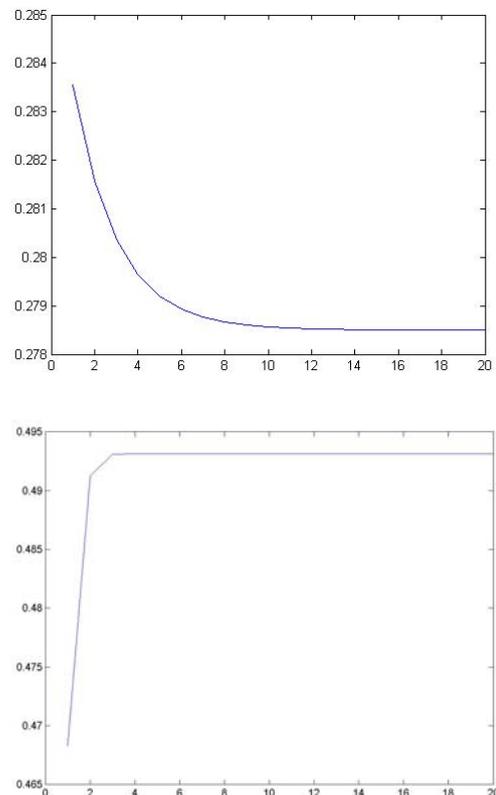


Figure 4. Forecast of the percentage of economic crops area (above) and food crops area (below)

As shown in fig. 4, we can predicate that if we keep the value of four parameters fixed, which means keeping the factors invariability, the  $\mu$  of the data will be in a stable situation in the future years, but the value of  $\mu$  is different between the percentage of economic crops area and that one of food crops area. It is predicated that they will be in a stable situation in 4~10 years. And the value of percentage of food crops area will be at around 49.3% when the value  $\mu$  is 1.973, and the economic crops area will be at around 27.85% when the value of  $\mu$  is 1.386.

#### 4. DISCUSSIONS

The experiment above is preliminary for the prediction of crops area. We try to use multi-target Logistical evolutionary model to analyze the tendency of the percentage of crops area of different types. The result of the experiment shows that the percentage of crops area will be in a stable situation in 4~10 years. However, we should emphasize some problems, such as how to fix the value of  $\mu$  in the stable period, how to determine the value of the factors, how to consider the threshold of factors according to  $\mu$ , and get the range of  $\mu$ . etc. Excluding the four factors above, some other factors such as the psychology of the Chinese farmers should be taken into account. At the same time, we have to consider much more uncertain things, such as LM, which is the most important factor to the change and MK which is also the factor that is becoming more and more important. The change of Chinese economic construction, national food security and income of people has an indirect impact on the change of both species through the factors such as MK, SM, etc.

According to the records, the Chinese experts prospected that, after the rural reform policy from 1978, the crop areas in China reduced by more than 6.67 million hectares, though the crop yield is rising instead of dropping, which make us not to worry about the food crisis now. But now we have to face the new issues and crisis of food security. The farmland reduction is a key problem. If only we make the study objectively, with cautious attitude and continuous observation that can we get the rule of the tendency of the percentage of crops area.

#### 5. CONCLUSIONS

As discussed before, we are in the exploring stage of precious agriculture by using DM, and we have a long way to go to realize the scientific management and fully automated operation of agriculture. The research of land-use is taken as a complicated task, according to their characteristics such as spatial self-organization, long-term nonlinear driven behavior, etc. The first thing we have to do is to find the rule of the data, and then make experiments to fit the real world approximately. We conclude the difficulties in the following:

1. The uncertainty of agricultural data and geographic data are obvious and inevitable. The data we get is always disorder, redundancy and inconsistency, and the property of agricultural is empirical. , so we should make a guidance for data mining, derive some meaningful information hidden behind the record data, solve the problem of "rich in data, lack in knowledge" in field of agriculture.
2. The ability to solve practical problems is poor. The real world is constantly changing; the factors we considered in the paper are always changing. Our model is designed to adopt the dynamic situation, to acquire knowledge continuously which will be used in the following knowledge discovery phase. This process will be constantly repeated

until the predication is done. However, the results should be verified in the next stage.

#### REFERENCES

- Wang,S.L.,et al., 2005, Attribute Uncertainty in GIS Data, Springer, Berlin, Vol.3614, pp. 614-623.
- Kyung-Joong Kim.,et al, 2005, Robust Inference of Bayesian Networks Using Speciated Evolution and Ensemble, Spinger,Berlin, Vol.3488, pp. 92-101.
- Sheng-sheng Wang., et al.,2006, Spatial Reasoning Based Spatial Data Mining for Precision Agriculture, Spinger, Berlin, Vol.3482, pp. 506-510.
- SHA Zong-yao. BIAN Fu-ling, 2003, "3S" Application in Agriculture and Precision Farming, Bulletin of Surveying and Mapping,6, pp. 29-31.
- LI Bao-qian., et al., 2004, Research advancement of precision agriculture, Journal of Henan Agriculture University, Vol.38-4,12, pp. 477-481.
- T.Waheed,R.B.Bonnell\*, et al., 2006, Measuring performance in precision agriculture: CART -- A decision tree approach, Agricultural Water Management, 84, pp. 173-185.
- M Gandah, A.Stein.,et al., 2000, Dynamics of spatial variability of millet growth and yields at three sites in Niger, West growth and yield at three sites in Niger, West Africa and implications for precision agriculture research, Agricultural Systems,PII:S0308-521X(99)00076-1, 63, pp. 123-140 .
- MENG Jun., 2007, Application of data mining on agricultural production decision support system, agricultural network information, 1, pp. 63-64.
- Fei Nao., Yuye.zhu., 2007, The application of Data Mining technology in agricultural, Journal of Anhui, Sci, 35(13), pp. 4053,4082.
- HU Wei-qun, ZHU Li-li, ZHENG Ke-feng, YE Shao-ting, ZHU Xu-bin, 2006, Study on the Mining of the Data with the Attribute of the Agricultural Resources., SCI- TECH INFORMATION DEVELOPMENT & ECONOMY, 16(15), pp. 207-208.
- Long,Tengfang., 2005, The application of Data Mining in Agriculture, micro-computer information, 21(8), pp. 42-43,27.
- Xing Pingping., Shi Pengfei., Xiong Fanlun, 2001. Efficient Application of Data Mining in Agriculture Databases, computer engineering and application,2, pp. 4-6.
- Baisen Zhang, Ian Valentine, Peter D. Kemp, 2005, A decision tree approach modelling functional group abundance in a pasture ecosystem. Agriculture, Ecosystems and Environment, 110, pp. 279-288.
- Hai H.Dam,Chris Lokan, Hussein A. Abbass,Evolutionary Online Data Mining: An Investigation in a Dynamic Environment, chap7.
- Wang Zhenyu, Xiong Fanlun, Hang Xiaoshu. 2002, A New Algorithm for Mining Sequential Patterns and Application in Agriculture, 622-628 in Proceedings of the World Congress of Computers in Agriculture and Natural

Resources, eds. F.S. Zazueta and J. Xin. Publication data March 13, ASAE Publication Number 701P0301.

J. Liu, C.E. Goering, L. Tian, 2001, A Neural Network for Setting Target Corn Yields, American Society of Agricultural Engineers, Vol. 44(3), pp. 705-713.

Haykin, S. 1994. Neural Networks: A Comprehensive Foundation. New York, N.Y.: Macmillan College Publishing.

A. Ribeiro, B. Diaz, M.C. Garcia-Alegre., 2003, Extracting Fuzzy Rules to Describe Weed Infestations in Terms of Soil Factors. IEEE International Conference on Fuzzy Systems, pp. 1032-1037.

Dilip Datta, Kalyanmoy Deb et al., 2005, Multi-Objective Evolutionary Algorithm for Land-Use Management Problem. KanGAL Report Number, 005, pp. :1-24.

<http://www.nbstats.gov.cn/tjnj/2004njbg.htm>

<http://www.nongye.ningbo.gov.cn/infoview/infoshow.jsp?iid=4000>

Ricardo Lopez-Ruiz, Daniele Fournier-Prunaret, 2008, Periodic and chaotic events in a discrete model of logistic type for the competitive interaction of two species. CHAOS SOLITONS & FRACTALS, 01, 015

LIU Yansui, LU Dadao., 2003, The basic Trend and Regional Effect of Agricultural Structure Adjustment in China. ACTA GEOGRAPHICA SINICA, 58(03), pp. 381-389.

LIU Xinwei, 2007, Farmland protection in the period of agriculture reconstruction, Land Resource Information, 11, pp. 44-49.

#### ACKNOWLEDGEMENT

The authors would like to give their thanks to the people who made great help for this paper. Dong Lingyan and Fu Dongjie gave their help for modeling in Matlab. Zhang Yuansheng and Hui Guangyu gave their advices on the modification of this paper.