Prediction Model of Household Planting based on Case-based Reasoning based on Cellular Automata

Bing Jiang[1], Hongbing Luo[2], Zhengyi Ji[1]

[1]School of Business, Sichuan Agricultural University, Chengdu, 611830, China

[2]School of Civil Engineering, Sichuan Agricultural University, Chengdu, 611830, China

Keywords: Case based reasoning; cellular automata; prediction; farmer's planting

Abstract: A novel model of farmers planting prediction has been applied to solve the existing problems of agricultural industry planning by using the combination between the case reasoning technology and cellular automaton algorithm for the verification analysis based on investigated data. Results show that this model has higher prediction accuracy of the local planting industry development, and has well predicated results compared with the investigated data. The average ratio of the predicated farming households and the actual farming households was 87%, and the average accuracy of predicated farming households and the actual farming households was 81.5%. Further, with the development, the larger the number of local farming households is, the higher the prediction accuracy is. It concludes that this model has a strong application value, and has a promising application to provide a strong support for the governmental policies of local agricultural industry development according to local market changes.

1. Introduction

The above study is the research of the macro planting characteristics to local farmers, The most of the studies are the perspective of government, which formulate corresponding policies to guide farmers planting, And the conclusions of the study have certain local characteristics, the above research mainly from the government and the farmers' perspective, The research is still in the initial stage between farmers and farmers for crop farming influence, there are few scholars to study in this field. In this paper, a case based reasoning cellular automata model is introduced. Cellular (CA, cellular automata) is a dynamic system which is discrete in time, space and state. A large number of cellular systems follow the same rules, And the local effect make cellular synchronous update, The system taking the dynamic evolution, CA structure mode of “bottom-up” fully reflects the individual behavior of complex systems generated in the overall pattern of order, the CA has ability to simulate complex systems that it changing as Time variation, it have a self evolution characteristic. Its application is also gradually expanded to many fields, CA contains three parts, which are the cell and transformation rules, the cell has “state” attribute, it as shown in figure 1:

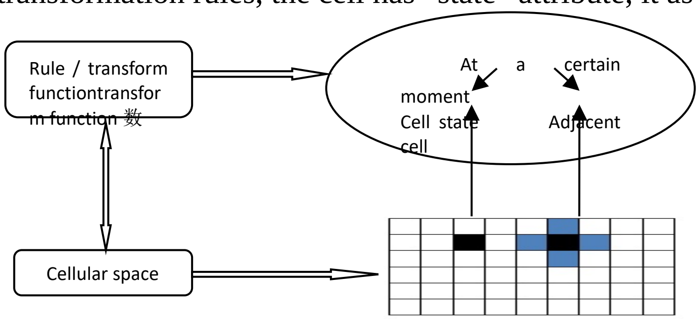

Figure 1. the composition of cellular automata

In general, It is key that make the rules of transformation to build cellular automata. Many scholars use artificial intelligence algorithms to formulate the transformation rules, but they often fall into the local optimum. This paper uses case based reasoning (CBR) to specify the conversion rules, CBR algorithm is a solution to reuse previous successful cases, CBR contains four major

  DOI: 10.25236/isafb.2019.010

parts, they are case retrieval, case revision, case reuse and case storage. It is important that make success rate high for case retrieval, it is necessary to ensure that there are enough cases in the case base, and the setting of the retrieval threshold is also very important.

2. CBR- cellular automaton model

The transformation model function is established, the weight of each cell transformation factor is considered, and the similarity values by calculation are converted to the [0,1] interval. Because the factor are so much for decision farmers planting crop, There are not an absolute decision factor for the farmers planting, so the model can only come to a certain probability of farmers planting crops,The threshold is set to determine the transition for the farmers planting condition,Such as the state changed from 0 to 1.The transition probability of PN, which is determined by the characteristic attribute factor, A farmer is a single cell, the transformation probability can be calculated according to the following formula:

$$PN = \frac{\sum_{i=1}^{m_1}(1 - \frac{1}{1+e^{1/d_i}})}{m} \qquad (1)$$

The d_i is Euclidean distance,which represent a cell with state not changing and a cell with state changed in characteristic attribute spaces. We can seen from the formula, the smaller the d_i value and the greater the PN value, It also means that the similarity is height between the cell and the m_1 case in case library. So the transition probability is big to the case,which state change into 1 in the future. The euclidean distance have not been calculated for the case that state is 0 in case database, because the case that state is 0 is unknown in next time. Therefore, PN is used to represent the probability of the transformation of cellular states. In formula (1), the formula of d_i can be expressed as follows (2):

$$d_i = \sqrt{\sum_{j=1}^{n} w_j \left|(x_j - m_{ij})\right|} \qquad (2)$$

w_j is the weight of the first j attribute In the formula(2); the x_j is the j attribute for the cell that state is undetermined; m_{ij} is the first j attribute of the i case in the case library; n is the total number of attributes of a cell. The weight of the w_j attribute can be calculated according to the standard deviation, because the properties changing of j and cellular state relationship is very large, the corresponding cellular that state is 1 and cellular properties j should be less than the standard deviation to all cellular attributes corresponding to the standard deviation j in the research area, at this time, w_j should be more big. The calculation process is as follows:

$$sd_j = \sqrt{\frac{1}{m_1}\sum_{i=1}^{m_1}(x_{ij} - \mu_j)^2} \qquad (3)$$

sd_j is the standard deviation of the j attribute in the cell that state is 1 in case library. x_{ij} is the j attribute value of the first i case, is the number of the cell ,which state is 1, μ_j is the average value of the j attribute for the number of m_1 cell. The standard deviation is calculated for the first j attribute that total number is m cases,the formula as follows:

$$SD_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_{ij} - \mu_j *)^2} \qquad (4)$$

SD_j is the standard deviation of the first j attributes of the adjacent m cells in the case library, x_{ij} is the j attribute value of the first i case, $\mu_j *$ is the average value of the j attribute values of the adjacent m cells. The calculations are as follows:

$$w_j = \frac{1-(sd_j / SD_j)}{\sum_{j=1}^{n}(sd_j / SD_j)} \quad (5)$$

In the formula, n is the total number of cell attributes, and the cell is affected by the number of cell states in the neighborhood of 1, as well as the number of cell states in the domain of 1. If a cellular automata n state, denoted as R(i)(i∈[1,n]), if the i cellular transformation for growers, R(i)=1, or R(i)=0. Using the conversion probability, the formula is as follows:

$$PNe = \frac{\sum_{i=1}^{n}(\delta_i \times R(i))}{n} \quad (6)$$

$\delta_i \in [0,1]$ is the star,which is cellular index for cellular that state is 0, The cellular that the state is 1, the third party agricultural products inspection agencies are introduced, The farmers are evaluated for index of star by inspection of agricultural products and comprehensive quality. In this way, each production cycle of agricultural products will form an evaluation ranking for farmers, which will promote the continuous improvement of the production process of agricultural products, and constantly improve their quality, and form a virtuous circle. The government policy guidance is also very important guiding factors in regional, because the crop have a problem of supply chain, if supply exceeds demand, it will cause the corresponding farmers changing from planting to non planting state; if the market demand is greater than supply, the government guide farmers to more plant according to the preliminary investigation. Therefore, the government will introduce the relevant guidance policies in timely and according to the status of the supply chain of agricultural products. It will also affect the transformation of the cells state, It can be reflected by an adjustment function P_{adj} . the final cellular state affected by PN,Pne, $P_{adj.}$ It is calculated as follow:

$$P_{ch} = PN \times PNe \times P_{adj} \quad (7)$$

To summarize the above model, the government can investigate the market and the local comprehensive environment, and then direct the farmers' planting. According to the model assumptions, the current state after several cycles, the conversion of 1 farmers for a total of T, based on the current state of undetermined farmers transition probability ranking, and then adjust the transition probability, the effective control of the total conversion. The structure flow chart of the model is shown in figure 2:

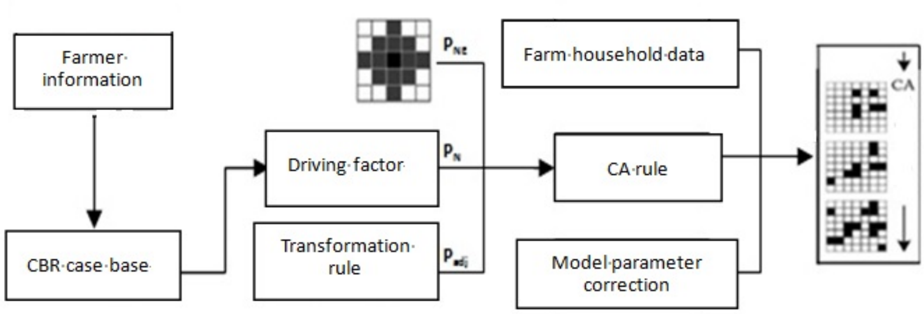

Figure 2. structure flow chart of CBR- cellular automaton model

3. Empirical analysis

There are two type variables,they are Socioeconomic variables and natural variables, Social economic variables: Social Policy (if the current development orientation and policy, The value is 1 in regional area, the value is 0.5 outside the area), the market heat index (the variables need to be analyzed by which specialized personnel engaged in market research after the investigation came out of the year the market heat index, the index value in the [0,1] range) family labor index (family labor divided into 18-30, 31-45, 46-60, 61-80, the corresponding index were 1, 0.8, 0.6, 0.4, and the family member to get the sum divided by the total number of numerical index value, the value in the [0,1]).The natural variables contains land area (mu), organic matter (g/kg), total nitrogen (g/kg) and phosphorus (mg/kg) and available potassium (mg/kg), the main reference index of the soil testing Fertilization Expert Group Guidance listed in the main index.

This paper takes a village 2014-2016 a total of 3 years about the soil and land data to simulate data from the provincial land survey and statistics data, the data are shown in Table 1:

Table 1. Indicators related to land information in a village

Farmer-ID	Land area (mu)	organic matter (g/kg)	total nitrogen (g/kg)	Olsen-P (mg/kg)	K ppcm (mg/kg)	Guiding policy	Market heat index	Labour index
0001	3.2	22	1.8	11	98	1	0.7	0.87
0002	4.3	33	1.2	15	112	1	0.7	0.76
0003	3.6	26	0.9	13	109	1	0.7	0.66
0004	2.8	29	1.6	15	122	1	0.7	0.75
…	…	…	…	…	…	…	…	…

Table 2. The data of Kiwifruit planting for 3 villages in 2014 year

Village name	The total number of households	The number of households planted	Non plantation households
Village-1	1236	36	1200
Village-2	998	32	966
Village-3	2658	55	2603

According to the survey data of land 1 villages and related variables forecasting model, according to the village of 1 2014-2016 years Kiwi planting industry development data, with the village of 3 years of data as training data, the model formula, training up to 96.73% accuracy rate of Pch=0.756 as non farmers into converted growers threshold. Then to simulate the development of the village 2 and village 3 kiwifruit planting industry, the simulation data are as follows:

Table 3. the data of kiwifruit planting between village 2 and village 3 in 2014-2016 year:

Year	The statistics of Village-2	The forecast of Village-2	The statistics of Village-3	The forecast of Village-3	The real rate Village-2	The real rate Village-3
2014	32	25	55	42	0.781	0.763
2015	44	50	68	63	0.88	0.926
2016	56	60	73	78	0.933	0.935

As can be seen from the data above, the ratio of the predicted value to the actual value is getting higher and higher, getting closer to the actual value. This is also with farmers around the cultivation of Kiwi more and more farmers, the farmers into the growing probability of farmers, therefore, the ratio of pre - village is also increasing year by year. According to the data, the distribution of kiwifruit planted by farmers from 2014 to 2016 is as follows:

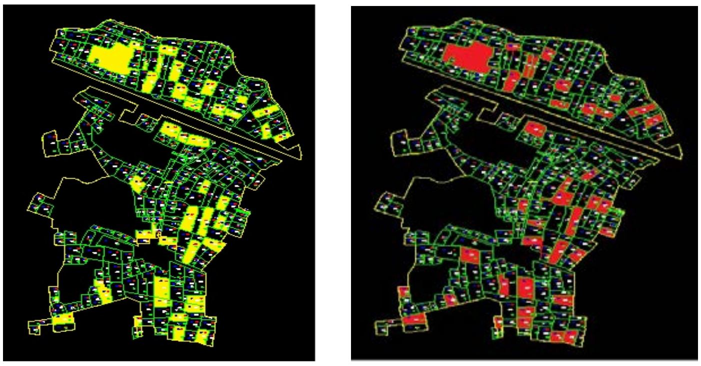

Figure 3. The data comparison map of village 2 in 2016 year

Above the picture on the left is 2 Village real field planting plans, be filled with red plots planted said kiwi, no field filled said no planting kiwifruit, right above the village 2 Figure prediction may grow, be filled with yellow cultivated field said the kiwi, no field filled no planting kiwifruit. The forecast and statistical data accuracy Table is shown in Table 4:

Table 4. The data Table of statistics forecast accuracy for village 2 in 2014-2016 year

Year	The number of actual statistics farmers for village-2	The number of forecast exact farmers for Village-2	The accuracy rate of Prediction for village-2
2014	32	22	0.687
2015	44	30	0.682
2016	56	48	0.857

Figure 4. The comparison map data of village 3 in 2016 year

The figure above is the same as in Figure 3, and its prediction and statistics calculation accuracy Table is shown in Table 5:

Table 5. The Table of data statistics forecast accuracy for village 3 in 2014-2016 year

Year	The number of actual statistics farmers for village-3	The number of forecast exact farmers for Village-3	The accuracy rate of Prediction for village-3
2014	55	38	0.69
2015	68	58	0.852
2016	73	67	0.917

4. Conclusion

This paper will combine CBR and cellular automaton algorithm farmers planting prediction is a new and valuable attempt to apply to CBR farmers as a case study for similar research, neighboring farmers, peasant household is regarded as a cellular automaton, using the algorithm to simulate farmers plant changes, the simulation process determines the cellular state change rule is the key, a lot of research often adopts the design of transform function or artificial intelligence algorithms to determine the conversion rules, the results tend to fall into the local optimal and can not adapt to dynamic changes.In this case, the CBR technique is introduced to determine the transition threshold by comparing the similarity between the driving variables of the cell and the experimental data, thus determining the state change of the cell. The algorithm can dynamically adapt to the changes of cellular state data, the dynamic simulation of farmers planting changes, related to the village through the empirical data show that the reliability and accuracy of the prediction algorithm can reach more than 80%, the algorithm for the local government to formulate relevant policies, guide farmers planting is of great value.

Acknowledgement

The research is supported by the Youth Foundation of Sichuan Education Department (Project number: 14ZB0005), It is also supported by the Double support plan of Sichuan agriculture

university.

References

[1] Amit Kumar Srivastava , Cho Miltin Mboh, Thomas Gaiser, Heidi Webber, Frank Ewert.Effect of sowing date distributions on simulation of maize yields at regional scale – A case study in Central Ghana, West Africa,Agricultural Systems, 147 (2016) 10–23.

[2] Bai Li, Zhang Runqing, Zhao Banghong. Behavior decision analysis of farmers participating in different industrial organization models- Taking edible fungus growers in Hebei Province as an example , agricultural technology and economics, 2015 (12): 42-51.

[3] Borja Ponte,José Costas,Julio Puche, David de la Fuente, Raúl Pino.Holism versus reductionism in supply chain management:An economic analysis,Decision Support Systems, 86 (2016) 83–94.

[4] Cai Mei, Cao Jie. A fuzzy case-based reasoning method for emergency decision making based on knowledge management. ,soft science, 2015,29 (9): 135-139.

[5] dragon fuyuhira, Li Tongsheng, Rui Yang, Ma Liyang. - Kiwi Zhouzhi County of Shaanxi Province planting demonstration village as an example ,economic geography, technology supply response to different patterns of behavior characteristics of farmers 2015, 35 (5): 136-143.

[6] E.M.Suárez-Rey, M.Romero-Gámez, C. Giménez, R.B.Thompson, M.Gallardo. Use of EU-Rotate_N and CropSyst models to predict yield, growth and water and N dynamics of fertigated leafy vegetables in a Mediterranean climate and to determine N fertilizer requirements,Agricultural Systems, 149 (2016) 150–164.

[7] F. Fdez-Riverola, D. Glez-Peña.A case-based reasoning system for aiding detection and classification of nosocomial infections, Decision Support Systems, 84 (2016) 104–116.

[8] Gong Jian, Yang Jianxin, Li Yafang. Simulation of rural residential land change based on case-based reasoning cellular automata ,resources science, 2015,37 (9): 1797-1806.

[9] H.J. Gómez-Vallejo , B. Uriel-Latorre , M. Sande-Meijide , B. Villamarín-Bello, R. Pavón,

[10] H.Navarro-Hellin,J.Martinez-del-Rincon,R.Domingo-Miguel,F.soto-Valles.A decision support system for managing irrigation in agriculture, Computers and Electronics in agriculture,124(2016) 121-131.

[11] high Shan, Huang Xianjin bell, too, Chen Zhigang. The commercialization of agricultural products to affect farmers' planting structure -- Investigation on ,resources science, Shanghai, Jiangsu and Anhui Based on households 2014, 36 (11): 2370-2378.

[12] John M. Antle , James W. Jones, Cynthia E. Rosenzweig.Next generation agricultural system data, models and knowledge products: Introduction, Agricultural Systems,2016.

[13] Li Hong Mei, Li Juanjuan. Analysis of influencing factors of farmers' planting behavior change in South China based on a survey of typical irrigation districts in Hunan province ,agricultural modernization research, 2015, 36 (4): 617-623.

[14] Wang Jianying, Chen Zhigang, Huang Zuhui, The relationship between land productivity and the scale of farmers' management during the transition period. Review, management world, 2015 (9): 65-81.