ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 165

International Conference on Modeling, Simulation, Optimization and Numerical Techniques (SMONT 2019)

Modeling of Household Financial Behavior Types in China Based on Complex Network Theory

Ying Liu[1,2,*], Jun Li[2], Junchan Zhao[1], Mei Wu[1] and Liuxi He[1]

[1]Mathematics and Statistics school of Hunan University of Commerce, Hunan province, China

[2]Business School of Hunan Normal University, Hunan province, China

[*]Corresponding author

***Abstract*—From the perspective of complexity science, the influencing factors of household financial behavior in China are presented, then the current situation and development characteristics of household financial behavior also given. The complexity characteristics of household financial behavior in China, such as openness, dynamics, non-periodicity and self-organizing criticality are probed. By using the method of fuzzy clustering and random forest algorithm model, the types of household financial behavior in China are classified and an empirical simulation analysis is made. The simulation results show the validity and accuracy of the model.**

Keywords—complexity perspective; influence factor; household financial behavior; fuzzy clustering; random forest

I. INTRODUCTION

Since the 1980s, developed countries have experienced a period of unusually active household financial behavior. Both the scale of household investment and the debt have increased dramatically, which has attracted extensive attention of scholars [1] [2]. Especially since Campbell, president of the American Financial Society, delivered a keynote speech at the AFA annual meeting in 2006, household finance has gradually become an important research field and focus of emerging finance [3]. Existing research shows that there are many differences in macroeconomic environment, structure, consumption concept and traditional cultural concept between China and developed countries abroad, which makes the study of family financial behavior in China heterogeneous and unsolved. Studying the relationship between household income level, household structure, education level, social interaction, risk preference and household financial asset allocation plays an important role in product development and design of financial institutions [4]. A thorough understanding of the formation rules and selection characteristics of household financial behavior in China not only improve the theory of asset selection and provide guidance for healthy financial behavior, but also point out the direction for financial market reform and financial innovation in our country. Therefore, under the current economic and social background, it is very important to study the financial behavior of Chinese households.

II. THE COMPLEXITY OF HOUSEHOLD FINANCIAL BEHAVIOR

Family financial behavior can be proved to be an open dynamic and complex system. The reasons for the complexity of family financial behavior of Chinese residents are mainly due to some basic characteristics of our family and society.

Firstly, large numbers of families interweave into complex social networks. China has a large population and a large number of households. According to the national population sampling survey data of the National Bureau of Statistics in 2016, there are 364 431 households in China by the end of 2016, which is about 10.46 times the number of households in 2000. The large population and the influence of traditional relative network in China have intertwined to form a complex social network among families. Therefore, the research object of household financial behavior itself is a complex and huge system. Family financial behavior is an open system [5][6]. A large number of families do not engage in economic and financial activities independently, but are closely linked with other families through social networks, as shown in the Figure I.

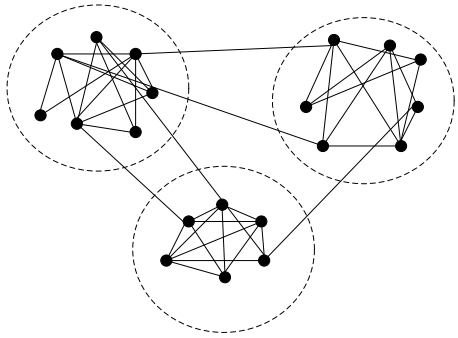

FIGURE I. DIAGRAM OF FAMILY SOCIAL NETWORK

Secondly, the external environment for the exchange of funds, information and materials with the family financial system is also change dynamically, and the impact on the household financial behavior system is not simple and superimposable. The ever-changing social and economic environment, especially the development of China's financial market, has prompted families to face more choices in investment and financial management, and further complicate their financial behavior. Diversified family types and individual

traits of decision makers are also an important reason for the complexity of household financial behavior in China.

Because of the large number of families in our country and the complex intertwined huge system, influenced by the external environment changes and the internal forces of family decision-makers, the household financial behavior of our resident shows obvious complexity characteristics.

III. MODELING OF HOUSEHOLD FINANCIAL BEHAVIOR TYPES

Dimension reduction and classification techniques are needed in the definition of family financial behavior. Common methods including: factor analysis, clustering analysis, discriminant analysis, decision tree, random forest algorithm, etc. These methods can be divided into linear and non-linear methods according to the nature of the model, single equation and multi-equation methods according to the number of models, and some methods have high requirements for data types and data distribution. Due to the large number of data indicators and complex data structure of household financial behavior, the relationship between variables tends to be non-linear, and the phenomenon of over-fitting is prone to occur when using one-equation model. Therefore, this paper combines the fuzzy clustering method with the random forest algorithm model to define and predict the types of household financial behavior.

A. Fuzzy C-Mean Clustering Modeling Method

Fuzzy class mean clustering analysis is a kind of fuzzy clustering algorithm which divides high-dimensional mixed data into specific categories [7][8]. The main purpose of this method is to calculate the membership degree of sample points to class centers and to judge the category of sample points so as to realize high-dimensional mixed data classification.

Divide the original n cases sample set $X=\{x_1,x_2,\cdots,x_n\}$, into $k(k \ll n)$ categories, which are $G=\{G_1,G_2,\cdots,G_k\}$. The basic calculation steps of fuzzy C-means clustering are as follows as shown in Figure II., including:

(1) Constructing membership matrix $R=\left[r_{ij}\right]$, i=1,2, …, n; j=1,2,,…, k. Where r_{ij} satisfies the constraints:

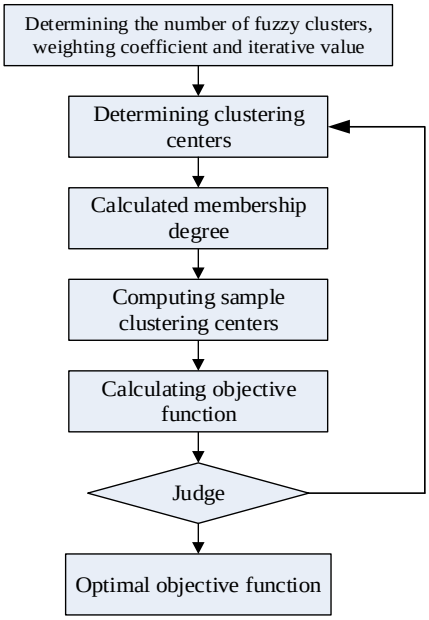

FIGURE II. FUZZY MEAN CLUSTERING FLOW CHART

$$s.t.\begin{cases} 0 \le r_{ij} \le 1 \\ \sum_{j=1}^{k} r_{ij} = 1 \\ 0 < \sum_{i=1}^{n} r_{ij} < n \end{cases} \tag{1}$$

(2) Fuzzy weighted index $h(h \ge 1)$ is set to control the influence of membership degree, and the maximum number of iterations $m_{\max}$, iterative termination threshold ε, defined g_j as the class center of class j data, then the clustering objective function is:

$$\min Z\left(R,X,G\right)=\sum_{i=1}^{n}\sum_{j=1}^{k} r_{ij}^{h}\left\|x_i-g_j\right\|^2 \tag{2}$$

(3) Computing the mean clustering centers of fuzzy classes:

$$g_j=\frac{\sum_{i=1}^{n} x_i\cdot\left(r_{ij}\right)^h}{\sum_{i=1}^{n}\left(r_{ij}\right)^h} \tag{3}$$

(4) Update the membership matrix of the fuzzy cluster center according to the calculation result of formula (3).

$$R=\left[\left(\sum_{j=1}^{k} d_{ij} \Big/ d_{ij}\right)^{\frac{2}{h-1}}\right]^{-1} \tag{4}$$

Where, d_{ij} is the distance between case x_i and class center g_j. Generally, Euclidean distance can be used.

(5) Comparing the membership matrix R^m with R^{m+1}, if the condition $\left\|R^{m+1}-R^m\right\|<\varepsilon$ or $m=m_{\max}$ is satisfied, the iteration is stopped, otherwise return to step (3).

B. Random Forest Modeling Method

Random forest is a combined classifier composed of multiple single-class decision trees. Compared with single model method, it has higher prediction accuracy and avoids the over-fitting problem of single classification model effectively. Therefore, it has been widely used in biology, computational science, informatics, management and other fields [9] [10]. The basic idea of random forest classification and prediction is to first use bootstrap method to extract n_0 samples from n original training samples in a playback way, and then establish a decision tree classification model for each sample, and then combine the prediction of multiple decision trees to obtain the final prediction results. The basic steps of building a random forest model are as follows:

(1) n_0 samples were randomly sampled from sample set X in a playback manner, and k times were sampled to form a training sample set k.

(2) A decision tree $t_k=\left[X,\Theta_{n_0}\right]$ is constructed for each training sample set, a total of k decision trees can be obtained. Each decision tree has one vote to select the optimal classification result.

(3) The decision tree sequence $\{t_1,t_2,\cdots t_k\}$ is formed by multi-round training, and the final classification decision is obtained by counting the votes of k classification results.

$$T(x)=\arg\max_Y \sum_{i-1}^{k} I\left(t_i(x)=k\right) \tag{5}$$

Where Y represents the classification result variable and $I(\bullet)$ is the demonstrative function.

The introduction of two randomness makes the random forest not easy to fall into over-fitting, and has a good anti-noise ability. For large sample data, because the trees are parallel to each other, the random forest algorithm can easily be used as a parallel method, so it has a fast processing speed.

In order to classify the types of household financial behavior in China, we need to use high-dimensional large sample mixed type data, and there is a certain proportion of missing data, so it is very suitable to use random forest algorithm to build the model.

C. Models Analysis and Simulation Results

1) Selection of research indicators

According to the idea of modeling, this paper first distinguishes and identifies the types of household financial behavior in China from the data of china family panel studies of the year 2016(CFPS2016). According to the research contents, the main indicators used in the process of model construction are divided into three categories: household financial assets investment behavior indicators, household debt management behavior indicators and household real estate investment behavior indicators, as shown in Table II. .

The CFPS 2016 survey data lacks the type variables of financial products invested by households. In the model construction, it may be assumed that the types of financial products invested by households have not changed from 2014 to 2016. Therefore, the type data of financial products are matched by households in the CFPS 2014 database, and other data are from CFPS 2016.

A total of 20 positive indicators are selected for modeling, which have different measurement methods such as quantity, presence or absence, category and so on. Therefore, data standardization is needed before modeling and analysis.

2) Fuzzy c-mean clustering simulation analysis

Taking the financial behavior data of households as samples, the sample set is analyzed in 22-dimensional space by means of the fuzzy clustering algorithm, and the fuzzy category identification is carried out according to the general steps of the fuzzy clustering analysis in the previous section. Supposed the fuzzy weighting index h=2 , maximum iterations $m_{\max}$=100 , iterative termination threshold ε=0.00001 . The membership matrix is calculated by 61 steps iteratively to achieve stability. All effective families are divided into five types of behavior. Each category is given a unique label to indicate the behavior type represented by the fuzzy cluster. Specific classification results as shown in Figure III.

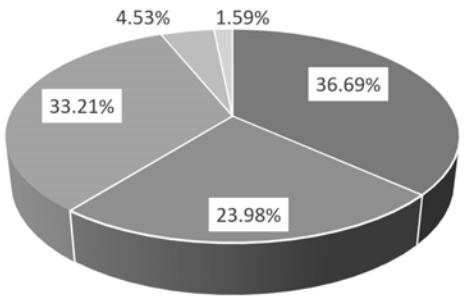

FIGURE III. FUZZY MEAN CLUSTERING RESULT DISPLAY GRAPH

The average value of silhouette, which is used to test the clustering effect, is 0.6, indicating that the clustering effect is good. The ratio of the largest class to the smallest class is 23.1, which is slightly larger. This may be due to the high uniformity of some types of family financial behavior in our country.

3) Random forest simulation analysis

On the basis of the label of investment behavior types of family financial assets obtained from the results of fuzzy c-

mean cluster analysis, the random forest algorithm is used to further establish the prediction model in order to improve the accuracy of the prediction results as much as possible. According to the basic process of building random forest algorithm in section above, families to be classified first calculate the classification results of leaf nodes falling into each decision tree, and then vote on the classification results of all trees in the random forest to get the final prediction results of family financial investment behavior types.

More sub-trees can make the model have better performance, but also take longer time. All household data were divided into 70% data as training set and the remaining 30% data as test set to train random forest prediction model. From the observation of Figure IV, it can be found that with the increasing number of trees, the classification accuracy presents an inverted U-shaped shape, which increases first and then decreases. Therefore, 32 sub-trees are selected to construct a random forest. The whole training process lasts about 0.256 seconds, and the completion speed is very fast. The error probability outside the bag of the model is also analyzed. Results show that with the increase of the number of trees, the error probability converges rapidly: around 10 trees, the error probability outside the bag has nearly converged to zero.

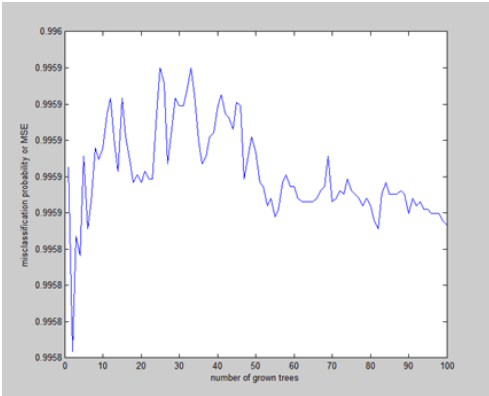

FIGURE IV. RELATIONSHIP BETWEEN THE NUMBER OF TREES AND THE PREDICTION

By comparing the linear single model discriminant classification (DC) method, the non-linear single model decision tree and the non-linear multi-model combination random forest algorithm, the data in Table I show that the random forest algorithm has higher discriminant stability.

TABLE I. STABILITY COMPARISON OF MODEL CLASSIFICATION

Model category	DC analysis	Decision tree	Random forest
Stability ratio (%)	73.6	84.5	98.2

TABLE II. A LIST OF INDICATORS FOR THE ANALYSIS OF FAMILY FINANCIAL BEHAVIOR TYPES

Index name			**Measurement**	**Unit**
Investment behavior of family financial assets	Holding situation	Household cash and total deposit	Amount	Yuan
		Total household time deposit	Amount	Yuan
		Total price of household financial products	Amount	Yuan
		Types of financial products held	Possess/None	-
	Effectiveness	Profit from household financial products investment	Amount	Yuan
Family debt management behavior	Lending channel	Preferred borrower	Category	-
		Bank loan	Have/not	-
		Non-governmental loans	Have/not	-
	Lending scale	Total bank loans	Amount	Yuan
		Total relatives loans	Amount	Yuan
		Total non-government loans	Amount	Yuan
Household real estate investment behavior	Holding scale	Current market price of residential property	Amount	Yuan
		Market price of other property	Amount	Yuan
		Total house purchases	Category	-
	Mortgage burden	Building and decoration loans	Have/not	-
		Building and decoration borrowing	Have/not	-
		Loan quota for house purchase, building and decoration	Amount	-
		Borrowing quota for purchase, building and decoration	Amount	Yuan
	Effectiveness	Rent income	Amount	Yuan
		Real estate value added ratio	Amount	%

Based on CFPS 2016 data, a total of 14033 households were surveyed. And 12417 valid household samples were obtained, accounting for 88.48% of all households surveyed. According to the random forest model, CFPS 2016 households were divided into five typical household financial behavior: Traditional, normative, active, speculative and balanced.

The ratio of maximum class to minimum class is 14.24, which is lower than the corresponding ratio of 23.6 of the fuzzy c-mean clustering method shown in Figure III. It shows that the random forest algorithm has been optimized on the basis of the fuzzy c-mean clustering method, and the prediction result of the model is more reasonable.

IV. CONCLUSIONS

Household financial behavior is a complex system. It has the characteristics of dynamic openness, non-linear tight coupling and self-organizing criticality. Simulation results show that, from the perspective of complexity network theroy, the combination of fuzzy cluster method and random forest algorithm model is effective and accurate in defining and predicting family financial behavior types.

ACKNOWLEDGMENT

This work is supported by National Social Science Fund project (NO. 18BTJ025).

REFERENCES

[1] Lynsey K. Romo, "A review of family financial communication", The Handbook of Financial Communication and Investor Relations, First Edition, 2017, pp:399-408.

[2] FKoropp. C, Kellermanns. FW, Grichnik. D, et al. "Financial decision making in family firms: An adaptation of the theory of planned behavior". Family Business Review, 2014, vol:27, issue:4, pp:307-327.

[3] Julie Birkenmaier. "Promoting financial credit-building for financially vulnerable families". Families in Society: The Journal of Contemporary Social Services, 2018, vol:99, issue:2, pp:93-99.

[4] Smith, Thomas E., Richards, Kristin V., Panisch, Lisa S.,et al. "Financial therapy with families". Families in Society, 2017, vol:98, issue:4, pp: 258-265.

[5] Dj Watts, Sh Strogatz. "Collective dynamics of small-world networks". Nature, 1998, vol:393, issue:6684, pp:440-442.

[6] Al Barabasi, R Albert. "Emergence of scaling in random networks". Science, 1999, vol:286, issue:5439, pp: 509-512.

[7] Jian-Ping Mei, Yangtao Wang, Lihui Chen, et al. "Large scale document categorization with fuzzy clustering". IEEE Transactions on Fuzzy Systems, 2017, vol:25, issue:5, pp: 1239–1251.

[8] Lun Hu, Keith C. C. Chan. "Fuzzy clustering in a complex network based on content relevance and link structures". IEEE Transactions on Fuzzy Systems, 2016 , vol:24, issue:2, pp: 456 – 470.

[9] Erwan Scornet. "Random rorests and kernel methods". IEEE Transactions on Information Theory, 2016, vol:62, issue:3, pp:1485 – 1500.

[10] Youqiang Zhang, Guo Cao, Xuesong Li, et al. "Cascaded random forest for hyperspectral image classification". IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018 , vol:11, issue:4, pp:1082 – 1094.