IOSR Journal Of Humanities And Social Science (IOSR-JHSS)
Volume 27, Issue 7, Series 3 (July. 2022) 01-07
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Research on Retailer customer group of new energy vehicle industry based on Neural Network

ChuFang
College of Economics and Management,Zhaoqing University, Zhaoqing City, Guangdong, China

Abstract: Information technology advances have allowed researchers to collect a large number of individual-based customer data, customer databases are becoming increasingly large and complex, making the technical capacity to analyze the data fiasco. To solve this problem, this study uses neural networks (ANNs) as a way to split the retail database, and the Hopfield neural network of networks, to construct a Hopfield-Kagmar (HK) clustering algorithm, then compared the Hopfield-Kagmar (HK) clustering algorithm and the K means method and the advantages and disadvantages of mixed-mode algorithm. The results indicated that for retailers, market segmentation by HK clustering approach is more useful because it can provide more homogeneous segments and less sensitive to initial solutions.

Key Word: **C**luster analysis; new energy automobile industry ; data mining technology

---

Date of Submission: 01-07-2022 Date of Acceptance: 11-07-2022

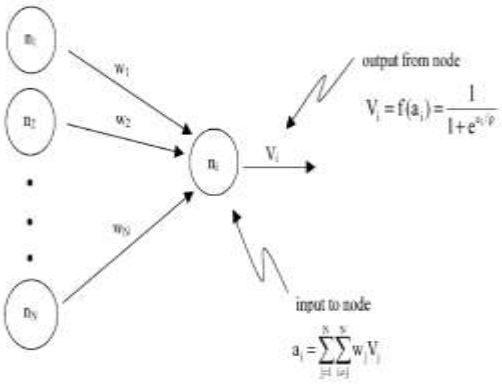
I. Introduction

Retailers pay more and more attention to the specific needs and preferences of different customers. However, through a large number of customer personal data that is not integrated, it will make effective customization more difficult to implement. In recent years, electronic data collection methodologies such as customer loyalty program or personal purchase plan are faced with the problem of excessive data, which makes market segmentation and market target too complex. Therefore, how to effectively process, analyze and interpret customer information will be the main challenge for retailers in order to implement the marketing strategy of specific cluster market. In order to effectively implement the target marketing strategy, retailers must segment their market correctly, so that the individuals within the group are highly isomorphic, while different groups are highly different, so as to avoid inappropriate segmentation, reduce the effect of each interval strategy and waste marketing resources and costs.

Therefore, in order to get the right market segmentation, retailers must choose the appropriate customer clustering method. For example, previous studies have found that non hierarchical clustering is more suitable than hierarchical clustering in segmenting large amounts of data, and reasonable information is needed to obtain better results. In this study, the Hopfield network of ANN is used to verify the effectiveness of HK clustering method in the segmentation of retailer database.

Neural network, which uses a large number of simple connected artificial neurons to simulate the ability of biological neural network. Artificial neuron is a simple simulation of biological neurons. It obtains information from the external environment or other artificial neurons, and outputs the results to the external environment or other artificial neurons with very simple operation, so as to be used for estimation, prediction, decision-making and diagnosis.

The basic Ann pattern consists of an input layer with multiple nodes and an output layer with one node (as shown in the figure below).



n_j and w_j are the nodes and weights, respectively, for the $j = 1, 2, \ldots, j \neq i$, N nodes in the network

$$a_i = \sum_{j=1}^{N} \sum_{i \neq j}^{N} w_j V_j$$

Where node input value formula ; Excitation function

$$V_i = f(a_i) = \frac{1}{1 + e^{a_{pi}/\rho}}$$

Usually, there will be one or more hidden layers in the architecture, between the input layer and the output layer. In the feedback network, the output data of the output layer will be fed back to the input layer or the hidden layer in front of it by means of backward transmission, and the weight value of the optimal solution can be obtained through repeated calculation.

The basic structure of neural network: there are three basic architectures of neural network, which are "operands", "layers" and "networks". The three levels are explained one by one

Processing element (PE): the basic unit of neural network is the "operands", also known as neurons or processing units. The structure of the operands is like the soma of a biological neuron. It is the main core, which processes the input and output signals.

Through the operation of the operands, the input signals to the operands are processed and converted, and the converted signals are output as the input signals of other operands.

A layer is composed of several sets of operands with the same function. The processing unit of the input layer is used to input the information of the external environment, and the processing unit of the output layer is to output the signal to the external environment or other neurons. At the same time, the neural network also contains a hidden layer, which provides the ability of neural network to represent the interaction between processing units and the internal structure of problems.

Network is composed of single or multiple layers with different functions. There are two general operation modes of neural network, which are learning process and recalling process.

In the learning process of neural network, when there is a deviation between the actual value and the correct value of the network, the learning algorithm of the root network adjusts the link weight value of the network by learning from examples to reduce the error between the actual value and the correct value.

Recall after the end of the learning process, the network will determine the output data of the network according to the recall algorithm, which is called "recall process". After learning the program, the learning process of neural network will be memorized in the link weight of the network, so the knowledge learned in the past can be quickly obtained through the link weight to process the input data and calculate the output data of the network.

At present, there are not many researches using ANN as market segmentation, and the results are mostly inconsistent. For example, mangiamili, Chen, and West (1996) compared Kohonen's (1989) self-organizing map networks (SOMs) with traditional hierarchical clustering techniques, and found that SOMs were relatively good, while Balakrishnan, Cooper, and Jacob, And Lewis (1994) compared SOMs with k-means method, but found SOMs were relatively poor; recent studies pointed out that feedback Ann combined with cluster analysis and differential analysis can produce good group results. Therefore, relevant researchers pointed out that the lack of network architecture of ANNs is the reason for the inconsistent results, which shows that ANNs can effectively segment the market, but the choice of network architecture is the main key to determine the degree of success.

II. Research model and hypothesis

Unlike other neural networks, each node in the Hopfield network is connected to each other, and the information can flow through each other. The Hopfield network has three advantages. 1. Unlike MLPs and Kohonen networks, the input weights of Hopfield networks do not change, and the optimal solution is to minimize the objective cost function. 2. The threshold excitation level of each node is fixed, and the difference can be shown when the new solution is obtained. 3. MLPs and Kohonen network need training, while Hopfield network needs no training.

Kagmar Parsi et al. (1990) developed the HK clustering algorithm based on Hopfield network and proved that it is better than MLPs and Kohonen networks. An outline of this algorithm is shown below.

Stage 1: network initial design

(1) A K × n matrix is established, where k is the number of preset clusters (obtained by using prior knowledge) and N is the number of customers in the data set.

(2) It is randomly assigned to each element in the matrix, where VPI represents the strength of group members of customer I in group P.

(3) A small subset of customers is randomly selected to determine the initial group center of gravity.

Stage 2: network execution

(4) Each n = k × n nodes in the network can accept the information from each node as the input value, where API represents the value input to node PI, which is a function of VPI, and vqi represents the strength of customer I assigned to groups outside group P. RPI is the square of Euclidean distance between customer I and the center of gravity of group P.

(5) In the T stage, after the first node in the network produces the result of VPI, it uses VPI to carry out two-stage program.

(a) The API is calculated by mathematical formula.

$$\frac{\mathrm{d}a_{pi}}{\mathrm{d}t} = -A\sum_{q\neq p}^{K} V_{qi} - B\left(\sum_{q=1}^{K} V_{qi} - 1\right) - CR_{pi}V_{pi}$$

Theformula is

(b) The API value is used to update the VPI through the excitation function formula.

$$V_i = f(a_i) = \frac{1}{1 + e^{a_{pi}/\rho}}$$

The formula is

(6) The center of gravity of group P was recalculated by using the updated average weight of VPI, and RPI was recalculated.

(7) In the T + 1 phase, the second node in the network updates VPI and recalculates the center of gravity and RPI of group P, as described in steps (5) and (6).

(8) At t + 2, t + 3, t + 4 In the stage, all the remaining nodes in the network repeatedly update VPI and recalculate RPI, as mentioned in steps (5) and (6).

(9) The network is continuously calculated until the output of each node remains unchanged. In this case, VPI (T stage) = VPI (T + T stage), the optimal solution can be obtained.

In HK mode, the development of Hopfield network is controlled by the following mathematical formula.

$$S = \frac{A}{2}\sum_{i=1}^{N}\sum_{p=1}^{K}\sum_{q\neq p}^{K} V_{pi}V_{qi} + \frac{B}{2}\sum_{i=1}^{N}\left(\sum_{p=1}^{K} V_{pi} - 1\right)^2 + \frac{C}{2}\sum_{i=1}^{N}\sum_{p=1}^{K} R_{pi}V_{pi}^2$$

Among them, a, B and C denote coefficients greater than 0, while the first mathematical item restricts the same customer to be divided into two groups, the second item restricts each customer to belong to a certain group, and the third item restricts RPI (Euclidean distance between customer I and the center of gravity of group P) to be the minimum. Therefore, when the minimum solution of this formula is obtained, it is the best solution to obtain the clustering result.

There are four advantages of HK cluster method

1. Based on the minimization of the variation in the group, the HK clustering method adjusts the number of groups by repeating the calculation process, and the HK clustering method uses partial redistribution to perform the calculation.
2. The initial solution of HK clustering method is more flexible and does not need prior knowledge.
3. In HK cluster method, Ann is used to determine the next group members, and then adjust the group members and variation.
4. HK cluster method does not need prior reasonable information.

III. Empirical Research

This study is to verify the HK clustering algorithm through two sets of data. Firstly, the actual data is used to compare the HK clustering method and the traditional clustering algorithm, and then the artificial data is used to compare the HK clustering method and the traditional clustering method.

Actual data

A total of 4317 customer observation data were collected from the customer database of large retailers. Six purchase behavior variables were used as input variables. The following table shows the group variables and narrative statistics for the dataset.

Variable	Mean	Standard deviation	Maximum	Minimum
Number of credit cards used	1.16	0.65	5	0
Account age (months)	39.34	11.22	69	19
Days since first purchase	698.05	211.74	1095	184
Days since last purchase	546.65	255.47	1004	2
Total number of orders	1.81	1.19	9	1
Total dollars spent	172.84	162.07	1468.70	9.00

Manually generated data

According to the Research Report of Milligan and Cooper (1985) and Carmone, Kara, and Maxwell (1999), the artificially generated data set is to use three factors to establish four groups containing 2000 observations. These three factors can show the data complexity of the real world and serve as the basis for influencing the clustering, so as to test the validity of various clustering methods.

For example, in clustering studies, a large number of complex and related unstructured data sets often contain noise variables, outliers and clusters of different sizes. Therefore, in this study, the first factor is the number of variables, which can be divided into two levels. The first level represents 8 noiseless variables and 1 disturbance variable, while the second level represents 8 noiseless variables and 2 disturbance variables. The second factor is group density, which means that groups of different sizes are also divided into two levels: one group with the same size contains 60% data, while the other groups have the same size. The third factor is the outliers of each group, which also has two levels: no outliers and 20% outliers. Through $2 \times 2 \times 2$ experimental designs and program development, 8 sets of data sets can be generated, showing the external differences and internal isomorphism of the groups.

Comparative group analysis technique

At present, there are more than 50 kinds of clustering techniques, but none of them is suitable for various situations. Therefore, it is necessary to verify that clustering method has higher validity through comparison. In this study, k-means method and mixed mode clustering method are used as the comparison standard of HK clustering method.

Dependent variable

In this study, intra group differences and Hari (Huber Arabie adjusted Rand index) were used as dependent variables. Among them, the difference value within the total group is to verify the group isomorphism in the actual data, and Hari is to test the correctness of the group results obtained from the artificial data.

Program

Jmp4.0.2 software is used to implement k-means method and mixed mode, while HK cluster rule is a UNIX Environment calculus constructed by C program. In the calculation of real data, the initial input variables of three clustering methods are obtained by means of hierarchical clustering of average method, and 5% sub samples are randomly selected as another input value. Finally, ANOVAs was used to test the difference value, input value type and their interaction within the total group. In the artificial data calculation, we compare the HK clustering method, k-means method and mixed mode with Hari and the difference within the average total group, and finally test the Hari with ANOVAs.

IV. Result

Real data results

The following two tables are the results of 4317 real data operations.

Segment means by algorithm and seed-type for real world data set

Segment	Number of credit cards used	Account age (months)	Days since first purchase	Days since last purchase	Total number of orders	Total dollars spent	Segment size
(a) HK random seeds (minimum variation solution)							
One	1.6	39.8	797.2	319.2	4.0	425.67	603
Two	1.2	54.8	824.0	659.4	1.6	137.02	648
Three	0.9	31.7	541.0	452.5	1.4	113.53	1173
Four	1.5	48.0	454.4	335.2	1.6	144.16	645
Five	1.0	33.8	858.3	795.8	1.4	139.85	1248
(b) K-means rational seeds							
One	1.1	56.8	712.1	591.9	1.5	133.00	684
Two	1.0	35.0	480.3	417.8	1.4	115.83	1296
Three	0.9	34.7	839.2	794.7	1.4	138.48	1383
Four	2.0	38.2	787.2	333.8	2.4	207.74	605
Five	1.4	41.5	765.6	322.3	4.7	536.43	351
(c) Normal mixtures rational seeds							
One	1.5	40.5	754.8	273.2	4.3	429.05	404
Two	3.1	45.9	725.3	557.4	1.6	163.77	141
Three	4.0	51.5	871.0	188.5	6.0	787.52	12
Four	1.0	38.9	689.4	577.6	1.5	136.20	3729
Five	1.2	41.6	809.7	483.0	4.9	1044.47	31

The rational seeds for *K*-means and normal mixtures were based on the centroid locations obtained from hierarchical clustering using the average method.

Total within-segment variation for real world data set using standardized variables (N=4317)

Seed-type	Clustering algorithm		
	HK	*K*-means	Normal mixtures
Rational (based on hierarchical clustering using average method)	19.29	19.66	29.21
Random (mean across 250 analyses)	19.23	20.47	27.31

According to the results of the above table, there is no significant difference between the HK clustering method and the k-means method when the input value types are different, but the results of rational seeds are better for the mixed mode. When using random seeds, HK cluster method is better than k-means method and mixed mode, while using rational seeds, HK clustering method is the same as k-means method and better than mixed mode. Therefore, on the whole, the HK means method is better than the k-means method and the mixed model no matter what kind of data is used.

There are several reasons for the difference. 1. The data in the real world are too large, complex and disturbed, and lack sufficient definition. 2. In the process of calculation, the HK clustering method and the mixed mode use the partial assignment of group members as the operation criterion, while the k-means method uses all assignments. 3. The hybrid mode is based on the error when adjusting the group members, while the HK cluster rule is to adjust the group members only after each node is updated.

Manual data results

Algorithm	Cluster density		Outliers		Variables and noisy variables	
	Equal	0.60	None	0.20	8+1 Noisy	8+2 Noisy
(a) Average HARI by clustering algorithm and data complexity for N= 72 artificial data sets						
HK	0.92	0.74	0.93	0.72	0.84	0.82
K-means	0.80	0.87	0.92	0.75	0.82	0.85
Normal mixtures	0.66	0.73	0.92	0.48	0.68	0.72
(b) Average total within-segment variation by clustering algorithm and data complexity for N= 72 artificial data sets						
HK	21.82	24.76	20.29	26.31	21.39	25.20
K-means	23.87	29.13	20.34	32.66	26.24	26.77
Normal mixtures	43.95	43.64	20.40	67.20	42.17	45.42

The results of the above table can be divided into two parts. Firstly, it can be seen clearly in part (a) that when the group density is the same, the result of HK cluster method is better, but when there is a certain group density of 60%, the result of K-means method is better. When there is no outlier in the group, the results of the three clustering methods are good, but when the outlier is 20%, the values of the three methods decrease, especially in the mixed mode. In addition, if there is only one disturbance variable, the solution of HK cluster method is similar to that of K-means method and is better than that of mixed mode. After adding another disturbance variable, the difference between the three methods will be reduced.

In part (b), because the total difference between groups is the smallest, the smaller the results of the three clustering methods are better, and the results of comparison (a) and (b) are similar.

The results show that the HK cluster method is better than the k-means method and the mixed model in customer clustering of retailers. Because the HK clustering method can provide higher isomorphism within the group, and is less sensitive to the initial solution of the calculation process, it can obtain good results no matter using random or rational seeds, so it is more flexible in actual operation. When we use HK cluster method to cluster market customers, we can avoid the complex operation of hierarchical clustering. The disadvantage of HK clustering method is that when the density of a group is too large, the k-means method can get better results.

Because improper selection and application of market clustering technology will seriously affect the company's financial structure, and the more correct clustering results can save the company's resources and money, so when making marketing strategy for customers, we must select a more suitable clustering technology for customer clustering, and then make marketing strategy according to the characteristics of each different group of customers, so as to achieve the most effective strategy result.

References

[1]. Gurley, J. W. The one Internet metric that really matters. Fortune. 2000, 2, 141-392.
[2]. Bickerton, P. 7 technologies that are transforming the hospitality industry. Hosp. Mark. 2015, 1, 14-28.
[3]. Pui, L. T.; Chechen, L.; Tzu, H. L. Shopping motivations on Internet: A study based on utilitarian and hedonic value. Technovation. 2007, 27, 774-787.
[4]. Lee, E. Y.; Soo, B. L.; Yu, J. J. Factors influencing the behavioral intention to use food delivery Apps. Soc. Behav. Pers. 2017, 45, 1461-1474.
[5]. Armstrong, G.; Kotler, P. Marketing, Prentice-Hall: Englewood Cliffs, America, 2000; pp. 1–98.
[6]. Ozkara, B. Y.; Ozmen, M.; Kim, J. W. Examining the effect of flow experience on online purchase: A novel approach to the flow theory based on hedonic and utilitarian value. J. Retail. Consum. Serv. 2017, 37, 119-131.
[7]. Park, J.; Ha, S. Co-creation of service recovery: Utilitarian and hedonic value and post-recovery responses. J. Retail. Consum. Serv. 2016, 28, 310-316.
[8]. Anderson, K. C.; Knight, D. K.; Pookulangara, S.; Josiam, B. Influence of hedonic and utilitarian motivations on retailer loyalty and purchase intention: A facebook perspective. J. Retail. Consum. Serv. 2014, 21, 773-779.
[9]. Chiu, C. M.; Wang, E. T.; Fang, Y. H.; Huang, H. Y. Understanding customers' repeat purchase intentions in B2C E-Commerce: The roles of utilitarian value, hedonic value and perceived risk. Inf. Syst. J. 2014, 24, 85-114.
[10]. Lin, K. Y.; Lu, H. P. Predicting mobile social network acceptance based on mobile value and social influence. Internet Res. 2015, 25, 107-130.
[11]. Huang J. H.; Yang, Y.C. Gender difference in adolescents' online shopping motivation. Afr. J. Bus. Manage. 2010, 4, 849-857.
[12]. Grandom, E.; Mykytyn, P. Theory-based instrumentation to measure the intention to use electronic commerce in small and medium sized businesses. J. Comput. Inf. Syst. 2004, 44, 44-57.
[13]. Chan, T. K. H.; Cheung, C. M. K.; Shi, N.; Lee, M. K. O. Gender differences in satisfaction with Facebook users. Ind. Manag. Data Syst. 2015, 115, 182 -206.
[14]. Meyers-Levy, J.; Sternthal, B. Gender differences in the use of message cues and judgments. J. Mark. Res. 1991, 28, 84-96.
[15]. Jo¨reskog, K.G.; So¨rbom, D. LISREL 8: User's Reference Guide, Chicago: Scientific Software International, America, 1996; pp. 1–98.
[16]. Arnold, M.J.; Reynolds, K.E. Hedonic shopping motivations. J. Retail. 2003, 79, 77-95.
[17]. Schwab, D. P. Research methods for organizational studies. NJ: Lawrence Erlbaum Associates, America, 2005; pp. 1–30.
[18]. Byrne, B. M. Structural equation modeling with AMOS: Basic concepts, applications, and programming. NJ: Lawrence Erlbaum Associates, America, 2001; pp. 1–28.
[19]. Swaminathan, V.; Lepowska, W. E.; Rao, B. P. Browsers or buyers in cyberspace: An investigation of factors influencing electronic exchange. J Compur-Mediat. Comm. 1999, 5, 224-234.

Acknowledgements : The research result of this subject is "the 2022 subject of Guangzhou philosophy and Social Sciences Planning" and the subject No.: 2022GZGJ88.

ChuFang "Research on Retailer customer group of new energy vehicle industry based on Neural Network." *IOSR Journal of Humanities and Social Science (IOSR-JHSS)*, 27(07), 2022, pp. 01-07.