SSOAR

Open Access Repository

www.ssoar.info

Segmentation of life insurance customers based on their profile using fuzzy clustering

Jandaghi, Gholamreza; Moradpour, Zahra

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Segmentation of life insurance customers based on their profile using fuzzy clustering

[1]Gholamreza Jandaghi, [2]Zahra Moradpour

[1]Professor, Faculty of Management and Accounting, Farabi College, University of Tehran, Iran

[2]MA, Faculty of Management and Accounting, Farabi College, University of Tehran, Iran

Corresponding Author: jandaghi@ut.ac.ir

ABSTRACT. In the current competitive environment, companies will be able to adjust business strategies, they use market segmentation based on practical ways rather than using traditional approaches or incomplete and impractical mass marketing. In recent years, mining has gained attention and popularity in the business world. The goal of data mining projects is to convert the raw data into useful information. Clustering can also be used to explore differences in attitudes and intentions of the clients. In this study, we used fuzzy clustering on 1071 life insurance customers during March to October 2014. . Results show that the optimal number of clusters was 2 which were named as "investment" and "life safety". Some suggestions are presented to improve the performance of the insurance company.

1. INTRODUCTION

Today, competition among insurance companies has been tightened. New customer acquisition is more difficult than before and its acquisition cost is higher (HOSSEINI, et al., 2013). A strategy of "customer-oriented" has become a critical element of a successful business environment. The purpose of CRM is to build deeper relationships with customers, and is able to change corporate behavior with respect to personal customer (Bae, et al., 2005). CRM are business processes and technologies that represent management's efforts to understand customers and designed by a company. The main core of CRM includes understanding customer profitability and retention of profitable customers. Many organizations start making customer value to retain customers and maximize potential profit in the management of their own. Organizations use market segmentation to identify the client. Data mining techniques can be used to explore the hidden information and identify patterns and associations that have useful applications (Liang, 2010). An effective strategy to achieve this target is to study the customers' profiles and their past behavior and future needs and demands (Lin, et al., 2012). The main idea of segmentation or clustering is to group similar customers. A cluster can be thought as a set of customers who have similar characteristics of demographics, attitudes, values, etc. (Hiziroglu, 2013). Clustering is a special class of statistical methods for understanding customer behavior. When clustering of customers, behavioral data is used. Each cluster includes customers with a specific pattern that is associated with the cluster centers (Bose & Chen, 2015). The insurance sector is primarily dependent on the client (Umamaheswari & Janakiraman, 2014). Giving service to customers is an integral part of the insurance. therefore, they need to identify the key success factors in the life insurance industry for customer satisfaction. The insurance industry faces some key challenges, including increased competition in the market, expectations for customer service and the need to build competencies to remain in the insurance industry. All these efforts need to share information across the business units that insurance companies do a better job (RAMANATHAN, 2012). And insurance companies with massive data analysis at their disposal to attract new customers and retain previous customers and to take measures to further enhance. This paper presents a segmentation model for customers of Pasargad life insurance and then categorize them based on fuzzy clustering. In the current study we answer the following questions:

1. How can we segment life insurance customers using the fuzzy clustering?

2. What is the status of Pasargadae life insurance customers?
3. What are the key variables in life insurance customer segmentation?

2. LITERATURE REVIEW

Data mining was introduced in late 2009 and for the first time in 2010 in the subject of medicine. While working in the field of statistics and machine learning, data mining primarily is an interdisciplinary science. In 2001 data mining was identified as one of the ten world-changing technologies by MIT University. One of the definitions that are mostly used is that data mining is the analysis of (often large) observed data sets to discover hidden relationships and summarize the data in ways that are understandable (Yoo et al., 2012). Today, data mining is ready to support the three technologies. These three technologies are:

• Collecting and storing large amounts of data
• Powerful computer processing
• Advanced data mining algorithms (Behrouzian-Nejad, et al., 2012).

Data mining techniques are used to improve decision-making in various fields. Articles summarizing data mining techniques are classified into seven main categories, Classification, clustering, regression, association (association) (Umamaheswari & Janakiraman 2014), Discovering outliers, visualization, prediction (Sithic & Balasubramanian, 2013). Market segmentation is classification of customers who can be a marketing tool for the discovery of naturally grouped according to accessible information (Hong, 2012). Another way to understand market segmentation is to identify homogenous subgroups in a heterogeneous market (Casabayó, et al., 2015). Cutler (2003) classified market segmentation variables into four major areas of geographic, demographic, psychological and behavioral (Hiziroglu, 2013).

Customer segmentation is grouping customers according to their needs to be able to give more services. Each customer will be given a list of products, but each customer has different needs and demands. The classification of customers is easy with a variety of patterns (Singh & Rana, 2013). Life insurance is a type of insurance operations liabilities related to the function of human life (Karimi, 1391: 481). Clustering in data mining and pattern recognition is a fundamental method for the classifying of the observed cases into clusters (groups) which satisfies the following main criteria:

• Each group or cluster is homogeneous. The objects that belong to the same group are identical.
• Each cluster should be different from other clusters. Clusters of objects belonging to a cluster should be different from other objects.

In fuzzy clustering, each cluster is a fuzzy set of all models. Greater membership values indicate higher confidence in allocation pattern to cluster (Thomas & Nashipudimath, 2012).

• ghazanfari et al (1389) studied segmenting customers in the export of garments based on clustering. The purpose of that research was segmentation of countries based on the value of garment exports performed during the period of 14 years. They used k-means clustering algorithm to cluster the countries. After segmentation of countries was segmented into six parts: loyal customers, the average value, the high value, new, and turned off. Using RFM model to determine the value of each cluster and policies tailored to each part is explained.

• Izadparast et al (1391) in their study focused on "Using data mining techniques to predict the damage level of car insurance customers". The purpose of that research was to predict risk in each category.

• Lim and colleagues (Lim et al., 2013) in a study on "segmentation using fuzzy clustering Luxury buyers" dealt with 500 consumers of luxury brands in the transaction on the basis of demographic variables (age, income, marital status, employment status) behavioral characteristics (affordable trends in luxury shopping, buying satisfaction, loyalty to the brand in the luxury brand, Objectives of buy in the future). Customers were classified into four clusters that include bargain hunters, buyers please talented, active and buyers royal affordable luxury buyers. The study used clustering software and EM algorithm to search through the 2329 questionnaire survey.

• Lee and Jia Yan (Yan-li & Jia., 2012) examined credit card data of bank customers using neural networks to explore the 141,562 records to study 2180 credit customers on a one-year period. Assess variables such as the ID card number (No.) final card, money transactions, the settlement, the serial number, system serial number, date of transaction, the expiration date, settlement date, the number of business, commercial enterprise, date of birth, gender, age , marital status, education, occupation, amount of the transaction is conducted monthly income.

3. METHODOLOGY

The method of this study, in terms of purpose is applied and in terms of data collection method is descriptive. Data was collected from the database of Pasargad life insurance. There were 1071 life insurance contracts in the period from April to October 2014. To this end, the proposed general steps include defining variables, data collection, data cleansing, and data analysis and model design is more fully described on.

Step 1: Definition of variables: one of the most important factors in making the model very carefully, choosing the right parameters. The first and most important step is to identify segmentation variables. For this study, the library study was used to identify variables affecting life insurance and then based on the insured's profile, characteristics that are used for the purpose were specified.

Step 2: Data collection: for every customer the data was extracted in this phase as a set of rows and columns of data into the Excel software. So that the columns represent attributes (as variables). The number of rows is insured persons.

Step 3: Data preprocessing: data processing one of the important steps in data mining so that the accuracy of the results has a strong association with how to do it. In this study we deleted wrong data and converted all amounts of text data into numerical values.

Step 4: Design and construction of model: First, the specifications obtained from the modeling and, ultimately, the analysis showed that clusters used to separate customers. Here FCM method is used to build the model.

FCM is one of the most popular clustering techniques. FCM clustering involves two main steps: 1. Compute cluster centers and the allocation of points to the center using the Euclidean distance. The process is repeated continuously to stabilize the cluster centers. The algorithm has a membership value to items for clusters in the range of 0 to 1 and a fuzzy parameter in the range [1, n] which determines the degree of fuzziness of the cluster. FCM clustering algorithm provides a method that enables data items belonging to two or more clusters, and most of this plan can be used in pattern recognition. This is based on minimizing the objective function:

$$J_{m_f} = \sum_{=1}^{N} \sum_{k=1}^{c} \mu_{jk}^{mf} \left\| x_j - c_k \right\|^2$$

Where

$1 \leq m_f < \infty:$ *a real number*

μ_{jk} *the membership* deg *ree of* x_j *to* $k - th$ *cluster*

x_j $j - th$ *sample*

C_k *center of* $k - th$ *cluster*

In FCM the U matrix follows the below condition:

$$\sum_{j=1}^{c} \mu_{jk=} 1, \forall \; k = 1, \dots , n$$

Fuzzy segmentation is done through optimal iterations to update u_{jk} and cluster centers c_k.

$$u_{jk} = \cfrac{1}{\sum_{p=1}^{c}\left[\cfrac{\|x_j - c_k\|}{\|x_j - c_p\|}\right]^{\frac{2}{m-1}}}$$

and

$$c_k = \cfrac{\sum_{j=1}^{N}\mu_{jk}^{fm}.x_k}{\sum_{j=1}^{N}\mu_{jk}^{m}}$$

Steps in FCM algorithm: (Grover 2014)

Step 1: Create a membership matrix U

Step 2: Calculation of fuzzy cluster centers

Step 3: Calculation of objective function and doing iterations to reach the optimum

Step 4: Calculate the new membership matrix U

Step 5: Stop the procedure if $\quad \| U^{k+1} - U^{k} \| < \varepsilon$

4. INTRODUCING ATTRIBUTES (VARIABLES)

To select appropriate characteristics in this study, first we examined the characteristics of each contract. At this stage a number of characteristics such as (name, surname, place of birth, etc ...) personal data and behavioral characteristics were determined. The features that are not effective against the target variable was deleted according to the interviews and advice of insurance experts. Table 1 shows some of variables used in the analysis.

Table 1: Influencing variables in segmentation

Variable	Values
Age	Different age groups
Gender	Male, Female
Children	Number of children
Marital Status	Single, Married
Job	Student, Jobless, Worker, Self-employed, physician, Government jobs
relationship between insured and insurer	Self-insured, Spouse, daughter or son, other
Insurance term	5 to 30 years
Payment method	monthly, 3-month, 6- month, yearly
Premium	The value of premium in Rials
Number of Supplementary Coverage	1,2,3,4

5. IMPLEMENTATION MODEL USING FCLUST FUNCTION

In this study, we used FCLUST function in version 3.2.2 of R statistical software on 1071 cases. In this section the results of fuzzy clustering algorithm is investigated. The purpose of fuzzy clustering is dividing m subjects into k clusters by calculating and determining the probability (factor) of membership of each subject to different clusters. Thus generalized optimum value for the customer clustering should be monitored carefully. To run the algorithm in R the measure of Euclidean distance was used with a fuzzy parameter ranging from 1.1 to 4. The number of clusters is determined by trial and error.

We started with 21 variables and ended up with 12 effective variables. The optimum number of clusters by trial and error was set to 2. Table 2 shows some cases and their membership value to the two clusters.

Table 2: The membership degree of insured people in clusters using Fclust Program

	Clus 1	Clus 2
Obj 1	0.36	0.64
Obj 2	0.31	0.69
Obj 3	0.52	0.48
Obj 4	0.44	0.56
Obj 5	0.66	0.34
Obj 6	0.39	0.61
Obj 7	0.69	0.31
Obj 8	0.43	0.57
Obj 9	0.64	0.36
Obj 10	0.63	0.37
Obj 11	0.35	0.65
Obj 12	0.34	0.66
Obj 13	0.42	0.58
Obj 14	0.33	0.67
Obj 15	0.29	0.71
Obj 16	0.59	0.41
Obj 17	0.43	0.57
Obj 18	0.48	0.52
Obj 19	0.66	0.34
Obj 20	0.69	0.31

It is obvious that for each object(case) the sum of membership values for all clusters is n1. i.e.

$$\sum_{j=1}^{c} \mu_{jk} = 1, \forall \ k = 1, \ldots, n$$

Based on our analysis the profiles of the 2 clusters are as follows:

Profile of clients in 2- cluster analysis:
First cluster: investment policy

•The cluster contains 36 percent of the cases.
• 60 percent of children under 9 years of age.
• 51% of them are women.
• 89% of them are single.
• 94% of people do not have children.
• 47% of them are students.
• 77% have the insurance period of 30 years.
• 56 percent are on monthly payment method.

• 65 percent of the premium is under 99000 Rials.
• 25% of the final capital 100 to 499 million Rials
• 85 percent of people insured by policyholders are children.

Therefore, the first cluster consists of customers that are students who do not have any source of income to pay the premium. This cluster is called investment policy.

Second cluster: life security

•This cluster contains 64 percent of the cases
•41 percent have age between 20 to 29 years
•67 percent are married
•51 percent have no children
•44 percent are self-employed
•67 percent have 1 to 3 supplement coverage
•70 percent have an insurance term of 30 years

So, the second cluster of clients consists of married men who are self-employed and 70% have the insurance period of 30 years. The final capital outlay is high. This cluster is called life security.

Table 3: Cluster centers in fuzzy clusters output

	family	illness	weight	height	user	cover	fire	inv.life	annual	life
Clus 1	0.09	0.16	-0.52	-0.48	0.45	-0.39	0.01	-0.19	0.03	-0.07
Clus 2	-0.04	-0.10	0.38	0.35	-0.35	0.31	-0.03	0.16	-0.04	0.08

	The.premium	payment	insurance	Term.	job	Ratio	marry	children	gender
Clus1	-0.01	0.04	0.02	-0.53	0.55	-0.38	-0.23	0.15	
Clus 2	-0.01	-0.05	0.01	0.41	-0.42	0.30	0.17	-0.14	

	Age	location	final.capital
Clus1	-0.49	-0.03	0.00
Clus 2	0.36	0.02	-0.01

6. DISCUSSION AND CONCLUSION

The computational results based on the results of the implementation of the model, can be considered to improve the segmentation of customers. Using segmentation techniques to identify clients if justified by the results of the classification can be effective in the profitability of the insurance industry. In other words, segmentation should have the performance that forms homogeneous groups of customers to facilitate the process of identifying and extracting variables and ultimately improve sales guarantee. The results of this phase showed that in addition to the desired characteristics such as the ratio of insured to policyholders, investors final capital, the number of supplementary coverage, demographic segmentation of customers have a major impact. This analysis results in some implications as follows:

• Organizations can distinguish different groups and different customers in the form of simple, effective and reasonable for detailed information about potential customers and existing customers offering value assessment. From the perspective of customer management, such an approach would not only help in crisis management, but also can promote the customer value and enable significant savings in cost effective customer service is using data mining for the life insurance industry.

• Another interesting topics in the field of fuzzy logic and fuzzy look is applied to life insurance customers and other research studies, less fuzzy look insured customers in industry and

classical clustering each customer belongs to one cluster But the cluster of fuzzy membership degree of customer can belong to cluster the results carefully examined.

  • Using clustering to study a larger number of variables at the same time, the impact of market segmentation is provided. In other words, the confidence in the results for decision making in real terms is higher.

  • Individual marketing assert that the marketing strategy must be focused on the customer's individual specifications. In accordance with the same subject, most insurance companies will use experience-oriented model, because they have a lot of customers, as well as many of their properties to be considered. This is what Izadaparast et al (2011 and Hanafizadeh and rastkhyz (2012) discussed in their study. The research was done in life insurance. It is suggested that the clustering system may be used financial services, including auto, property and liability insurance companies and public and private businesses. Due to the remarkable results obtained from the use of fuzzy clustering, it can be applied in market researches.

References

[1]   Iyzadparast S. M., Farahi A., Fathnejad F. and Pourteimour B. (2011) Using data mining techniques to predict the damage level of car insurance customers. Journal of Information Processing and Management (former Information Science and Technology), 27(3):699-722.

[2]   Hanafizadeh P. and Rastkhyz Paydar N. (2012) Comparison of two methods of data mining segmentation car insurance customers by risk (Case Study: Insurance Company nations). Industrial Management Studies, 11(30): 77-97.

[3]   Ghazanfari M., Malek Mohammadi S., Alizadeh S. and Fathollah M. (2009) Segmentation of customers in the export of garments based on clustering. Journal of Business Research, 14(56): 59-86.

[4]   Karimi A.yatollah (2012) General insurance (second edition). Tehran, Islamic Republic of Iran Central Insurance, Insurance Institute.

[5]   Bae S. M, Ha S. H., and Park S. Ch. (2005) A web-based system for analyzing the voices of call center customers in the service industry, Expert Systems with Applications, 28(1): 29-41.

[6]   Bose I. and Chen Xi. (2015) Detecting the Migration of Customers of Mobile Services Using Fuzzy Clustering, Information & Managment 52(2): 227-238.

[7]   Casabayó M., Agell N. and Sánchez-Hernández G. (2015) Improved market segmentation by fuzzifying crisp clusters: A case study of the energy market in Spain". Expert Systems with Applications, 42(3):1637-1643.

[8]   Grover N.(2014) A study of various Fuzzy Clustering Algorithms. International Journal of Engineering Research (IJER), 3(3): 177-181.

[9]   Hiziroglu A. (2013) Soft computing applications in customer segmentation: State-of-art review and critique, Expert Systems with Applications, 40(16): 6491-6507.

[10] Hong Ch. W. (2012) Using the Taguchi method for effective market segmentation, Expert Systems with Applications, 39(5): 5451-5459.

[11]   Mirza Hosseini H., Mahmoudi Maymand M., Karimii O. and Ahmadinejad M. (2013) Predicting the bank customer switching based on data mining technique, Spectrum: A Journal of Multidisciplinary Research, 2(10):8-17.

[12] Liang Y. H. (2010) Integration of data mining technologies to analyze customer value for the automotive maintenance industry, Expert Systems with Applications, 37(12): 7489-7496.

[13] Lim Ch. M., Kim Y. K. and Runyan R. (2013) Segmenting luxe-bargain shoppers using a fuzzy clustering method, International Journal of Retail & Distribution Management, 41(11/12):848-868.

[14] Lin J. B., L. T. H. and Lee Y. G. (2012) Mining Important Association Rules on Different Customer Potential Value Segments for Life Insurance Database, Granular Computing (GrC), 2012 IEEE International Conference on Granular Computing, 283-288.

[15] RAMANATHAN, K.V. (2012) A study on policyholder's satisfaction with reference to Life Insurance Corporation of India, Thanjavur division, Doctor of Philosophy in Commerce, Bharathidasan University, Tiruchirappalli.

[16] Singh A. and Rana A. (2013) Mining of Customer data in an Automobile Industry using Clustering Techniques, International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS). 5(3): 251-258.

[17] Sithic H. L. and Balasubramanian T. (2013) Survey of Insurance Fraud Detection Using Data Mining Techniques, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2(3): 62-65.

[18] Thomas B. and Nashipudimath M. (2012) Comparative Analysis Of Fuzzy Clustering Algorithms In Data Mining, International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), 1(7): 221-225.

[19] Umamaheswari K. and Janakiraman S. (2014) Role of Data mining in Insurance Industry, COMPUSOFT, An international journal of advanced computer technology, 3(6): 961-966.

[20] Wu R. Sh. and Chou P. H. (2011) Customer segmentation of multiple category data in e-commerce using a soft-clustering approach, Electronic Commerce Research and Applications, 10(3): 331-341.

[21] Yan-li Z. and Jia Z.(2012) Research on Data Preprocessing In Credit Card Consuming Behavior Mining, Energy Procedia, 17: 638-643.

[22] Yoo I., Alafaireet P., Marinov M., Pena-Hernandez K., Gopidi R., Chang J. F. and Hua L. (2012) Data Mining in Healthcare and Biomedicine: A Survey of the Literature, Journal of medical system, 36(4): 2431-2448.