

Overview of Data Mining techniques for CRM management at Insurance Broker

Fethi ATA^{1,*} and Lyazzat Atymtayeva²

¹ Arel University, Istanbul, Turkey

² Suleyman Demirel University, Almaty, Kazakhstan

Received: 2 Nov. 2017, Revised: 24 Dec. 2017, Accepted: 28 Dec. 2017

Published online: 1 Jan. 2018

Abstract: In today's world, hard conditions in the market lead the companies to find new ways to be competeable. With the intensive global competition and rapidly changing technological environments, meeting customers' various needs and maximizing the value of profitable customers are becoming the only viable option for many contemporary companies. Together with technological developments, companies and institutions constantly store customer and sales data. By applying data mining techniques, the companies may obtain valuable, meaningful, potentially useful and valuable information from the data analysis, which is unknown beforehand. Among the data mining techniques we can distinguish the clustering and associative rules mining as the most used efficient techniques for data based analysis. This paper is devoted to research and overview of the techniques and methods for the development of data mining application in an insurance brokerage company based on effective analysis of the customer relationship management activities meanwhile the customer master data and sales transactions can be converted to meaningful information. In this concern, data mining application can be developed to segment processes among customers and products, and to find links between them.

Keywords: Customer relationship management, insurance brokerage, data mining, clustering analysis, associative rules

1 Introduction

Insurance Industry has been subjected to fundamental changes the last decade. The insurance industry has shifted from traditional insurance to next generation insurance. In the insurance industry, fixed products were offered to the customers in the past, but nowadays this technique does not work effectively. The reason is in changing of marketing offers and that customers are looking for special products, so the companies have to propose the offers which customer really needs. Therefore, well knowing of the customers base in order to propose the suitable products becomes very crucial issue for any insurance company. In this case the customers segmentation becomes one of the efficient techniques that include the information about the relations between customers and products as well as relations between products and products.

The most efficient techniques for customers segmentation can be data mining techniques that include the clustering analysis and associative rules mining. Clustering analysis is a method of data mining that is

used to group similarities within the data, while Associative Rules identify the rules for similar events. Clustering analysis is an exploratory analysis that tries to identify structures within the data. Clustering analysis is also called segmentation analysis or taxonomy analysis. More specifically, it tries to identify homogenous groups of cases if the grouping is not previously known. Associative rule mining is a procedure which is directed to findings of frequent patterns, correlations, associations, or causal structures from the data sets that are found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories. By using the clustering analysis we can easily make segmentation of customers. And the associative rules can give us good opportunity to find out the relation between the products.

This paper is devoted to the description of main data mining techniques for customer segmentation and overview of researches in this area. The next sections give some analysis of different issues at the area of insurance companies services improvement, applying of data mining techniques, and getting valuable results.

2 Problem statement

As it's known, for the insurance industry, the premiums of customers are the main profit sources and the settlement of claims are the major expenditures, so the premium profit and the claim risk are the two important aspects for insurance companies for evaluating customers [2]. But for a insurance broker we should not take claim risk that is not an issue of insurance broker's bussiness claim risk. Insurance broker's task is just only find the right customers for the right insurance products. For solving these tasks the modern technologies allow us using the data mining techniques.

Data mining techniques can help the insurance firms for taking crucial business decisions. The insurance sector is primarily dependent on customer's base. The most appropriate scenario of any insurance firm is based on the effective management of customer data. Data mining techniques are intended for the support of the market specialists in the decision making process. With the help of data mining techniques, the customer data may be processed more accurate. Insurance industry companies usually collect huge amount of data about their customers. Due to protective regulations, the information extracting process from the database occupies a lot of time. In this situation, data mining technique is very helpful for easy data access [3].

Based on the researches of Dr. S. Janakiraman [3] the following techniques can be used for report generation. They provide better decision making for large volume of customer database by using the visualization tools. This solution may provide more functionality in business decision making. For solving the business problems and making decision, this data mining techniques can help to select the appropriate techniques for the organization.

Data mining techniques takes a vital role in the entire applications domain especially in insurance sector. It can be highlighted the importance and role of data mining techniques are useful in insurance sector for managing the customer data and gain business advantage. It can enhance an insurance business process.

The customers and products segmentation as a part of data mining techniques can help to process the data more efficiently.

The most challenging part in segmentation research, however, is the identification of appropriate variables to make reliable predictions of peoples' future buying behavior. For instance, author argues that demographics such as an age and an income are not optimal since a correlation between these variables and the motivations behind the decision of purchase might be very fuzzy [6]. Author tries to offer insights on how customers are clustered, which parameters belong to customers used in these clustering processes, and which clustering algorithms are used based on their value. The researchers have used the different methods to calculate the value of individual customer to make rankings of individual clients or segments or even predictions of the value, as it

Table 1: How data mining techniques will add value to insurance data [3]

Data Mining Techniques	Patterns
Clustering	Customer with the similar characteristics.
	Analysis of customer attribution in the insurance sector.
	Policy that is the most likely to be used, the most unlikely to be used.
	Segments that are related to policy.
	Predicting consumer behavior.
Classification & Prediction	Predicting the likelihood of policy success.
	Classifying the historical customer records.
	Prediction of what type of policy most likely to be retained, most likely to be left.
	Predicting insurance product behavior and attitude.
	Predicting the performance progress of segments throughout the performance period.
Association	Prediction to find what factors will attract new avenues in Insurance sector.
	Classify trends of movements through the organization for successful /unsuccessful customer historical records.
	Discovery of such association that promotes business technique
Summarization	Provides summary information.
	Various multidimensional summary reports.
	Statistical summary information.

can be found in Identification of customer groups in the German market: a benefit segmentation [6]. We can find the different using of segmentation techniques in the following studies:

- Customer profitability forecasting using Big Data analytics: A case study of the insurance industry [2]
- Customer segmentation with purchase channels and media touchpoints using single source panel data [5]
- Customer segmentation of multiple category data in e-commerce using a soft-clustering approach [8]
- How does customer self-construal moderate CRM value creation chain? [4] and at the other papers.

3 Related studies overview

This section describes the findings and research methodologies of different studies in the area customers segmentation techniques for the needs of insurance companies.

Florian Schreiber [6] in his paper “Identification of customer groups in the German term life market: a benefit

segmentation” have been focused on the suggestion that a consequent alignment of important strategic decisions related to product innovations, pricing, and distribution channels to the certain identified consumer segments enables incumbents to maintain a stable and sustainable market share and profitability. This suggestion confirms the crucial role of the segmentation and its value in the process of data mining.

The authors Kuangnan Fang, Yefei Jiang and Malin Song in their paper “Customer profitability forecasting using Big Data analytics: A case study of the insurance industry” in September 2016 [2] have also been claimed that the method which considers historical purchasing behavior and the foreseeable future cash flow can measure the real insurance customer contribution more effectively.

Satoshi Nakano and Fumiyo N. Kondo in their studies “Customer segmentation with purchase channels and media touchpoints using single source panel data” [5] emphasize that customers can be clustered according to the their purchase channels such as online stores, media touchpoints of PC, mobile, and social media, and psychographic and demographic characteristics.

The researchers Rounq-Shiunn Wu and Po-Hsuan Chou in their paper “Customer segmentation of multiple category data in e-commerce using a soft-clustering approach” [8] have summarized the analysis of online consumers segmentation into multiple categories that may contribute to a better understanding and characterization of purchasing behavior in the electronic commerce market. They have considered the fact that online shopping databases consist of multiple kinds of data on customer purchasing activity and demographic characteristics, as well as consumption attributes such as Internet usage and satisfaction with services. Therefore, information about customers covered by segmentation enables to company administrators to establish good customer relations and refine their marketing strategies to match customer expectations.

The authors Jia-Yin Qi, Qi-Xing Qu and Yong-Pin Zhou in the study “How does customer self-construal moderate CRM value creation chain?” [4] have found that the most of the existing literature on CRM value chain creation has been focused on the impact of customer satisfaction and customer loyalty to customer profitability. However, a little has been studied about the individual customer level in the CRM value creation chain and the role of self-construal level (i.e., independent self-construal level or interdependent self-construal level) in such a chain. They aim in their research to construct the chain from customer value to organization value (i.e., customer satisfaction, customer loyalty, patronage behavior and etc.) and investigate the moderating effect of self-construal.

By summarizing the overview of researches in the field of data segmentation for Customer relationship management by the needs of insurance companies we can distinguish the following Descriptive Sample Characteristics for customer attributes that were defined

in some of the papers. The most meaningful customer attributes were highlighted as the following:

- Age
- Gender
- Smoker
- Marital Status
- Claim amount
- Place Residence - Region
- Level of Education
- Net Household Income
- Occupation
- Insurance Premium
- Payment Type
- Number of Family
- Number of Childeren

Besides these sample characteristics the authors of this paper can define the following list of special additional attributes that have a value in terms of their using by the certain Insurance Brokerage Company:

- Customer Age
- Sales Repeatition
- Product Group
- Sales Repeatition for a product

Table 2 shows how often the mentioned characteristics have been found in the reviewed publications.

4 Proposed Methods and Data Mining tools

Based on the researches analysis mentioned above as the most appropriate Data Mining tools for the solving of the segmentation problem we can consider two methods:

- Clustering Analysis
- Associative Rules

Clustering Analysis can show how customer may have similar characteristics and how customer segments are related to the common policy of segmentation. The Associative Rules Mining may demonstrate the relation between insurance products as well as may find out the association that can promote business technique.

The main methodologies in Clustering Analysis include K-Means, X-means, EM, and other methods. The Associative Rules consists of different algorithms. Among them are Appriori, HP(Direct Hashing and Pruning), FP-growth algorithm and others.

Cluster analysis is an exploratory analysis that tries to identify structures within the data. Cluster analysis is also called segmentation analysis or taxonomy analysis. More specifically, it tries to identify homogenous groups of cases if the grouping is not previously known. Because it is exploratory, it does not make any distinction between dependent and independent variables [7].

Association rule mining is a procedure which is intended to find the frequent patterns, correlations,

Table 2: Descriptive Sample Characteristics vs publications

	Authors	[6]	[2]	[5]	[8]	[4]
	Age	X	X	X	X	X
Descriptive Sample Characteristics (Attributes)	Gender	X	X	X	X	X
	Smoker	X				
	Marital Status	X			X	
	Claim amount		X			
	Place					
	Residence - Region	X				
	Level of Education	X		X		
	Net Household Income	X			X	
	Occupation		X			X
	Insurance Premium Payment Type	X				
	Number of Family	X		X		X
	Number of Children	X		X		
	Customer					
	Age					
	Sales Repetation				X	
Some additional Attributes (have been added in the current researches)	Product Group		X			
	Sales Repetation for a product					

associations, or causal structures from data sets that were found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories.

Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness.

Based on the concept of strong rules, Rakesh Agrawal, Tomasz Imielinski and Arun Swami [1] introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule that was found in the sales data of an insurance company would indicate that if a customer buys car insurance service and house insurance services together, they are likely to also buy health insurance service. Such information can be used as the basis for

decisions about marketing activities such as, e.g., promotional pricing or product placements.

In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usage mining, intrusion detection, continuous production, and bioinformatics. In contrast with sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions.

5 Discussion

Based on the analysis of techniques above as well as on the own researches we can propose some hypotheses for the solving of segmentation problems in an insurance brokerage company.

The first hypothesis is to discover some associative rules in sales data, and then make some promotion campaigns about these insurance products.

The second hypothesis is to make clustering analysis of the customers of brokerage insurance company according to some their attributes or items. Then we can predict the cluster for the new customer as well as we can suggest some insurance products related to its own cluster.

The third hypothesis is to find the most valuable customer for the company and contacting with him at the first order. This allows to track such customers and don't let the valuable customers get away.

All hypotheses that are described above can increase customer loyalty and satisfaction. All these efforts aim to provide customer lines continuity and increase the number of customers in the long term. At the same time it is for creating new sales opportunities by discovering connections between customers and products.

In the future, as the technology progresses very fast, the data collected by intelligent chipers in vehicle and article fuses will be analyzed more accurately, and more accurate estimations will be made to calculate the correct contribution share of the person or property. For example, with a chip placed on a vehicle or an application installed on a person's phone, the risk of accidental recording of that person's driving behavior can be accurately measured. Those who have a high risk of accidents will have their insurance policy costs too low for those who have the right to do so correctly or those who have a low accident risk.

6 Conclusion

In this paper the clustering and customer segmentation in the insurance sector and the studies on attributes, data mining methods used in the past and customer segmentation and clustering were investigated. As a result of this study we have discovered that there have not been

held the previous researches for an Insurance Broker Company in this area. We had a cooperation with Sigortayeri insurance broker company in their decision to use the clustering analysis and customer segmentation that can be done by adding additional customer attributes to make efficient the process of optimal segmentation. These newly added attributes can be defined as very important parameters in determining the loyalty and value of the customer.

In all these studies we had the purposes to segment the existing customers in the system, clustering according to the newly added attributes in addition to the preused attributes and to get a better recognition of the customer. For better known customers, it is extremely important and vital to be able offering the suitable products in order to better recognize the customer-product relationship and to offer them special offers, campaigns and having additional sales opportunities.

Acknowledgement

This research was supported by Sigortayeri Insurance Brokerage Company, subsidiary of Aktif Investment Bank. We would like to thank colleagues from Information Technology Team of Aktif Investment Bank who provided insight and expertise that greatly assisted the research.

The special thank to Mr. Cüneyt Tanrıverdi, general manager of Sigortayeri Insurance Brokerage Company. We also want to thank Managers, Mr. Mehmet Eligür, Managing Director of Dijital Channel Application Development Department, Dr. Uğur Gökhan Özdiç, Information Technologies and Operations Executive Vice President in Aktif Investment Bank for comments that greatly improved the manuscript.

References

- [1] Dogan, B. (2014). Sigortacilik Sektorunde Musteri iliskileri Yontemi icin Birliktelik Kurali Kullanilmasi. Marmara Fen Bilimleri Dergisi, 107.
- [2] Fang, K., Jiang, Y., and Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. Computers and Industrial Engineering, 2.
- [3] Janakiraman, D. S., and Umamaheswari, K. (2014). Role of Data mining in Insurance Industry. An international journal of advanced computer technology-Compusoft-Issn:2320-0790, 1.
- [4] Qi, J.-Y., Qu, Q.-X., and Zhou, Y.-P. (2014). How does customer self-construal moderate CRM value creation chain? Electronic Commerce Research and Applications, 1.
- [5] Satoshi Nakano, F. N. (2018). Customer segmentation with purchase channels and media touchpoints using single source panel data. Journal of Retailing and Consumer Services, 1.
- [6] Schreiber, F. (2017). Identification of customer groups in the German term life market: a benefit segmentation. Springer Science+Business Media New York, 2.
- [7] StatisticsSolutions. (2018, 04 17). /cluster-analysis-2/. Retrieved from StatisticsSolutions: <http://www.statisticssolutions.com/cluster-analysis-2/>
- [8] Rounq-Shiunn, W., and Po-Hsuan, C. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. Electronic Commerce Research and Applications, 2.



Fethi Ata received the master's degree in Computer and Control Systems at Technical Education Faculty of Marmara University in Istanbul, Turkey. Now, He is studying Engineering Management at Industrial Engineering Department of Arel Univeritesi in Istanbul, Turkey. His research interests Machine Learning, Artificial Intelligence and Data Mining.



Lyazzat Atymtayeva received the PhD and Doctor of Science degree in Mechanics, Mathematics and Computer Science at al-Farabi Kazakh National University, Kazakhstan. Her research interests are in the areas of mechanics, applied mathematics and computer science including the numerical and rigorous mathematical methods and models for mechanical engineering and computer science, intelligent and expert systems in Information Security, Project Management and HCI. She has published research papers in reputed international journals of mathematical and computer sciences. She is reviewer and editor of international journals in mathematics and information sciences.