A New Approach of Using Association Rule Mining in Customer Complaint Management

Behrouz Minaei-Bidgoli, Elham Akhondzadeh

[1] Department of computer engineering, University of Science and Technology
Tehran, Iran

[2] Department of industrial engineering, Shahed University
Tehran, Iran

Abstract

A new approach of using data mining tools for customer complaint management is presented in this paper. The association rule mining technique is applied to discover the relationship between different groups of citizens and different kinds of complainers. The data refers to citizens' complaints from the performance of municipality of Tehran, the capital of Iran. Analyzing these rules, make it possible for the municipality managers to find out the causes of complaints, so, it leads to facilitate engineering changes accordingly. The idea of contrast association rules is also applied to identify the attributes characterizing patterns of complaints occurrence among various groups of citizens. The results would enable the municipality to optimize its services.

Keywords: *association rule, customer complaint management, e-government, data mining.*

1. Introduction

With the advent of information and communications technology (ICT), many governments have been promoting using information systems to deliver services and information to citizens effectively [1], [2]. Some researchers have studied the adoption of IT and technological innovations in public sector [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. Using such electronic means in governments is defined as e-government that facilitates providing services and helps governments to be more citizen-oriented [1], [2].

One of the topics that is often included in this domain, is the adoption of Customer Relationship Management (CRM) in public sector. King (2007) [14] introduced CRM as a key e-government enabler that uses ICT to collect and store data which can then be used to discover valuable knowledge about customers. Schellong (2005) [15] introduced the concept of CiRM as a part of new NewPublicManagement (NPM) that is included in the area

of e-government [13]. CiRM applies CRM in the private sector focusing on citizens and uses IT to improve relationship with them [1], [13], [15], [16], [17].

Call centers and citizens' voice databases are among the common ways to connect between the government and citizens [1], [13], [16]. Furthermore, the databases could be analyzed in order to improve government policies [1]. One of the effective techniques that can be used in analyzing the customer data is data mining that have been widely used in this domain [18]. Many researchers have focused on using data mining tools in CRM and according to Ngai et al. (2009) [18], the research on it will increase in the future.

In this paper, focusing on CiRM, we use data mining tools to extract useful information from a citizens' complaints database in Iran. Customer complaint management is one of the elements of customer retention that is crucial for successful management of customers' needs [18]. According to a survey research [18] on using data mining techniques in CRM, only a few studies (two papers) have examined the application of data mining tools in complaint management. These two papers which were conducted by Bae et al. (2005) [19] and LarivieRe and Poel (2005) [20], discussed the use of clustering and sequence discovery respectively. Compared to the other elements of customer retention, the number of papers regarding the application of data mining in complaint management (2 out of 54 papers) is low and further researches are needed to be carried out on this field [18].

In this paper, a new approach of using data mining tools in complaint management is presented by applying the association rule techniques. The apriori algorithm is implemented to discover the associations between different variables and the type of complaints.

Several researchers believe that behavioral variables are better predictors in comparing with the non-behavioral

variables [21], [22], [23]. Some kinds of these variables such as demographic and cultural characteristics are also considered in this study. Citizens are grouped according to the variables and the idea of contrast association rule mining is applied to compare the citizens' complaints according to their behavioral characteristics.

The remaining of this paper is organized as follows: in section 2, the association rule and the apriori algorithm is discussed. The application of association rule and results are shown in section 3. Finally, section 4 includes the concluding remarks.

2. Methodology

In this section, we discuss the association rule and apriori algorithm that are used in analyzing the customer data.

2.1. Association rule

The concept of association rule is originated from the market basket analysis which aims to identify frequent itemsets in a supermarket shopping [24], [25], [26].

The main idea is to find products that are frequently bought together for improving marketing strategies [24]. Besides the market basket analysis, cross selling programs are also introduced as another common use of association rule mining. This technique is ranked after neural network and decision tree among the data mining tools which have been used for CRM [18].

An association rule is presented in the form A → B, where $A \cap B = \phi$. A is the antecedent of the rule and B is its consequent. It means that when A occurs, B occurs with certain probability. The strength of association rules is often evaluated by means of the support and confidence measures. These two measures are defined as follows: [27].

$$Support(A, B) = Count(A, B) / The\ total\ number\ of\ records \quad (1)$$

$$Confidence(A \rightarrow B) = Count(A, B) / Count(A) \quad (2)$$

2-2- Apriori Algorithm

Apriori is one of the common tools [18] and the basic algorithm for finding frequent itemsets [28]. This algorithm was proposed by Agrawal and R. Srikant in 1994 [28] and is the first technique that controls the exponential growth of candidate itemsets. It uses the support for pruning candidate itemsets [27].

This algorithm is composed of two major steps, the generating frequent itemsets and finding the association rules from found frequent itemsets. At the first phase, it finds the itemsets that have the minimum support threshold. At the second step, the minimum confidence threshold is applied to select the rules [27].

2.3. Implementation of association rule mining in complaint management

In this section, the mining process is described. Firstly, we extract the association rules among the geographical location in which citizen lives, time fields and the subject of complaint in order to find out the correlation between these fields and type of complaint. By analyzing these associations, the municipality managers can find the primary subjects and the causes of complaints in different regions. The analyzing process will be more explained in section 3.

Next, we apply the idea of contrast association rule mining in the data in order to compare the type of complaints of different citizen groups according to their demographical and cultural characteristics. The idea of contrast association rules was presented by Minaei-Bidgoli et al. (2004) [29]. The main purpose of mining such rules is to discover interesting differences between segments of a population. In this study, citizens are categorized according to gender and education as demographical variables and cultural characteristics. For example, to find out the impact of gender of citizens on the type of complaint, the dataset is divided into male and female subsets. After that, we try to find those rules with the same antecedent and consequent but notable difference in support and confidence measures. By investigating these rules, we can thus understand similarities and conflicts between the requirements of males and females.

Both of educational and cultural characteristics could be examined in a similar manner using the same approach. The dataset is divided into graduated and non-graduated subsets in the case of education variable while it is categorized into 9 groups according to experts' opinions in the case of cultural characteristics.

3. Case study

In this section, the association rule mining is applied to discover the relationship between different citizens and the type of their complaints in the case of Tehran municipality.

3.1. Research data and preprocessing

The citizens' data of a call center system in Iran are analyzed. The system belongs to the Tehran municipality. Its mission is to evaluate the performance of municipality. Citizens can call to this center and thank, criticize, complaint or give their opinions about the performance of managers, employees and different operating units of the municipality.

IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010
ISSN (Online): 1694-0814
www.IJCSI.org

144

The data includes 190135 records of citizens' messages from spring 2008 to spring 2010. An example of the data is shown in table 1. Each row of this table provides information of a particular message. The definition of fields is represented in table 2. The initial analysis of the data implies that 90 percent of the citizens' calls relate to their complaints which we concentrate on in this study.

After handling the duplicate data, missing values and outliers, the "season" feature was created.

Table 1: an example of data

Subject ID	Message Time	Citizen Region	Citizen Gender	Citizen Education
301	2008-11-25 15:38:50	4	Male	Master
299	2008-10-13 13:57:13	22	Male	Bachelor
1040	2009-10-20 08:40:00	2	Female	Student
1038	2010-02-24 11:12:00	1	Male	Student

Table 2: the definitions of fields

ID	Attribute	definition
1	Subject ID	The subject of citizen call
2	Message Time	The time of citizen call
3	Citizen Region	The geographical location of where the citizen lives
4	Citizen Gender	Male or Female
5	Citizen Education	The degree of citizen education such as BSc, MSc, PhD.

3.2. Association rule mining

In this section, we discuss the results of the presented mining process.

3.2.1. Region, season and type of complaint

As discussed before, the association rules between 'citizen region', 'season' and 'subject' have been extracted to find out the correlation between geographical location, time and the type of complaint. The confidence level is set to 5%. Choosing a very low minimum confidence make it possible to find those kinds of complaints that occurred with a low confidence. By comparing these rules with the ones that have a high confidence, ranking the subjects based on their importance is possible. Analyzing these rules, we can find those kinds of subjects that cause complaint more than the others or the ones that are not important for citizens.

In analyzing the rules in the above manner, if a subject occurred in a particular region with a confidence more than 70%, it can be notable as the primary factors that cause complaints in that region.

Examples of the association rules are shown in table 3. CR, Se and S in the antecedent and consequent explanation stand for "citizen region", "season" and "subject" respectively. The fourth and fifth columns show the support and confidence metrics, respectively.

For example, the first three rules represent the occurrence confidence of three kinds of complaints in region 10 which are more important for citizens living in this region. The first rule implies that more than 70% of citizens' complaints in this region are against the urban service management system of the municipality. The occurrence confidence of this subject is more than 70% and can be noted as the primary factor that causes complaints in this region. The administrative-financial and delinquency of employees and managers are detected as the second priority subjects. The 7% remaining complaints refer to other issues which have been occurred with a confidence less than 5.

The next three rules state the occurrence confidence of complaints in region 11. As shown in the rows 4 and 5, most of the complaints are related to the administrative-financial matters but with a little difference with the urban service management system. 10% of the complaints refer to the contractors of the municipality. Obviously, the 33% remaining complaints occurred with a low confidence and are not important for citizens in this region.

In a like manner, the rules 7-11 show the occurrence confidence of different kinds of complaints in region 22 in spring.

Table 3: the results of association rules between "citizen region", "season" and "subject" attributes

ID	Antecedent	Consequent	S (%)	C (%)
1	CR = 10	S = the urban servic management system	70.7	7.77
2	CR = 10	S = Administrative-Financial matters	12	1.32
3	CR = 10	S = delinquency of employees and managers	10.3	1.14
4	CR = 11	S = Administrative-Financial matters	30.8	0.73
5	CR = 11	S = the urban servic management system	27	0.62
6	CR = 11	S = Contractors	10.3	0.24
7	CR = 22 and Se = spring	S = Administrative-Financial matters	50	0.18
8	CR = 22 and Se = spring	S = delinquency of employees and managers	25	0.1
9	CR = 22 and Se = spring	S = Transportation and traffic	8.3	0.03
10	CR = 22 and Se = spring	S = ICT	8.3	0.03

3.2.2. Demographical characteristics and type of complaints

In this section, we want to answer the question "is there any conflict between males and females' complaints or graduated and non-graduated citizens' complaints? As described in section 2, we answer to this question using contrast association rule mining.

Some of the rule which is obtained in this manner is presented in table 4. SE in the antecedent explanation stands for "citizen education". The other variables are defined as before.

A minimum confidence of 5% was chosen to obtain as many rules as possible.

Table 4: contrast rules in males and females

ID	Gender	Antecedent	Consequent	S (%)	C (%)
1	females	CR = 13 and CE = BS	S = social-cultural	1.4	80
2	males	CR = 13 and CE = BSc	S = social-cultural	0.31	15.6
3	males	CR = 22	S = Transportation and traffic	1.2	43
4	females	CR = 22	S = Transportation and traffic	1	9.6

The first two rules indicate that females with a Bachelor of Science (BSc) education degree living in region 13 complain on the social-cultural subjects with a high confidence, but males of this region with similar education degree feel displeasure on this subject with a low confidence. The support of the first rule is also more than second one. These two rules imply that there is a close correlation between gender and the complaints which relate to social- cultural subjects.

Analyzing the third and fourth rules make it possible to compare the importance of transportation and traffic subjects between males and females.

These rules are not unexpected and can validate the results to some extent.

In a like manner, the dataset has been divided into graduated and non- graduated subsets and the contrast rules have been extracted. Table 5 lists some of these rules. CG stands for citizen gender. Other variables are defined as before.

Based on the rules 1 and 2, there is a contrast between graduated males and non-graduated males' complaints in region 14. These rules show that the "delinquency employees and managers" subject for graduated males is more important than non-graduated males in this region, with a notable difference.

Table 5: contrast rules in graduated and non-graduated citizens

ID	Education	Antecedent	Consequent	S (%)	C (%)
1	Graduated	CR = 14 and CG = male	S = delinquency of employees and managers	0.89	66.7
2	Non-graduated	CR = 14 and CG = male	S = delinquency of employees and managers	0.89	10.8
3	Non-graduated	CR = 12 and CG = male	S = the urban service management system	0.89	80
4	Graduated	CR = 12 and CG = male	S = the urban service management system	0.21	15

There is a high difference between confidences of the rules 3 and 4. Non-graduated males complaint on the performance of urban service management system of municipality with a high confidence, but graduated males complaint with a lower confidence. This rule is neither expected nor unexpected and should be investigated by municipality experts.

3.2.3. Cultural variables and type of complaint

In order to investigate the correlation between cultural characteristics and type of citizens' complaints, dataset is segmented into 9 groups based on the experts' opinions. The segmentation is done by dividing geographical regions into different groups according to cultural characteristics.

Similar to the previous section, the association rules have been extracted. These rules are illustrated in table 6.

Analyzing these rules give us an idea about the type of complaints in different citizens' groups. For example rules 1-9 show that the most important subject for males with a BSc education degree in segments 3, 4 and 1 is the performance of urban service management system, while males with a BSc education degree in other segments complaint on this subject with a very low confidence.

More than 5000 contrast rules are obtained in this section. With such a great number of rules and different kinds of attributes, it is too difficult to check all the rules. But in the case of males with a BSc education degree, we can say that the occurrence patterns of different kinds of complaints are alike in segments 3, 4 and 1 with a high difference to other segments which are similar to each other.

IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010
ISSN (Online): 1694-0814
www.IJCSI.org

146

Table 6: contrast rules in different citizens' groups

ID	Segment	Antecedent	Consequent	S (%)	C (%)
1	3	CE = BSc and CG = male	S = The urban service management system	9.01	35
2	4	CE = BSc and CG = male	S = The urban service management system	6.44	34.2
3	1	CE = BSc and CG = male	S = The urban service management system	3.18	21.4
4-9	2,5,6,7,8,9	CE = BSc and CG = male	S = The urban service management system	Very low	Less than 1

4. Conclusion

Using association rule mining in customer complaint management is the main purpose of this paper. The data of citizens' complaints on Tehran municipality were analyzed. Using this technique made it possible to find the primary factors that cause complaints in different geographical regions in different seasons of the year.

The idea of contrast association rules was also applied to discover the variables that influence complaints occurrence. In order to accomplish this objective, citizens were grouped according to the demographical and cultural characteristics and the contrast association rules were extracted.

The results show that there is a strong relationship between citizen gender and education and patterns of complaints occurrence. Given have been focused on cultural characteristics, it is notable that some segments are alike, while other segments are similar. Applying this approach for CiRM shows an understandable way for clustering the data using the feature impact on finding similar clusters.

Acknowledgments

References

[1] T. Sasaki, , Y. A.Watanabe and K. Minamino,K, "An Empirical Study on Citizen Relationship Management in Japan", in PROC. PICMET conference of IEEE Xplore., 2007, Portland, Oregon – USA.

[2] SH. Paul, "A case study of E-governance initiatives in India", The International Information & Library Review, Vol.39, 2007, pp.176–184.

[3] J. L. Brudney and S. C. Selden, "The adoption of innovation by smaller local governments: The case of computer technology", The American Review of Public, Administration, Vol.25, No.1, 1995, pp.71−86.

[4] J. L. Perry and K. L. Kraemer, Technological innovation in American local governments: The case of computing, New York: Pergamon Press, 1979.

[5] D. Bugler and S. I. Bretschneider, Technology push or program pull: Interest in new information technologies within public organizations. In B. Bozeman (Ed.), Public management: The state of the art, San Francisco: Jossey-Bass, 1993.

[6] C. C. Hinnant and J. A. O'Looney, "Examining pre-adoption interest in online innovations: An exploratory study of e-service personalization in the public sector", IEEE Transactions on Engineering Management, Vol.50, No.4, 2003, pp.436−447.

[7] N. Al-Qirim, "Personas of e-commerce adoption in small businesses in New Zealand", Journal of Electronic Commerce in Organizations, Vol.4, No.3, 2006, pp.17−45.

[8] F. Damanpour, "Organizational innovation: A meta-analysis of effects of determinants and moderators", Academy of Management Journal, Vol.34, No.3, 1991, pp.555−590.

[9] J. L. King, V. Gurbaxzni, K. L. Kraemer, F.W. McFarlan, K. S. Raman and C. S. Yap, "Institutional factors in information technology innovation", Information Systems Research, Vol.5, No.2, 1994, pp.139−169.

[10] K. Layne and J. Lee, "Developing fully function E-Government: A four stage model", Government Information Quarterly, Vol.18, No.1, 2001, pp.122−136.

[11] L. G. Tornatsky and M. Fleischer, The process of technological innovation, Lexington, MA: Lexington Books, 1990.

[12] M. M. Brown, "Understanding E-Government benefits", American Review of Public Administration, Vol.37, No.2, 2007, pp.178−197.

[13] C. G. Reddick, "The adoption of centralized customer service systems: A survey of local governments", Government Information Quarterly, Vol. 26, 2009, pp. 219–226.

[14] S. F. King, "Citizens as customers: Exploring the future of CRM in UK local government", Government Information Quarterly, Vol.24, 2007, pp. 47–63.

[15] A. Schellong, "CRM in the public sector: Towards a conceptual research framework", in Proceedings of the 2005 national conference on digital government research (Atlanta, Georgia, May 15–18, 2005.

[16] A. Schellong and T. Langenberg, "Managing citizen relationships in disasters": Hurricane Wilma, 311 and Miami-Dade County. In Proceedings of the 40th annual Hawaii international conference on system sciences, January 03–06, 2007.

[17] R. Silva and L. Batista, "Boosting government reputation through CRM", The International Journal of Public Sector Management, Vol. 20, No.7, 2007, pp.588-60.

[18] E.W.T. Ngai, L. Xiu and D.C.K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", journal of Expert Systems with Applications, Vol. 36, 2009, pp. 2592–2602.

[19] S. M. Bae, S. H. Ha and S. C. Park, "A web-based system for analyzing the voices of call center customers in the service industry", Expert Systems with Applications, Vol.28, 2005, pp.29–41.

[20] B. LarivieRe and D. V. D. Poel, "Investigating the post-complaint period by means of survival analysis", Expert Systems with Applications, Vol. 29, 2005, pp.667–677.

[21] C. M. Heilman, F. Kaefer and S. D. Ramenofsky, "Determining the appropriate amount of data for classifying customers for direct marketing purposes", Journal of Interactive Marketing, Vol.17, No.3, 2003, pp.5–28.

[22] P. E. Rossi, R. E. McCulloch and G. M. Allenby, "The value of purchase history data in target marketing", Marketing Science, Vol.15, No.4, 1996.

[23] A. Prinzie1, D.V.D. Poel, "Constrained optimization of data-mining problems to improve model performance: A direct-marketing application", Expert Systems with Applications, Vol.29, 2005, pp.630-640.

[24] J. Ahn and S. Young, "Customer pattern search for after sails service in manufacturing", Journal of Expert Systems with Applications, Vol.36, 2009, pp.5371–5375.

[25] R. Agrawal, and R. Srikant, "Fast algorithm for mining association rules", In Proceedings of the Twentieth International Conference on Very Large Databases, 1994, pp. 487–499.

[26] Q. Yang, L. Tianyi, and K. wang, "Building Association-Rule Based Sequential Classifiers for Web-Document Prediction", Data Mining and Knowledge Discovery, Vol.8, 2004, pp. 253–273.

[27] P. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining", Addison Wesley, ISBN: 0-321-32136-7, 2006.

[28] J. Han and M. Kamber, Data Mining: "Concepts and Techniques, Second Edition", Morgan Kaufman Publisher, 2006.

[29] B. Minaei-Bidgoli, P. Tan and W. Punch, "Mining interesting contrast rules for a web-based educational system", In International conference on machine learning applications, Los Angeles, California, 2004, pp.1–8.

Dr.Behrouz Minaei-Bidgoli obtained his Ph.D. degree from Michigan State University, East Lansing, Michigan, USA, in the field of Data Mining and Web-Based Educational Systems in Computer Science and Engineering Department. He is working as an assistant professor in Computer Engineering Department of Iran University of Science & Technology, Tehran, Iran. He is also leading at a Data and Text Mining research group in Computer Research Center of Islamic Sciences, NOOR co. developing large scale NLP and Text Mining projects for Persian and Arabic languages.

Elham Akhondzadeh received the M.Sc. degree in industrial engineering from university of Shahed. She is interested in data mining studies and its applications. Her current researches focus on using data mining in citizen relationship management.