

A Hybrid Data Mining Approach for Citizen Satisfaction Analysis

Amir Albadvi^a and Elham Akhondzadeh^{b*}

Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran

^aalbadvi@modares.ac.ir ^be.akhondzadeh.n@gmail.com

Abstract

As citizens' needs and expectations increase from citizen-centric governments, they have to plan special policies and services to improve citizens' satisfaction. In this way, many governments have been focused on citizen relationship management (CiRM) which applies the concept of CRM for the private sector to the public sector. The main purpose of CiRM is to change government-oriented management into citizen-oriented one in local governments. Furthermore, identifying citizens' needs and improving citizens' satisfaction are introduced as two important concepts in citizen relationship management. In this paper, focusing on the topic of the citizen satisfaction analysis, a hybrid data mining framework is presented and applied on the database of Tehran municipality. This data was gathered from the citizens' feedback on provided services and implies the citizens' satisfaction about the quality and the speed of services by the municipality. The methodology is composed of the association rule and clustering techniques. Firstly, a new approach of association rule mining is applied to quantify the value of citizen satisfaction. Secondly, a clustering approach is implemented on the satisfaction data to extract the subjects that cause dissatisfaction. The results show that the proposed approach can be beneficial in improving citizens' satisfaction. This approach could also be used in other cases in the domain of public sector.

Keywords: Citizen Relationship Management, Satisfaction analysis, Data mining, clustering and association rule.

1- Introduction

With the advent of citizen-oriented concept in the administration of the public sector, there has been a rise in citizen expectations. As citizens' needs and expectations increase, citizen-centric governments have to plan special policies and services to improve citizens' satisfaction. But, how the governments can implement the citizen-centric management and improve their services? Citizen Relationship Management (CiRM) responses to this question and enables the governments to be more effective and citizen-oriented (Kannabiran et al., 2004; Reddick, 2009 & Sasaki et al., 2007). CiRM, as a part of New Public Management (NPM), applies the concept of CRM in the private sector to the public sector to improve the citizen-government relationship (Schellong, 2005; Schellong & Langenberg, 2007; Silva & Batista, 2007, Reddick, 2009 & Sasaki et al., 2007). As mentioned by Sasaki et al. (2007), the main purpose of CiRM is to change government-oriented management into citizen-oriented one in local governments. Furthermore, identifying citizens' needs and improving citizens' satisfaction are introduced as two important concepts in citizen relationship management. Accordingly, many researchers have focused on the topic of citizen satisfaction management and the ways to improve it (Magoutas & Mentzas, 2010; Charnley & Engelbert, 2005; Hwang et al., 2005; Garcia & Cao, 2005 & Sasaki et al., 2007)

Most of these scientists used statistical methods and a few applied data mining tools. The capability of data mining has been proven in extracting hidden knowledge of customer data (Ngai et al., 2009) and it could be also useful in analyzing citizen data and citizen satisfaction management. A research done by Ahmadvand et al. (2010) shows that using data mining techniques provide managers with knowledge that helps them to develop citizens' satisfaction. In details, they used association rule mining to find the factors that affect the rate of satisfaction. However, the number of researches in the domain of data mining and citizen satisfaction management is very low.

In this paper, focusing on the topic of the citizen satisfaction analysis, a hybrid data mining framework is presented and applied on the data of citizens' satisfaction. This data was gathered from the citizens' feedback on provided services by the Tehran municipality and implies the citizens' satisfaction about the quality and the speed of providing services.

The rest of this paper is organized in the following. In section 2, we present the proposed methodology. In section 3, the implementation of the methodology on the database of the Tehran municipality is discussed. Finally, section 4 concludes the paper.

2 -Methodology

In this section, a hybrid data mining approach is presented in order to analyze citizen satisfaction. The methodology is composed of the association rule and clustering techniques. Firstly, a new approach of association rule mining is applied to analyze the citizens' satisfaction and to quantify the value of citizens' satisfaction. Secondly, a clustering approach is implemented on the satisfaction data to extract the items that cause dissatisfaction. A schema of the methodology steps is shown in Fig1. This methodology could be also used in customer satisfaction analysis in other cases in the domain of public sector. A brief discussion of the steps is as follows.

Step 1 is the preprocessing operation which includes data selection, data cleaning and handling the missing values.

Step 2 is the association rule mining. It contains a new approach of using association rule technique in order to analyze the citizens' satisfaction and to quantify the satisfaction values. In details, the frequent patterns between the fields "Category", "Quality ID" and "Speed ID" are extracted by the Apriori algorithm and a primary satisfaction analysis is done. The values of satisfaction, is also quantized based on the results.

Step 3 relates to the clustering of citizens' needs based on the "Quality value" and "Speed value" which show the value of citizens' satisfaction about the quality and the speed of providing services by the municipality. The main purpose of this stage is to find the subjects that cause dissatisfaction.

Finally, in step 4, the interpretation of the clusters is presented and the results are analyzed. In the next section, the implementation of the methodology is discussed step by step.

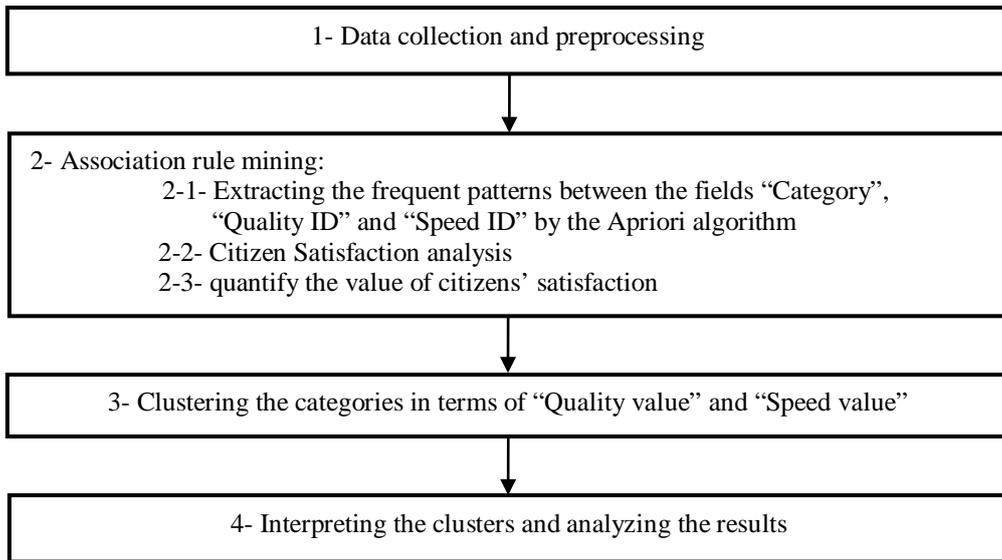


Fig1: the research methodology

3- Case study

In this section, the implementation of the methodology on the satisfaction data of Tehran municipality is discussed. The result of each step is as follows:

Step1: As mentioned before, the data includes the information on citizens' satisfaction about the quality and the speed of services provided by the municipality. In details, this data contains the feedback of 201108 citizens and shows the levels of citizens' satisfaction. A sample of the data is shown in Table1. The fields of Quality ID and Speed ID indicate the level of satisfaction. These attributes vary from 1-3 with respecting values of “low”, “medium” and “high”. The “category ID” field indicates different types of citizens' needs such as “cleaning”, “transportation”, “parks and green lands”, “pathways and roads”, “washing the urban equipments” and so on.

The process of cleaning data of duplicate, excess, redundant and noise was done in this step. The missing values were also eliminated. Finally, the total number of records decreased to 175838.

Table1: an example of data

feedback ID	Category ID	Quality ID	Speed ID
156	28	2	1
45	5	3	2
13456	7	3	3

Step2: As mentioned before, the association rules between the fields “Category”, “Quality ID” and “Speed ID” were extracted by the Apriori algorithm and a primary satisfaction analysis was done based on the results. The minimum confidence was set to 1% and the minimum antecedent support for each category was identified according to the opinion of experts of the municipality. It is notable that a distinct minimum antecedent support is given for each category.

The antecedent support metric determines the fraction of the feedbacks on each category respect to the total number of records. Accordingly, the minimum antecedent support indicates the least necessary feedbacks in order to gain a fairly idea about the citizens' opinion

of being satisfied or not. The confidence metric determines how often items in the consequent part appear in transactions that contain the antecedent part (Tan & Steinbach, 2006). Accordingly, in this case, the confidence index indicates the fraction of citizens that their opinion was equal to the value of the consequent part.

An example of the rules is also shown in Table2. For example, the first three rules show the opinion of citizens about the quality of providing services which refers to the category 14 that relates to the “parks and green lands” needs. According to the above explanation about the confidence, it is obvious that the level of satisfaction for 57.524% of citizens is equal to high. Respectively, 27.306% and 15.17% of citizens evaluate the quality at “medium” and “low” levels. In other words, the confidence of being “high”, “medium” and “low” levels for the “parks and green lands” category is 0.57524, 0.27306 and 0.1517 respectively.

Based on the previous results, a new approach is presented to quantify the value of satisfaction. Accordingly, the values of confidence are multiplied to the respective values of “Quality ID” and sum of them is defined as the average score of satisfaction. So, a mean value of satisfaction could be calculated for each category. For example, the value of citizens’ satisfaction about the quality of services that relate to the “parks and green lands” is equal to 2.42. Respectively, the “Speed value” is equal to 2.53. So, the “Quality ID” and “Speed ID” are substituted with “Quality value” and “Speed value”. It is obvious that these two new variables have a minimum of 1 and a maximum of 3.

Table2: An example of the association rules

ID	Antecedent	Consequent	Antecedent Support	Confidence
1	category = 14	Quality ID = 3.000	0.0469	0.57524
2	category = 14	Quality ID = 2.000	0.0469	0.27306
3	category = 14	Quality ID = 1.000	0.0469	0.1517
4	category = 14	Speed ID = 3.000	0.0469	0.64563
5	category = 14	Speed ID = 2.000	0.0469	0.24272
6	category = 14	Speed ID = 1.000	0.0469	0.11165

Step3: in this stage, different categories of citizens’ needs were clustered in terms of “Quality value” and “Speed value”. These two variables were normalized by the min-max normalization method. The K_means algorithm was implemented in order to cluster the items and the Silhouette index was used to investigate the validity of the results. The value of this index for different values of K was calculated. Based on the results, when K changes to 3, the value of the Silhouette index is maximized. So, the initial number of clusters was set to 3. The schema of clusters is shown in Fig2.

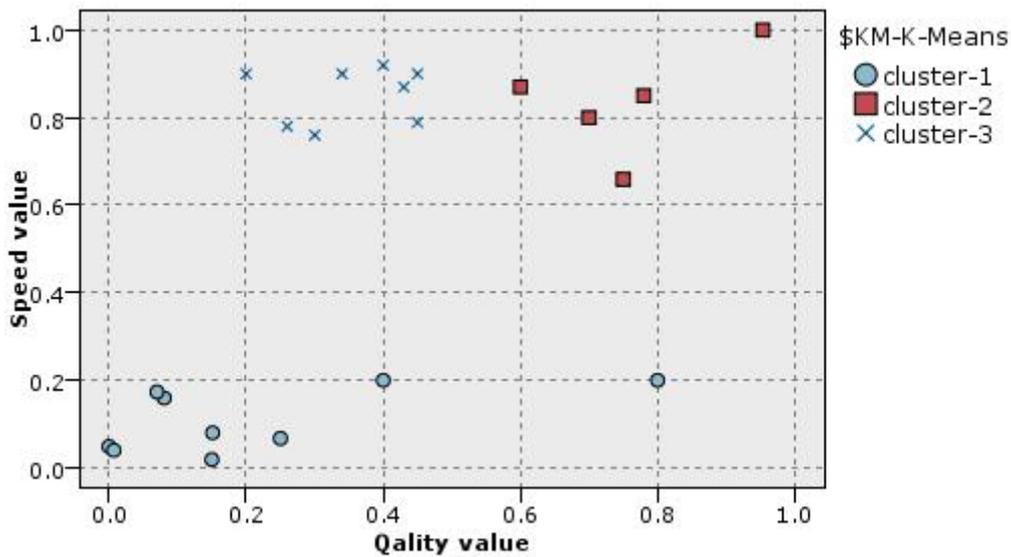


Fig2: The schema of clusters

Step4: According to Fig2, it is obvious that there is a group of needs that has a low value of satisfaction for the both of quality and speed of providing services. It seems that citizens feel satisfied with the services of cluster2, because the value of both quality and speed is high. Finally the items of cluster3 have a high value of speed but a low degree of quality. So, the municipality should have the top priority to the items of cluster1 and improve the quality and speed of providing such services. After that, improving the quality of items of cluster3 is necessary. According to the results, the first nine items that should be considered by the municipality are the “cleaning”, “sikes and streams”, “building and construction”, “pathways and roads”, “washing the urban equipments”, “asphalt”, “animals”, “municipality's employees” and “safety” subjects.

4- Conclusion

A hybrid approach of the association rule and clustering techniques is presented in this paper in order to analyze the citizens’ satisfaction. Firstly a new approach of association rule mining is proposed to analyze the citizens’ satisfaction and to quantify the value of satisfaction. Then, different kinds of citizens’ needs are clustered according to the value of quality and speed of providing services by the municipality. The results of clustering shows that there is a group of services that citizens fell dissatisfied with the both quality and speed of providing services and the municipality should give the top priority to these items. The results of the association rule mining and clustering phases could be applied by the municipality to improve citizens’ satisfaction. The proposed approach could also be used in other cases in the domain of public sector.

References

Ahmadvand, A.M., Minaei B., B. & Akhondzadeh, E. (2010). A hybrid data mining model for effective citizen relationship management: A case study on Tehran municipality, in

- Proceedings of 2010 International Conference on e-Education, e-Business, e-Management and e-Learning, Sanya, China, 22– 24 January 2010(277-281). IEEE Computer Society.
- Charnley, S. & Engelbert, B. (2005). Evaluating public participation in environmental decision-making: EPA's superfund community involvement program, *Journal of Environmental Management*, 77, 165–182.
- Garcia, V. & Cao, L. (2005). Race and satisfaction with the police in a small city, *Journal of Criminal Justice* 33, 191– 199.
- Hwang, E.G., McGarrell, E.F. & Benson, B.L. (2005). Public satisfaction with the South Korean Police: The effect of residential location in a rapidly industrializing nation, *Journal of Criminal Justice* 33, 585–599
- Kannabiran, G., Xavier, M. J., & Anantharaaj, A. (2004). Enabling E-Governance through citizen relationship management-concept, model and applications. *Journal of Services Research*, 4(2), 223–240.
- Magoutas, B. & Mentzas, G. (2010). SALT: A semantic adaptive framework for monitoring citizen satisfaction from e-government services, *Expert Systems with Applications*, 37, 4292–4300.
- Ngai, E.W.T., Xiu, Li & Chau. D.C.K., (2009). Application of data mining techniques in customer relationship management: A literature review and classification, *Expert Systems with Applications*, 36, 2592–2602.
- Reddick, C. G., (2009). The adoption of centralized customer service systems: A survey of local governments, *Government Information Quarterly*, 26, 219-226.
- Sasaki, T., A.Watanabe, Y. & Minamino K., (2007). An Empirical Study on Relationship Management in Japan. in *Proceedings of 2007 PICMET conference of IEEE Xplore.*, Portland, Oregon-USA.
- Silva, R., & Batista, L. (2007). Boosting government reputation through CRM. *The International Journal of Public Sector Management*, 20(7), 588-60.
- Schellong, A. (2005). CRM in the public sector: Towards a conceptual research framework. In *Proceedings of the 2005 national conference on digital government research*, 15–18 May 2005, Atlanta, Georgia.
- Schellong, A. & Langenberg, T., (2007). Managing citizen relationships in disasters: Hurricane Wilma, 311 and Miami-Dade County. In *Proceedings of the 40th annual Hawaii international conference on system sciences*.
- Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining* Addison Wesley, ISBN: 0-321-32136-7.