

EFFECTUAL CITIZEN RELATIONSHIP MANAGEMENT WITH DATA MINING TECHNIQUES

P. Pushpa Latha¹, R. Madhuri², K. Prasad Rao³, M. Vijaya Bharathi⁴

^{1, 2, 4}Asst.professor, CSE Department, GMR Institute of Technology, A.P, India,

³Asst.professor, CSE Department, Aditya Institute of Technology, A.P, India

pushp.mtech@gmail.com, ravi.madhuri5@gmail.com, prasadrao.karu@gmail.com, vijayabarathi.m@gmrit.org

Abstract

Effective knowledge sharing and learning is an important source of organizational performance. Citizens are a valuable source for any organization in terms of planning, providing and performance management. In this paper, we have developed a hybrid data mining model for providing effective citizen relationship with the government with respect to municipality services. The complaints of the citizens are taken from the Visakhapatnam Municipal Office as an input database in this paper. We developed a hybrid data mining model using clustering and association rules in combination. For finding high frequent complaints we used clustering technique K-Means and to determine which factors affect the satisfaction rate in those kinds of complaints we used Apriori algorithm.

Index Terms: CiRM, ICT, K-Means algorithm, Apriori algorithm

-----***-----

1. INTRODUCTION

Citizen Relationship Management (CiRM) draws from the commonly known concept of Customer Relationship Management (CRM) in the private sector, and is a cluster of management practices, communication channels and technological solutions to handle issues, problems, concerns, and demands of the citizen.

Since the numbers of governments who realize the importance of becoming more citizen-centering are increasing nowadays; in addition they discover that they must deliver authentic citizen knowledge across multiple governments departments and to all citizen touch points. This results in citizen relationship management.

The governmental service is one of the most primitive and predominant service domains in any community with a wide array of services. As such, urban services and the concept of citizen relationship management (CiRM) are currently gaining importance in this domain.

1.1 Definition

CiRM is defined as “a strategy and set of management practices, enabled by technology with a broad citizen focus, to maintain and optimize relationships and encourage new forms of citizen participation.” The main purpose of CiRM is to understand the needs of different citizen groups and to provide respective services for each group accordingly. In this way, many governments are actively promoting the use of ICT.

The main issues of this approach are "how e-governments can manage effectively" and "be more citizen-oriented". In other

words, ICT can be strategically significant in promoting e-government effectiveness through understanding the citizen requirements. In this domain, the application of data mining tools seems to be useful. Appropriate data mining tools, which are good at extracting and identifying useful information and knowledge from enormous customer databases, are one of the best supporting tools to obtain a deep understanding of citizen's characteristics and needs.

1.2 Information and Communication Technology (ICT)

ICT is an umbrella term that includes any communication device or application, encompassing: telephone, computer and networks .It also encompasses various services and applications associated with them, such as videoconferencing. ICT is a strategically significant in promoting CiRM through e-government services. ICT provides better e-government services by understanding citizen need.

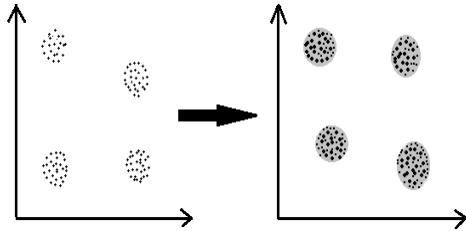
1.3 Customer Relationship Management

Citizen Relationship Management is the application of customer relationship management (CRM) in an e-governmental context and establishment guiding principles on how customer relations can be effectively strategized in the public sector. Specifically, it talks about the differences in CRM practices between public and private institutions.

1.4 Clustering

Clustering is the process of grouping a set of data objects into groups such that objects from the same cluster are more

similar and the objects from different clusters are more dissimilar.



1.5 Association

Association rule mining falls under the descriptive category. Association rules aims in extracting important correlation among the data items in the databases. An association rule is an implication expression of the form $X \rightarrow Y$ where X and Y are disjoint itemsets. In general, association rule mining can be viewed as a two-step process.

- (i) Generating all item sets having support factor greater than or equal to, the user specified minimum support.
- (ii) Generating all rules having the confidence factor greater than or equal to the user specified minimum confidence.

2. PROBLEM DESCRIPTION

The present using public sector systems are not able to provide citizen-oriented services. These systems are poor data system, difficulty in accessing data in the database and storing data in database. Using these systems we face difficulty in determining major complaint and frequent complaints. Using this approach government doesn't know either citizen is satisfied or not.

In this paper, we want to illuminate how data mining tools are useful to manage citizens' requirements and process which collect, analyze, reflect, and evaluate their needs. The main purpose of this approach is "how e-governments can manage effectively" and "be more citizen-oriented".

For solving this problem we consulted Visakhapatnam Municipal Office and enquired about the procedure for solving complaints posted by the citizens. After receiving adequate information from them we choose some of the basic services to develop.

We considered municipality services like complaints posted by the citizens. The complaints of the citizens are taken as our database. Then we have developed a hybrid data mining model, a combination using clustering and association rules for providing effective citizen relationship with the government.

We first find out the high frequent complaints in the database using clustering technique using K-Means algorithm. We took two factors which are as follows:

Frequency of complaints:

Frequency refers to the number of complaints of each subject during a season, for example 100 times during the winter.

Time interval between complaints:

Time interval refers to the interval between the first and the latest time that a complaint of each subject happened during a season, for example 22 days.

After finding high frequent complaints, we applied association rule mining using Apriori algorithm to determine which factors affect the satisfaction rate and in those kinds of complaints which occur permanently during the season with a high frequency.

3. ALGORITHMS

K-Means Algorithm

The k-means algorithm takes the input parameter, k , and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's centroid or center of gravity.

The k-means algorithm proceeds as follows. First, it randomly selects k of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster. This process iterates until the criterion function converges.

Algorithm

Input: The number of clusters k , and a database containing n objects.

Method: The k-means algorithm is implemented as follows:

- 1) arbitrarily choose k objects from D as the initial cluster centers;
- 2) repeat
- 3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster
- 4) update the cluster means, i.e., calculate the mean value of the objects for each cluster
- 5) until no change

The k-means method, however, can be applied only when the mean of a cluster is defined. This may not be the case in some applications, such as when data with categorical attributes are involved.

Apriori Algorithm:

For finding frequent patterns, associations, correlations, or causal structures among sets of items in transaction databases,

relational databases, and other information repositories we use an influential algorithm known as Apriori Algorithm.

Apriori algorithm is used in Basket data analysis, cross-marketing, catalog design, Loss-leader analysis, clustering, classification etc. In order to express the quality of an association rule one uses measures such as

Support: This metric determines how often a rule is satisfied in the transaction database. It is obtained by dividing the support count for $X \rightarrow Y$ by the total number of transactions.

Confidence: This metric determines how often items in Y appear in transactions that contain X .

Algorithm

Join Step: C_k is generated by joining L_{k-1} with itself

Prune Step: Any $(k-1)$ -itemset that is not frequent cannot be a subset of a frequent k -itemset

Pseudo-code

C_k : Candidate itemset of size k .

L_k : frequent itemset of size k .

$L_1 = \{\text{frequent items}\};$

for $(k = 1; L_k \neq \emptyset; k++)$ do begin

$C_{k+1} = \text{candidates generated from } L_k;$

for each transaction t in database

do

increment the count of all candidates in

C_{k+1} that are contained in t

$L_{k+1} = \text{candidates in } C_{k+1}$ with min support

end

return $\cup_k L_k$;

Generate Candidates

Suppose the items in L_{k-1} are listed in an order

Step 1: self-joining L_{k-1}

insert into C_k

select $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-1}$

from $L_{k-1} p, L_{k-1} q$

where $p.item_1 = q.item_1, \dots, p.item_{k-2} = q.item_{k-2},$

$p.item_{k-1} < q.item_{k-1}$

Step 2: pruning

For all itemsets c in C_k do

For all $(k-1)$ -subsets s of c do

if (s is not in L_{k-1}) then delete c from C_k

4. RESULTS

For our paper we considered municipality services like complaints posted by the citizens. The complaints of the citizens are taken as our database. Then we have developed a hybrid data mining model, a combination using clustering and association rules for providing effective citizen relationship with the government.

In this paper we have developed a hybrid data mining model which is the combination of clustering and association rules. We first find out the high frequent complaints in the database using clustering technique using K-Means algorithm. After finding high frequent complaints, we applied association rule mining using Apriori algorithm to determine which factors affect the satisfaction rate and in those kinds of complaints which occur permanently during the season with a high frequency.

The following are some of the results what have obtained:

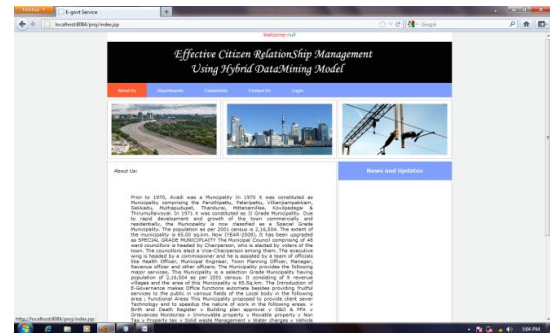
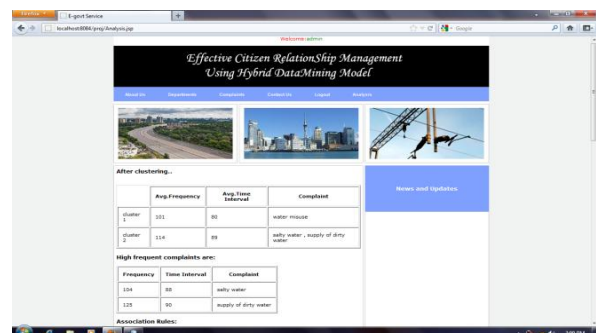


Figure 4.1 Home Page(Home page for CiRM)



Figure 4.2 Complaints Page(complaints received from citizens)



Antecedent	Consequent	Support%	Confidence%
salty water and member=13 and unit=4	satisfied	17.307692307692307	100.0
salty water and member=13 and unit=5	dissatisfied	10.576923076923077	91.66666666666666
salty water and member=13 and unit=6	satisfied	12.5	91.25
salty water and member=13 and unit=4	dissatisfied	11.538461538461538	100.0
salty water and member=13 and unit=5	satisfied	7.6923076923076925	98.88888888888889
salty water and member=13 and unit=6	satisfied	8.653846153846153	100.0
salty water and member=13 and unit=4	satisfied	10.576923076923077	91.66666666666666
salty water and member=13 and unit=5	dissatisfied	8.653846153846153	99.99999999999999
salty water and member=13 and unit=6	dissatisfied	4.807692307692308	81.25333333333334
supply of salty water	satisfied	17.538461538461538	100.0

Figure 4.3 Hybrid Data Mining Model

CONCLUSIONS

Governments have to consider convenient channels and services to connect between the governmental managers and citizens. Furthermore, data mining tools are very useful to manage citizens' requirements and process which collect, analyze, reflect, and evaluate their needs.

In this paper, we have used clustering and association rules on the data of the urban service management system in Visakhapatnam to find the subjects that cause complaint and the factors that affect the rate of satisfaction. Analyzing the rules, make it possible to understand the impact of factors such as time and responsible units on the rate of satisfaction. Besides, units with a perfect or worse performance in providing services and handling complaints are identified.

The results of the research are very beneficial in providing improved urban services and the development of citizens' satisfaction. This study could be notable as one of the first studies on using data mining tools in CiRM.

REFERENCES

- [1]. Ali Mohammad Ahmadvand, Behrooz Minaei Bidgoli, Elham Akhondzadeh "A hybrid data mining model for effective citizen relationship management: a case study on Tehran municipality".
- [2]. Ahn, J. and Young, S., "Customer pattern search for after-sales service in manufacturing". Journal of Expert Systems with Applications, vol.36, pp.5371–5375, 2009.
- [3]. Pan,S., Tan, C., Lim, E., "Customer relationship management (CRM) in e-government: a relational perspective", Decision Support Systems, vol. 42, pp. 237–250, 2006.
- [4]. Han, J. and Kamber, M., Data Mining: Concepts and Techniques, Second ed., Morgan Kaufman Publisher, 2006, pp. 383-407.
- [5]. Alexander Schellong and Thomas Langenberg, "Managing Citizen Relationships in Disasters: Hurricane Wilma, 311 and Miami-Dade County" Proceedings of the 40th Hawaii International Conference on System Sciences – 2007.

[6]. R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules, In Proc. 20th Int. Conf. Very Large Data Bases, VLD, Santiago, Chile, 1994, pp. 487-499.

BIOGRAPHIES



Pushpa Latha Palli, is M.Tech in Computer Science from Andhra University, A.P., India. She has vast experience in Computer Science and Engineering areas. She is presently working as Assistant Professor in Department of Computer Science and Engineering, GMRIT, Rajam, A.P, India. Her area of research includes Data Mining, Web Technologies and Emerging Technologies.



Madhuri Ravi, is M.Tech in Computer Science from JNTU, A.P., India. She has vast experience in Computer Science and Engineering areas. She is presently working as Assistant Professor in Department of Computer Science and Engineering, GMRIT, Rajam, A.P, India. Her area of research includes Data Mining, Computer Networks and Computer Architecture



Prasad Rao Karu, is M.Tech in Computer Science from JNTU, A.P., India. He has vast experience in Computer Science and Engineering areas. He is presently working as Assistant Professor in Department of Computer Science and Engineering, AITAM, Tekkali, A.P, India. His area of research includes Data Mining, Databases, Web Technologies.



Vijaya Bharathi Manjetti, is M.Tech in Computer Science from JNTU, A.P., India. She has vast experience in Computer Science and Engineering areas. She is presently working as Assistant Professor in Department of Computer Science and Engineering, GMRIT, Rajam, A.P, India. Her area of research includes Software Engineering, E-Commerce.