



AENSI Journals

## Advances in Environmental Biology

ISSN-1995-0756 EISSN-1998-1066

Journal home page: <http://www.aensiweb.com/AEB/>



### Provide a Framework for Text Mining in Systems Management Utilities to Run on the 137 System Tehran Municipality

Hourieh Abedpour and Dr. Abbas Toloie Eshlaghy

Department of Management and Economics, Islamic Azad University, Science and Research Branch, Saveh. Iran.

#### ARTICLE INFO

**Article history:**

Received 26 September 2014

Received in revised form 20 November 2014

Accepted 25 December 2014

Available online 2 January 2015

**Keywords:**

text mining, Data mining in texts and HTML documents, opinion mining, feeling mining, comment mining, Tehran Municipality.

#### ABSTRACT

This study presents a framework for text mining in systems management utilities to run on the 137 system for Tehran Municipality deals. Data needed for the investigation of the 30 recorded messages were collected on the 137 system. To identify the best model for text mining algorithms were used based on a system of 137 municipal experts and faculty. Many systems offer a simple prediction techniques are used. Another widely used technique is also called correlation-based methods. To explore what problems occur, the first way that comes to mind is using association rules. In this study, our aim is to get problems that occur together in a day in the winter. At this stage, the problem that we face is converting data into a format for many fields. Here we propose a way to use the a priori algorithm is a two stage process. The goal is too especially on a day to investigate problems that have occurred in the winter time, the issues with which are most likely to occur in winter. Data Analysis was done using MATLAB software, and Rapid Miner. The results of the analysis based on the problems of concurrency, most average problems linked to the 5 & 6 District 1. Major problems related to improper construction and non-construction of such principles, violation of privacy passages and crossings, as well as problems related to the park. On the other hand, the problems are related to each area of distribution. That's an area at a given date, there are many reports of problems.

© 2014 AENSI Publisher All rights reserved.

**To Cite This Article:** Hourieh Abedpour and Abbas Toloie Eshlaghy., Provide a framework for text mining in systems management utilities to run on the 137 system Tehran Municipality. *Adv. Environ. Biol.*, 8(25), 571-575, 2014

#### INTRODUCTION

To measure audience satisfaction and citizens some tools are used, such as interview techniques, questionnaire design, and like it. Although these methods can be useful, but it cannot cover all the comments folks. Due to the nature of theological opinion, not simply to analyze them as human error and personal interpretation is very effective. All these aspects into a uniquely powerful analysis tool makes it imperative for all audiences. To monitor and measure through semi-structured data (HTML), and unstructured (verbal) are also considered. Various methods and tools for text-analysis have been developed. Since the implementation of such methods requires the analysis of real samples as a case study, the system has 137 municipalities of Tehran. Sure dense text and verbal reports and appeals the judgment of this thesis is not only a test run on a specific area of the data, attempted to break the ambiguities and uncover the real needs and opinions, and they will be satisfied.

The main objectives of this study are as follows:

- A) The purpose of the research ideal optimization measures to detect, measure and monitor the level of citizen satisfaction with municipal services
- B) General Purpose: Provides information technology and text mining-based framework for monitoring citizen satisfaction from utilities

In this study, the main questions are:

- 1) What are the environmental parameters vary with environment? How significant is this?
- 2) How predefined indicators of increasing / decreasing the fact that the measures are effective? In other words, how we did ignore the facts of the use of such indicators, or contrary to its lower complexity?
- 3) What is the significant difference in the level of public satisfaction with the cleanliness of the streets of Tehran on the night there?

**Corresponding Author:** Hourieh Abedpour, Department of Management and Economics, Islamic Azad University, Science and Research Branch, Saveh. Iran.

- 4) 4) What types of problems vary from one neighborhood to another neighborhood?  
 5) What are the main problems in the area? What is the importance and priority of each?

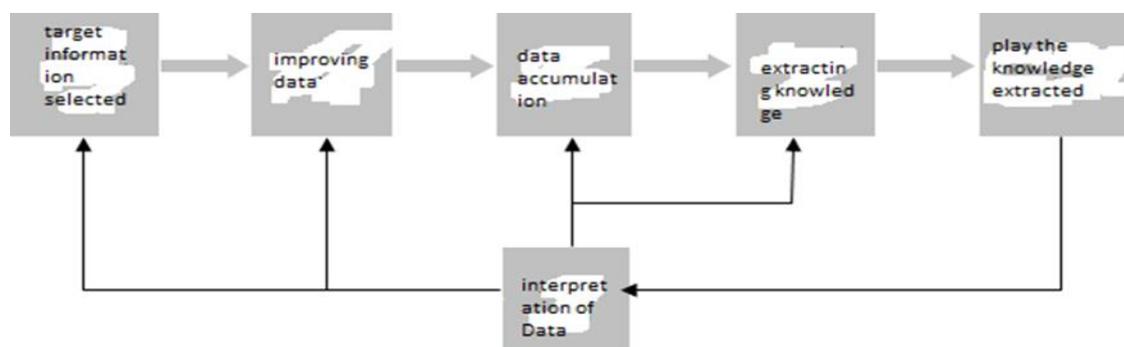
This research is done on the basis of library studies and practical, the researcher first reviewed the literature and specific techniques used in the field to be identified. The application of these methods in the study will determine the best ways in which they doubted. Then he needs the help of software tools based on experimental data carried out the study. Finally, the actual data collected from the 137 system Tehran Municipality, should be taken into account and the results of an empirical study on the real data are reported.

During the following steps are necessary for the successful implementation of this study:

- ✓ Identification Requirements
- ✓ Defining the problem
- ✓ Data collection and data
- ✓ Analysis of alternatives
- ✓ Actual implementation on real data

The main variables in this study can be named as follows:

- Identification and evaluation of recorded conversations and comments
- Categories identified and categorized on the basis of experts' texts
- Identified categories of classified documents based on methods text mining
- Identify current needs and how to respond to them.



**Fig. 11:** The data mining process.

The population included in this study of the recorded conversations text and comments submitted by the citizens of Tehran Municipality will provide a transcript of the 137 system. Based on the criteria that will be mentioned in the research and based on the nature of the problem and the study will attempt to determine sample size, sampling is random, but the performance will be determined.

Data mining software stored communications and transactions stored patterns based on the analysis of end-user queries. Various kinds of analysis and dental applications are available:

- 1-statistical
- 2-machine learning
- 3-neural network

*Mainly four types of connection are recommended:*

- 1-classes: the data stored in the application data can be assigned to predefined groups. For example, a restaurant chain could be given to making the customer buying mine when a customer comes into a restaurant, what kind would you order.
- 2- clusters: data items can be grouped in a logical connection or customer preferences. For example, the data can be used to define sections of the mine shop.
- 3- Association: The data can be used to define the association are, mine is.
- 4- Sequential pattern: the behavioral patterns of behavior can be predicted to be Mine.

The system database contains 137 useful content about the services provided to citizens and can be an important resource for identifying problems and needs of urban areas should be studied and explored. In this regard, this study tried to use a priori algorithm of association rules in time / space exploration of the occurrence of problems. Also using a two-stage technique of the algorithm, problems together and at the same time are identified in a specified time interval. The rules derived can be used in identifying the needs of citizens and improve the management of urban services.

$$x \Rightarrow y(\text{support}, \text{confidence})$$

$$\text{Confidence}(y | x) = \text{support}(x \cup y) / \text{support}(x)$$

Algorithms used in text mining:

To identify the best model for text mining algorithms are based on a system of 137 municipal experts and faculty were used. Many systems offer a simple prediction techniques are used. Another widely used technique is based on the correlation method is called. Some algorithms have recently been proposed for the system proposed, a statistic and machine learning are used.

- 1- correlation-based methods
- 2- Bayesian network model and a Bayesian Classifier
- 3- Neural network with feature reduction techniques
- 4- of Unison-CF
- 5- Material Content-Boosted Collaborative Filtering
- 6- Based collaborative filtering method is proposed algorithms item
- 7- Methods used to reduce the dimensions of the submitter systems
- 8- Ways to combine content-based and collaborative filters in an online newspaper
- 9- Items for Collaborative Filtering approaches
- 10- Association Rules

(2) Rules of classification algorithms

```

1   $F_1 = \{\text{large 1-rule items}\};$ 
2   $CAR_1 = \text{genRules}(F_1);$ 
3   $prCAR_1 = \text{pruneRules}(CAR_1);$ 
4  for ( $k = 2; F_{k-1} \neq \emptyset; k++$ ) do
5     $C_k = \text{candidateGen}(F_{k-1});$ 
6    for each data case  $d \in D$  do
7       $C_d = \text{ruleSubset}(C_k, d);$ 
8      for each candidate  $c \in C_d$  do
9         $c.\text{condsupCount}++;$ 
10       if  $d.\text{class} = c.\text{class}$  then  $c.\text{rulesupCount}++$ 
11     end
12   end
13    $F_k = \{c \in C_k \mid c.\text{rulesupCount} \geq \text{minsup}\};$ 
14    $CAR_k = \text{genRules}(F_k);$ 
15    $prCAR_k = \text{pruneRules}(CAR_k);$ 
16 end
17  $CARS = \bigcup_k CAR_k;$ 
18  $prCARS = \bigcup_k prCAR_k;$ 

```

Implemented fields

|            |                   |   |
|------------|-------------------|---|
| Subject ID | Number of message | 1 |
| Region     | Region            | 2 |
| Data       | Date              | 3 |
| Subject    | Subject           | 4 |
| Section    | Section           | 5 |

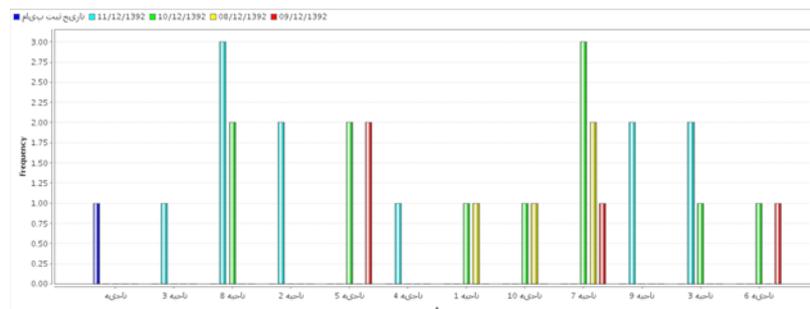
B → Message Subject

C → Address

D → Number of message

E → Date of Registration

Distribution histogram-based registration problems:



**Table 4-4:** Training set to extract the user association.

| Subject_ID | U1: satisfy | U1:dissatisfy | U2: satisfy | U2:dissatisfy | Target: satisfy |
|------------|-------------|---------------|-------------|---------------|-----------------|
| 100        | 1           | 0             | 0           | 1             | 0               |
| 201        | 1           | 0             | 1           | 0             | 0               |
| 202        | 0           | 1             | 1           | 0             | 1               |
| 300        | 1           | 0             | 0           | 1             | 0               |

[U1: dissatisfy], [U2: satisfy] → [target: satisfy].

**Table 4-5:** Set of training data to extract the Subject association.

| UserID | [S1:S] | [S2:S] | [S3:S] | [S4:S] | [S5:S] | [Target_S:S] |
|--------|--------|--------|--------|--------|--------|--------------|
| 1      | 1      | 0      | 0      | 0      | 1      | 0            |
| 2      | 1      | 0      | 1      | 1      | 0      | 1            |
| 3      | 0      | 1      | 0      | 1      | 0      | 0            |
| 4      | 0      | 1      | 0      | 0      | 1      | 1            |

S1-S2-S3S4-S5 → Subject 1-Subject2-Subject3-Subject4-Subject5

S → Satisfy

[S2: Satisfy] AND [S5: Satisfy] → [target\_S: Satisfy]

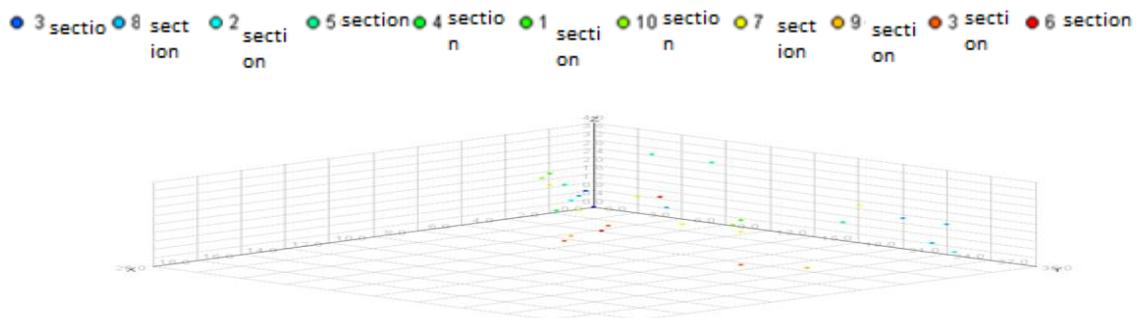
Score<sub>rule</sub> = Confidence<sub>rule</sub> \* Support<sub>rule</sub>

Coincidence detection of problems:

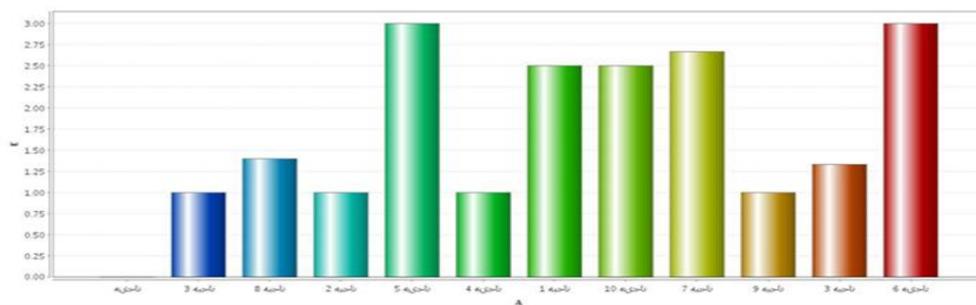
X → region

Y → message content

Z → Date of registration



Out of synchronization problems in different areas based on the date and Message Subject



**Table 4- 6:** Synchronization of problems in one day.

| Row | Problems that occur together in a day  |
|-----|--|
| 1   | "Unauthorized construction at the site will be done" and "«Joy to the dredging is needed"                                |
| 2   | "At this point we need to dry the waste tank" and «It will be tabulated stream restoration and repair"                   |
| 3   | "Untimely excavation of construction has been making a nuisance" and "Due to the lack of safety signs is likely to fall" |
| 4   | "There is a power to stop the material, the narrow street And "Plant and Tree Planting"                                  |

**Results:**

The results of the analysis based on the problems of concurrency, are most average problems linked to the 5 & 6 District 1. Major problems related to improper construction and non-construction such principles, violation of privacy passages and crossings, as well as problems related to the park. On the other hand, the problems are

related to each area of distribution. That's an area at a given date, there are many reports of problems. For example, in the winter of 2013, in the third issue on 11 "patch up holes in the roadway" and "unauthorized construction" are done that have little to do with each other. On the other hand, some problems can occur simultaneously in a given day, for example, the "untimely excavation of construction is burdening" and "there are no signs of safety could fall" happened in one day.

#### *Conclusion:*

In recent years a growing research in the field of data mining has been done by the researchers. From the point of view of electronics, data mining has the potential to reduce costs and gain competitive advantages for all the stakeholders: citizens and businesses. While the increasing use of data mining can provide many data management tools, organizations that take the initiative and use the limitations of the abuse and neglect of its policy implementation and mistakes are vulnerable. In other words, while data mining initiatives are evolving, issues and challenges related to the implementation and evaluation are political communities and the government to prevent harmful risks to be aware of them. Data mining can be applied on quantitative data, text, or multimedia. Its applications include the following (Two Crows Corporation, 1999):

- Association rules: a model in which there is an item indicating the existence of another item,
- Classification: assign patterns to a small set of pre-defined classes by finding some relationships between features,
- Clustering: Grouping customers or set of patterns that have similar characteristics,
- Prediction: logical patterns to predict the future,
- Path analysis and ordinal patterns: patterns where one event leads to another event.

#### *Suggestions:*

- 1- Since the model was a good choice of authentic and reliable, it is suggested that urban managers of the usefulness and ease of use as the citizens like the Technology 137 Strategies to improve communication with citizens and introduce a system.
- 2- One of the findings identified during the research process, are citizens of the pale with little knowledge of the system and also the diversity of services is recommended that are associated with pervasive propaganda efforts, especially must be performed through the mass media. Police rewarding experience in the preparation and distribution of national media advertising animation can introduce more systems to Tehran Municipality and in addition to introducing the variety of services offered to citizens and even for communicating with their education system, particularly through new technologies such as the internet and proprietary web systems is helpful.
- 3- Due to a growth in call statistics in past years, it is recommended that the Center is organized for comprehensive cooperation in relation to coordination with other service agencies in urban Administration and distribution of electricity, water and sewage, office supplies ... which makes the facility and the relief service of the city center and a telephone number (137) and conducted an Internet portal for Urban citizens, especially in times of disaster relief without worrying about remembering multiple phone just by calling 137 and describe their requests without delay. It is evident that in this case, the lower the costs of duplication and parallel working in the field of urban services will be reduced.
- 4-To draw attention to the increasing popularity of IT-137 system, it is suggested that the performance of reporting system of the type and amount of services performed periodically is provided and targets to be inclusive to all citizens savings rate and the results were due to the use of urban management system 137 which in turn leads to further growth of the city information. One way to increase understanding of the usefulness of a system 137 for citizen action appear to be advertising.

### **REFERENCES**

- [1] Moghimi, S.M., 2009. Organization and management research approach, Tehran: Termeh publication, Co.
- [2] Khaki, G.R., 2010. Methods to approach the dissertation, Tehran: the reflection.
- [3] Naderi, E. and M. Seifnaraghi, 2007. The research and how to evaluate it in the Humanities with an emphasis on behavioral sciences, Tehran: Badr publication, Co.
- [4] Zhuge, H. *et al.*, 2004. An Automatic Semantic Relationships Discovery Approach. The 13th International World Wide Web Conference (WWW2004), New York, USA.
- [5] Hearst, M.A., 1999. Untangling text data mining. In Proceedings of the ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics. University of Maryland, pp: 20-26.
- [6] <http://www.gate.ac.uk/sale/tao/index.html>.
- [7] Tan, P., M. Steinbach, V. Kumar, 2006. Introduction to Data Mining, Addison Wesley, ISBN: 0-321-32136-7.