

IMPROVING QUALITY OF FEEDBACK MECHANISM IN UN BY USING DATA MINING TECHNIQUES

Mohammed Alnajjar¹, Prof. Samy S. Abu Naser²

¹Faculty of Information Technology, The Islamic University of Gaza, Palestine

²Information Technology Department, Al-Azhar University, Gaza, Palestine

ABSTRACT

Data mining consists of evolving set of techniques that can be used to extract valuable information and knowledge from massive volumes of data. This paper highlights the data mining techniques applied to mine a group of received staff petitions through staff complaint mechanism in UN organization in Gaza.

Staff complaint mechanism (feedback mechanism) is a mechanism that supports the staff members. The employees send complaints / petitions that face them by filling the requested details including the request type (complaint, question, suggestion, others, beneficiary or not categorized).

The response unit receives the requests and categorize it, and finally prepares the reply then sends it back to the employee.

The important thing is: after the employee receives the reply, the employee should indicate whether he/she is satisfied with it or not. So the period between request date and reply date may affects the feedback since the regular reply period must not exceed 7 days after receiving any request.

In this paper we attempted to study and analyze the data received within 18 months. This data sample contains more than 6000 records that cover different employee requests types. Thus, we try to know the effects of delay of reply on the feedback received by complainant staff; furthermore, we need to know what kinds of staff that always ask questions and try to find the relationship between staff salary (according to grade) and nature of the request.

Thus this paper gives insight on how data mining can be applied in finding the consequences of staff requests and needs and hence try to minimize these requests over the time.

KEYWORDS

Feedback Mechanism, Complaint mechanism, data mining techniques.

1.INTRODUCTION

According to large number of records we have in the database, it has become increasingly necessary for users to utilize automated tools to find the desired information, and to track and analyze their usage patterns. These factors requires using intelligent systems that can effectively mine for knowledge.

Data mining can be defined as the discovery and analysis of useful information from the database that not known yet and can lead to useful actions that support goals of the organization as client satisfaction and quality of service improvement[6-10].

So to enhance the quality of services of staff complaint mechanism in the organization, data mining techniques can be used effectively to discover the rules those are unknown before and to improve the quality of services and decrease the cost.

Staff complaint mechanism supports the staff members with their complaints, since the employee send his petition that face him/her in the work or have a suggestion, the employee fills the request details and selects the request type: complaint, question, suggestion, or others. Once the response unit receives the request and categorize it to be: financial, social, managerial, leave, etc., a proper reply will be prepared and finally send back to the employee. The important thing is after receiving the reply, the employee set his feedback by indicating if the employee is satisfied or unsatisfied with the reply received. So, the period between request date and reply date may have an effect on the feedback.

In this paper we made an attempt to study and analyze the sample data received by response unit. The data sample contains large number of records that covers different employee types of requests. Thus, we tried to know the effects of delay of reply on the feedback of the staff; furthermore, we needed to know what are the kinds of staff that always ask questions and tried to find the relationship between staff salary (according to grade) and nature of the request.

Also we tried to discover useful relationships from that data that can help us to know the staff interests and try to minimize the questions and complaints.

Our goal of this analysis for the data is to :

- optimize the quality of service provided by the staff response unit in UNRWA – Gaza through discovering new knowledge from the existing data that can help us to know the factors that affect the quality
- discover new rules that help in prediction of staff needs and requests.
- know the limitations of the current feedback mechanism by discovering the errors and inconsistent data by using some data mining methods as outlier analysis.

To achieve these goals, a discussion of the different data mining methods are provided in this paper. Classification is used to build classification model to help us classify the new requests according some target classes as feedback and grade of employee. Association rules are used to discover some interesting rules that help us to take some useful actions. Clustering is used to discover the related data. The seldom pattern and errors are discovered by using outlier analysis.

2.RELATED WORKS

There are a few papers talking about using data mining techniques to mine the data of feedback mechanism, as:

In [1] , D. Jatin Das & S. Arun Kumar used K-mean clustering algorithm to know the groups of data and identify cluster quality, and also search the data in single site for long period of time by using any algorithm automatically to generate rating for that site.

In[2], Bettina Berendt used the data mining techniques to analyze the huge number of records that represent questions covers thousands of issues and described some of the problems found in the data set.

3.DATA SET

The dataset has 6062 record of the Employees petitions data received through the feedback mechanism in UNRWA institute – Gaza in the period between 2010 – 2011. It contains several cases of staff complains . In table1, it includes a description of selected field for the mining process and details of each attribute (description possible values).

Attribute Name	Attribute Description	Role
Serial	Request Id	ID
Subject	Title of the request	Regular
ReqType	Request type (complain, question, suggestion, others, beneficiary or not categorized)	Regular
Status	Request Status (pending – Replied)	Regular
ReqDate	Request Sent Data	Regular
RepDate	Request Replied Data	Regular
Hide name	For privacy, mark the request as hidden	Regular
ToDepName	The request related to specific department	Regular
ToUnitName	Nature of request (finance, medical , employment ..)	Regular
Emp Gender	Gender of the employee (F,M)	Regular
Emp Grade	Grade of the Employee (1 to 20) ,1: is lowest Grade , 20 : is the heights grade	Regular
Emp step	Steps count for the employee	Regular
Emp DOB	Date of birth for the employee	Regular
FeedBack	Feedback by the employee (Satisfied – Unsatisfied)	Label
Period	The period of receiving the reply	Regular

Table 1: Data Set Attributes

4.DATA MINING

This section, we describes some data mining method that we applied on the sample data and explain the results. We used pre-processing methods, association rules methods, classification methods, clustering methods and outlier analysis methods.

4.1 Data Pre- Processing

The data set contain more than 6000 records. It has some missing data attributes in about 210 records. Thus, we had to use a data cleaning method to handle the missing values. We used the Rapid miner program to import the dataset and used the following process to fill in the missing values in the data set:

- **Data Cleaning (Replace missing value):**

- We have replaced the missing values for the attributes by using **(average)**.

- **Data Discretization :**

- Because the period of reply (number of days) is important attribute, we used discretize by user specification to convert the values of it into two ranges:

- * ***InTime*** : when the period is less than 7 days

- * ***WithDelay***: when the period is more than 7 days

- **Data Discretization by Binning :**

For the grades 1..20, we divided them into three groups

- **Use Select Attributes:**

Remove un-useful attributes as EmpSteps, entry date

- **Data Reduction (Sampling)**

Because the records of the data set are large number, we need to make sampling to reduce the data set records from 6062 records to 2500 records.

- **Data Transformation:**

- **Nominal to binominal :**

We used this method to transform attributes (all attributes) from nominal to binominal.

- **Nominal to numeric:**

We used this method to transform attributes from nominal to numeric.

4.2 Association Rules:

We tried to discover new patterns from the available data set that we have, and tried to select the most interesting rules from the results that can lead to good actions contributes to improve the performance of the feedback mechanism.

Before applying association rule methods, we have applied some pre-processing for the data as discretization to convert the values to groups, and doing Data Transformation (Nominal to binominal) to transform all attributes from nominal to binominal to be suitable for association rules methods.

And then we used FP Growth method with min support = 0.95, calculate all frequent items sets from the data set. And then applied Create Association Rule operator with min confidence = 0.8 to generate a set of association rules for a given frequent item sets as shown in figure 3.

One of the resulted rules is [Emp Department = "education", Emp Grade = range2 [7.333 - 13.667], type = question] --> [FeedBack = Satisfied] (confidence: 0.809). That means most of request of type question for educational issues got FeedBack = Satisfied.

One of these rules is [Emp Department = "finance"]--> [Emp Grade = range2 [7.333 - 13.667]] (confidence: 0.818).

That means most of the employees in medium grades 7.. 13 interested in financial questions due to their salaries, since some of these question about getting loans or deductions and pay slip.

One of these rules is [Period = WithDelay] --> [FeedBack = Satisfied] (confidence: 0.822).

That means most of the request that take more than 7 days for reply got unsatisfied Feedbacks.

4.3 Classification

We have applied two methods to classify our data, these methods are: K-NN classification method & Naive Bayes classification method.

For K-NN classification method, to choose the best classification method, we used split validation to split data set into Training and Testing phases with split ratio = 0.5 and stratified sampling type.

We used For Training: K-NN with K = 4, and for Testing we used apply model & performance evaluator.

We got the best accuracy when K = 4, and split ratio = 0.5 and sampling type = stratified sampling, the accuracy was 77.67 % of our classification model.

Also we applied Naïve Bayes classification Method with split ratio = 0.6 and stratified sampling type, the resulted accuracy of our classification model was 77.50 %

Also we applied Decision tree method on our data , we got the tree as shown in figure 1.

We have noticed that the employees with grades more than 14 are not satisfied with the reply of their questions that sent to health department, that means they are not convinced with the quality of reply, but we noted the feedback of lower grades are sometimes satisfied that means that their questions may be simple questions .

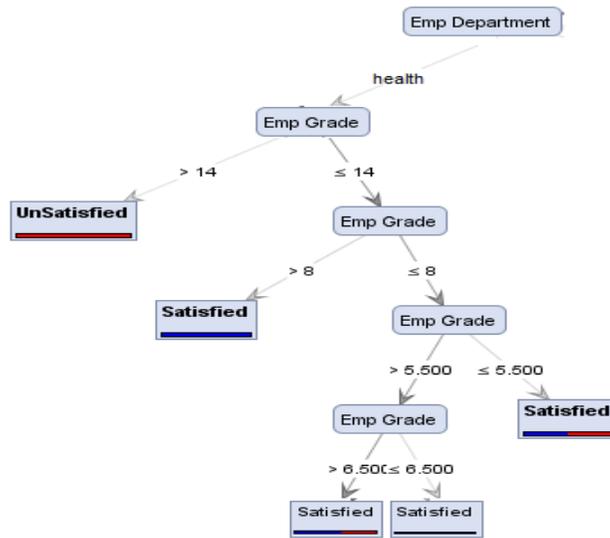


Figure 1: Decision Tree

4.4 Clustering

To know the groups of related data, we applied clustering on our data by using Clustering method[3,4] called K-Mean Method with $k = 4$, and max runs= 10, and max optimization steps = 100. We had a cluster model with four clusters. The data presented by using the Singular Value Decomposition (SVD) with two dimensions. In the result of SVD, we can view the result as Plot View with scatter plotter and the x-axis is svd_1 and the y-axis is svd_2 and color column by cluster as shown in figure 2.

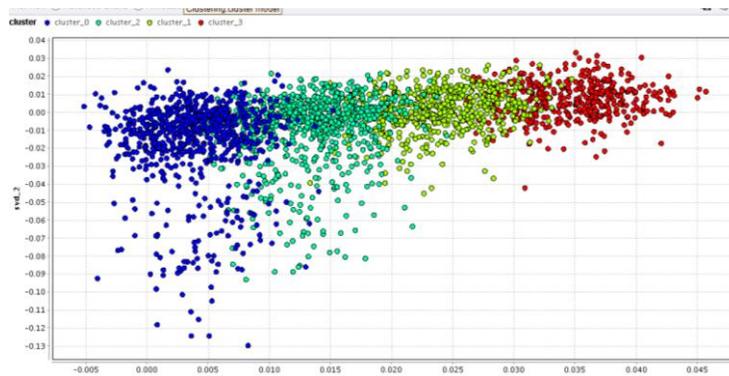


Figure 2: Resulted Clusters

We got four clusters that represent the groups of received petitions.

4.4.1 Explaining the results:

Where Cluster_0 presents the pending requests that still not replied for period between 12 days and more than 29 days for different type of questions. Cluster_1 presents the Replied requests, when the period of reply is less than 6 days and feedback is at most positive feedback (satisfied). Cluster_2 presents the Replied requests, when the period of reply is more than 14 days and feedback is at most negative feedback (unsatisfied). Cluster_3 presents the Replied requests, when the period of reply is more than 18 days and feedback is at most negative feedback (unsatisfied).

We can note from the resulted clusters that the feedback of the staff affected by period of reply. We think that the delay of reply leads to un-satisfaction feedback of the employee, and there are a large group of the requests still pending (not replied) and waiting for more than 6 days, that may leads to delay in response and thus un-satisfaction of the petitioners.

4.5 Outlier Analysis

We need to know the strange values in our data that may be an error or may be new discovered value to help us to take useful actions to improve the quality of our feedback mechanism. We applied a distance based outlier detection method[5] with number of neighbors equal to 10, and 10 outliers, with Euclidian distance function. The data presented by Using the SVD with two dimensions. In the result of SVD we can view the result as Plot View, by applying it we have the data with outliers as shown in figure 3.

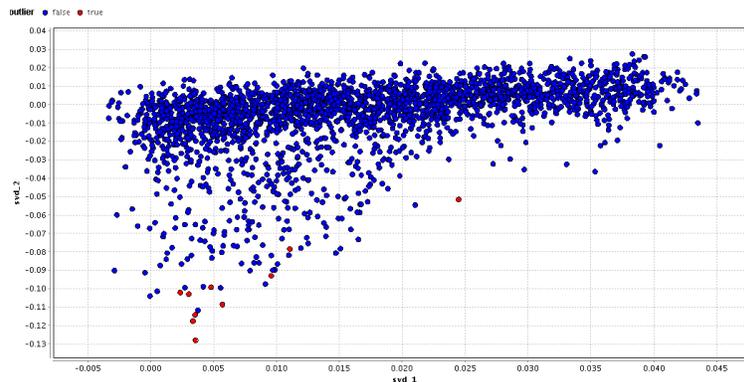


Figure 3: Results Of outlier Analyses

As a result of this method, the outlier represents the requests that are waiting for long time that is unusual, since some of these requests are waiting for more that 150 days and still pending not replied.

By using the outlier analysis method, we discovered that there are requests still pending for long and unacceptable period(150 days) of time, while the usual period must be within 7 days, this give us an indication to conclude that there is negligence of response unit that responsible for making the replies.

5.CONCLUSION

In this paper, we applied the data mining techniques to mine a group of received staff petitions through Staff complain mechanism (feedback mechanism) in UNRWA - Gaza, each of these methods provided us with new knowledge that helped us to take some actions to improve the quality of service provided by the response unit in UNRWA, so we have discovered important results as importance of reply period that have an effect on the staff feedback. Thus, we have noted that the long period of reply lead to un-satisfaction of the requesters. We have noted the interests of staff groups as the staff member with lower grades are interested in question about salaries and getting loans. Furthermore, we noted the staff with higher grades need detailed answers about their questions.

So according to the discovered knowledge, we have some suggestion (actions) that can contribute in improving the quality of service provided by the response unite:

The suggested actions are:

- Try to reply on the received questions within short time, not exceeding 7 days and in case of complex cases, send a notification message to the requester telling him that his request needs time more than 7 days.
- According to the results, we noted that there are a large group of staff ask similar questions, so we suggest adding these questions in FAQ library, it will minimize the received questions by the staff and speed up the process.
- Try to provide the staff with detailed answers to preserve satisfaction of the requesters.
- For unsatisfied staff member, send clarification message to them to know the reasons of their negative feedback.
- To ensure the reply process done within short time, develop alert system to notify the response unit with the requests that are waiting for more than 5 days.

REFERENCES

- [1] Prof. D. Jatin Das, S. Arun Kumar, B. Ramakantha Reddy, S. Shiva Prakash ,Web Data Refining Using Feedback Mechanism and k-mean Clustering (2011).
- [2] Bettina Berendt ,Data Mining for Information Literacy (2011)
- [3] Xindong Wu, Chengqi Zhang, and Shichao Zhang, Efficient Mining of Both Positive and Negative Association Rules, ACM Transactions on Information Systems (2004).
- [4] Han, J. and Kamber, M , Data Mining: Concepts and Techniques, 2nd edition. The Morgan Kaufmann Series in Data Management Systems, (2006).
- [5] Douglas Fisher, Iterative Optimization and Simplification of Hierarchical Clusterings, Journal of Artificial Intelligence (2007)
- [6] Abu Naser, S.Al-Masri A., Abu Sultan Y., and Zaqout I. (2011) . A Prototype Decision Support System for Optimizing the Effectiveness of E-learning in Educational Institutions, International Journal of Data Mining & Knowledge Management Process (IJDKP),1(4)
- [7] R. Agrawal and R. Srikant, Fast Algorithms for Mining Association Rules (1994) Proc. 20th Int. Conf. Very Large Data Bases, VLDB-94.
- [8] P. Domingos, MetaCost: a general method for making classifiers cost-sensitive, KDD-99, Proceedings of the 5th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM Press, 1999.
- [9] Fayyad, Piatetsky-Shapiro, Smyth, and Uthurusamy, Advances in Knowledge Discovery and Data Mining, (Chapter 1), AAAI/MIT Press 1996.

- [10] [MPSM96] C. Matheus, G. Piatetsky-Shapiro, and D. McNeill, Selecting and Reporting What is Interesting: The KEFIR Application to Healthcare Data, in Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996