

SENTICOMPOSITE COMPLAINT PORTAL USING SUPERVISED APPROACH

¹Suryia M,²Dr Valarmathi. P, ³Radhika Devi

^{1,2} Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, Tamil Nadu.

³ Department of Computer Application's, Aurora's PG College, Uppal.

Abstract: Reporting women and children harassment problems has no longer been an easy process for the citizens. They have to undergo a long procedure and formalities to report such problems like robbery, sexual assault, rape, domestic violence etc or in short everything that comes under the surveillance of public places. There is still no guarantee that the reported grievances would be addressed by the concerned authority. Most of the time complaints go unheard, unanswered and unresolved usually because the company is too large to worry about one little complaint from a single person. To facilitate this complaining procedure, an online web application is implemented that the citizens can report problems to relevant authority. So whenever people come across any defects in city's infrastructure, emergency situations or any daily life disturbances, they can share, discuss and get resolved the problems by concerned authority by means of this online web portal. And it also provides emergency code via a mobile application. Online Complaint Management is a management technique for assessing, analyzing and responding to customer complaints. Complaints management software is used to record resolve and respond to customer complaints, requests as well as facilitate any other feedback. These feedbacks are analyzed using text mining algorithms to provide credit scores to particular authorities.

Keywords: Sent composite, data mining, text mining algorithm

I. Introduction

Data Mining is the process of posing queries to large amounts of data sources and extracting patterns and trends using statistical and machine learning techniques. It integrates various technologies including database management, statistics and machine learning. Data mining has applications in numerous disciplines including medical, financial, defense and intelligence. Data mining tasks include classification, clustering, making associations and anomaly detection. For example, data mining can extract various associations between people, places or words. During recent years there have been many developments in data mining. The process of digging through data to discover hidden connections and predict future trends has a long history. Sometimes referred to as "knowledge discovery in databases," the term "data mining" wasn't coined until the 1990s. But its foundation comprises three intertwined scientific disciplines statistics, artificial intelligence and machine learning.

As data mining technology keeps evolving to keep pace with the limitless potential of big data and affordable computing power. Various data mining techniques have been developed. These include techniques for extracting associations, neural networks, inductive logic programming, decision trees, fuzzy logic and rough sets. Furthermore, data mining has gone beyond mining relational databases to mining text and multimedia data.

Also, data mining is being applied to areas such as information security and intrusion detection. While there have been many practical developments, still have major challenges. One of the most important challenges is scalability. If data mining is to be useful need to mine very large databases. Therefore, it is critical that to understand the limitations of the data mining algorithms. There are techniques such as inductive logic programming and rough sets that have underpinnings in logic and mathematics.

II. Architecture Of Data Mining

The major components of any data mining system are data source, data warehouse server, data mining engine, pattern evaluation module, graphical user interface and knowledge base.

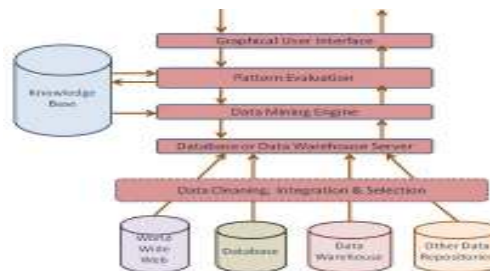


Figure .1 Architecture of Data Mining

a) Data Sources

Database, data warehouse, World Wide Web (WWW), text files and other documents are the actual sources of data. It

needs large volumes of historical data for data mining to be successful. Organizations usually store data in databases or data warehouses. Data warehouses may contain one or more databases, text files, spreadsheets or other kinds of information repositories. Sometimes, data may reside even in plain text files or spreadsheets. World Wide Web or the Internet is another big source of data.

Different Processes

The data needs to be cleaned, integrated and selected before passing it to the database or data warehouse server. As the data is from different sources and in different formats, it cannot be used directly for the data mining process because the data might not be complete and reliable. So, first data needs to be cleaned and integrated. Again, more data than required will be collected from different data sources and only the data of interest needs to be selected and passed to the server. These processes are not as simple. A number of techniques may be performed on the data as part of cleaning, integration and selection.

b) Database or Data Warehouse Server

The database or data warehouse server contains the actual data that is ready to be processed. Hence, the server is responsible for retrieving the relevant data based on the data mining request of the user.

c) Data Mining Engine

The data mining engine is the core component of any data mining system. It consists of a number of modules include association, classification, characterization, clustering, prediction, time-series analysis for performing data mining tasks.

d) Pattern Evaluation Modules

The pattern evaluation module is mainly responsible for the measure of interestingness of the pattern by using a threshold value. It interacts with the data mining engine to focus the search towards interesting patterns.

e) Graphical User Interface

The graphical user interface module communicates between the user and the data mining system. This module helps the user use the system easily and efficiently without knowing the real complexity behind the process. When the user specifies a query or a task, this module interacts with the data mining system and displays the result in an easily understandable manner.

f) Knowledge Base

The knowledge base is helpful in the whole data mining process. It might be useful for guiding the search or evaluating the interestingness of the result patterns. The knowledge base might even contain user beliefs and data from user experiences that can be useful in the process of

data mining. The data mining engine might get inputs from the knowledge base to make the result more accurate and reliable. The pattern evaluation module interacts with the knowledge base on a regular basis to get inputs and also to update it.

III. Architectural Overview

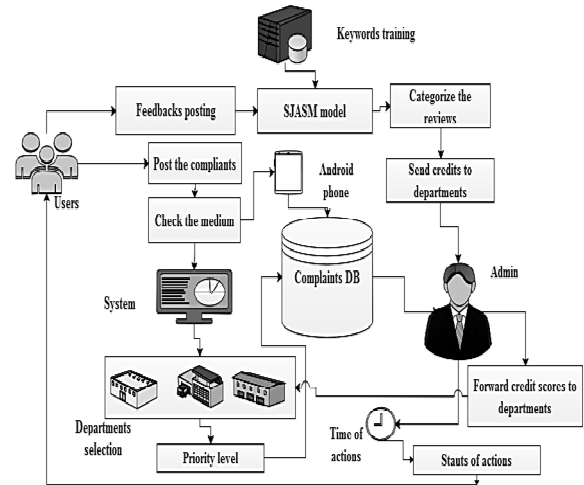


Figure .2 System Architecture

It is used to make complaints easier to coordinate, monitor, track and resolve, to provide company with an effective tool to identify and target problem areas, monitor complaints handling performance. In this architecture, provides application in an online way of solving the problems faced by the public by saving time and eradicates various complaints and designone web aplicaci3n systemtohandlevariouscomplaints. Finallyused to resolve records and respond to customer complaints within proper time period. Sentimentanalysis can be implementedtoanalyzefeedbacks.

IV. Modules Description

A. User GUI

The main purpose of the project is to help the public who are facing different problems in the localities by this online application. It has the potential to reduce the gap between people and Govt. It can control unethical work of bribe and even it can reduce the processing time. The Complaint Management system is web based application and it is designed to keep track of complaints registered by the peoples, so this system need to have distributed platform independent web application. In this module, user registers their basic details such as name, phone number, address, contact and so on. Then post the compliants through mobile or web site. If it is android means, send emergency code to server otherwise post compliants based on priorities and also departments such as EB, Water, Transportation, Women and children problems and so on.

Priorities includes immediate or normal. The women and children problems are provide highest authorities.

B.Admin GUI

The proposed model acts as the platform for the users to address the problem regarding any issue and which should be handled carefully. Admin can read the complaints which are posted by users and check the priorities of complaints. If the priorities are high means, action may be taken immediately. Then information is forward to appropriate departments. The privileged user has access with both admin level and customer side. It can be able to view the tasks, requests, complaints, login details of both the user and admin. The privileged user has to login into the system first, and then access the data as he wants. Privileged user can be able to view the complaints reported by the user and also the managed complaints by admin.

CAAlertAnd Notification System

It is in real time as SMS alert system. Admin can send notification to user about status of complaints and deadline about the complaints which are resolved. In the managed complaints it checks for the solved and unsolved complaints. If there are any unsolved complaints it will take actions towards the problem. If the complaints are not resolved means, automatically forward to higher authority. Otherwise, send notification to user about status of complaints. The admin of the portal manages all the complaints and passes those complaints to the respective complaint handling departments. If the complaint is not solved within the time provided by the system, the complaint will be automatically sent to the municipal commissioner who is the head for all departments. The users are made satisfied regarding their problems.

V. Algorithms And Techniques

Supervised joint aspect and sentiment model (SJASM)

A. Information Retrieval

Information retrieval (IR) deals with the storage, representation, organization and access to information items, the representation and organization of which provides the user with easy access to the information in which he is interested. In other words, IR is finding material of an unstructured nature that satisfies an information need from within large collections. IR systems identify the documents in a collection which matches a user's query and thus narrow down the set of documents that are relevant to a particular problem thereby speeding up the analysis considerably by reducing the number of reviews to be analysed.

There are three major information retrieval techniques

1. Scraping reviews from feedbacks

2. Collecting data sets

3. By web API's

B. Preprocessing and Cleaning

Data preprocessing is a technique that involves transforming raw data into understandable format by eliminating incomplete, noisy and inconsistent data. Online informal text requires more sophisticated methods to clean noise in raw text to perform sentiment analysis. Therefore equal importance should be given to preprocessing along with classification. 'Bag of words' is required to identify opinion targets (features) from this pure textual information. This is also known as complaint feature extraction.

C. Classification

In machine learning terms, classification is the problem of identifying to which of a set of categories a new observation belongs. This is decided on the basis of a training set of data containing observations whose category membership is known. The stochastic algorithm can be stated as follows

Input: A review collection D , its attached ratings R , its attached emoticons E , a user set U and Complaints V

Output: Recommend the positive feedbacks P

Step 1: Initialize D_i, R_i, E_i

Step 2: Set $P_p = 0$ and $P_n = 0$ and $K_i = 0$

Step 3: Read ratings R , reviews D and emoticons D of each feedbacks

Step 4: If the rating R_i is higher than 5 star or 4 star, consider as positive P_p , otherwise consider as negative P_n

Step 5: Read the words from reviews datasets D_i , Consider words as keywords K_i

Step 6: Match the keywords K_i with training datasets

Step 7: Labeled the review D_i as "positive" and also labeled review D_i as negative based on training words

Step 8: Read the emotiocons E_i from datasets

Step 9: If the symbol is happy symbol labeled as positive P_p , otherwise labeled as negative P_n

Step 10: Combine rating, review, emoticons labels, Feedbacks are stated as positive or negative

Step 11: Update each user reviews for each itemsets V

Step 12: Recommend positive label products P_p

The above pseudocode stated to calculate the polarity of the review, rating, emoticons for analyzing the polarity of each individual sentence needs to be calculated. Aggregation is finding out the polarity of each review,

rating, and emoticon to conclude if it falls in the positive class or negative class. However, to find out the overall response about the product, an evaluation of all the reviews is required. Finally recommend the positive feedbacks to admin

VI. Flow Diagrams

A. User GUI

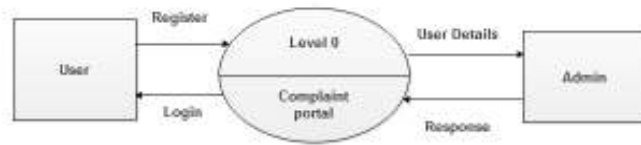


Figure 3 User GUI

B. Admin GUI

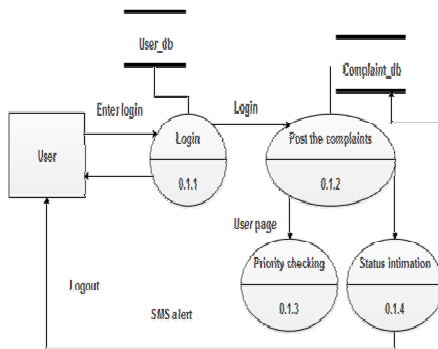


Figure .4. Admin GUI

VII. Conclusion

The system has been designed in the view of present and future requirements and made very flexible. The goals that are achieved by the software are Instant access, improved productivity, Optimum utilization of resources, Efficient management of records, Simplifications of the operations, Less processing time and getting required information, User friendly, Portable and flexible for further enhancement. The system has the benefits of easy access because it is developed as a platform independent web application which may be access anywhere. Feedback analysis is used to analyze the performance of each department. Android based system can be used at the time of emergency.

References

[1] Dermouche M, Velcin J, Khouas L, and Loudcher S, "A joint model for topic sentiment evolution over time," IEEE International Conference on Data Mining, Dec 2014, pp. 773–778.

[2] Hu X, Tang L, Tang J, and Liu H, "Exploiting social relations for sentiment analysis in micro blogging," Sixth ACM International Conference on Web Search and Data Mining, 2013, pp. 537–546.

[3] Jo Y and Oh A H, "Aspect and sentiment unification model for online review analysis," fourth ACM international conference on Web search and data mining, 2011, pp. 815–824.

[4] Lin C, He Y, Everson R and Ruger S, "Weakly supervised joint sentiment-topic detection from text," IEEE Transactions on Knowledge and Data Engineering, vol. 24, Jun 2012, pp. 1134–1145.

[5] Liu S, Cheng X and Li F, "Tasc: topic-adaptive sentiment classification on dynamic tweets," IEEE Transactions on Knowledge and Data Engineering, vol. 27, June 2015, pp. 1696–1709.

[6] Maas L, Daly R E and Pham P T, "Learning word vectors for sentiment analysis," 49th Annual Meeting of Association for Computational Linguistics 2011, pp. 142–150.

[7] Tang D, Qin B, and Liu T, "Learning semantic representations of users and products for document level sentiment classification," 7th International Joint Conference on Natural Language Processing Beijing, China, July 2015, pp. 1014–1023.

[8] Yang and Cardie C, "Context-aware learning for sentence-level sentiment analysis with posterior regularization," 52nd Annual Meeting of Association for Computational Linguistics, 2014, pp. 325–335