INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Voice based Public Opinion Mining and Automatic Category System

Arunkanth. C.P[1], Balasubramaniyan. D[2], Godson Prabhu. S[3], Umanesan. R[4]

[1, 2, 3]UG Scholar, Department of Information Technology, Valliammai Engineering College, Chennai-603203,INDIA
[4]Assisstant Professor, Department of Information Technology, Valliammai Engineering College, Chennai-603203, INDIA

Abstract: Nowadays, central and state governments allocates/spends heavy budget for people beneficial schemes. Even though there is no maintenance been carried and the schemes are not beneficial to the public hence people face several problems in day to day life. The People being unaware about the norms and procedure to approach the respective government departments that is responsible for the specific issue. Because of the above problem public and government negligence several lives are taken off.
To overcome the considerable serious problems, voice based public opinion mining application is developed. In the proposed system the public can give their opinions in the web application with voice recognition module using their smartphones. For speech to text conversion, a Hidden Markov Models (HMMs) are used and the text is fed to the sentimental analysis module, in which pre-processing, part of speech, unigram analysis and fuzzy logic (machine learning technique) is used to forward the user opinion/issue to the concerned government department. The system provides two methods, Automatic e-mail alert to the government officials and a Report generation, where public and both government officials can view the issue which are opened and closed. The system tries to provide problems faced by the public in region / area wise, so that with the aggregated data the government would know which areas / region needs more attention. Thus the proposed system created a close relationship and communication mode with the government representatives elected by people of India.
Keywords: Hidden markov model, k-means, Stop word removal, Report generation, E-mail alert.

I. INTRODUCTION

The main motivation of this project is to classify the user reviews which are given in the form of speech and then sentimental analysis is performed to analyse the category of the comment and provide an automatic update to the respective government department. It specifically focuses on how public opinions towards democratic, participatory, and transparency models across a range of government contexts and their day to day problems with the government. It proposes data mining solutions to mine the user opinions and e-mail to the respective local municipalities to make their decision support mechanism easier. The intention is to acquire smart solutions related to local government services from past data and to estimate the future activities.

A. Overview

Public opinion mining towards government schemes, maintenance and public problems find out how government schemes are reached public and how the public is reacting towards the scheme and their grievance. Opinions of others highly influence the human behaviour and are central to almost all decision making activities. The major part of our information gathering process is to find out what others thin. This new research domain is called opinion mining. The area of opinion mining, also called sentiment analysis is concerned with analysing people's opinions. The main motivation of this project is to classify the user problem that are given in the form of speech and then clustering algorithm is performed to analyse the category of the comment and provide an automatic update to the respective government department.

B. Benefits

It can be used by all common public peoples to report their general public issues.
The issues are forwarded to the respective government official if they fail to finish the problem .It can be sent to the collector/public advocate directly by the user.

C. Challenges

No voice recognition developed . The public complaints are noticed only by the hand written document. If it is note safe to restore the complaints and no one cannot view the complaints every day. Regarding the problem communication was poor.

Linguistic is the major drawback of the system though the people in rural areas cannot communicate easily.

II. RELATED WORK

In [1], The aim of this paper is to present an outline for discussion upon a new Research Challenge on Opinion Mining and Sentiment Analysis. This research challenge has been developed in the scope of project **CROSSOVER** "Bridging Communities for Next Generation Policy-Making" in the view of the definition of a new Research Roadmap on ICT Tools for Governance and Policy Making, building on the model and the research roadmap developed within the scope of the CROSSROAD, but with a stronger focus on governance and policy modelling .

In[2],Electronic Participation (e Participation), both in its traditional form and in its emerging Web 2.0 based form, results in the production of large quantities of textual contributions of citizens concerning government policies and decisions under formation, which contain valuable relevant opinions and knowledge of the society, however are exploited to a limited only extent. It is of critical importance to analyse these contributions in order to extract the opinions and knowledge they contain in a cost-efficient way.

In [3], Mining opinion data that reside in online discussions is a way to track opinions of people on specific subjects. Many of the existing techniques model a discussion as a social network of users and they represent it with a user-based analysis. We combine Social Network Analysis and Opinion Mining in order to give structure to a discussion. Such techniques have not been combined until now. We propose the use of an opinion-based graph whose vertices contain message objects and its «reply-to» edges are labelled with opinion polarities. We compare the opinion based with the user-based graphs and we analyse the different information that can be extracted from them.

In [4], The purpose of this course is to provide a general overview of the practices of public relations and to provide exposure to a number of selected topics in the field of public relations such as history, law, ethics, and the international nature of public relations. The course will provide an overview of public relations' role in organizations and society; of how public relations has been developed as a discipline; and the contemporary roles of public relations in organizations and society. Students will discuss public relations issues and trends as well as international and ethical considerations involved in the practice of public relations.

In [5], Many maturity models have been used to assess or rank e-government portals. In order to assess electronic services provided to the citizens, an appropriate e-government maturity model should be selected. This paper aims at comparing 25 e-government maturity models to find the similarities and differences between them and also to identify their weaknesses and strengths. Although the maturity models present large similarities between them, our findings show that the features included in those models differ from a maturity model to another. Furthermore, while some maturity models are covering some features and introducing new ones, it seems that others are just ignoring them.
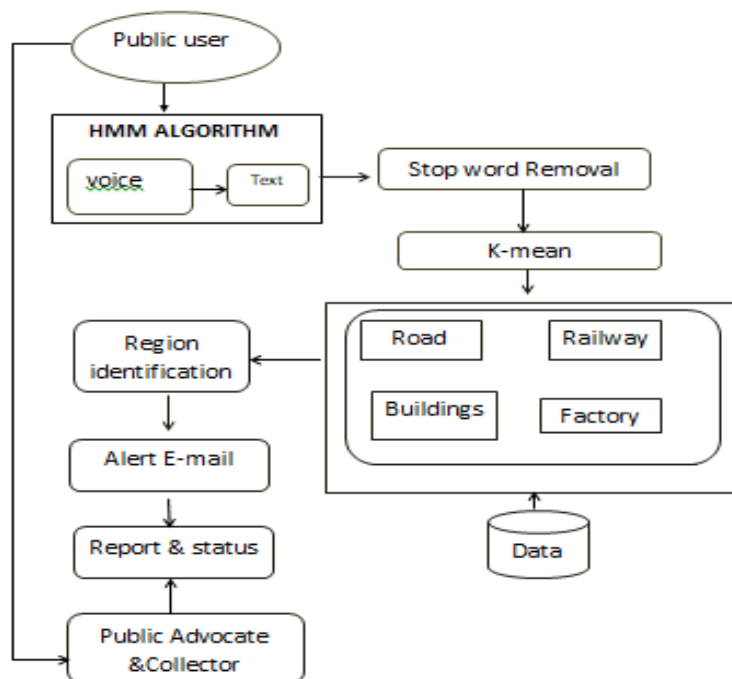

Figure (1) Architecture diagram

III. PROPOSED SYSTEM

We have proposed a voice based public opinion mining application. In this project, the public can give their opinions in the web application. Few modules are developed for speech to text conversion we have proposed Hidden Markov Models (HMMs), stop word removal is a predefined set of process which is used to remove the undefined words from the dictionary base, clustering analysis is used to categorize the problem automatically ,an automatic location identification is also developed and the problem is forwarded to the government officer directly. If the government officer fails to report the status user can forward the problem directly to collector/public advocate.

## IV. MODULES DESCRIPTION

A. Speech to Text conversion

The Java applet provides the interface for speech recognition users, records the users' voice, transmits speech data to the speech recognition server, and displays the recognized text to the users. There are a few protocols between the recognition server and the applet that nee -d to be maintained. For example, before an applet transmits the speech data to the server, it has to tell the server how much data is going to be transmitted. This is important because otherwise the server would not know where the end of the speech data is. The server keeps reading the socket, and blocks it.

The speech recognition server is also responsible for invoking the speech recognizer after it receives all the speech data and stores it to the server's hard disk. This is realized by a system call. The speech recognition server and the speech recognition processes communicate with each other by reading and writing files in the server's local hard disks. Java-based speech recognition system, we used a small-vocabulary, isolated word recognition task, and in particular the digit recognition task. The core of the recognizer was based on continuous density HMMs.

Semantic orientation (SO) is a measure of subjectivity and opinion in text. It usually captures an evaluative factor (positive or negative) and potency or strength (degree to which the word, phrase, sentence, or document in question is positive or negative) towards a subject topic, person, or idea. When used in the analysis of public opinion, such as the automated interpretation of on-line product reviews, semantic orientation can be extremely helpful in marketing, measures of popularity and success, and compiling reviews.

In our proposed system, we have implemented speech to recognition module. In this user can provide their opinions as a speech in our application. The reviews will be processed in a lexicon approach.

The public should provide the complaint first the person choose the category and give their complaints as voice note. The server collect all public complaints about their issues in five different category. The public complaints are analyzed and categorized into specific conditions. The voice notes are convert into text with help of HMM process. The text should be retrieved as parts of speech, unigram, pre-processing process.
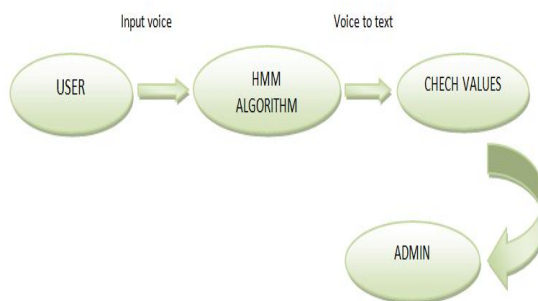
Figure(IV.I) speech to text conversion

B. Stop Word Removal

A dictionary based approach is been utilized to remove stop words from document. A generic stop word list containing 75 stop words created using hybrid approach is used. The algorithm is implemented as below given steps.

Step 1: The target document text is tokenized and individual words are stored in array.

Step 2: A single stop word is read from stop word list.

Step 3: The stop word is compared to target text in form of array using sequential search technique.

Step 4: If it matches, the word in array is removed, and the comparison is continued till length of array.

Step 5: After removal of stop word completely, another stop word is read from stop word list and again algorithm follows step 2.The algorithm runs continuously until all the stop words are compared.

Step 6: Resultant text devoid of stop words is displayed, also required statistics like stop word removed, no. of stop words removed from target text, total count of words in target text, count of words in resultant text, individual stop word count found in target text is displayed.
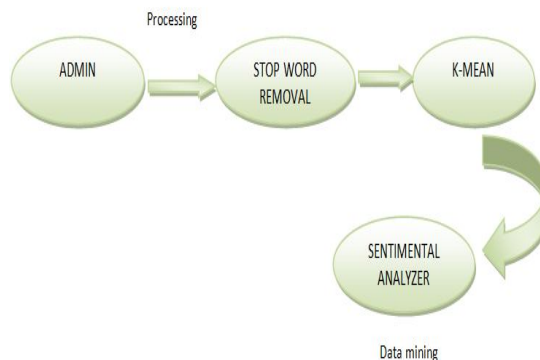

Figure(IV.II) stop word removal

C. Clustering Algorithm

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as bary centers of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroidS. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an *objective function*, in this case a squared error function.

D. Categorization

In our project we consider 5 categories that is Road, Railway, Buildings, Factory Sports etc. The public complaints are categorized into five categories that is Road, Railway, Buildings, Factory Sports etc. The complaints are analysed and separate with their categories these work done by server . the road complaints are separated as well as all categories. . The voice note are converted to text then the texts document are analysed for sentimental analysis and finally polarity / category to which the public opinion is indicated would be identified. Those complaints are analysed and viewed by server.
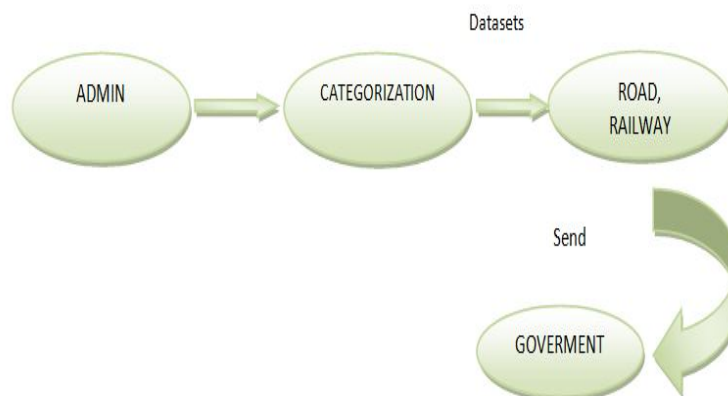
Figure(IV.IV) clustering algorithm

E. Mailing Complaints

Finally the server authorized to send the public complaints to respected faculty members in all government sector through mail. The respected faculty should analyse the public mail and recover their problem as soon as possible. By using our application able to retrieve the written document. By our application we should store public complaints in one storage system but in existing there is a possibility to miss the documents. By generating e-mail complaints, the complaints are viewed day by day by the authorized persons.

F. Report Generation

The public can raise the issue of a particular service with regard to the application or banking service using the ticketing system. The ticketing system is an open type comment box where users can express all their problems with regard to the service in words. Once the issue is been submitted, ticket ID is generated. The public can track the current progress of the ticket raised using the ticket ID. The user can view all the tickets raised.

The government officials validates the tickets and users. After the validation, the open tickets are been assigned to the respective technical team. If the tickets are in general, the government officials itself tries to resolve the tickets. If the ticket cannot be resolved by the government officials, the tickets are been assigned to the technical team. Once the technical resolves the issue, the staff validates and close the ticket from their end.
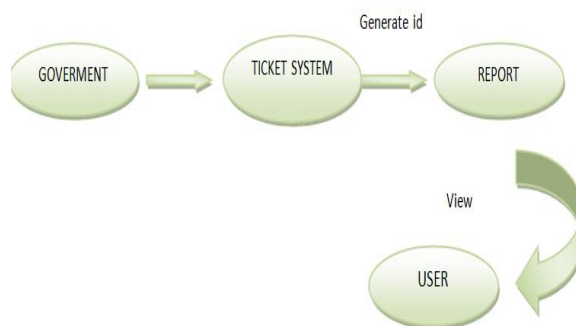

Figure (IV.VI) Report generation

V. CONCLUSION

This project is concentrated on mining public opinion for government officials regarding the issues faced by the public every day. Also through this platform, the public can raise their issues directly to the respective government officials and track the status.

REFERENCE

[1] Alomari M. K., Sandhu, K., Woods, P, "Measuring Social Factors in E-government Adoption in the Hashemite Kingdom of Jordan", International Journal of Digital Society (IJDS), Volume 1, Issue 2, June, 2010.

[2] Dwivedi, S. K., Bharti, A. K., "E-Governance in India – Problems and Acceptability", Journal of Theoretical and Applied Information Technology, 2010.

[3] GozdeBakili, DeryaBirant, erolmutlu, AlpKut,Leventdenktals and Dish Cetin.(2012) "Data mining solutions for Local municipalities", IEEE electronic journal of e-Government 2012.

[4] Guoshi Guan, Liuyang Zhou2, and Pengcheng Tang "Government Public Relations and Social Media: An Analysis of the Perceptions and Trends of Social Media Use at the Local Government", 2012.

[5] Hanaa, M. Said1, Mohamed Hamdy, Rania El Gohary And Abdel-Badeeh M. Salem E-government Maturity Models: A Comparative Study (International Journal of Software Engineering and Applications (IJSEA), Vol.5, No.3, May 2014.

[6] Krishna Kumar Verma, NavitaShrivastava Exploring Role and Associated Challenges of Data Mining Technique and its Implementation in E-Governance Volume 4 Issue 8, August 2015.

[7] Kumar Sanyal, Sudhangsu Das, SajalKantiBhadra, "E-District Portal for District Administration in West Bengal, India: A survey to identify important factors towards citizen's Satisfaction", Procedia Economics and Finance, 510 – 521, 2014.

[8] Manas Kumar Sanyal, Sudhangsu Das, SajalKanti, Bhadra EDistrict Portal for District Administration in West Bengal, India: A survey to identify important factors towards citizen's satisfaction, Electronic Journal of e-Government Volume 8 Issue (2014).

[9] Missy Graham and Elizabeth Johnson Avery, "Government Public Relations and Social Media: An Analysis of the Perceptions and Trends of Social Media Use at the Local Government", Public Relations Journal, Vol. 7, No. 4, 2013.

[10] Mrs.Sangeetha, G, Dr. L. ManjunathaRao. "A Review on Contribution On E-governance Frame Work", Journal of International Journal of Engineering Research and General Science, 2015.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⓢ (24*7 Support on Whatsapp)