International Journal of Research in Advent Technology, Vol.3, No.4, April 2015
E-ISSN: 2321-9637

Public Opinion Mining For Effective Policy Making

Mr. AmitVasant Pandhare[1], Dr. Abrar Alvi[2],
[1]*Student ,Computer Science and Engg. Prof.RamMeghe Institute Of Tech. and Research, Badnera, Amravati, Maharashtra, India.*
[2]*Professor, Computer Science and Engg. Prof.RamMeghe Institute Of Tech. and Research, Badnera, Amravati, Maharashtra, India.*
Email: amitvpandhare@gmail.com[1] ,abrar_alvi@rediffmail.com[2]

Abstract: In today's world millions of people are using internet. Internet provides the opportunity to the people to post their opinion online, they can also share their views on various topics. This opinion can be useful for effective policy formulation. The main aim of this paper is to propose a forum in which government introduced new policy for example Ganga rejuvenation, separate Vidarbha state, direct money transfer scheme etc. on this forum and people post their opinion on this topic. From this public opinion government can understand what is the opinion of the people of this country it is positive or negative? As different age group of people can have different opinion so it is important to categorize this opinion. In this paper we are propose method which shows our output in age group and gender wise. This helps government for effective policy making. This will increase the efficiency of the government scheme as people are directly participating in policy formulation. We are using Natural Language Processing tool to find out opinion polarity.

Index Terms: opinion mining, Natural Language Processing, E-Democracy, Sentiment analysis, Machine Learning

1. INTRODUCTION

As we all know to millions of people use internet. Comment on Facebook, write their thoughts on twitter or on blogs. If we process this large amount of data and find out what is the opinion of the masses for the particular thinks then that opinion is very useful for impartment i.e. if we want to purchase any mobile phone then we first asked our friends and relative if they are using that mobile phone then what is there feedback about that mobile and if the feedback is positive then we purchase that phone otherwise we look for another one. But many times just taking the feedback from one or two people is not sufficient so we need more feedback from internet sides are useful. But this feedback is not in structural format so it is very difficult to mine it and find out the opinion of the people therefore we used opinion mining technique.

Opinion mining is a branch of data mining which deals with the computational techniques for extracting, classifying, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated content Sentiment analysis is often used in opinion mining to identify sentiment affect, subjectivity and other emotional states in online text [1].

Opinion mining is not only used in feedback analyses or movie review but also used for constructive purpose like effective policy making. As we all know in many country people are not directly participate in creating or framing policy there representative involve in framing or creating policy. But history shows us that many times when creating policy representative think about their vote bank their own interest. In that case policy making is not become effective one. To make policy formulation more effective opinion of the people is important. History shows us that if the people of the country participated in policy making directly then that country grows fast as compare to other this is called direct democracy. Switzerland has direct democracy in their country and hence it is in developed nation category. As Switzerland is small it is possible for them to take opinion of the people directly from the people. But for big country like India were population is about 1.24 billion it is not possible to take opinion of each person directly like Switzerland. In that case opinion mining is effective tool to take opinion of the people by using web 2.0 technologies.

In this paper section one is all about introduction i.e. it tell us about opinion mining and direct democracy. Section two is about the background of the paper i.e. why this topic is selected; section tree is about different previous paper review; what are the existing technologies available for opinion mining technology. Section four is about propose technology and architecture of propose technology and the last section is about future scope of the paper.

2. BACKGROUND

Finding out the opinion of the people is very challenging tasks. But it is very important for any system, institution or company for their future point of view. As use of internet is growing now each and every company want their website on internet and they also want feedback from the customer for their product and services they provide. So initially they focus on just take a feedback from the customer in the form of YES or NO. But if we add too many question for the customer to answer the they will not interested in giving feedback. So opinion mining is very effective tool to get feedback from people and analysis it.

Opinion mining was first M.Hu and B. Liu [2] in his research paper. They focus on taking feedback from the customer and mine it so that new customer can understand the feedback of the old customer how has already bought and used it so that new customer can understand what the quality of that product is. It is benefited to company also because is there product not satisfied the customer then they improve the quality of that product. Sentiment analyses are previously used for online shopping websites where different author applied different methods for detecting the polarity of product reviews and movie reviews respectively. This work is at the document level. A different method for determining sentiment is the use of a scaling system whereby words commonly associated with having a negative, neutral or positive sentiment with them are given an associated number on a -10 to +10 scale (most negative up to most positive) and when a piece of unstructured text is analysed using natural language processing, [1][3][4][5] .

Public participation in policy making of any government plays an important role. If we want to make "Good Governance" in our country then public participation is very important. In particular, societies become more and more heterogeneous and pluralistic in terms of culture, values, concerns and lifestyles, and this makes government decision and policy making problems 'wicked', i.e. lacking clear and widely agreed definition and objectives, and having many stakeholders with different and heterogeneous problem views, values, concerns and objectives[6].Countries like India where population is about 1.24 billion and which contain different cast, religion, languages good governance is challenging task. But automated software can be useful here to take opinion of the people and participate in policy making.

3. PREVIOUS WORKDONE:

In business and e-commerce application Micro blogging such as facebook and twitter websites are very important for marketing and other purpose but it can also be important because people express their own views may be positive or negative about any think they see on internet. This data or opinion of the people which express on facebook, twitter or feedback of product on online shopping website are very important for product improvement.

M. Hu and B. Liu[9],first introduces the concept of opinion mining in his paper. He basically focus on feature and the opinion of the people about that feature. Taking this forward Hsinchunchen and David Zimbia [1] improve his idea in "AI and Opinion mining" and apply Natural Language Processing technique to that opinion mining to improve their results.

Aspect-based summarization is usuallycomposed of three main tasks: aspect identification,sentiment classification, and aspectrating[12]. Aspect identification is focusedon extracting the set of aspects or productfeatures from the source collection.The word aspect is intended to representthe opinion or sentiment targets, which arealso referred to as product features [5] whenthe collection of posts—typically, customerreviews—is about products or services. Forexample, given the sentence, "The bed wascomfortable" in a review about a hotelroom, the aspect being referred to is "bed"and the opinion is positively expressedby means of the opinion word "comfortable."The sentiment classification taskconsists of determining the opinions aboutthe aspects and/or their polarities, whereasaspect rating leverages the relevance of aspectsand their opinions to properly presentthem to users [12].

Julian Padget introduce the concept of e-participation in his paper "E-Government and E-Democracy in Latin America"[13] . In this he introduce the concept of e-democracy were people can participate in democratic process through electronic communication. i.e. He introduce the concept of establishment of electronic communication between citizen to Governmental functions.

Taking forward we improve this idea of e-democracy with the use of Natural Language Processing and in that we can take age, gender, cast as important factor which we will take from user at the time of registration. In Indian democracy these thinks are very important. They playsa important role in decision making process.

4. PROPOSED METHEDOLOGY

People can express opinions on any target entity—products, services, individuals,

organizations, or events. In this context, the term object is used to denote the target entity that has been commented on. An object can have a set of components (or parts) and a set of attributes (or properties), which we collectively call the features of the object [1].

For example, a particular scheme i.e. direct money transfer scheme each person is an object and It has a set of components (such as Gas subsidy, Minimum price grain) which are all called features (or aspects). An opinion can be expressed on any feature of the object and also on the object itself.

Suppose in response with scheme a person say " I like this scheme but it is too small" so our system will take this two line as a separate lines and find sentiment analysis of this line i.e. one is good and second one is not good i.e. bad this positive and negative sentences are calculated i.e. how many positive and how many negative sentences are present for each G.R. and according to that result will be calculated.

ARCHITECHURAL DESIGN OF PROPOSED SYSTEM:

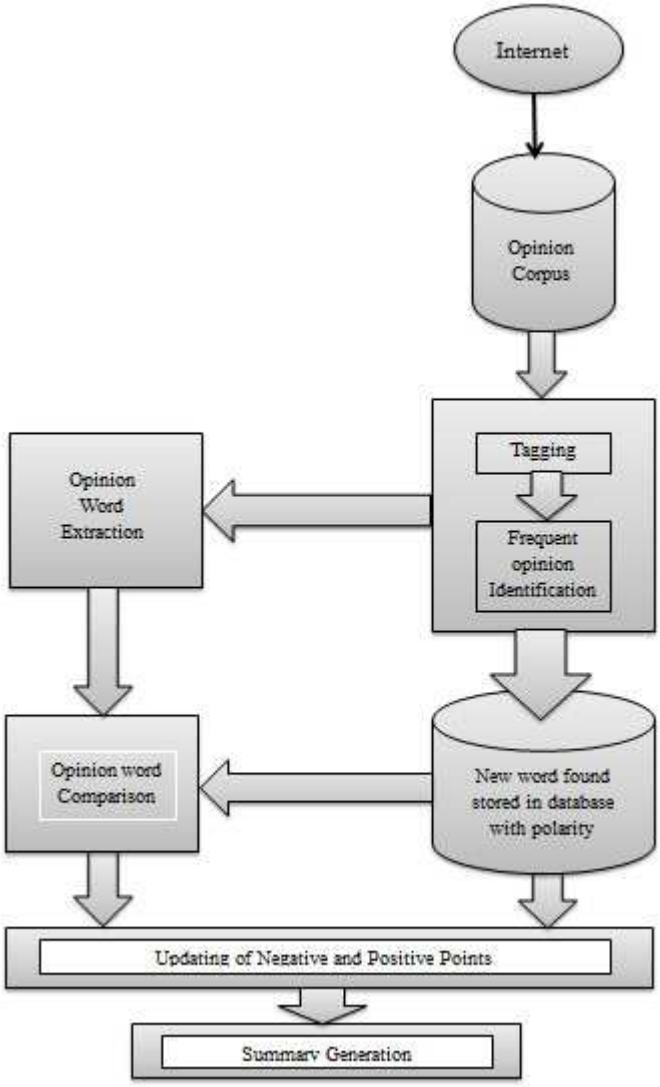

Fig 1 Architectural design of proposed system

We first asked user to sign up on our website so that we can get there age, gender, place where he live and stored it in database so that we can use this information when we want to display result we can easily classify our result according to age, gender etc. After that when sign up get complete the he can post his opinion on the textbox which will be provide there and express his view after reading topic and we stored his or her opinion in database which technically called corpus. This corpus will be used for sentiment analysis purpose.

Sentiment Analysis:

Sentiment analysis technology is needed to extract citizen sentiment because manual collection would be impractical. We use sentiment analysis technology to automatically classify unstructured reviews as positive or negative and then identify citizens sentiment as either toward

creating policy making or they do not want that policy will be created. To identify this there are two approaches for sentiment analysis. First one is machine learning approach and other is lexicon base approach. But in our proposed system we use both the approaches i.e. machine learning approach as well as lexicon base approach.

1. Machine Learning Approach:

In this approach there is no need to store all positive and negative word in database, we can simply add basic word with its polarity to our own database when a new word found then it will first check with existing word and if that word is not present in database then it is search in dictionary for which we are using thesaurus online dictionary.

2. Lexicon Base Approach:

The basic process of the lexicon approach includes pre-treatment, word segmentation, part of speech (POS) tagging polarity (Positive or Negative) tagging combining and result output [0675664]. In this approach we first take input from user i.e. its opinion about related topic and stored it in database i.e. opinion corpus database. After that each word is tagged with Part Of Speech which is also called POS tagging. Stanford university is constantly doing Research and Development work on Natural Language Processing have developed Stanford CoreNLP tool which is very effective tool for part of speech tagging.

In sentiment analysis Part Of Speech tagging is very important because in sentiment analysis we focus on positive sentiment and negative sentiment and this sentiment can be find out with the help of keyword like good, bad etc. and this keyword can be separated with the help of Part Of Speech tagging only.

In our system we are using both the approaches. When user first writes opinion about policy we take this unstructured text as an input to system which is called opinion corpus. We take this opinion corpus and transfer to tokenization phase for tagging purpose, for that we are using Stanford CoreNLP tool and tagging it. This token is then pass to the function which separates useful keyword which is used for sentiment analysis form non useful keyword. Then useful keyword are with tagging i.e. token is pass to Frequent Opinion Identification phase which shows in diagram. In identifying useful keyword if word found then it transfer to other phase of system, if the word is not found in database then it first check in dictionary and then it transfer to our existing database with its polarity value. It is also called Automatic Machine Learning Approach.

Now the word which is from the user are then compare with existing word in database and its polarity is check and its polarity is added to a variable. If it is toward positive then it is added to positive variable if it found negative then it is transfer to negative variable and at the end polarity is calculated as

$$V_P(O) = \sum(P(O)*Polarity(P(O)))$$

Where $V_P(O)$ = Positive polarity value

$P(O)$ = Positive Opinion of the people

Polarity (P(O)) = Polarity of that word

Similarly negative polarity can be calculated as

$$V_N(O) = \sum (N(O)*Polarity(N(O)))$$

Where $V_N(O)$ = Negative polarity value

$N(O)$ = Negative Opinion of the people

Polarity (N(O)) = Polarity of that word

From $V_P(O)$ and $V_N(O)$ we can calculate the result as if $V_P(O) > V_N(O)$ then citizen are toward policy making and if $V_N(O) > V_P(O)$ then peoples are toward Negative i.e. they don't want that policy will be created. In this way our system will work and summary will be generated in graphical format so that it can be easily understand and it is separated by age, gender so that data will be useful for policy decision.

5. CONCLUSION

To make democracy a successful democracy participation of people is important and if people are taking part in actual decision making process then that democracy is called true democracy. India is large country there are 1.24 billion people live in this country. So it is not possible to take actual voting on each and every decision making. With this technique i.e. with the help of this website we can take opinion of the people. In India Computer literate people are increasing rapidly. They are using social side post their views on different topic. We can motivate this people to post their opinion on this site so that we can find out the opinion of the people is positive or negative. This opinion is also very import for government point of view as it gives government data in classified format i.e. Government can easily

identifies what is the opinion of male, female of different age group. It will not only help to designing better policy but also use to understand the problem of different age group. Its make our democracy a true democracy.

6. FUTURE SCOPE:

With this opinion mining technique we can find out the opinion of the people and make the as a part of decision making process. But in India there are many languages spoken and not all the people know English so they may express their opinion in their own language. As this software only process English corpus other language corpus will not process. To do this this project can be further extended with Indian Languages with the help of Hindi, Sanskrit or Marathi WordNet which is available at Centre For Indian Language Technology, IIT Bombay.

REFERENCES:

[1] Hsinchun Chen and David Zimbra "AI and Opinion Mining", published in Intelligent Systems IEEE ,Vol.25 , Issue. 3 ,pp. 74 – 80, May-June 2010.

[2] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD), ACM Press, 2004, pp 168 – 177.

[3] "Retrieving Product Features and Opinions from Customer Reviews" LisetteGarcía-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori

[4] B. Liu, "Sentiment Analysis and Subjectivity," Handbook of Natural Language Processing, 2nd ed., N. Indurkhya and F.J. Damerau, eds., Chapman & Hall, 2010, pp. 627–666.

[5] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, nos. 1–2, 2008, pp. 1–135.

[6] YannisCharalabidis, EuripidisLoukis, LefkotheaSpiliotopulou, VasilikiDiamantopoulou, "A framework for utilizing web 2.0 social media for participative governance" European, Mediterranean and Middle Eastern Conference on Information Systems, 2013.

[7] Paltoglou, G., Theunis, M. ;Kappas, A. ; Thelwall, M. "Predicting Emotional Responses to Long Informal Text" Published in Affective Computing, IEEE Transactions on (Volume:4 , Issue: 1) pp. 106 – 115, Jan.-March 2013.

[8] J. Wiebe et al., "Learning Subjective Language," Computational Linguistics, vol. 30, Sept. 2004, pp. 277–308.

[9] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD), ACM Press, 2004, pp. 168–177.

[10] "Sentiment Analysis of Twitter Data" ApoorvAgarwalBoyiXie Ilia Vovsha Owen Rambow Rebecca Passonneau

[11] "Retrieving Product Features and Opinions from Customer Reviews" LisetteGarcía-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori

[12] N. Jindal and B. Liu, "Opinion Spam and Analysis," *Proc. Conf. Web Search and Web Data Mining* (WSDM), ACM Press, 2008, pp. 219–230.

[13] Julian Padget"E-Government and E-Democracy in Latin America" published in IEEE INTELLIGENT SYSTEMS, IEEE Press, 2005, pp. 1541-1672.

[14] Hsinchun Chen "Business and Market Intelligence 2.0, Part 2" published in Intelligent Systems IEEE,pp. 1541-1672, March/April 2010

[15] Yingcai Wu, Furu Wei, Shixia Liu, Norman Au, Weiwei Cui, Hong Zhou, and HuaminQu "OpinionSeer: Interactive Visualization of Hotel Customer Feedback" published IEEE transactions on visualization and computer graphics VOL. 16, NO. 6, November – December 2010

[16] C. Strapparava and A. Valitutti, "WordNet-Affect: An Affective Extension of Wordnet," Proc. Fourth Int'l Conf. Language Resources and Evaluation, vol. 4, pp. 1083-1086, 2004.

[17] "Graph-Based Methods for Natural Language Processing and Understanding—A Survey and Analysis" ", published in Intelligent Systems IEEE Man and Cybernetics : Systems,Vol.44 , Issue. 1 , pp. 59 – 71, January 2014.

[18] Jingbo Zhu, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, Matthew Ma "Aspect-Based Opinion Polling from Customer Reviews" Published by the IEEE Computer Society Transactions On Affective Computing , VOL. 2, NO. 1, January - March 2011.