

Analysis of Users' Comments on Political Portal for Extraction of Suggestions and Opinion Mining

Swati Verma

CDAC-Noida

+91-9430649412

swati.verma15feb@gmail.com

Aparna Ramamurthy

CDAC-Noida

+91-9560709931

aparna.ramamurthy@gmail.com

ABSTRACT

Online political portals contain a huge volume of users' comments. These comments express whether writer's attitude is positive, negative, or neutral towards the subject. In most of the comments, we find that there are suggestions or feedback also hidden along with the opinion. These suggestions or feedbacks from the mind of different people on various topics can potentially coin new fruitful idea or provide solution to a given problem. So, it is beneficial to determine the opinion and filter out the suggestions from comments. This helps greatly in social issues' analysis and development of people centric governance.

Till date, most of the researches on sentiment analysis and suggestion mining have been done for comments related to product or services. Very less work has been done for analysis of comments related to social issues. Even recent algorithms give accuracy of 60 to 65 percent for opinion mining of social issues and a maximum of 73% accuracy for suggestion extraction.

This paper focuses on evaluating the polarity of sentiment and extracting actionable key suggestions in users' comments on political portal. Current work gives 88% accuracy for suggestion extraction and approx 86% accuracy for Opinion Mining.

CCS Concepts

• *Information systems~Information retrieval*

General Terms

Algorithms

Keywords

Opinion mining, sentiment analysis, suggestion extraction, social issues, NLP, POS tag

1. INTRODUCTION

We are living in the era where, we check for reviews; be it a matter of any political portal like mygov.in, purchasing any product, hotel booking or college selection, name the field and people opinion is one factor that influences actions today.

Analysis of users' comments helps in decision making. We look for peoples' opinion and then decide the best possible solution. It is a common saying that "every head can coin a different idea." So, by analyzing different ideas and suggestions we can come out with the best possible solution for any problem.

As we know that any political portal has a huge volume of comments. These comments express people's opinion and suggestion/ feedback about any issue. Analysis of these comments helps in decision making and implementation of people centric government. But the volume of such data is so huge that manually reading them, finding out sentiment inside the comment and filtering out suggestions or analyzing critical inputs is next to impossible given the large volume of data.

We have developed an automated tool which filters the suggestions from users' comments and finds out the overall polarity of comment.

2. RELATED WORK

Sentiment analysis and suggestion mining are still in very primitive phase. Most of the works have been done on analysis of products. Very few people have worked on analyzing social issues. Also, currently sarcasm is handled with accuracy of maximum of 55 to 60 percent. Suggestion extractions are still at a very nascent stage.

We did a detailed literature analysis to understand the concept of Opinion Mining and Suggestion Extraction. [3] appreciates the use of shallow parsing to speed up the sentiment analysis process. [4] reached at a conclusion that the analysis of product and social issues for fetching opinion needs opinion word with different Part of Speech to be taken care of. Product's opinion mainly depends upon Adjective and Adverb whereas social issues take Verb into account. [7] goes by creating syntactico-semantic patterns to filter suggestion of customer reviews for product. [2] discusses method to differentiate the subjective and non-subjective i.e., factual sentences. [6],[1],[5] discuss various sentiment analysis techniques.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission from Permissions@acm.org

AICTC'16, August 12–13, 2016, Bikaner, India

© 2016 ACM. ISBN978-1-4503-4213-1/16/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2979779.2979870>

3. OUR APPROACH TO SUGGESTION EXTRACTION AND OPINION MINING

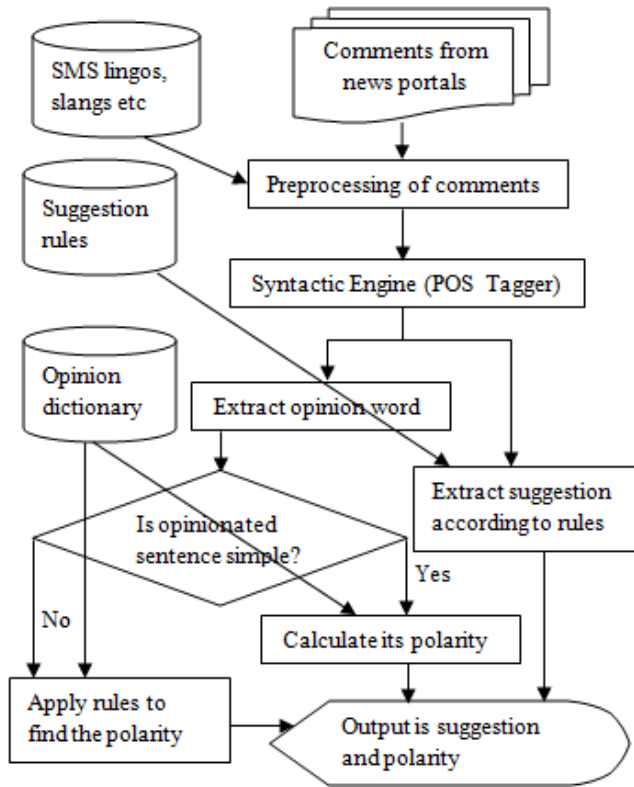


Figure 1. Flowchart for suggestion extraction and opinion mining

As shown in the above figure, first of all we extract comments from the political portal. We have created a small API (works with MyGov.in blog) which takes the URL and the total number of pages required to be loaded from user and gives total comments present in those pages as an output.

After extraction of comments, the next task is to pre-process those comments. Pre-processing includes spell check, sentence extraction, tokenization and POS tagging in the subsequent order. We have used Stanford's sentence extractor, tokenizer and tagger. Once the pre-processing is done, we POS tag the comments.

With the help of POS tagged comments two tasks are being performed. We first extract the suggestions and then calculate the overall polarity of the comment. The final result is the suggestion and polarity of the comments.

Our suggestion extraction algorithm relies on the observation that sentences containing suggestions are generally in the future tense. As for example, "I think our Govt. should launch an App for Incredible India with every detail of the tourist place including map." Words like should, shall, could etc are tagged as _MD in their POS tagged representation. So, we tried filtering such sentences on the basis of _MD. We were able to filter suggestions with an accuracy of approximately 80 percent.

But, suggestive sentences like "Try reducing the cost of Cancer related medicines" do not contain _MD in the tagged token. So, a more precise filtering was needed. We analyzed that such

sentences use base form of verb in their tagged token. Like "Try" in the above sentence. By filtering sentences containing base form of verb we are able to filter those sentences also which we were filtering with _MD previously because in the sentence, "The price of cancer medicine should be reduced"; "be" is the base form of verb.

Base form of verbs is tagged as _VB in their POS tagged representation. We have created a suggestion dictionary which contains suggestive entities like suggest, request, propose etc.

So, the idea is to search for an _VB in tagged comment. If the sentence doesn't contain _VB then check for the presence of suggestive entities.

We analyze tagged sentences based upon above rule and as we find a match we mark it as suggestive sentence. We then apply discourse and co-reference relation to identify whether the previous or next sentence of the suggestive sentence is related to the suggestive sentence or not. This analysis is done to capture complete information.

Algorithm 1 (For suggestion extraction):

Step 1: Start

Step 2: From the preprocessed and POS tagged comments for each sentence find the presence of _VB POS tag in the sentence.

Step 3: If _VB is found, mark it as suggestive sentence else search for suggestive entities from the dictionary and based upon suggestion rules identify suggestive sentences.

Step 4: Stop

After filtering the suggestions we have to check for the opinion of the comment. All the remaining sentences, which are not suggestions, qualify for sentiment analysis task. First of all stop word removal and lemmatization is performed on remaining sentences. For stop word removal we have used our own customized stop word dictionary and code. Lemmatization is performed with the help of Stanford's Lemmatizer. We have mainly focused on finding the overall sentiment of the comment, how discourse relation affects the sentiment and if there is any sarcasm hidden inside the comment or not.

After suggestion extraction, we are left with all those sentences which are not suggestive and are candidate for opinion mining. We check all pairs of consecutive sentences for the presence of discourse connectors like But, Additionally etc. These connection heads change the polarity of comment in a different way.

As for example, if we have a comment about the current government that "Mr. Modi is a good person. But, the government is bad." If we do not consider the effect of connector head "But" and do a simple arithmetic by adding polarity of words good (+3) and bad (-3) then the result comes out to be zero which means that the comment is neutral about the government. But actually it is not. It is negative about the government.

To handle such cases we have created a discourse connector head dictionary which contains such words along with their nature i.e., how they alter the opinion of overall comment. We analyse all consecutive pair of sentences and if there is discourse connector head present in a particular pair then we compute the polarity of both sentences and calculate the polarity based upon the rule. As

for example, in similar cases as mentioned above, the overall polarity will be the polarity of the second sentence.

Algorithm 2(For handling discourse connector heads in Opinion Mining):

Step 1: Start

Step 2: For every pair of sentences, if the second sentence starts with any discourse connection head, perform step 3.

Step 3: If the connection head is negative polar, the overall polarity of the pair of sentences is the polarity of the second sentence.

Else if the connection head is positive polar, the overall polarity of the pair of sentences is the sum of the polarity of both sentences.

Step 4: Stop

The next task is to identify the sarcastic sentences. We found that if there is a consecutive pair of sentences and the second sentence starts with a pronoun and if the polarity of first sentence and the second sentence are opposite then there is a sarcasm.

As for example, I am not a supporter of the current government. I am just a fan and blind supporter!!

So, in such cases we identify polarity and conclude results accordingly as sarcastically negative or sarcastically positive.

Algorithm 3 (For handling sarcasm in Opinion Mining):

Step 1: Start

Step 2: For every consecutive pair of sentences, if the second sentence starts with pronoun (tagged as _PRP), perform step 3.

Step 3: If the polarity of the first sentence is highly positive and the polarity of the second sentence is highly negative, mark the pair as sarcastically negative.

Else if the polarity of the first sentence is highly negative and the polarity of the second sentence is highly positive, mark the pair as sarcastically positive.

Step 4: Stop

At the end, we are left with some more simple sentences which are either opinionated or non opinionated. We analyse them independently with the help of our opinion dictionary and calculate overall score of the comment.

Output of our analysis is the suggestion in the comment and overall polarity of the comment.

4. CONSTRAINTS

The accuracy of our approach is very much dependent upon the Suggestion Dictionary and the Opinion Dictionary.

5. CURRENT STATUS

We have implemented the system and are now in the testing phase. We have tested the implemented system on comments extracted from two blogs of MyGov.in and found approx 88% accuracy for suggestion extraction and approx 86 % accuracy for opinion mining.

6. RESULT ANALYSIS

Ours is a novel approach for suggestion extraction in comments related to social issues. Existing papers discuss about products with the accuracy of maximum of 73 % for suggestion. Our accuracy percentage for suggestion is 88 % or more for social issues.

Very less work has been done for analysis of sentiment of social issues with the accuracy of maximum of 65 percent. Current work is an improvement over previous as it filters out suggestions from comments and applies discourse, negation and sarcasm and thus gives accuracy of approximately 86 %.

We have shown the comparative analysis of results of our previous approach i.e., with the use of _MD tagged token and the new improved approach i.e., with the help of _VB tagged token.

Table 1. Result of myGov blog: Communicating and changing the discriminatory social construct against the girl child (for 50 comments)

Total number of suggestions identified correctly	Total number of sentiments identified correctly	Suggestion percentage	Sentiment percentage
_MD _VB	_MD _VB	_MD _VB	_MD _VB
38 44	39 43	76% 88%	78% 86%

Table 2. Result of myGov blog: Let the World See #IncredibleIndia through Your Own Eyes (for 125 comments)

Total number of suggestions identified correctly	Total number of sentiments identified correctly	Suggestion percentage	Sentiment percentage
_MD _VB	_MD _VB	_MD _VB	_MD _VB
99 110	98 107	79% 88%	78% 86%

7. CONCLUSION AND FUTURE WORK

Sentiment analysis and suggestion mining is relatively new field of research area. Most of the works have been done on the product and services related reviews. Very less work has been done for analysis of comments related to social issues or news.

The proposed approach analyses comments available on news portals. It evaluates the polarity of the opinion expressed in the comment and extracts suggestions or feedback if present in the comment. This approach analyses complex and sarcastic sentences because it consults special rule base for these sentences.

Analysing comments related to social issues present on different news portals will help greatly in analysing people's opinion on any topic. The extracted suggestions can be analysed for potential solution of any problem. This will help establishing people centric governance.

Our future work is to test the implemented system on more of comments and improve the accuracy percentage. We will analyse more complex and sarcastic sentences for opinion mining. Also, we will focus on suggestion extraction of sentences which neither contains _VB in their tagged tokens nor composed of any suggestive entity.

8. REFERENCES

- [1] Bhadane, C., Dalal, H., and Doshi, H. Sentiment Analysis: Measuring Opinions. in *International Conference on Advanced Computing Technologies and Applications (ICACTA)*, (Mumbai, 2015), 808-814
- [2] Biyani, P., Caragea, C., and Mitra, P. Predicting subjectivity orientation of online forum threads. in *CICLing'13 Proceedings of the 14th international conference on Computational Linguistics and Intelligent Text Processing* (Berlin, 2013), 109-120
- [3] Desai, J. M., and Andhariya, S. R. Sentiment analysis approach to adapt a shallow parsing based sentiment lexicon. in *International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, (Coimbatore, 2015), IEEE, 1-4
- [4] Karamibekr, M., and Ghorbani, A. A. Sentiment analysis of social issues. in *International Conference on Social Informatics (SocialInformatics)*, (Lausanne, 2012), IEEE, 215-221
- [5] Liu, B. *Sentiment analysis and opinion mining*. Morgan and Claypool Publishers, USA, 2012
- [6] Madhoushi, Z., Hamdan, A. R., and Zainudin, S. Sentiment analysis techniques in recent works. in *Science and Information Conference (SAI)*, (London, 2015), IEEE, 288 - 291
- [7] Viswanathan, A., Prasanna V., Vasudevan B., Balakrishnan R., and Shastri L. Suggestion Mining from Customer Reviews. in *Seventeenth Americas Conference on Information Systems*, (Detroit, 2011), AMCIS