

Construction of Twitter Poll Matrix on Sentiment Analysis and Opinion Mining

Akkireddi Vara Prasad ¹, V.Satya Aruna ²

¹ Assistant Professor, GITAM CDL, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India.

varaprasad.akkireddi@gmail.com

² Associate Professor, Lingaya's Institute of Management and Technology, Vijayawada, Andhra Pradesh, India.

satya.aruna.v@gmail.com

Abstract

Now a days in the real world there is a gradual change in the trends and issues. Various social media platforms are used to survey on different areas and fields. Twitter polling based on metadata is considered for performing sentiment analysis. This paper discusses about the structure of the twitter poll from twitter micro blogging sites and classifies different opinions like positive, negative and neutral with the help of survey poll matrix that can analyze on different surveys in one area. This paper also provides the design model for the twitter poll metadata which can further enhance real world surveys to get an understanding of deep opinion mining.

Keywords: Twitter Poll matrix, Classification of Data, Metadata, Sentiment Analysis.

I. INTRODUCTION

Sentiment Analysis (SA) and Opinion Mining (OM) are subfields of machine learning. They are very important in the current situation to perform computational study on people's opinions and sentiments.[1] The entity can represent attributes and events of each individual. The challenges in sentiment analysis in one aspect the opinion may be positive at the same time it may be negative in another aspect.[9]

The more informal medium like twitter or blogs, where people will express different opinions, obviously those are very easy to read and to understand, but it is very difficult to parse to the computer[1][2]. Hence, many blog users also facing problem frequently to understand the context of the message or statement due to lack of clarity. Detection, Sentiment Prediction are the fields those are based upon Sentiment Summarization and Text summarization for the sentiment analysis.[5]. Contractive viewpoint Summarization, Product Feature Extraction, Detecting opinion spam based upon Subjectivity Detection is a task of determining whether text is opinionated or not in the opinion mining. Sentiment summarization provides sentiment summary in the form of star ratings or scores for features of the product. Text summarization generates a few sentences that summarize the reviews of a product. Contrastive viewpoint summarization puts an emphasis on contradicting opinions. Product Feature Extraction is a task that extracts the product features from its review. Detecting opinion spam is concerned with

identifying the fake opinions from reviews. Sentiment analysis can be done at Document level, Sentence level, and Aspect or Feature level. In Document level the whole document is classified either into positive or negative class. [3][4].

The current research paper explains about the analysis of the contents related to enormous areas which are gradually developing day by day with huge number of websites. These sites are dedicated to analyze on specific area or field with different queries along with different opinions in the same context. On the same topic different queries posed to get opinion. Depending upon the various factors like geological, environmental and social elements the opinions may differ for the same area or field. All the opinions are deeply analyzed to report about the exact result or verdict given by the public. Under this process we can concentrate on unstructured data like user reviews, feedbacks on specific products as opinions. These can be considered to convey the opinion to generate exact verdict or the satisfactory judgment.

The features of source data can be extracted by the feature level class. These are classified into three common levels positive, negative and neutral. Manual training can be used to solve sentiment analysis problem, at the same time a fully automated system also requires manual training to get fruitful solutions at satisfactory level[3][6]. The approach towards machine learning uses classification techniques to classify text. Lexicon based approach towards sentiment dictionary with opinion words and matches them with the data to determine polarity[6][7]. Opinion words are assigned by scores of the sentiment describing how Positive, Negative and Objective the words contained in the dictionary. The objective of this paper is to construct a relational matrix to denote sentiments and opinions in the various areas and fields. Natural Language and its techniques widely used in this concept. Other sections in this paper show different levels of opinion mining and sentiment analysis also.

II. EXISTING SYSTEM

In the existing system, normally tweets are retrieved by using twitter API based on the query. The preprocessing technique will be applied to extract tweets which are related to the source information.[9]. Later We can apply the supervised learning algorithm on the stored data.[10]. This algorithm used to perform the sentiment classification task and also considers the sentiment classification accuracy. The results of the algorithms i.e. the sentiment classification will be represented in various types of pictorial notations like bar charts, pie charts. The proposed system is more effective than the existing one. This is because it can demonstrate more accurate results. we will be able to know how the statistics determined from the representation of the result can have an impact in a particular field from the selected area.[11].

III. PROPOSED SYSTEM

In the proposed system twitter data will be collected from each part of the area and field. This data may be either in the textual form ,statements , queries and notations etc. Each specific area contains finite number of subfields which can produce clarity on the overall data areas and policies. This will be shown by constructing twitter poll matrix which can help as supporting tool. This poll matrix contains **m** number of rows and **n** number of columns for different areas and different opinions on the survey poll. N number of

opinions generally develops three types like positive, negative and neutral. Each row represents the number of choices of the whole area, these rows are verified by n number of opinions, there after the maximum obtained result will be highlighted and considered.

In this paper we will take exactly three considerations positive, negative and neutral in the matrix.

- The following flow chat depicts extraction of tweets from source information.
- Transformation of information from field to sub fields.
- Classify opinions in each area.

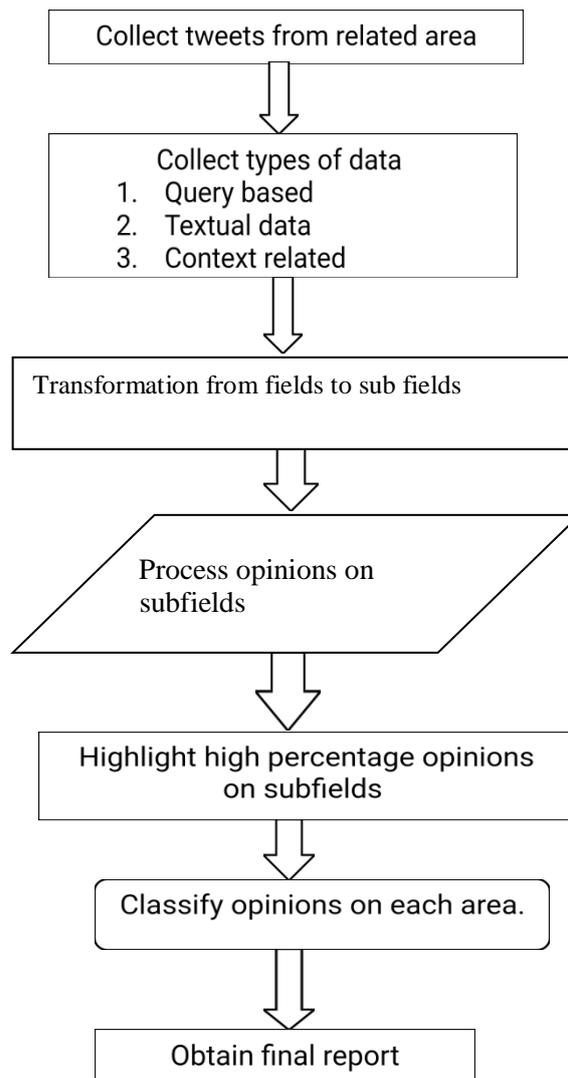


Figure 1. Process diagram for the Sentiment Analysis to get opinion and to produce accurate results.

Matrix Notation:

Each poll in the survey related to different areas contains different machine vectors[12]. In the initial stage these vector elements of each area will be taken each row. These will be compared and matched with survey opinions positive, negative and neutral

continuously. The subarea which is having more positives will be highlighted in the first row, negatives in the second row, neutral in the third row.

Chosen area categorized in to sub fields denoted as M

M

Sub fields : 1 2 3 n

positive	1	{	11 12 13	}	1n
negative	2	{	21 22 23	}	2n
neutral	3	{	31 32 33	}	3n

m x n Where m=3

Table 1. Comparison of Existing System and Proposed System

S. No	Existing System	Proposed System
1.	Existing system takes a stored dataset on a particular topic into consideration.	Proposed system will allow collecting data from formal and informal medium.
2.	It fails to determine the impact of the results might or will have in the respective field.	Here, it will construct a relational matrix to obtain results.
3.	Existing system does not allow the retrieval of data based on the query entered by user.	Proposed system allows retrieval of metadata based on the environmental and socio ethical issues also.
4.	Existing system does not provide accurate feature selection.	Proposed system requires feature selection and lexicon based methods.

IV. CONCLUSION:

The main goal of this paper is to generate a structured methodology for the sentiment analysis and opinion mining. For this classification of data in structured and unstructured is analyzed. Further, the development of construction of matrix can be extended to different areas and fields along with their opinions are collected to justify the final verdict on a particular context. This matrix in the extension can produce fuzzy sets using fuzzy logic for the deep analysis on the various opinions on structured and unstructured data. In the future scope analysis to be focused on the fuzzy based systems and sets. These fuzzy tones will ensure more accurate results to produce final verdict.[7] [13].

And also combining with supervised learning approach provides better accuracy. It can be concluded that the constructive methodology provide better accuracy even on unstructured data.

V. REFERENCES

- [1] A. Barhan, A. Shakhomirov, "Methods for Sentiment Analysis of Twitter Messages", 12th Conference of FRUCT Association, 2012.
- [2] Duyu Tang et al., "Coooolll: A Deep Learning System for Twitter Sentiment Classification", SemEval, 2014. Show Context [CrossRef](#) [Google Scholar](#).
- [3] Khairullah Khan, Baharum B. Baharudin, Aurangzeb Khan, Fazal-e-Malik, Mining Opinion from text Documents: A Survey, 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009.
- [4] Hsinchun Chen and David Zimbra, AI and Opinion Mining, IeeeInTeLLIGenTSySTeMS 2010.
- [5] K. Ahmed, "Sentiment Analysis over Social Networks: An Overview," in 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2015, no. October, pp. 2174–2179.
- [6] David Alfred Ostrowski. Sentiment Mining within Social Media for Topic Identification IEEE Fourth International Conference on Semantic Computing 2010.
- [7] Nilesh M. Shelke, Shriniwas Deshpande, Vilas Thakre, Survey of Techniques for Opinion Mining, International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.
- [8] V. M. Pradhan, "A Survey on Sentiment Analysis Algorithms for Opinion Mining," Int. J. Comput. Appl., vol.133, no. 9, pp. 7–11, 2016.
- [9] ANA SUFIAN, RANJITH ANANTHARAMAN, Social Media Data Mining and Inference system based on Sentiment Analysis 2011,
- [10] Johan Bollen, Huina Mao, Xiao-Jun Zeng, "Twitter mood predicts the stock market", Journal of Computational Science, 2011.
- [11] B. O'Connor, R. Balasubramanyan, B. R. Routledge, N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", N.A, 2010.
- [12] KENNETH BLOOM, Sentiment Analysis Based On Appraisal Theory And Functional Local Grammar 2011.