Lexicon Based Sentiment Analysis Algorithm for Web Reviews

Umang

Department of Information Technology, SSJ Campus, Kumaun University Almora, Uttarakhand, India

ABSTRACT: Sentiment analysis is the process of determining emotional tone for gaining knowledge of attitudes, opinions, and decisions expressed on online forums, social networking sites, etc. It is one of the major parts of natural language processing. Now a day's web sentiment analysis is common practice to expand the business, perform market research, built reputation, doing market analysis, developing public relations, obtaining feedback from the public for the electoral campaign, and a lot more things by using reviews/ comments on social media platforms. Sentiment tools can be designed based on machine learning, lexicon, and hybrid approaches. In the present study lexicon-based tool has been proposed to determine positive and negative sentiments from the reviews and comments available on the web datasets. The tool can use paragraphs, break them into sentences, and further into words. The output will be a total count of reviews, positive and negative sentiments among total.

Key words: Algorithms, Natural Language Processing, Web Data Mining, Sentiments Analysis.

I. INTRODUCTION

Attitudes and experiences about an object reveal a distinct picture of its perception in the common man. Now a day's web sentiment analysis is common practice to expand the business, perform market research, built reputation, doing market analysis, developing public relations, obtaining feedback from public or customers, [1] movie reviews and a lot more things by using reviews/ comments on social media platforms. [2] Large-scale sentiment analysis for news and blogs and [3] opinion mining was conducted by various researchers. Due to the wide domain of social media text; related analytics tools like sentiment analysis tools are in great demand so that one can draw conclusions and get directions to adopt favorable policies.

Sentiment Analysis tools help to monitor social networking sites, reviews on e-commerce applications, and [4] tweets on microblogging sites. These tools help to improve online shopping experiences, improve retailing services, monitoring quality and consistency for customer services.

The sentiment analysis is possible with the availability of tools that can process natural language text so that the positive and negative sentiments in comments and reviews can be analyzed. [5] One such tool based on C++ that works at sentence level was designed to extract positive and negative emotions [6] Such tools are based on three approaches first one is the Machine Learning approach where we train a dataset by using deep learning methods, the second method is lexicon-based where a variety of words are annotated by polarity score, to decide the general assessment score of given content. The strongest point of this technique is that it does not require any training data, while its weakest point is that a large number of words and expressions are not included in sentiment lexicons or arrays that we create the third and last method is using a hybrid approach by using [7] applied data structure algorithms. Hybrid approach is the combination of machine learning and lexicon-based approach. [8] Number of techniques and complex algorithms used to command and train machines to perform sentiment analysis with pros and cons to each. But, when used in hybrid manner can provide exceptional results.

In the present study, an effort has been made to develop one such lexicon-based algorithm for analyzing affirmative and pessimist/negative sentiment by implementing it using java programming language considering its wide features for processing files and strings.

II. METHODOLOGY

The web data can be imported via any software in text files. The text files are pre-processed based on sentences delimited by a full stop or newline character. An array of sentences is created. Like in the below example:-

1. Awesome product and information.

2. Worth every penny.
3. No value for money.
4. I have visited this site.
5. Looks good…..is good.

Analysis Conclusions:-
Sentence 1 expresses a positive review of the product and related information in the blog. Sentence 2 expresses a positive review of the product and its cost. Sentence 3 expresses a negative review about products' price. Sentence 4 expresses neutral review and it doesn't highlight any sentiment. Sentence 5 expresses a positive review about product delivery service.

Out of context, the words 'Awesome 'worth' and 'good 'could be read as positive, and the third sentence is a negative comment. Using sentiment analysis, computers can automatically process text data and understand it just as a human would, saving hundreds of employee hours.

The stepwise flow of information is as –
Step 1: Define Affirmative or satisfactory or Positive or Good words array [] and Negative or dissatisfactory or Bad words array [].
Step 2: Accept input of reviews from the dataset in the form of text files. Read the text files, pre-process it for extracting sentences
Step 3: Sentences are divided into two arrays of words.
Step 4: Match each word to positive words [] or negative words [] datasets/array and increment the count of the related dataset and total reviews/ sentiments.
Step 5: Calculate to the polarity of reviews/ sentiments by using formula
Polarity = (Positive Reviews sentiments)/Total no. of sentiments
Step 6: Displays the polarity results to the user, which represents total reviews.
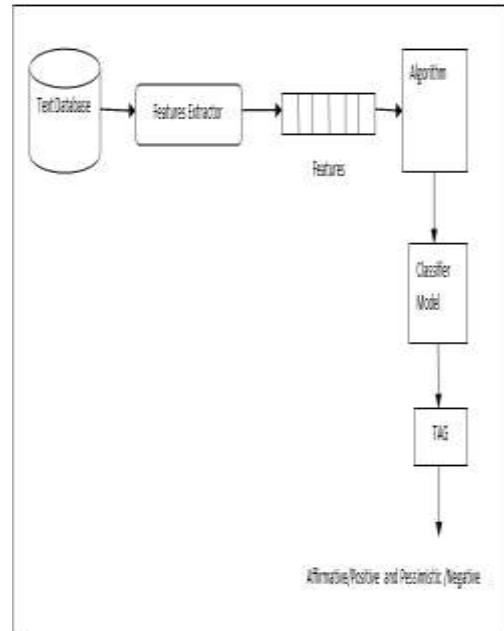

Figure 1 Steps showing extraction of words from datasets

III.   IMPLEMENTATION
The above algorithm has been implemented in java by using the following functions-
1. Function defined to fetch text files in data. Then sentences were extracted with the help of delimiter like full stop and newline character and stored in the array.
2. Function defined to categorize positive and negative sentences.
3. Accept input from the text array
4. Two arrays to store affirmative and negative word and match sentences.
5. Classifier function to represent polarity of statements.

Polarity Calculation:

$$Polarity = \frac{(Positive\ Reviews\ sentiments)}{(Total\ No.\ Of\ Sentiments)}$$

6. Function defined to displaying output to the user in form of total reviews, positive and Negative reviews

IV.   CONCLUSION
Sentiment analysis tools are having a wide domain nowadays. This tool will help in improving business analysis, public opinion, improve customer support, and automate tasks with fast turnarounds thus will be saving lots of time and efforts to reach actionable insights and make the right decisions. In the future, this algorithm can be implemented in any object-oriented programming

language and can be used with the hybrid approach.

REFERENCES

[1]   K.Sarvakar,   U.Kuchara,   June   2018, "Sentiment Analysis of movie reviews: A new feature-based sentiment classification", IJSRCSE, vol.6,pp. 8-12.

[2]   N.Godbole, M. Srinivasaiah and S. Skiena, 2007, "Large-scale sentiment analysis for news and blogs", ICWSM,07.Godbole et al.

[3]   Bing Liu, 2012, "Sentiment analysis and Opinion Mining", ASEE Journals, April 22, PP 30-43.

[4]   A.Palve,R.Sonawane   and   A.Potgantwar, 2017, "Sentiment Analysis of Twitter Streaming Data for Recommendation using, Apache Spark", IJSRNSC, vol. 5,pp.99-103.

[5]   JE. Rajput and MB. Patil, 2020, An Algorithm to perform sentiment analysis of web reviews using C++.IJSE, vol 8 issue 4.38-39.

[6]   https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html

[7]   A. Levitin, 2011, "Design Analysis of Algorithm", Pearson Publication, New York, pp. 41-94,

[8]   https://monkeylearn.com/blog/sentiment-analysis-machine-learning/#:~:text=Sentiment%20analysis%20is%20a%20machine,detect%20sentiment%20without%20human%20input.