

Sentiment Analysis for running text in Email conversation

PRAVEEN GUJJAR J

Research Scholar,

Visvesvaraya Technological University, Belagavi- 590 018, Karnataka, India.

Email: gujjarpraveen@gmail.com

Dr. Prasanna Kumar H R

Professor and Head, Department of ISE

PES Institute of Technology and Management, Shivamogga-577 204, Karnataka, India.

Email: hrpbhat@gmail.com

Abstract: Evolution in the field of web technology has made an enormous amount of data available in the web for the internet users. In this web opinions expressed by a user are highly heterogeneous and unstructured this paper focuses on analyze the email conversation through sentimental analysis. Opinions may be negative, positive or neutral. This paper, provide a result for aforesaid data using lexicon analyzer and Natural language toolkit. The proposed method can also be used in the identifying emotions of the user involved in short message service as well with other conversation in any other social network. This paper also discussed general procedure for sentimental analysis.

Keywords: Sentimental Analysis, Email Conversation, Lexicon analyzer

I. Introduction

The Internet has changed the manner in which individuals express their perspectives, by the introduction of web 3.0 user's can have conversation about the product or service with manufacturer or service provider. It is for the most part done through blog entries, online discussions, item survey sites, internet based life, and so on. Users expressing their views, emotions and sentiment through social network sites like Google Plus, Twitter, Facebook etc. Social network is producing a huge volume of notion rich information as tweets, reviews, comments, discussion, blog entries, and so forth. Social media network gives a chance to businesses by giving a stage to interface with their target customer for advertisement. A user on the most of the cases depends widely on other user's generated content for decision making about the product available in the online. Because of the huge content generated by the users in a daily basis it becomes a difficult job for the ordinary users to analyze the content. Hence there is a great demand to automate the users review. In this context Sentimental Analysis (SA) has a very vital role to play. Sentiment analysis enlightens the customer whether the data concerning the product is attractive or not before they purchased it. Advertisers and firms utilize this data to comprehend about their product or administrations so that it very well may be offered according to the user's prerequisites. Hence in the context of analysis one can use textual Information retrieval techniques. It mainly focuses on searching, preparing or analyzing the genuine information present. But, there is some other textual information which may express subjective features. This information's are primarily focused on sentiments, attitudes, opinions, emotions and appraisals, which may form the centre of SA. Sentimental analysis offers many provoking chances to grow new applications, essentially because of the enormous development of accessible data on online sources like sites and social network.

II. Sentiment Analysis

Sentiment analysis is a process that computerizes mining of attitudes, mentalities, suppositions, perspectives and feelings from content, tweets and database sources through Natural Language Processing (NLP). SA includes classifying suppositions in content into classifications like "neutral" or "negative" or "positive". SA is also called as opinion mining, subjectivity analysis or appraisal extraction. An example for terminologies for Sentiment Analysis is as given below,

<SENTENCE> = "Your paper has been accepted. Congratulations

<OPINION HOLDER> =<author>

<FEATURE> = <conference>

<OBJECT> = <paper>

<OPINION >=

<POLARITY> = <positive>

SA is a term that incorporates numerous undertakings, for example, conclusion extraction, assessment characterization, subjectivity arrangement, rundown of suppositions or spam detection. SA plans to examine individuals' assumptions, attitudes, mentalities, conclusions feelings, and so forth towards items, people, subjects, associations, and administrations etc.,

Let's focus on terminologies of SA

Opinion holder: The holder of a supposition is the individual or association or an element that communicates the opinion.

Feature: Describe the object concerning which assessment is made.

Object: An object which can be an, individual, occasion, item, product, service, association, or subject.

Opinion or polarity: The direction of a sentiment on a feature which describes the feeling is neutral, negative or positive.

In the recent past wide number of a great deal of work has been done in the field of sentimental analysis by number of scientists and scholars. In its beginning period it was expected binary classification to classify either positive and negative. In the following section let's understand different technique used for sentimental analysis. Pak and Paroubek [1] proposed techniques to classify the twitter tweets as neutral, positive or negative. They were used twitter API to collect the tweets and for the analyses used Naïve based algorithms.

Parikh and Movassate [2] used Naïve based algorithm and Entropy model. It was found that Naïve based algorithm doing better compared to entropy model.

Go and L.Huang [3] developed a solution for sentiment analysis in social network for twitter data by using distant supervision. They provided a model using naïve based algorithm and support vector machine algorithm.

Barbosa et al. [4] proposed a method for classifying tweets for sentimental analysis by using two phase automatic method. Method gives the direction of a sentiment on a feature which describes the feeling is neutral, negative or positive.

Bifet and Frank [5] designed a solution for sentimental analysis for twitter streaming data using Firehouse API. They were used multinomial naïve Bayes algorithm along with stochastic gradient descent technique. The result shows that stochastic gradient technique was better compared to multinomial naïve Bayes algorithm to classify the twitter tweets as neutral, positive or negative.

Agarwal et al. [6] proposed a model to classify the twitter tweets as neutral, positive or negative. They have used tree kernel based model, unigram model along with a feature based model. Results shows that The tree kernel based model was better compared to rest of the model for to classify the twitter tweets as neutral, positive or negative. Davidov et al. [7] designed a solution for the SA for the classifying the emotions of the tweets. This proposed solution uses K-Nearest Neighbor algorithm to allocate a sentimental labels with the help of support feature vector for all the training data set.

Po-Wei Liang et.al. [8] Proposed a solution for SA with reference to twitter data. These data are collected using the Twitter API. The training data is assigned a label as neutral, negative and positive. This proposed solution made use of Naïve based model and used Chi square feature to eliminate useless features in order to describe the emotion as Either Neutral, negative or positive.

Kamps et al. [9] utilized the lexical database WordNet to decide the passionate substance of a word along various measurements. They built up a separation metric on WordNet and decided semantic extremity of adjectives.

Liu [10] analyzed that there are various challenges and issues with respect to opinion mining in proposing emotions of the people from the data available in the twitter. Further they quoted that varying languages makes opinion mining as challenging task.

Lexicon based method [11] utilizes slant lexicon with sentiment words and match them with the information to decide extremity. They allocates emotional scores to the sentiment words portraying how Positive, Negative and Objective the words contained in the lexicon are. This approach related to Lexicon is focuses on sentimental lexicon Piryani et.al [12] analyses the citation pattern of the paper with the help of sentimental analysis. Bordbar et.al [13] Uses emotional implications for classifying sentimental emotions. For the classification purpose Fuzzy logic was used. Here the major challenge is finding the fuzzy rules.

Phuong et.al [15] uses Apriori algorithm for finding the patterns. These patterns are applied in stock market, market basket analysis and so on.

Praveen Gujjar et.al [16] argues that sentimental analysis place a very important role in the decision making of the product by the customer.

III. Proposed Method

The proposed method includes the following steps:

1. Collection of email conversation
2. Use of NLTK dictionary for extracting token
3. Identifying probability of Positive and negative statement in the email conversation using VADER.

VADER is an acronym for Valence aware dictionary for sentiment reasoning. This mode is used for analyzing sentiment available in the text. For example good and excellent both can be classified to the polarity positive but the excellent is more positive than good hence excellent should have more VADER score than good. In the context of the paper VADER is used to describe the sentiment in an effective manner. The pictorial representation of the proposed method is shown in fig 1

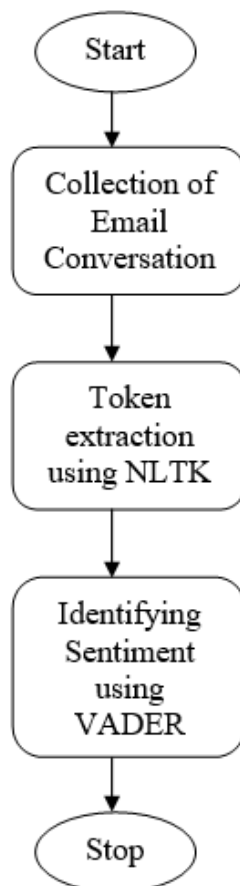


Fig. 1 Proposed Methodologies

Aforesaid proposal has been implemented in Python Language and Pseudocode for the aforesaid method is given below

```

sid = SentimentIntensityAnalyzer()
tokenizer=nlTK.data.load('tokenizers/punkt/english.pickle')
for sentence in sentences:
    print(sentence)
    scores = sid.polarity_scores(sentence)
    for key in sorted(scores):
        print('{0}: {1}, '.format(key, scores[key]), end="")
    print()
  
```

IV. Results and Discussion

The email text probability of positive and negative statement is as shown below

- message_text = "Your paper has been accepted. Congratulations"

Result shows:

- Your paper has been accepted.

Compound: 0.2732, neg: 0.0, neu: 0.656, pos: 0.344,

- Congratulations

compound: 0.5994, neg: 0.0, neu: 0.0, pos: 1.0,

From the above said result compound score is 0.732 and 0.5994 which is the sum of the lexicon ratings. The standard lexicon score varies between -1 and +1. Aforesaid result shows 0.59 which is positive. Neg, neu and pos represents negative, neutral and positive category of the statement used in the email.

Consider one more example for the email text, the probability of positive and negative statement is as shown below

- message_text = "Sorry, your paper is not approved"

Result shows:

- Sorry, your paper is not approved

compound: -0.3252, neg: 0.318, neu: 0.682, pos: 0.0,

From the above said result compound score is -0.3252, in this context text is tending towards negative and more towards neutral.

Consider one more example for the email text, the probability of positive and negative statement is as shown below

- message_text = "it seems bad"

Result shows:

- it seems bad

compound: -0.5423, neg: 0.636, neu: 0.364, pos: 0.0,

From the above said result compound score is -0.5423, here emotion is tending towards negative with the score 0.636 and neutral with the score 0.364.

Comparing VADER and TextBlob

The TextBlob can also be used as an alternative for the VADER but using Text Blob identifying the valence of the statement is bit challenging. Result of the TextBlob is as shown below

- text=TextBlob("Amazing product for a least cost")

The sentiment of the statement using TextBlob is as given as,

Sentiment (polarity=0.15, subjectivity=0.65)

The above said result sentiment polarity score is 0.15 and subjectivity is 0.65.

Since polarity is 0.15 it clearly states that statement is positive and subjective score is 0.65.

TextBlob may be used for identifying the sentiment either subjective or objective but VADER may be used for identifying the valence of the statement.

V. Conclusion

A practical, computationally cost-effective sentimental analysis is proposed in this paper for email conversation. It may help the email user to identify emotions of the user involved in email conversation. The proposed method can also be used in the identifying emotions of the user involved in short message service as well in any other social network.

REFERENCES

- [1] A.Pak and P. Paroubek. „Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326
- [2] R. Parikh and M. Movassate, "Sentiment Analysis of User- Generated Twitter Updates using Various Classification Techniques", CS224N Final Report, 2009.
- [3] Go, R. Bhayani, L.Huang. "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper, 2009
- [4] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010: Poster Volume, pp. 36-44.
- [5] Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer, 2010, pp. 1-15.
- [6] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011 Workshop on Languages in Social Media, 2011, pp. 30-38

- [7] Dmitry Davidov, Ari Rappoport." Enhanced Sentiment Learning Using Twitter Hashtags and Smileys". Coling 2010: Poster Volume pages 241-249, Beijing, August 2010
- [8] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social MediaData", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3 - 6, 2013, pp 91-96, ISBN: 978-1-494673-6068-5,
- [9] J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using wordnet to measure semantic orientations of adjectives," 2004.
- [10] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1-167, 2012.
- [11] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., &Stede, M.."Lexicon based methods for sentiment analysis". Computational linguistics, 2011:37(2), 267-307.
- [12] R. Piriyani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000-2015," Information Processing & Management, vol. 53, no. 1, pp. 122-150, 2017.
- [13] Samira Bordbar and Pirooz Shamsinejad, "A New Opinion Mining Method based on Fuzzy Classifier and Particle Swarm Optimization (PSO) Algorithm", Cybernetics and Information Technologies, Volume 18, Number 2, pp 36-50, 2018
- [14] Truong Duc Phuong , Do Van Thanh and Nguyen Duc Dung, "Mining Fuzzy Sequential Patterns with Fuzzy Time-Intervals in Quantitative Sequence Databases", Cybernetics and Information Technologies, Volume 18, Number 2, pp 3-19, 2018
- [15] Praveen Gujjar and T Manjunatha, "Parameter Evaluation of Indian Information Technology Companies Using DuPont Model", International Journal of Management and Social Sciences (IJMSS), 8 (2.5), pp 67-68, 2019