

Plagiarism Scan Report

Summary

Report Generated Date	30 Mar, 2018
Plagiarism Status	88% Unique
Total Words	983
Total Characters	6296
Any Ignore Url Used	

Content Checked For Plagiarism:

1 INTRODUCTION

Sentiment Analysis (also referred as opinion mining) is the study of affective states and sub-

jective information in the customer data (such as reviews and survey responses, online and social media) by using natural language processing and data mining techniques [1].

Sentiment

analysis aims to determine the attitude of a subject with respect to some topic or the overall

contextual polarity or emotional reaction to some object, such as a document, interaction, or

event. The attitude may be a judgment or evaluation, affective state, or the intended emotional communication.

For opinion mining or sentiment analysis some methods are applied like - Naive Bayes Machine Learning Classifier, Sentiwordnet, Support Vector Machine. Here we have used Lexicon

based approach of Sentiment Analysis. Sentiment lexicon is used in the lexicon based approach.

Sentiment lexicon is a collection of known and defined words. A specific sentiment is assigned

to each word in the collection. The lexicon based approach is divided into dictionary based approach and corpus based approach[2].

Sentiment analysis task is divided into three categories; Aspect level, Sentence level, Document level [3]. Aspect level analysis deal with the aspects of items. It can also be considered

as phrase level analysis. In Sentence level, each sentence is considered as an entity.

Summa-

tion method is used to provide overall result of the document. In document level, the whole document is considered as a single entity.

1.1 Motivation

Twitter Sentiment Analysis was thoroughly dealt by Alec Go, Richa Bhayani and Lei Huang, Computer Science graduate students of Stanford University. They used various classifiers, including Naive Bayes, Maximum Entropy as well as Support Vector Machines to classify the tweets. The feature extractors used by them were both unigrams and bigrams combined.

Parts of speech tag was used because same word may have different meaning depending on its

usage. The data-set used by them was huge, comprising 1.6 million tweets divided equally into positive and negative classes.

We have chosen to work with twitter since we feel it is a better approximation of public senti-

ment as opposed to conventional internet articles and web blogs. The reason is that the amount

of relevant data is much larger for twitter, as compared to traditional blogging sites.

Moreover

the response on twitter is more prompt and also more general (since the number of users who

tweet is substantially more than those who write web blogs on a daily basis). Sentiment analy-

sis of public is highly critical in macro-scale phenomena like predicting the needs of tourist and

their opinions on the tourism spot. This could be done by analysing overall public sentiment towards the place with respect to time for finding the correlation between public sentiment and

the place of interest. The government can also estimate the changes to be made, facilities to be

provided to attract more tourists in the future and in which a negative response was registered

since twitter allows us to download stream of geo-tagged tweets for particular locations.

Other

applications of Sentiment Analysis includes the review of movies and products, popularity of

an event. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good

reflection of offline sentiment

1.2 Literature Survey:
Sentiment analysis has been studied in wide area of domain such as movie review, teaching

review [4], product review, e-learning, hotel review and many more. A small number of studies

have focused on applying machine learning techniques in the tourism sector.

A study [5] aimed to create a system that would assist users in understanding tourism opinions on the web by finding and extracting subjective information from reviews in tourism

websites. Aspect extraction was performed with the use of frequent nouns and the opinion was

determined.

Estela Marine-Roig et al.[6] addressed the problem of finding out the frequently occurring trends of different tourist places from tourist opinions. The authors proposed a trends

extraction framework that consisted of five phases i.e. semi automatic downloading, arrang-

ing, cleaning, debugging, and analyzing. Trends extraction framework is better than previous

method Liu (2011) in trends extraction because two extra phases of cleaning and debugging

has been added up to eliminate the noise present in the tourist's opinions. The limitations of the work are that i. Method does not classify the derived frequent trends into positive and negative trends ii. method extracts same trend in one opinion sentence multiple times that create the reputation of trends iii. Method extracts many irrelevant and meaningless trends during classification.

In another way to enhance the performance of opinion sentences extraction Shimada, K.. [7] used support vector machine for sentences classification. The authors addressed the problem to identify whether tweet on-site are more likelihood or tweet off-site. The authors proposed a method to evaluate on-site likelihood. Firstly, this method takes tweets and identifies tourism related tweets. Secondly, extracts tourism related tweets and deletes the remaining ones. Lastly classifies the extracted tweets on the basis of different features of tourist places using SVM. The finding of this paper is that classification has improved by applying the method of on-site likelihood filtering method. The same fact is shown in the results i.e. without applying this filtering method Recall=58.2% and Precision = 75.0% and after applying the filtering method Recall=65.0% and Precision=80.5% . If there is a location name at the start of any tweet then it is high onsite likelihood tweet. The limitations of the work are that i. mostly the comments of authors on tweets are more than any other person which are mostly positive or negative that create noise in sentiment analysis ii. method extracts some sentences in which no opinion about targeted tourist place is given that creates noise during classification of reviews.