International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Furnishing Efficient Opinion from Federation of Context and Emoticons

Priya C [1], Sankar Ganesh K [2]

[1]PG Scholar, Department Of Computer Science and Engineering, Kingston Engineering College

[2]Assistant Professor, Department Of Computer Science and Engineering, Kingston Engineering College

***Abstract**: The use of emoticons in text to express opinion and sentiment has become more and more common. So emoticons play an important role in conveying a text's overall sentiment. This project aims to extract the meaning of the emoticons used in context of the message and provide a more clear and accurate analysis of the sentiment of the text.*

Keywords: emoticons, opinion, sentiment analysis, context

1. Introduction

1.1 Webmining

Webmining is the process of obtaining useful information from the contents of web. Content data is the agglomeration of facts a web page is designed to contain. It may contains of text, images, audio, video, or structured records such as lists and tables.

1.2 Web Content Mining

Web content mining is basic process of extracting all the knowledgeable information from the web content. This mining varies in the form of text image, audio, video.

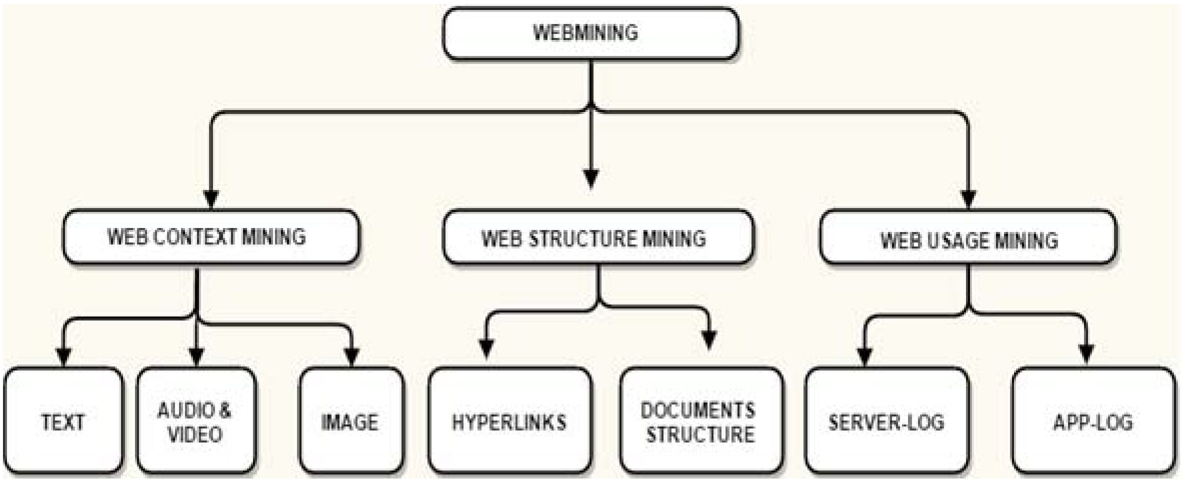

Figure 1.1: Web Mining Taxonomy

1.3 Web structure mining

Web structure mining is that providing the hyperlinks to the website to avoid the irrelevant searches.

1.4 Web usage mining

This web usage mining is that history or logs that stored in the database that which viewed in the web

2. Sentiment Analysis

Sentiment analysis is the process of data pre-processing, identifying and segregating opinions expressed in a context, especially in order to take decision whether the user's attitude towards a particularized topic, product, etc. is positive, negative, or neutral. But now a days it is becoming more and more difficult to identify the meaning and mood of the reviewer as the message becomes confusing due to vague statements, contradictory sentences, and sarcasm.

One such factor is the increasing use of emotions a representation of a facial expression such as a smile or frown to express emotions over digital media. An emoticon may signal the intended sentiment of a differently intensive statement, e.g., "It's too hot here ☹ ".Therefore, this projects analyses how emoticons are typically used to convey sentiment and detects the actual opinion of the person by examining the emoticons in context to the text.

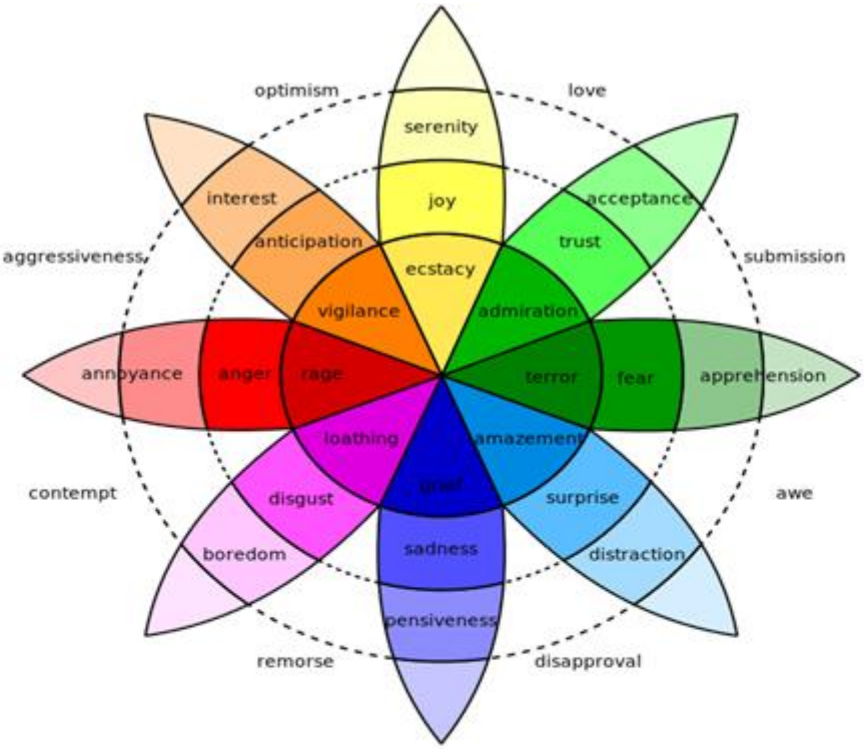

Figure 2.1: Variance of Sentiments

3. System Analysis

3.1 Existing System

Current sentiment analysis approaches typically do not consider emoticons. On the contrary, one of the first steps in most existing work is to remove many of the typographical symbols typically fix emoticons, thus preventing emotag from being detected at all. Even in systems that identify emoticons, they are considered as stand-alone entities and not analyzed in context to the message.

3.2 Problem Statement

As analyzed the existing system, the opinion mining from content based emoticon are not done up to the accuracy and be consistent. In this paper I focus on the identification of polarity in the text and the emoticon(e.g., this mobile is terrific to use ☺)I attempt to pinpoint the words with the polarity and classify them as happy, unhappy, neutral and mine the opinion in large amount of data in less time in good consistency.

3.3 Disadvantage

1) They have inconsistencies.
2) These dictionaries do not address the concept of polarity consistency of words/synset.
3) Takes more time to analyze in large number of data set.
4) At times gives false result

3.4 Proposed System

The proposed system will analyze the emoticons used in message and determine the sentiment of the text. First I identify the meaning of the emoticons. Then to understand it in the context of the text and I analysis where it is placed, how it affects the text, what it expresses and relate it with the message to provide a clear view of what the person is trying to convey.

Architecture Diagram

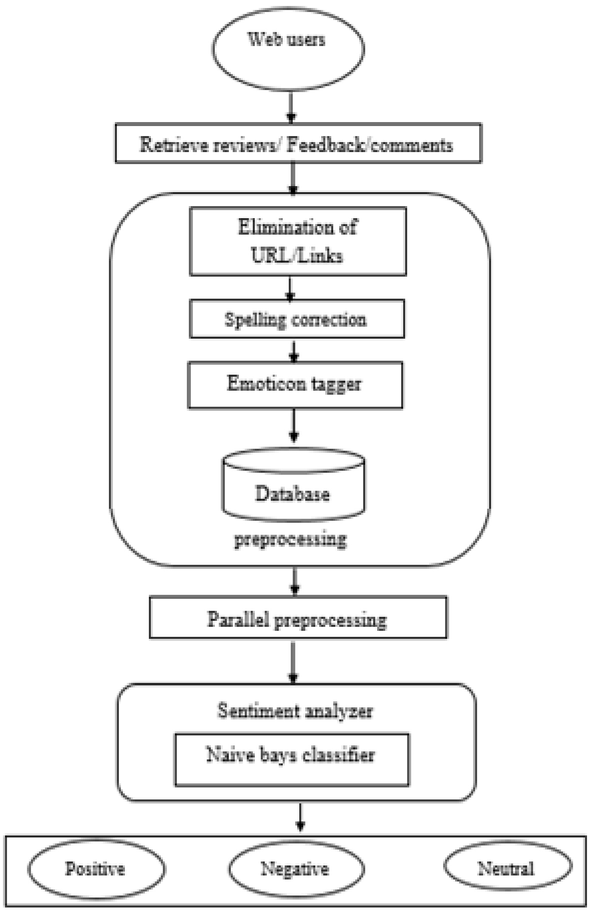

Figure 4.4: Architecture Diagram

4. Methodologies Description

4.1 Data preprocessing and clean up

This module is responsible for preprocessing, clearing and tokenizing the original texts and converting them into Document objects.Raw text is parsed by mining key words and irrelevant information. Key words are determined through a number of factors, including frequency within a document; frequency across all documents; and part-of-speech tagging. The main idea behind the pre-processing stage is to turn unstructured raw data into a shorter, structured format that maintains all the characteristic features of the original text without losing accuracy

4.2 Lexicon based text sentiment analysis

Lexicon based techniques use a dictionary to perform entity-level sentiment analysis. This technique uses dictionaries of words define with their meaningful orientation (polarity and strength) and calculates a score for the polarity of the document.
Naive Bayes classifier is used to classify the polarity of the text. It models that assign class labels to problem text, represented as vectors of countenance values, where the class labels are drawn from some finite set.

- **Common lexicon:** This contains data that would have the same semantic meaning or sense across different domains and categories.
- **Common or default sentiment word:** Positive and Negative sentiment words that have the same sentiment value or sense across different domains
- **Negation Words:** Negation words are the Words which reverse the polarity of sentiment. For example, “The battery life is not good” has negative sentiment.

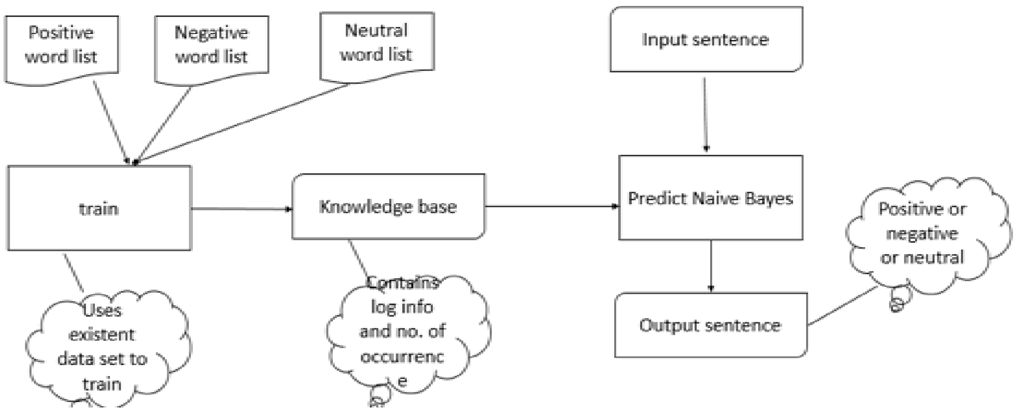

Figure 4.1: Naïve Bayes layout

Algorithm
Multinomial Naive Bayes is a specialized version of Naive Bayes that is designed more for text documents. Whereas simple naive Bayes would model a document as the presence and absence of particular words, multinomial Naive Bayes explicitly models the word counts and adjusts the underlying calculations to deal with in. This is the event model typically used for document classification, with events representing the occurrence of a word in a single document. If a given class and feature value never occur together in the training data, then the frequency-based probability estimate will be zero.

Multimonial Naivy Bayes Classifier
TrainMultinomialNB(C, D)
V—ExtractVocabulary (D)
N---CountDocs (D)
For each c € C
Do N_c---CountDocsInClass (D, c)
Prior[c] ---N_c/N
Text_{c---}ConcatenateTextOfAllDocsInClass (D, c)
For each t € V
Do T_{ct}---$T_{ct}+1/\sum_{t'} (T_{d'}+1)$
Return V, prior, condprob.

Chi Square Test:
A chi-squared test, also referred to as χ^2 test (or chi-square test), is any statistical hypothesis test in which the sampling distribution of the test statistic is a chi-square distribution when the null hypothesis is true. Chi-squared tests are often constructed from a sum of squared errors, or through the sample variance. Test statistics that follow a chi-squared distribution arise from an assumption of independent normally distributed data, which is valid in many cases due to the central limit theorem

$$X^2=\sum \textit{(observed value-expected value)}^2 / \textit{(expected value)}$$

4.3 Emoticon identification and analysis

Identify the emoticons used, its meaning and the sentiment conveyed by it. Emoticon sentiment lexicon, which is defined as a list of character sequences, representing emoticons, and their associated sentiment strength is used. These emoticons. May be standardized into emoticon synsets, which we define as groups of emoticons denoting the same emotion

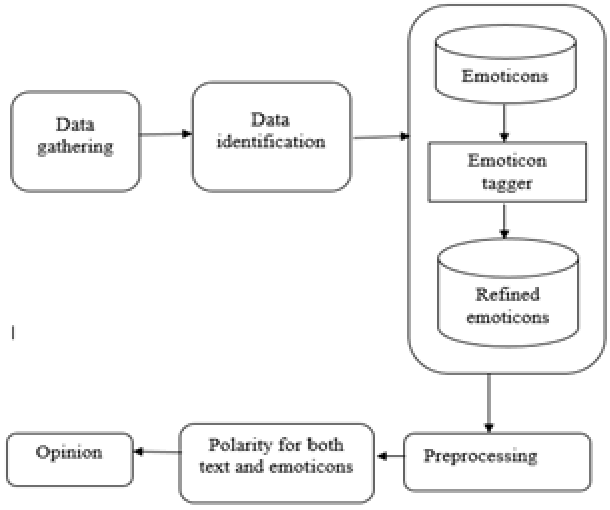

Figure 4.2: Emoticon tagger

4.4 Sentiment aggregation from context

The results from the text and emoticon analysis are examined together and aggregated. Using comparative analysis the context of the text is identified and the emoticon is examined in relation to the text in the document.Thus, the document's sentiment score is returned.

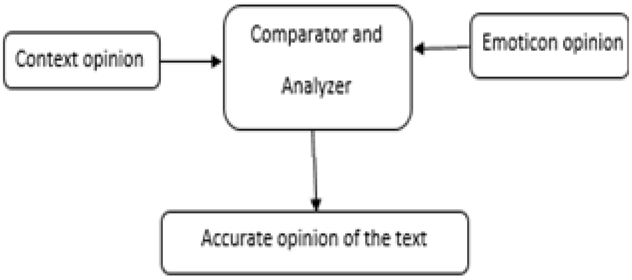

Figure 4.3: Text and Emoticon Comparator

4.5 Advantage

1) Provide consistency in classification
2) Identify exact meaning of the text and emotion
3) Provide accurate sentiment in context based text and as well as in emotion
4) Increase the efficiency and reduce the time

5. Conclusion

The proposed system extracts the review, comments, feedback from the customer. The reviews may contain emoticons also.Naive Bayesian algorithm is used to supervise the polarity in the context and identify the given context is positive or negative or neutral. The emoticon is classified using the emoticon classifier and identified that the emoji's is positive or negative or neutral .Now, the number of positive and negative and neutral comments from the context and the emoticons are identified and the results of the emoticon and the context are aggregated together and the opinion over the product or the particular topic is presented with the good and accurate opinion. This is proposed to get the exact opinion from the large amount of reviews without any inconsistencies and provide high efficiency and avoid the fault opinions.

References

[1] Soman, K. P., DiwakarShyam and Ajay V. "Insight Into Data Mining: Theory And Practice", PHI, 2009.
[2] Jiawei Han, MichelineKamber , Jian Pei, "Data Mining: Concepts and Techniques", Third Edition (The Morgan Kaufmann Series in Data Management Systems), 2012
[3] David J. Hand, HeikkiMannila and Padhraic Smyth "Principles of Data Mining" (Adaptive Computation and Machine Learning), 2005
[4] Mining Opinion from Text Documents: A Survey Khairullah Khan, Bah arum B.Baharudin, Aurangzeb Khan, Fazal-e-Malik 2010.
[5] An Opinion-Tree based Flexible Opinion Mining Model Juling Ding, Zhongjian Le, Ping Zhou, Gensheng Wang.
[6] An Approach Based on Tree Kernels for Opinion Mining of Online Product Reviews Peng Jiang1, Chunxia Zhang2, Hongping Fu1, Zhendong Niu1, Qing Yang1
[7] Opinion Mining and Social Networks: a Promising Match Krzysztof Jędrzejewski, Mikołaj Morzy,2011
[8] SentimentAnalysisforCantoneseOpinionMining, Jian Chen School of Software Engineering South China University of Technology Guangzhou, China,2012
[9] Tweet Segmentation and Its Application to Named Entity Recognition Chenliang Li, Aixin Sun, JianshuWeng, and Qi He, Member, IEEE,2015
[10] Polarity Consistency Checking for Domain Independent Sentiment Dictionaries Eduard C. Dragut, Member, IEEE, Hong Wang, Prasad Sistla, Clement Yu, and WeiyiMeng, Member, IEEE2015
[11] Language and Discourse Are Powerful Signals of Student Emotions during Tutoring Sidney K. D'Mello and Art Graesser2013
[12] CO-Extracting Opinion Targets and Opinion Words from Online Reviews Based On the Word Alignment Model Aditya Rane, SankalpRane, SailySawant, ShubhamSali Prof. Sunil Jadhav 2015.

Author Profile

Priya C received B. Tech in information technology and perusing M.E Computer Science and Engineering in Kingston Engineering College, Vellore

Sankar Ganesh K M.E., MISTE Assistant Professor, Department of Computer Science and Engineering, Kingston Engineering College, Vellore