Multi stratagem analysis of sentiments on twitter data using partial phrase harmonizing

T M Saravanan¹*, T Kavitha², S Hemalatha², Ankit Kumar³
¹Associate Professor, Department of Computer Applications
²Assistant Professor (Sr.G), Department of Computer Applications
³PG Scholar, Department of Computer Applications
Kongu Engineering College, Perundurai – 638 060, Tamil Nadu, India
*Email: drtmskgm@gmail.com

Abstract. Sentiment analysis is constructive in the application environment for business intelligence and suggests systems because it is a very easy medium for the two ends of the availability to communicate. Numerous strategies and schemes have been worn inside the sentiment analysis, such as language processing, polarity lexicons, machine learning, and psychometric scales which establish diverse types of analyzing sentiments as assumptions ended, scheme reveals, and corroborate data set. Since the internet has to turn into a commanding resource of retrospect the sphere of sentiment is moreover referred to as Sentiment Analysis or Opinion Mining. It has seen an enormous boost in academia over the decades. Analyzing sentiment to extract sentiments in different levels like word, sentence, and document provides articles’ feeling polarities. While well identified consumers’ sentiments articulated in sentences by opinion. Customary machine learning schemes cannot virtuously mirror the views of writers. This paper proposes a scheme called multi-strategy sentiments with semantic resemblance to disentangle the topic with partial phrase matching. Additionally, the Naïve Bayes classification is applied to search for the probability of the distribution of knowledge in different categories of knowledge set.

1. Introduction

Sentiment analysis is commonly employed in opinion mining for knowing sentiments, subjectivities moreover sensitive states in online texts. The process was accomplished on product evaluation by organizing the products attributes. At the present time, sentiment polarity analysis is utilized in extensive range of domains like in finance.

This concentrates on examining the direction-based text that involves text that contains statements or opinions. The process of sentiment classification investigates whether the specific text is subjective or objective or if the text constitutes both the feelings of positive or negative. This classification method has
much number of essential qualities that may include various process, jobs, techniques, attributes and also application domains.

There exists much number of jobs in the classification of sentiment polarity. There are three major characteristics of this classification are class, level besides assumption with respect to sentiment sources as well as targets. The distinctive two class problem incorporates the categorization of sentiments as positive or negative. Furthermore changes include organizing messages as subjective / objective. Sentiment analysis concentrates on the specification of user’s point of view with respect to specific area.

Analyzing sentiment is contextual text mining that recognizes and extracts subjective knowledge from source and allows a company to know its brand, product when tracking online discussions. However social media analysis is confined to basic analysis of sentiments and metrics.

It’s like scratching the surface and losing the important knowledge is searching for creative use of state of art AI techniques is also an important method for detailed analysis.

It is important to classify a few brands that support the following lines in the customer dialog:
1. Key product aspects and repair aspects of brand that concern customers.
2. The fundamental interests of users and their responses to these problems.

When used in combination, these basic concepts become a real important tool for analysis with human precision of many brand conversations. Intentional research improves sport by evaluating a message’s intention and figuring out how it applies to views, news, marketing, concerns, feedback, gratitude or inquiries.

Currently, the internet is a forum to express opinions and exchange experiences and it is not the source of information. Feedbacks are normally gathered within the network about the product tweeted by customers. Since it is an incredibly convenient communication platform for both ends of supply, believing analytics are useful in the setting for commercial intelligence and suggest systems. Various methods and techniques, such as machine learning, lexicons of polarity, natural language processing, and psychometric scales, have been used in feeling analysis, which analyse different kinds of sensation analysis, such as assumptions made, system reveals and validation datasets.

research generally takes place at three levels: word, term, and record level, where the majorities of recent studies usually use the term and document. However, the degree of the word is the fundamental and thus it is seldom considered to be more important and demanding. In fact, the short sentences of one or two Chinese characters in Chinese as one language are most frivolous. This function can not be mirrored in traditional machine learning schemes. This study therefore proposes a new hybrid sentiment analysis that uses the fluid set theory of the machine learning and the polarity lexicon approach fully.

Western thinkers began to understand emotions earlier. First they address the propensity of words or phrases to feel and calculate them as real values, which can be further used for deciding the propensity of phrases or paragraphs to feel. The pattern of feeling was examined. NB (Naive Bayes), M E (MaxEntro Maximum Entropy) and SVM (Support Vector Machine) are three key feel analysis algorithms for machine learning. For simplicity of analysis, we choose NB and SVMs.

Sentiment analysis is one of the complexes methods that consist of five important phases for examining sentiment data. The sequence of sentiment analysis process is shown in Figure 1.1. The phases includes,

i. Data collection,
ii. Preparation of text,
iii. Sentiment detection,
iv. Sentiment classification,
v. Displaying output

The various levels of sentiment analysis is depicted in the following figure 1,

![Levels of Sentiment Analysis](image)

**Figure 1. Levels of Sentiment Analysis**

### 2. Review of Literature

#### 2.1. System for Sharing Recommendations

Loren Terveen [1] stated that empirical findings support the feasibility of automatic recommendation recognition. First, Usenet messages are an overwhelming source of web resources recommendations: 23% of usenet messages relate to web resources, and 30% are recommendations. Secondly, machinery recognized instances of advice also have almost 90% accuracy. Third, quite a few resources are suggested by one person. The recommendations reported tend to be valuable resources for the respective community.

Finally, a reasonable indicator of resource quality is the number of independent resource recommenders. The more distinct recommenders a resource has, the more often it appears in the FAQs, is a comparison of the suggested services in FAQs (lists of Commonly Asked Questions compiled by human subjects specialists). Two main design principles: specification and reusable are differentiated from other recommending systems by PHOAKS.

What is recommended? What is important? The fundamental principle of collaborative filtration is that people suggest objects to each other at least. Usenet news readers know that this is also a conventional
newsgroup practice. Posters also offer their opinions and views, like web pages, on all kinds of subjects. You can tell what a page is good for and how useful it is: PHOAKS searches for site references (URLs) and takes a note as a suggestion if a number of tests are conducted.

The message must not be sent to so many newsgroups in the first place. Messages from a large number of groups are so generic that they are actually not related to any of the groups thematically. Second, whether the URL is a signature or signature file part of a document, it's not a recommendation. Third, if the URL happens in a previous message's quoted portion, it is not included. Fourthly, if the URL textual structure contains word markers which indicate that it is recommended and does not contain markers that indicate that it is marketed or promoted, it is listed as a recommendation. The categorization regulations have been quite complex and have introduced this fundamental technique to identify the different goals of web resources.

The future work includes the following things as mentioned by the author: Firstly, they continue the study of the relationship between Usenet messages' suggested tools and FAQs' sources. The temporal dimension of particular interest to them. So they can for instance assess the degree to which Usenet messages are a big FAQ content predictor. Second, FAQs are used to boost the recommendation data system. For example, one would be prepared to use the references to the resource in FAQs from a database. We plan to combine the best of recommendations that we immediately disregard (for example, timeliness) with ethical recommendations (for example, long-term significance and quality).

2.2. Exploiting Microblogging Social Ties for Sentiment Analysis

Xia Hu and Lei Tang [2] said Micro blogging has, like Twitter, become a popular human expression medium that allows users to easily generate news, public events or items. Mass feelings and thoughts about different topics can be a valuable resource for the vast number of micro blogging data. In general, this consideration constructs an aesthetic space to manage bright and short messages without the very fact that the micro-blogs are networked content.

Emotional theories of infectivity in supervised learning process and sparse earning in micro-blogging address loud texts. An observational analysis of two Real World Twitter data sets reveals the high performance of our short noisy tweets management system.

Micro blogging sites are commonly used in various fields for exchanging knowledge or opinions. As such a tool with an increasing abundance of opinion, it attracts a great deal of interest from those who seek to understand individual views or to measure the overall feeling of mass populations. For example, marketers may target users who want to start actively using a brand or product in social media. Agencies around the world continue to track developments before, during and after the crisis to facilitate recovery and to provide disaster relief.

Entire volume of knowledge in micro blogs poses opportunities and difficulties to study such short and noisy texts. Sentiment analysis for product and film reviews, which distinguish significantly from micro blogging results, was extensively studied. In micro blogging, the text is a few phrases or 1-2 sentences, as opposed to regular text with several terms that help collect statistics. Users can also use and invent novel acronyms which is rarely used in traditional documents when writing a micro blogging post.

Consider the example, messages such as "It’s cooool," and "OMG" are perceptive and common on micro blogs but some are not structured words. The semantic meanings of such messages are difficult for machines to precisely recognize, but they provide user friendliness in fast and instant communications for people.
One distinct feature of microblogging is that it is possibly connected via user connections that may contain useful semantic indices that cannot be found purely in text-based methods. Modern approaches do not use social relationship information when applied directly to microblogging data. It is well known in social science that emotions and feelings play an imperative role in our life pertaining to social media. When you feel feelings, you don't normally hold your feelings, you prefer to express them. Indivisible verbal and postural input, known in social science as emotional interference, often appear to take up emotions from others.

In personal relationships, it can be significant because emotional contagion "promotes convincing synchrony and monitoring of feelings of others, even if people do not directly listen to the details." The emotional contagion is the product of Fowler and Christakis recording the spread of joy in a social network.

![Figure 2.1: Data Representation of Message Content in Micro Blogging](image1)

![Figure 2.2: Data Representations of Social Relations in Micro Blogging](image2)

The figure 2.1 and 2.2 explain the phenomenon by two social processes, selection and influence: people who become friends or similar to their friends over time. Both explications show the possibility of similar behaviors or opinions being expressed by connected individuals. Inspired by these sociological findings, we speak about using social media information to encourage feelings research in the context of micro-blogging. The purpose of this paper is to provide a supervised approach to the study of microblogging feelings in order to understand the brilliant essence of message by learning the information related to social relations. They investigated, in particular, whether microblogging information contains social theories. They then talked about how social relations can be shaped and used for the supervised analysis of feelings.
3. Proposed Method

In proposed system, like existing system, data set is taken as records from Excel worksheet with category in second column. Preprocessing work is carried out. Then words combinations are found out and valid phrases are gathered.

These phrases conditional probability is found out among all categories which become Naïve Bayes Classification work. In addition, synonym words replacement is also made. Moreover, partial phrases like two words in one sentence and three words in other sentence are also treated as same phrases during naïve bayes classification.

The study of emotions is very critical and the task at word level is more difficult. The first step was therefore to construct a lexicon of feeling which would infer the polarities of feeling and words. There should be specified two types of emotional sentences: the fundamental and compound sentences as specified below: simple phrases which have two letters and no derogation or modifications. Composite emotion phrases are sentences with more than two characters or negative sentences or modifications.

The Naïve Bayes (NB) algorithm is widely used as a classification in document categorization. In a emotion analysis, Naïve Bayes addresses at first the labeled training corpus where every document knows the feeling polarities. The latter analyses the probability of a document that corresponds to different classes, provided the labels of function, which are then assigned to the higher probability groups. Every article is played and words of feeling are taken from the training corpus. Then the following probability is determined according to Equation (1) for each word of feeling and reported in a table of probability.

\[
P(c \mid d) = \frac{P(c) \ast P(d \mid c)}{P(d)}
\]

- \(P(c \mid d)\) - posterior probability
- \(P(c)\) - prior probability
- \(P(d \mid c)\) - likelihood
- \(P(d)\) – predictor of prior probability of class

So, if the problem is handling the classification of naïve bayes, then multi strategy analysis of sentiment on user reviews-based resting on semantic similarity.

4. Working Model
The figure 4.1 architecture diagram is discussed below,

4.1. Download Twitter Data

Twitter Data is downloaded using 'twitter' package, in which two or more search words such as tablet, mobile and laptop are given. In the files 'laptoptwitter.csv', 'tablettwitter.csv' and 'mobiletwitter.csv' all three contents are saved. The first column contains laptop, second has tablet and third column has mobile tweet posts.

4.2. Preprocess Twitter Data

In this phase Twitter Data is preprocessed using ‘tm’ package in which stemming, stop word removal and URL link removal is carried out. All the words are converted into lower case.

4.3. Sentiment Words File Creation

Here, a .csv file created in which sentiment phrase, category and sentiment value is being added as records. The category is one of laptop, tablet and mobile. The sentiment value is from -5 to +5 based on importance.

4.4. Two Adjacent Word Phrase Combination

At first, Twitter Data is converted into two words phrases such as first word and second word as one phrase, second word and third word as next phrase and so on for all tweets. These phrases are checked with sentiment value records taken from ‘sentimentvalues.csv’ created in previous module. If the phrase is matched with sentiment phrase then sentiment value of the corresponding category is taken and added. For all tweets, mobile category’s positive and negative score is found out and displayed. Likewise tablet and mobile categories are also prepared. Then conditional probability of these phrases in all the three categories are found out and displayed.

4.5. Three Adjacent Word Phrase Combination
Twitter data has been translated into three phrases of terms, namely first word, second word and the third word, for all tweets. The first sentence is a second, third and fourth letter. These phrases are checked with sentiment value records taken from ‘sentimentvalues.csv’ created in previous module. If the phrase is matched with sentiment phrase then sentiment value of the corresponding category is taken and added. For all tweets, mobile category’s positive and negative score is found out and displayed. Likewise tablet and mobile categories are also prepared. Then conditional probability of these phrases in all the three categories are found out and displayed.

4.6. Missed Word Phrase Combination

Here phrases are formed with middle word deletion from the previous module phrases. These phrases are checked with sentiment value records taken from ‘sentimentvalues.csv’ created in previous module. If the phrase is matched with sentiment phrase then sentiment value of the corresponding category is taken and added. For all tweets, mobile category’s positive and negative score is found out and displayed. Likewise tablet and mobile categories are also prepared. Then conditional probability of these phrases in all the three categories are found out and displayed.

5. Conclusion

A new approach proposed in this paper for measuring polarities and sentimental sentence strengths, which could be used also with partial sentence matched to evaluate the semantic similitude of sentences. It uses a probability value in contrast with traditional approaches and uses a normal value for the polarity of sentimental sentences. It proposes a multi-strategic sentiment analysis scheme focused on the polarities and strengths of certain words. It considers adverse conjunctures, particularly in the NB-based scheme. The system can be used to evaluate the documents' emotions. The approach was shown to be feasible and efficient. The shift will reflect in the future on how the photos of Emoticons and Unicode characteristics are close to those found.

References

[1] Terveen Loren, Hill William, Amento Brian, McDonald David, Creter Jos 1997 PHOAKS: A system for sharing recommendations Communication of ACM 40 59-62 10.1145/245108.245122.
[2] Hu Xia, Tang, Lei, Tang Jiliang, Liu Huan 2013 Exploiting social relations for sentiment analysis in microblogging WSDM 2013 - Proceedings of the 6th ACM International Conference on Web Search and Data Mining 537-546 10.1145/2433396.2433465.
[3] Tatemura Junichi 2000 Virtual Reviewers for Collaborative Exploration of Movie Reviews International Conference on Intelligent User Interfaces Proceedings IUI. 10.1145/325737.325870.
[4] Ku Lun Wei, Liang Yu Ting, Chen Hsin His 2006 Opinion Extraction, Summarization and Tracking in News and Blog Corpus AAAI Spring Symposium - Technical Report.
[5] König Arnd, Brill Eric 2006 Reducing the human overhead in text categorization 598-603 10.1145/1150402.1150474.
[6] Fang Y, Tan H, Zhang J 2018 Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness IEEE Access vol. 6 pp. 20625-20631 doi: 10.1109/ACCESS.2018.2820025.
[7] Saravanan T M, Prakash N, Selvambal K, Arthikha K, Monika S 2020 Multi Rated Sentiment Mining on Movie Reviews with User Preference Features International Journal of Advanced Science and Technology, 29(3) pp. 11053 - 11059.
[8] Pyingkodi M, Shanthi S, Thenmozhi K, Saravanan T M, Hemalatha D, Sudarshan Y 2020 Skin Cancer Classification Towards Melanoma Detection With Deep Learning Techniques International Journal of Advanced Science and Technology, vol. 29 No. 9 pp. 3911-3918.
[9] Pyingkodi M, Shanthi S, Saravanan T M, Thenmozhi K, Nanthini K, Hemalatha D, Muthukumaran M, Dhivya M 2020 Performance Study Of Classification Algorithms Using The Microarray Breast
Cancer Dataset. *International Journal of Future Generation Communication and Networking* **vol. 13** No. 2 pp. 1238-1245.

[10] Saravanan T M, Kavitha T, Monika S, Arthikha K, Dhevashree S Y, Dhanalakshmi M 2020 Reply Instance Based Best Possible Selection of Web Service for Music Recommendation *International Journal of Advanced Science and Technology* **29(05)** pp. 11166-11171.