

Performing Sentiment Analysis on the Product Brands using Tweets

Madhura G K

MTech 4th semester
Akshaya Institute of Technology,
Tumkur, Karnataka

Prof. Shivamurthy R C

Prof and Head: Dept. of Computer Science and Engineering,
Akshaya Institute of Technology,
Tumkur, Karnataka.

Abstract— In this recent years the concept of sentiment analysis of twitter data and semantic analysis with the enhancement of machine learning approaches is an important issue. In the past few years so many techniques have been proposed in the area of sentiment analysis to analyze the social media data and to provide a graphical presentation towards a particular brand. Sentiment analysis shows you the sentiment of people while typing a social media message about some product or brand. This is important information if you know that the sentiment of one person can influence other person about a company or their products. Sentiment analysis platforms are like all other online data mining systems, they are based on Support Vector Machine algorithms. In this case this algorithm which recognize certain words as ‘positive’ or ‘negative’, letting you know if your brand is being adored or floored. So here in this paper we first pre-processed the dataset, after that extracted the adjective from the dataset that have some meaning which is called feature vector, then select the feature vector list and thereafter apply machine learning based classification algorithms namely: Naïve Bays, Maximum entropy and SVM along with the Semantic Orientation based WordNet which extracts synonyms and similarity for the content feature. Finally we measured the performance of classifier in terms of recall, precision and accuracy.

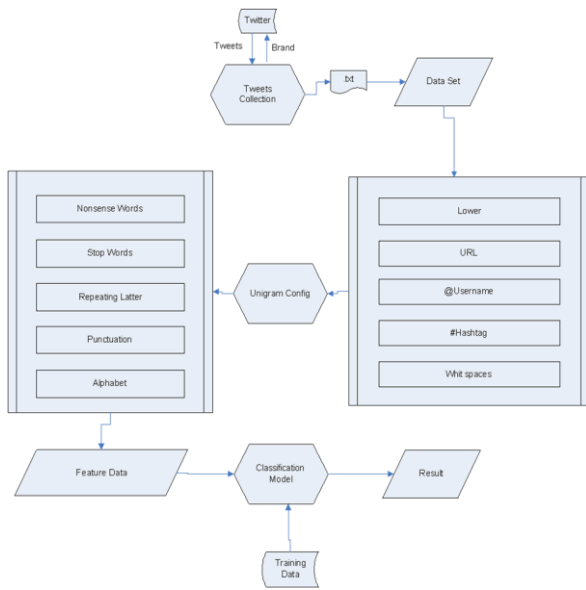
Keywords: *Machine Learning, Semantic Orientation, Sentiment Analysis, Twitter*

I. INTRODUCTION

The current research paper covers the analysis of the contents on the Web covering lots of areas which are growing exponentially in numbers as well as in volumes as sites are dedicated to specific types of products and they specialize in collecting user’s reviews from various sites such as Amazon, ebay etc. Sentiment analysis shows you the sentiment of people while typing a social media message about some product or brand. This is important information if you know that the sentiment of one person can influence other person about a company or their products. Sentiment analysis platforms are like all other online data mining systems, they are based on Support Vector Machine algorithms. In this case this algorithm which recognize certain words as ‘positive’ or ‘negative’, letting you know if your brand is being adored or floored. If someone tweets using the word ‘awful’, ‘disappointing’ etc. then the sentiment analysis will assign that post as ‘negative’. You can get problems though if someone uses sarcasm or irony as the tone... most tools, takes everything at face value. Sentiment analysis research involves giving a

better solution in enhancement of Vector Support Machine Algorithm in classifying sentiments towards a product brand. Sentiment analysis research is based on some Support Vector Machine Learning Algorithm to solve the classifying sentiments and positive or negative attitude of a customer towards a company’s certain products via any social media like Facebook or twitter. This research aims to improve the algorithm’s accuracy through choice of kernel and proper tuning of SVM hyper-parameters as core factors in contributing to SVM accuracy, having a huge amount of training sets in order to widen the hyper plane of vectors and strong support vectors. Sentiments will be gathered using twitter API and then they are sent to pre-processor and filter the unnecessary words. After that, the pre-processed sentiments will be converted to SVM format and use the formatted data as data Train. We know that sentiment is a complex combination of feelings and opinions as a basis for action or judgment. Generally these sentiments are what users have experienced about the products of a reputed company. Sentiment analysis involves the analysis of comments and suggestions left on social media sites such as blogs and social networks. Not only the words will be evaluated, all the feedbacks and sentiments of the users towards a brand will be analyzed. The words as variable such as context from twitter or Facebook status will be considered under our research. We will try to introduce an optimize solution to the machine learning algorithm.

II. ARCHITECTURE



III. ISSUES IN SENTIMENT ANALYSIS

Research shows that sentiment analysis is more difficult than traditional topic based text classification, despite the fact that the number of classes in sentiment analysis is less than the number of classes in topic-based classification. In sentiment analysis, the classes to which a piece of text is assigned are usually negative or positive. They can also be other binary classes or multi-valued classes like classification into 'positive', 'negative' and 'neutral', but still they are less than the number of classes in topic-based classification. The main reason that sentiment analysis is more difficult than topic-based text classification is that topic-based classification can be done with the use of keywords while this does not work well in sentiment analysis.

A. The Problem of Sentiment Analysis:

- **Object identification:** - The objects to be discovered in this blog are “moto” (Motorola) and “Nokia”. This problem is important because without knowing the object on which an opinion has been expressed, the opinion is of little use. The issue is similar to the classic named entity recognition problem. However, there is a difference. In a typical opinion mining application, the user wants to find opinions on some competing objects (e.g., products). The system thus needs to separate relevant objects and irrelevant objects. For example, “BestBuy” is not a competing product name, but the name of a shop.
- **Feature extraction and synonym grouping:**-In the example above, the phone features are “voice”, “sound”, and “camera”. Although there were attempts to solve this problem, it remains to be a major challenge. Current research mainly finds nouns and noun phrases. Although the recall may be good, the precision can be low. Furthermore, verb features are common as well but harder to identify. To produce a summary similar to the one, we

also need to group synonym features as people often use different words or phrases to describe the same feature (e.g., “voice” and “sound” refer to the same feature in the above example). This problem is also very hard. A great deal of research is still needed

- **Opinion orientation classification:**- This task determines whether there is opinion on a feature in a sentence, and if so, whether it is positive or negative. Existing approaches are based on supervised and unsupervised methods. One of the key issues is to identify opinion words and phrases (e.g., good, bad, poor, great), which are instrumental to sentiment analysis. The problem is that there are seemly an Visual comparison of feature-based opinion summaries of two cellular phones positive Picture , Battery, Camera, Size, Weight Cellular Phone 1 Cellular Phone 2. There are unlimited numbers of expressions that people use to express opinions, and in different domains they can be significantly different. Even in the same domain, the same word may indicate different opinions in different contexts [. For example, in the sentence, “The battery life is long” “long” indicates a positive opinion on the “battery life” feature. However, in the sentence, “This camera takes a long time to focus”, “long” indicates a negative opinion. There are still many problems that need to be solved.
- **Integration:** - Integrating the about tasks is also complex because we need to match the five pieces of Information in the quintuple.
- **Object and feature:** - The holder of an opinion is the person or organization that expresses the opinion. In the case of product reviews and blogs, opinion holders are usually the authors of the posts. Opinion holders are more important in news articles because they often explicitly state the person or organization that holds a particular opinion.
- **Opinion and orientation:** - An opinion on a feature f (or object o) is a positive or negative view or appraisal on f (or o) from an opinion holder. Positive and negative are called opinion orientations. With these concepts in mind, we can define a model of an object, a model of an opinionated text, and the Mining objective, which are collectively called the feature-based sentiment analysis model.
- **Direct opinion:**- A direct opinion is a quintuple (oj, fjk, ooiykl, hi, tl), where oj is an object, fjk is a feature of the object oj, ooiykl is the orientation of the opinion on feature fjk of object oj, hi is the opinion holder and tl is the time when the opinion is expressed by hi. The opinion orientation ooiykl can be positive, negative or neutral.
- **Comparative opinion:** A comparative opinion expresses a preference relation of two or more objects based on some of their shared features. It is usually conveyed using the comparative or superlative form of an adjective or adverb, e.g., “Coke tastes better than Pepsi”. Due to space limitations.

IV. RELATED WORK

In recent years a lot of work has been done in the field of “Sentiment analysis “by number of researchers. In fact work in the field started since the beginning of the century. In its early stage it was intended for binary

classification, which assigns opinions or reviews to bipolar classes such as positive or negative. Paper [3] predicts review by the average semantic orientation of a phrase that contains adjective and adverb thus calculating whether the phrase is positive or negative with the use of unsupervised learning algorithm which classifies it as thumbs up or thumbs down review. Some sentiment analyses are based on review of the user summarization system of the product e.g. [4]. In [4] the product feature uses latent semantic analysis (LSA) based filtering mechanism to identify opinion Paper [5] uses a comparison between positive and negative sentences. It extracts information from the Web and manually label the word set which requires a lot of unnecessary effort. Author in [6] has used a rule-based method, based on Baseline and SVM for sentiment analysis of Chinese document level, which extract the overall document polarity of specific words by a sentiment word dictionary, and adjust it according to the context information. In another work [7], the polarity of the word is being calculated by all the words in the sentence, which can either be positive or negative depending on the related sentence structure. Lakshmi and Edward [8] have proposed to preprocess the data to improve the quality structure of the raw sentence. They have applied LSA technique and cosine similarity for sentiment analysis. Basant Agarwal, et. al. [9] applied phrase pattern method for sentiment classification. It uses part of speech based rules and dependency relation for extracting contextual and syntactic information from the document. In [10] author intended to put forward aspect based opinion polling from unlabeled free form textual customer reviews which do not require customers to answer the questions. M. Karamibekr and A.A. Ghorbani [11] proposed a method based on verbs as an important opinion term for sentiment classification of a document belonging to the social domain. Paper [12] generates a sentiment lexicon called SentiFul which uses and enlarges it through synonyms, antonyms, hyponyms relations, derivation and compounding. They proposed method to distinguish four kinds of affixes on the basis of the role they play for sentiment features namely: propagation, weakening, reversing, and intensifying. These methods assign sentiment polarity which helps in expanding the lexicon to improve the sentiment analysis. A lot of work has also been done where researchers have explored and applied soft-computing approaches, mainly fuzzy logic and neural works for sentiment analysis. [13] And [14] are such examples of works which are based on the fuzzy logic approach. The main contribution of [13] is that it applied fuzzy domain sentiment ontology tree extraction algorithm. This algorithm constructs fuzzy domain sentiment ontology tree based on the reviews that includes extraction of sentiments words, features of the product and relation among features thus precisely predicting the polarity of the reviews. In [14] authors have designed a fuzzy inference system based on membership functions. By designing membership functions they formulated and standardized the process of quantifying the strength of reviewer's opinions in the presence of adverbial modifier. They applied the method for tri-gram patterns of adverbial modifiers.

V. RESEARCH ISSUES

The research in the field started with sentiment and subjectivity classification, which treated the problem as a text classification problem. Sentiment classification classifies whether an opinionated document (e.g., product reviews) or sentence expresses a positive or negative opinion. Subjectivity classification determines whether a sentence is subjective or objective. Many real-life applications, however, require more detailed analysis because the user often wants to know what the opinions have been expressed. For example, from the review of a product, one wants to know what features of the product have been praised and criticized by consumers. Let us use the following review segment on iPhone as an example to introduce the general problem (a number is associated with each sentence for easy reference):

“(1) I bought an iPhone 2 days ago. (2) It was such a nice phone. (3) The touch screen was really cool. (4) The voice quality was clear too. (5) However, my mother was mad with me as I did not tell her before I bought it. (6) She also thought the phone was too expensive, and wanted me to return it to the shop.

The question is: what we want to extract from this review? The first thing that we may notice is that there are several opinions in this review. Sentences (2), (3) and (4) express three positive opinions, while sentences (5) and (6) express negative opinions. Then we also notice that the opinions all have some targets on which they are expressed. The opinion in sentence (2) is on iPhone as a whole, and the opinions in sentences (3) and (4) are on the “touch screen” and “voice quality” features of iPhone respectively. The opinion in sentence (6) is on the price of iPhone, but the opinion/emotion in sentence (5) is on “me”, not iPhone. This is an important point. In an application, the user may be interested in opinions on certain targets, but not on all (e.g., unlikely on “me”). Finally, we may also notice the sources or holders of opinions. The source or holder of the opinions in sentences (2), (3) and (4) is the author of the review (“I”), but in sentences (5) and (6) it is “my mother”. With this example in mind, we can define sentiment analysis or opinion mining.

VI. APPROACHES TO SENTIMENT ANALYSIS

Many companies use opinion mining and sentiment analysis as part of their research. For instance, companies use opinion mining to create and automatically maintain review and opinion aggregation websites. Their systems continuously gather a wide array of information from the Web, such as product reviews, brand perception, and political issues. Other systems might also use opinion mining and sentiment analysis as subcomponent technology to improve customer relationship management and recommendation systems through positive and negative customer feedback. Similarly, opinion mining and sentiment analysis might detect and exclude “flames” (overly heated or antagonistic language) in social communication. Companies use sentiment analysis to develop marketing strategies by assessing and predicting

public attitudes toward their brand. Research and development focuses on designing automatic tools that crawl online Reviews and condense the information gathered. Numerous companies already provide tools that track public viewpoints on a large scale by offering graphical summarizations of trends and opinions in the blogosphere. Developing opinion-tracking systems is commercially important. Also, several tools already exist to help companies extract and analyze information from blogs about large-scale trends in customers.

- *Common Sentiment Analysis Tasks:* - The basic task of opinion mining is polarity classification. Polarity classification occurs when a piece of text stating an opinion on a single issue is classified as one of two opposing sentiments. Reviews such as “thumbs up” versus “thumbs down,” or “like” versus “dislike” are examples of polarity classification. Polarity classifications also identify pro and con expressions in online reviews and help make the product evaluations more credible. Agreement detection is another form of binary sentiment classification. Agreement detection determines whether a pair of text documents should receive the same or different sentiment-related labels. After the system identifies the polarity classification, it might assign degrees of Positivity to the polarity—that is, it might locate the opinion on a continuum between positive and negative. Also, it can classify multimedia resources according to mood and emotional content for purposes such as affective human-machine interaction; troll filtering, and cyber-issue detection. If the text doesn’t contain strong opinions or covers more than one issue or item, new challenges arise, such as subjectivity detection and opinion-target identification. Distinguishing between subjective and objective text helps classify the sentiment. Moreover, a piece of text might have a polarity without necessarily containing an opinion; for example, a news article can be classified into good or bad news without being subjective. Typically, a system performs sentiment analysis over on-topic documents using, for example, the results of a topic-based search engine. However, several studies suggest that managing these two tasks jointly might benefit overall performance. For example, a document’s off-topic passages might contain irrelevant affective information and create inaccurate global sentiment polarity about the main topic. Also, a document might contain information on multiple topics that interest the user. In such instances, it is important to identify topics and separate the opinions associated with each topic.
- *Evolution of Opinion Mining:* Currently, opinion mining and sentiment analysis rely on vector extraction to represent the most salient and important text features. We can use this vector to classify the most relevant features. Two commonly used features are term frequency and presence. Presence is a binary-valued feature vector in which the entries indicate only whether a term occurs (value 1) or doesn’t (value 0). Presence forms a more effective basis to review polarity classification and reveals an interesting difference: although recurrent keywords indicate a topic, repeated terms might not reflect the

overall sentiment. It’s possible to add other term-based features to the features vector. Position refers to how a token’s position in a text unit might affect the text’s sentiment. Further, we might consider presence n-grams—typically bigrams and trigrams—to be useful features. Some methods also rely on the distance between terms. General textual analysis uses part of speech (POS) information (for example, nouns, adjectives, adverbs, and verbs) as a basic form of word-sense disambiguation. Certain adjectives are good indicators of sentiment and guide feature selection to classify the sentiment. Also, selected phrases chosen by pre-specified POS Patterns, usually including an adjective or adverb, help detect sentiments. Some researchers have developed other text mapping techniques that assign labels to predefined categories or real numbers representing the degree of polarity. These approaches are strictly bound by domain and topic. Moreover, most research on sentiment analysis focuses on text written in English and, consequently, most of the resources developed (such as sentiment lexicons and corpora) are in English. Applying this research to other languages is a domain adaptation problem.

VII. CONCLUSION

In this paper, we proposed a set of techniques of machine learning with semantic analysis for classifying the sentence and product reviews based on twitter data. The key aim is to analyze a large amount of reviews by using twitter dataset which are already labeled. The naïve byes technique which gives us a better result than the maximum entropy and SVM is being subjected to unigram model which gives a better result than using it alone. Further the accuracy is again improved when the semantic analysis WordNet is followed up by the above procedure taking it to 89.9% from 88.2%. The training data set can be increased to improve the feature vector related sentence identification process and can also extend WordNet for the summarization of the reviews. It may give better visualization of the content in better manner that will be helpful for the users.

REFERENCES

- [1] J. Xie and B. K. Szymanski. Community detection using a neighborhood strength driven label propagation algorithm. In IEEE Network Science Workshop 2011, pages 188-195, 2011.
- [2] J. Xie and B. K. Szymanski. Towards linear time overlapping community detection in social networks. In PAKDD, pages 25-36, 2012.
- [3] J. Xie, B. K. Szymanski and X. Liu. SLPA: Uncovering Overlapping Communities in Social Networks via A Speaker-listener Interaction Dynamic Process. In Proc. of ICDM 2011 Workshop, 2011.
- [4] S. Gregory. Finding overlapping communities in networks by label propagation. New J. Phys., 12:103018, 2010.
- [5] K. Wakita and T. Tsurumi. Finding community structure in mega-scale social networks. In WWW Conference, pp. 1275-1276, 2007.
- [6] M. E. J. Newman and M. Girvan. Finding and Evaluating Community Structure in Networks. Phys. Rev. E, 69, pp. 026113, 2004.
- [7] V. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre. Fast Unfolding of Communities in Large Networks. J. Stat. Mech., 2008.
- [8] S. White and P. Smyth. A spectral clustering approach to finding communities in graphs. Proc. of SIAM International Conference on Data Mining, pp. 76-84, 2005.

-
- [9] I. Leung, P. Hui, P. Lio, and J. Crowcroft. Towards real-time community detection in large networks. *Phys. Rev. E*, 79:066107, 2009.
- [10] Y.Singh, P.Kumar Bhatia, and Omprakash Sangwan. "A review of studies on machine learning techniques." *International Journal of Computer Science and Security* 1, no. 1, 70-84, 2007
- [11] P.D.Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417-424. Association for Computational Linguistics, 2002.
- [12] L.Ramachandranand Edward F. Gehringer. "Automated assessment of review quality using latent semantic analysis." In *Advanced Learning Technologies (ICALT)*, 2011 11th IEEE International Conference on, pp. 136-138, 2011
- [13] J. Fernando Sanchez-Rada1, Marcos Torres1, Carlos A. Iglesias1, Roberto Maestre2, and Esther Peinado, "A Linked Data Approach to Sentiment and Emotion Analysis of Twitter in the Financial Domain", Retrieved 2014
- [14] Jalaj S. Modha* Prof & Head Gayatri S. Pandi Sandip J. Modha, " Jalaj S. Modha* Prof & Head Gayatri S. Pandi Sandip J. Modha", *International Journal of Advanced Research in Computer Science and Software Engineering*, 2013
- [15] H.Saif, Y. He, and H. Alani. "Alleviating data sparsity for twitter sentiment analysis." *CEUR Workshop Proceedings (CEUR-WS.org)*, 2012.
- [16] H. Saif, Y.He, and H. Alani. "Semantic sentiment analysis of twitter." In *The Semantic Web–ISWC 2012*, pp. 508-524. Springer Berlin Heidelberg, 2012.
- [17] A.Kumar and T. M. Sebastian, "Sentiment Analysis on Twitter", *International Journal of Computer Science Issues*, Vol. 9, Issue 4, No 3, 2012
- [18] P. Bo, L. Lee, and S. Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79-86. Association for Computational Linguistics, 2002.
- [19] A. Go, R. Bhayani, and L. Huang. "Twitter sentiment classification using distant supervision." *CS224N Project Report*, Stanford, pp.1-12, 2009
- [20] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu. "Combining lexicon-based and learning-based methods for Twitter sentiment analysis." *HP Laboratories, Technical Report HPL-2011 89*, 2011
- [21] A. Mudinas, D. Zhang, and M. Levene. "Combining lexicon and learning based approaches for concept-level sentiment analysis." In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, p. 5. ACM, 2012.