Volume 4, Issue 3, March 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Sentiment Analysis for Social Networking Messages Sentiment Detection

I.Hemalatha,
Assistant Professor,
Dept of IT, S.R.K.R. Engineering College,
Bhimavaram, India

Dr. G.P.S.Varma,
Professor, HOD
Dept of IT, S.R.K.R. Engineering College,
Bhimavaram, India

Dr. A.Govardhan
Professor& Director of SIT,
JNTU Hyderabad,
India

Abstract— The main objective of this technology is to prove the usefulness of Analyzing social media data. Twitter is an online social networking and microblogging service that enables users to send and read "tweets", which are text messages limited to 140 characters. The proposed approach combines the preprocessing and machine learning techniques in a system which collects Tweets from social networking sites, perform preprocessing techniques to remove noise in the data and applying machine learning techniques on those Tweets and thus provide some prediction for decision making to improve business intelligence. Results of specific issue analysis will be classified as Positive, Negative and Neutral.

Keywords— Pre-processing, Sentiment analysis, Classification

I. INTRODUCTION

Sentiment analysis has been an important topic for data mining, while the prevailing of social networking, more and more tweet analysis research focuses on social networking. Many people use Twitter as the media for sharing information, driven the wave of using Twitter as a communication tools, which makes sentiment analysis on Twitter become a valuable topic for further discussion. In this paper we introduce a sentiment analysis tool, it comprises three functions: sentiment analysis among Twitter tweets, finding positive, negative and neutral tweets from information resources. This tool focuses on analyzing tweets from those media sites, thus provide a way to find out technology trends in the future.

II. PAGE LAYOUT

A. Social Network Analysis

Social network analysis is a methodology mainly developed by sociologists and researchers in social psychology. Social network analysis views social relationships in terms of network theory, while individual actor being seen as a node and relationship between each node are presented as an edge. Social network analysis has been define in [1] as an assumption of the importance of relationships among interacting units, and the relations defined by linkages among units are a fundamental component of network theories. Social network analysis has emerged as a key technique in modern sociology. Afterwards, there are many scholars expanded the use of systematic social network analysis. Due to the growth of online social networking site, online social networking analysis becomes a hot research topic recently.

B. Twitter

Twitter is an online social network used by millions of people around the world to be connected with their friends, family and colleagues through their computers and mobile phones [3]. The interface allows users to post short messages (up to 140 characters) that can be read by any other Twitter user. Users declare the people they are interested in following, in which case they get notified when that person has posted a new message. A user who is being followed by another user need not necessarily reciprocate by following them back, which renders the links of the network as directed. Twitter is categorized as a micro-blogging service. Micro-blogging is a form of blogging that allows users to send brief text updates or other media such as photographs or audio clips. Among variety of microblogging include Twitter, Plurk, Tumblr, Emote.in, Squeelr, Jaiku, identi.ca, and others, Twitter contains an enormous number of text posts and grows quickly every day. Also, audience on Twitter varies from regular users to celebrities, company representatives, politicians [4], and even country presidents therefore provide a huge base for data mining. We choose Twitter as the source for trend analysis simply because of its popularity and data volume.

The approach contains two major parts. Pre-processing and applying supervised learning algorithm. In pre-processing, we remove the data to increase data consistency then we get higher accurate results. The supervised learning algorithm classifiers are Naive Bayes , Maximum entropy, Support vector machine and Semantic orientation of informal text.

3.1 PREPROCESSING

The pre-processing is necessary because there are some words or expressions in the review don't return any meaning and by the presence of those words we cannot get the correct sentiment analysis. So by doing pre-processing we get higher accurate results. In pre-processing we do Remove URLs, Remove Repeated Letters, Remove Special Symbols and Remove Questions.

3.2 PROPOSED SEMANTIC ORIENTATION ALGORITHM

The algorithm performs the tweets which are Assigned a weight to each word from the SentiWordNet dictionary and Sentence level polarity is calculated as consider the sentences to calculate the average score. check (sent_sentim_word + 3) and (sent_sentim_word - 3) for Modifier from modifier_dict if word found as modifier then calculate overall weight. If there is negation word (Not, Never, N't, Does'nt, Cannt, Nor, Don't, Would'nt, No) near the N, Check (N+3) and (N-3) then reverse its polarity. e.g. (OW=+0.8 → OM= -0.8) Check the modifier word in the sentence, if exists then recalculate the polarity referring the weightage dictionary the same process will be repeated that score of which opinion word will be effected. For e.g, in the sentence "the staff were very nice and cooperative", in this sentence the very is enhance the weight of the nearest opinion word i.e., nice Certain nouns affect the sentence polarity, so recalculate the polarity if such types of word occur. From the dictionary of the weights of words/terms, assign weights to each sentence accordingly.

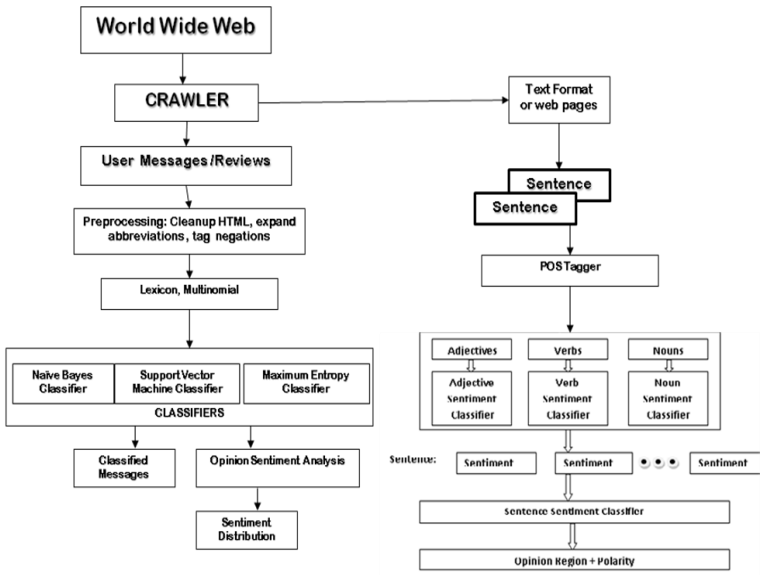

Fig 1: Architecture of Sentiment Analysis Using Machine Learning Algorithms

The steps of rule base system for contextual valance shifter is describes as below:

- if the modifier is a negation modifier then
- sentim_word_score:= Reverse the polarity of sent_sentim_word
- if the modifier is a intensifier then
- sentim_word_score:= sentim_word_score + modifier_weight
- if the modifier is a decelerator or enhancer or context shifter then sentim_word_score:= intensifying modifier_weight obtained from modifier_dict
- Calculate the final weights of each sentence and review to decide if it is positive, negative or neutral. So, the opinion strength for both sentence and feedback is calculated by assigning the combined opinion weight to the sentence and review using the Eq. 3 and4:
- Where, Score (Sen), are the positive or negative score of sentence Sen, Score(i) is the positive, negative score of ith word in sentence S. n is the total no. of words in Sen:
- Where, Rew(Score), are the positive or negative score of Review Rew, Score(Sen) are the positive, negative score of ith sentences in review. n is the total no. of sentences in the review.

$$SentenceScore(Sen) = \frac{\sum_{i=1}^{n} Score(i)}{n}, \quad ReviewScore(Rew) = \frac{\sum_{i=1}^{n} Score(Sen)}{n} \quad \text{....eq(1)}$$

III. EXPERIMENTAL DATA

The entire corpus in .txt format, structured in directories by product and with file names indicating positive and negative. Sentiment analysis is a kind of text classification task. The reviews were downloaded in 2004 from the opinions web site by Jack Grieve. They are divided in the following categories, with 25 positive and 25 negative reviews in each category. The classification into positive and negative was based on the "recommended" or "not recommended" tag that the reviewer provided. The collected corpus tweets are of Phones Data collection.

The results show that books features with positive and negative polarities. Therefore, we could improve the calculation performance is to find out the exact opinion words that were expressed on book features in one sentence. The other way to improve the calculation performance is to develop an algorithm to consider implicit book features, which were not expressed explicitly in the book reviews. To show that the usage of unigram and bi-grams could improve calculation performance, we computed results that used for evaluation of classifier performance Table 1 shows the computed results.

In the field of artificial intelligence, a confusion matrix is a visualization tool typically used in supervised learning (in unsupervised learning it is typically called a matching matrix). Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another).

Confusion Matrix	SVM		NB		ME		SOA	
Class	POS	NEG	pos	neg	pos	neg	pos	neg
Positive	151	31	168	38	171	29	179	31
Negative	37	147	59	101	70	96	39	117

Table 1: Total No of predictions for two classes for data base on Mobile Products

Algorithm	SVM			NB			ME			SOA		
Positive classified Instances	298	0.814	81.421	269	0.735	73.497	267	0.730	72.951	296	0.809	80.874
Negative classified Instances	68	0.186	18.579	97	0.265	26.503	99	0.270	27.049	70	0.191	19.126
Kappa Statistics	0.628			0.453			0.443			0.606		

Table 2: Stratified cross-validation summary

	NPV	TN RATE	TP RATE	FP RATE	FN RATE	Precision	Recall	F-Measure	ACC	F1 score
SVM	0.799	0.826	0.803	0.174	0.197	0.830	0.803	0.816	0.814	0.816
NB	0.631	0.727	0.740	0.273	0.260	0.816	0.740	0.776	0.735	0.776
ME	0.578	0.768	0.710	0.232	0.290	0.855	0.710	0.776	0.730	0.776
SOIT	0.750	0.791	0.821	0.209	0.179	0.852	0.821	0.836	0.809	0.836

Table 3: Detailed Accuracy By Algorithm's

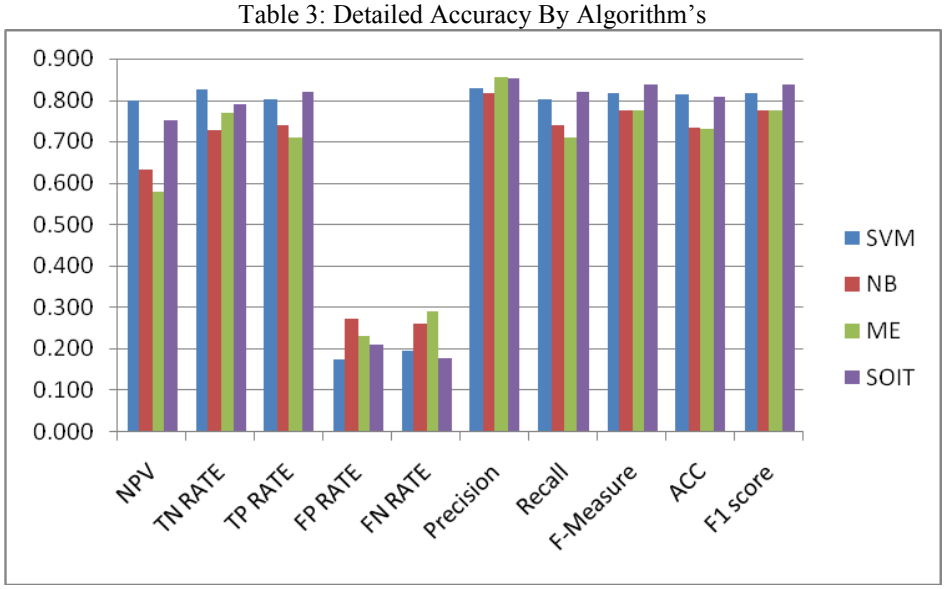

Figure 2: Detailed Accuracy graph by Algorithm's

IV. CONCLUSION

In this research paper, we have proposed a novel method for extracting the user opinions of products. One of the distinctive features of the proposed methodology is to pre-process the tweets while attempting to discover user opinions as well as characterize the classifiers. After the preprocessing phase, the cleaned and refined data is stored in a database meant to be used for the Machine learning process.For this determination, we had to offer a detailed Machine learning method, which is devoted to sentiment analysis. In the supervised learning, the objective was to calculate the sentiment score of product features by aggregating opinion polarities of opinion words around the product features. We considered several techniques that could improve the classification performance. Therefore, choosing the features that are directly related to sentiment analysis is important, because it can improve performance and time and space efficiency.

REFERENCES

1. Hemalatha, I., GP Saradhi Varma, and A. Govardhan. "Sentiment Analysis Tool using Machine Learning Algorithms."
2. Hemalatha, I., A. Govardhan, and G. P. Varma. "Machine Learning Methods in Classification of Text by Sentiment Analysis of Social Networks." *International Journal of Advanced Research in Computer Science* 2.5 (2011).
3. Hemalatha, I., GP Saradhi Varma, and A. Govardhan. "Preprocessing the Informal Text for efficient Sentiment Analysis." *International Journal* (2012).
4. G.P.Saradhi Varma,A.Govardhan, I.Hemalatha. "Sentiment Analysis Tool Using Machine Learning Algorithms." *Elixir International Journal, Elixir Comp. Sci. & Engg.* 58 (2013): 14791-14794.
5. Hemalatha, I., A. Govardhan, and G. P. Varma. "Machine Learning Methods in Classification of Text by Sentiment Analysis of Social Networks." *International Journal of Advanced Research in Computer Science* 2.5 (2011).
6. B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1{135, 2008.
7. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79{86, 2002.
8. Twitter Sentiment Classification using Distant Supervision by Alec Go, Richa Bhayani, and Lei Huang.
9. Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. Association for Computational Linguistics, 2005.

BIOGRAPHY

I.Hemalatha received her M.Tech degree from Andhra University, pursuing Ph.D in computer Science Engineering. A member of CSI, Co-ordinator for Microsoft Student Education Academy, Member in Infosys Campus connect Programme. Working as Assistant Professor in S.R.K.R. Engineering College, China-Amiram, Bhimavaram.

