Sentiment Analysis of Reviews for E-Shopping Websites

Dr. U Ravi Babu[1]

[1] Professor, Dept of Computer Science and Engineering,
Narsimha Reddy Engineering College,
Secunderabad, TS, INDIA
Uppu.ravibabu@gmail.com

Abstract: The *sentiment analysis is one of the popular research area in the field of text mining. Internet has become very popular resource for information gathering. People can share their opinion related to any product, services, events etc over internet. Websites like Amazon, Snapdeal, Homeshop18 etc are popular sites where millions of users exchange their opinions and making it a valuable platform for tracking and analyzing opinion and sentiments. "What other people thing" is being an important piece of information whenever we want to take any decision. Sentiment analysis is the best solution. This gives important information for decision making in various domains. Various sentiment detection methods are available which affect the quality of result. In this paper we are finding the sentiments of people related to the services of E-shopping websites. The main goal is to compare the services of different E-shopping websites and analyzing which one is the best. For this we use five large dataset of five different E-shopping website which contains reviews related to the services. "Sentiwordnet dictionary" is used for finding scores of each word. Then sentiments are classified as negative, positive and neutral. It has been observed that the pre-processing of the data is greatly affecting the quality of detected sentiments. Finally analysis takes place based on classification.*

Keywords: Sentiment analysis, opinion mining; E-shopping websites; classification.

1. Introduction

Sentiment analysis is one of the current research topics in the field of text mining. Opinions and sentiments mining from natural language are very difficult task. Sentiments are extracted from comments, reviews, feedbacks etc. "What other people think" has always been an important piece of information while taking any decision. Now days, before planning to go for movie, everyone what to know its reviews. Sentiment analysis is the best tool for finding whether the review is positive or negative. It helps people to find good quality product. It also helps companies by providing customers feeling related to their product. It also helps to analyze public sentiments related to political issues or political candidates. The main focus of the system is to analyze sentiments for E-shopping company services. The reviews are classified according to positive, negative and neutral score. These results can guide us to select particular site for e-shopping, based on maximum number of positive reviews.

So for detecting and analyzing sentiments, we need the streams of data generated from online sources. The first step is to collect the reviews related to services provided by company. We are using five dataset collected from online sources. We select five popular e-shopping website for our work: Amazon, Flipkart, Home Shop 18, Jabong, and Snapdeal. All these sites are very popular in present from where most of the people like to purchase.

In short, our work can be summarized as follows:

- Firstly we collect the five large E-shopping websites dataset which contains review related to the services of particular websites.
- Then we apply some preprocessing techniques on datasets for removing unwanted things and arranging data in proper manner.
- After that we use POS tagger for assigning tags to each word according to its role.
- Now finding the sentiments for the five large E-shopping website dataset. It uses "sentiwordnet dictionary" for finding score of each words.
- Then sentiments are classified as positive, negative and neutral.
- We analyze how preprocessing techniques and type of input data can affect the quality of topic detection method.

Finally analysis takes place based on classification. The analysis of the services according to positive and negative reviews can be shown in the graphical format.

The rest of the paper is prearranged as follows. Section II presents some previous research work and challenges of sentiment analysis. Section III presents the description of dataset, some pre-processing steps and sentiments analysis techniques which we use in this work. In Section IV, the Experimental results are shown and in section V conclusion is given.

2. Related Work

Shulong Tan et al., had proposed a two LDA based model (FB-LDA and RCB-LDA) for analyzing public sentiments variations and finding the possible reasons causing this variation. Their work mainly focused on tracking sentiments and interpreting sentiment variation. Proposed model also used for finding topic differences between two sets of document. [1] LIU Lizhen et al., had proposed a feature-based vector model and a novel weighting algorithm for sentiment analysis of Chinese product reviews. They classify reviews into positive, negative and neutral comments. They used supervised sentiment classification and a novel feature weighting algorithm [2].

Jalaj S. Modha et al., discussed about exiting approaches methods etc. for performing sentimental analysis on unstructured data available on web. Previously, Sentiment Analysis concentrated for subjective statements or on subjectivity and it just overlooked objective statements which

carry sentiment(s). They proposed a new approach which classifies and handles not only subjective but also objective statements for sentimental analysis. They evaluated their experimental results by using information Retrieval matrices such as precision, recall, f-measure and accuracy [3].

Subhabrata Mukherjee et al., presented a novel approach which identified feature specific expressions of opinion in product reviews with different features and mixed emotions. [4]

M. Thelwall et al., and Y. Tausczik et al., describe the SentiStrength tool which is based on LIWC sentiments lexicon. These two tools are used to assign sentiment labels for each tweet. [5] [6].

Pang *et al.,* presented a detailed survey of the existing methods on sentiment analysis. Sentiment analysis, also known as opinion mining which are widely applied to various document types, such as movie or product reviews. Online public sentiment analysis is gradually more popular topic in social network related research. There has been some research work focusing on assessing the relations between online public sentiment and real-life events. [7]

Pimpalkar, et al., had developed a system that shows the comments and feedbacks/reviews for products. They determined the polarity of sentiments for the products' reviews of the person. Finally they showed the prediction about product. This comparison leads to find the best product. The rule based and fuzzy logic approach was used to give the output. [8]

3. Sentiment analysis From E-Shopping Services reviews

Next, we describe all the components of our related work. In Section III-A, we describe the dataset. Then in Section III-B we explain the data preprocessing techniques. In Section III-C we used different sentiment analysis techniques that take preprocessed data as input and classifying them as positive, negative and neutral comments.

A. Datasets

We describe dataset used in this work. We prepared dataset from online e-shopping websites. We collected reviews related to services of these websites. We selected five popular e-shopping website for our work: Amazon, Flipkart, Home Shop 18, Jabong, and Snapdeal. All these sites are very popular in present from where most of the people like to purchase. We collected near about 2000 reviews for each site. Then we arranged our dataset in required format. Below table show the number of reviews collected for experimental work.

Table 1: Details of the Dataset

Category	Name of sites	No of comments
E-shopping website services related reviews	Amazon	2000
	Flipkart	1900
	Snapdeal	1800
	Home shop 18	1900
	Jabong	2000

B. Data Proccesing

User generated massages are very noisy. They are less formal and generally used non English words and symbols. If topic detection methods are applied on raw comments then it frequently get very poor performance. Therefore for removing noise and unwanted things we use different preprocessing techniques. They are very important for obtaining satisfactory results. Different preprocessing techniques we used are as follows:

Tokenization: Tokenization is the process of extracting bags of cleaner terms from raw comments by deleting stop words and punctuation, compressing redundant character repetitions and deleting IDs or name used in the text for messaging purposes. For removing stop word we maintain stop word dictionary which contain all stop words. We compared each word of comments with this dictionary and the matched word gets removed from comment.

Slang words translation: User generated comments often contains the slang words. Slang word translation means converting the slang words like lol, omg etc, into their standard form. We used the Internet Slang Word Dictionary for this and then add them to the comments.

Stemming: Stemming means a group of different words share the same meaning. It is the process of reducing words which share the same meaning. We used stem word dictionary for grouping all different words of same meaning.

URL removal: Many users include URLs in their tweets. These URLs make the sentiment analysis process more complex. So URLs are removed from the tweets

C. Sentiment Analysis

- **POS Tagging:** "POS (Parts of speech) tagging" means a type to which a word is assigned in according to its syntactic functions. In English language the main "parts of speech" are pronoun, noun, adjective, verb, adverb, etc. "POS tagging" means assigning the labels (tags) to words in sentence according to its function in the sentence. In our work for assigning a label (tag) to each word, we used "Stanford POS (Parts of Speech) tagger". A tag is allocated to each word, like, NNS, NN, JJS, JJ, RB, VB etc.

  - **Sentiments classification:** In this we calculate the polarity of the sentiments and classify them as positive, negative and neutral. We used "SentiWordNet dictionary" for calculating the score of each word. 'SentiWordNet" is a lexical resource in which each word is associated to positive, negative and objective score. So with the help of these score we find the positive, negative and objective score of each word in review comments. We use below simple logic for finding whether the word is positive, negative or neutral.

If Pos_Score of word > Neg_Score of word;
 We consider word is positive;
 Else If Pos_Score of word < Neg_Score of word;
 We consider word is negative;
 Else If Pos_Score == Neg_Score;
 Word is neutral;
 End If;

Then we perform summation of all positive score words and negative score words for knowing the status of whole comment whether it is positive or negative. After performing summation we calculate the final positive and final negative score with the help of below given logic.

Final_Pos_score = SPos / (SPos + SNeg);
Final_Neg_score = SNeg / (SPos + SNeg);

Where SPos is a variable which contains summation of all positive score words and SNeg contains summation of all negative score words

If (Final_Pos_Score – Final_Neg_Score) > 0.1

Whole comment is positive;

ElseIf (Final_Neg_Score – Final_Pos_Score)> 0.1

Whole comment is negative;

ElseIf Comment is neutral;

Finally obtain all positive, negative and neutral comments. The positive, negative and neutral feedbacks will be classified further for analysis.

4. Experimental Results

To detect the sentiments of people related to services of E-shopping websites we generate five large dataset which contains reviews for services. We collected these reviews from online sources. The selected dataset contains the reviews of most popular e-shopping websites: Amazon, Snapdeal, Home Shop 18, Flipkart, and Jabong.

Tables II show the result of preprocessing in which results of preprocessing methods on Amazon and Flipkart dataset are given. In comment field reviews related to Amazon and Flipkart are given. We stem all different words which share same meaning and output is given in After_Stemming field. After that we removed all unwanted things like symbols punctuation, stop words, removing mentions and compressing repeated words from the tweets and the output is given in After_Stoping field.

In Table III the results of POS tagger and SentiWordNet dictionary are shown. A tag is allocated to each word, like, NNS, NN, JJS, JJ, RB, VB, *etc* using POS tagger. To calculate score of an individual word SentiWordNet dictionary is used. Calculated score for each word are shown. Then we add all positive score and all negative score and find the total positive score and total negative score of whole comment.

In Table IV the result of classification of positive, negative and neutral comments are shown. From this we can analyze the views of people related to services of companies. How many are in favour and how many are in oppose related to services provided by particular company.

Table 2: Example of Preprocessing

Name	Comment	After_Stemming	After_Stopping
Amazon	Low prices fast and reliable delivery AND vast selection. Customers can now buy products any time. Realy Good product fair pricing overall amazon is good website to buy branded product on fair prices.	Low prices fast and reliable delivery AND vast selection. Customer can now buy product any time. Realy Good product fair price overall amazon is good website to buy brand product on fair price.	Low price fast reliable delivery vast selection Customer buy product time Realy Good product fair price overall amazon good website buy brand product fair price

| Flipkart | Flipkart is my all time favorite when it comes to online shopping. It has made shopping one of my favorite activities and I can indulge into this for hours and hours. The exchange policy and offers are good. Flipkart is good for the electronics devices. | Flipkart is my all time favorite when it come to online shopping. It has make shopping one of my favorite activitie and I can indulge into this for hours and hour. The exchange policy and offer are good. Flipkart is good for the electronic device. | Flipkart my time favorite come online shopping made shopping favorite activitie indulge hours hour exchange policy offer good Flipkart good electronic device |

Table 3: Example Of Pos Tagging Ans Sentiwordnet Dictionary

CompanyName	Comment	Noun_list	Verb_list	adj_list	pos_cnt	neg_cnt	final_posscr	final_negscr	result_comment
Amazon	One of the worst Service I ever...	movements se...	NULL	bad	0.125	0.625	.35	0.65	Negative
Amazon	But not refunding the value an...	refunding valu...	NULL	NULL	0.625	0	0.35	0.65	Negative
Amazon	Delivery response is so poor	delivery respon...	NULL	NULL	0.125	0	0.35	0.65	Negative
Amazon	I will never purchase any Item ...	purchase item ...	NULL	NULL	0	0	0.35	0.65	Negative
Amazon	Worst Sellers	sellers	NULL	worst	0.25	0.75	.35	0.65	Negative
Amazon	and worst Delivery	delivery	NULL	bad	0	0.625	.35	0.65	Negative
Amazon	Worst inService Support Infor...	inservice supp...	NULL	worst bad	0.625	1.25	.35	0.65	Negative
Amazon	Low prices fast and reliable del...	prices delivery ...	fast	low reliable vast	0.5	0.25	0.8148148	0.1851852	Positive
Amazon	Customers can now buy prod...	customers buy...	NULL	popular	0.125	0	0.8148148	0.1851852	Positive
Amazon	samsung	samsung	NULL	NULL	NULL	NULL	0.8148148	0.1851852	Positive
Amazon	voltas	voltas	NULL	NULL	NULL	NULL	0.8148148	0.1851852	Positive
Amazon	You can purchase any product...	purchase prod...	easily	NULL	0	0	0.8148148	0.1851852	Positive
Amazon	I decided to buy an adjustable ...	buy table hom...	NULL	adjustable	0.625	0	0.8148148	0.1851852	Positive
Amazon	because my room is to small t...	room fix table ...	NULL	small	0	0.375	0.8148148	0.1851852	Positive
Amazon	Realy Good product	product	realy	good	0.75	0	0.8148148	0.1851852	Positive
Amazon	fair pricing	fair pricing	NULL	NULL	0	0	0.8148148	0.1851852	Positive
Amazon	overall amazon is good websit...	overall amazon...	NULL	good	0.75	0	0.8148148	0.1851852	Positive
Amazon	Low prices fast and reliable del...	prices delivery ...	fast	low reliable vast	0.5	0.25	0.7619048	0.2380952	Positive

Table 4: Example of Sentiments Classification

main_com_id	CompanyName	Comment	result_comment
189	Amazon	One of the worst Service I ever hadI Order Items on th May but received 1 Item I followed 4-5 times with Amazon and Seller.But not refunding the value and item as well. Delivery response is so poor.I will never purchase any Item from Amazon. Worst Sellers.and worst Delivery.Worst inService Support Information Depth Content Timeliness DesignUsabilityWebsite Load TimeOne of the worst Service I ever had	Negative
190	Amazon	Low prices fast and reliable delivery AND vast selection.Customers can now buy products from popular brands like sony,samsung,voltas,You can purchase any product easily from any where any time.I decided to buy an adjustable table for my home to put the laptop.because my room is to small to put fix table in my room and I got from amazon online shopping.Realy Good product,fair pricing,overall amazon is good website to buy branded product on fair prices.	Positive
191	Amazon	Low prices fast and reliable delivery AND vast selection.Customers can now buy products from popular brands like sony samsung voltas ect.You can purchase any product easily from any where any time.I decided to buy an adjustable table for my home to put the laptop. because my room is to small to put fix table in my room and I got from amazon online shopping. Realy Good product fair pricing overall amazon is good website to buy branded product on fair prices.	Positive
		In my point of view Amazon is one of the best online shopping site among all available shopping site in India and this is only due to its one of the most trustworthy and its on time delivery.Their main concern is about customer ease and satisfaction.Customer care service is also good as compared to others.They have variety of items and they never	

The figure 1 shows the graphical analysis of the proposed method.
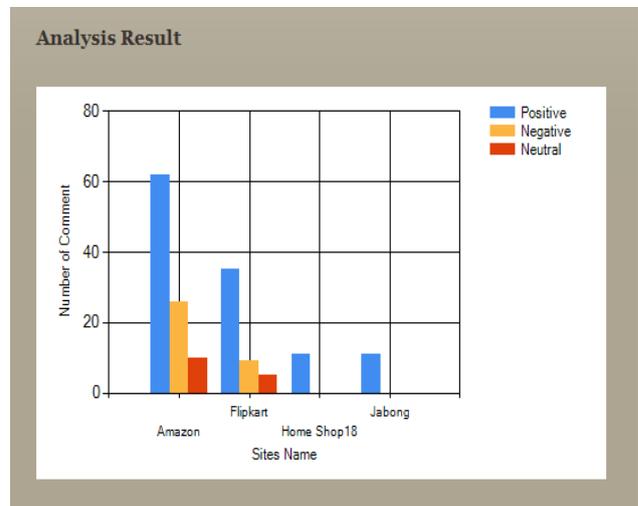
Fig.1. Result of graphical analysis.

5. CONCLUSIONS

In this paper we analyze sentiments of reviews related to the services provided by e-shopping companies. For this we select top five most popular e-shopping sites. We collect reviews from online sources. Sentiment detection from social media streams is a difficult task. We used different preprocessing techniques for removing unwanted things from reviews. It has been observed that the pre-processing of the data and sampling procedure are greatly affecting the quality of detected sentiments. We find the sentiments for the five large e-shopping dataset. It uses "Sentiwordnet dictionary" for finding score of each words. Then sentiments are classified as positive, negative and neutral. This gives the best performance, thus being more reliable. The analysis of the services according to positive, negative and neutral reviews can be represented graphically in experimental results section. These results can guide us to select particular site for e-shopping, based on maximum number of positive reviews.

References

[1] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen and Xiaofei He, Interpreting the Public Sentiment Variations on Twitter, *IEEE Transactions On Knowledge And Data Engineering,* Vol. 6, No. 5, May 2013.

[2] Liu Lizhen, Song Wei, Li Chuchu, Wang Hanshi and Lu Jingli, "A Novel Feature-based Method for Sentiment Analysis of Chinese Product Reviews," *Proceedings of ICT Management, China Communication (ICTM-2014),* pp. 154-164, March 2014.

[3] Jalaj S. Modha, Sandip J. Modha and Gayatri S. Pandi, "Automatic Sentiment Analysis for Unstructured Data," vol. 3 issue 12, pp. 91-97, December 2013.

[4] Subhabrata Mukherjee and Pushpak Bhattacharyya, "Feature Specific Sentiment Analysis for Product Reviews," Dept. of Computer Science and Engineering, IIT Bombay, 2011.

[5] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, Sentiment strength detection in short informal text, *J. Amer. Soc. Inform. Sci. Technol.*, Vol. 61, No. 12, pp. 2544–2558, 2010.

[6] Y. Tausczik and J. Pennebaker, The psychological meaning of words: Liwc and computerized text analysis methods, *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54, 2010.

[7] B. Pang and L. Lee, Opinion mining and sentiment analysis, *Found. Trends Inform. Retrieval*, vol. 2, no. (1–2), pp. 1–135, 2008.

[8] Amit Pimpalkar, Tejashree Wandhe, Minal Kene and Swati Rao, "Review of Online Product using Rule Based and Fuzzy Logic with Smiley's," *International Journal of computing and technology (IJCAT),* vol. 1, issue 1, pp. 39-44, February 2014.