International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Volume 1 Issue 10 (November 2014) www.ijirae.com

A Comparative Study on Sentiment Classification and Ranking on Product Reviews

C.EMELDA
Research Scholar,
PG and Research Department of Computer Science,
Nehru Memorial College, Putthanampatti,
Bharathidasan University, Trichy

Abstract— With the rapid development of e-commerce, most customers express their opinions on various kinds of entities, such as products and services. Reviews generally involves specific product feature along with opinion sentence. These reviews have rich source of information for decision making and sentiment analysis. Sentiment analysis refers to a classification problem where the main focus is to predict the polarity of words and then classify them into positive, negative and neutral feelings with the aim of identifying attitude and opinions. This paper presents a comparison of a sentiment analyzer with classifiers. The sentiments are classified based on the keywords, emotions and SentiWordNet. This paper also proposed review ranking of product reviews based on the features. The results are compared to machine learning tool weka.

Keywords— Review Mining, Information filtering, Sentiment Analysis, SentiWordNet.

I. INTRODUCTION

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features. Most of the reviews are stored either in unstructured or semi-structured format, if the reviews could be processed automatically and presented in a summarized form highlighting the product features and users opinions would be a great help for both customers and manufacturers.

The product reviews are classified using keywords, emotions and SentiWordNet. In a keyword based classification, the sentiments are classified based on the list of positive (super, efficient) and negative keyword (bad, cheating). The emoticons based, classification is done on the basis of emoticons. It uses regular expressions to detect presence of emoticons which are then classified into positive or negative using a rich set of emoticons which are manually tagged as positive (:-)), :'D) or negative (:-(, =( ). SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity.

This paper proposes a review ranking that automatically identifies the important aspects of products from numerous consumer reviews.

The remaining paper is structured as follows. Brief review of the existing sentiments classification systems is represented in section 2. Section 3 presents detail of the proposed system. Section 4 explains Result and discussion of the paper. Finally Section 5 concludes the discussion with possible enhancements to the proposed system.

II. RELATED WORKS

There are numerous techniques used for sentiment analysis. This section describes some of the techniques in sentiment analysis.

The author Liu [1] focuses on two important tasks in opinion mining, i.e., opinion lexicon expansion and target extraction. They propose a propagation approach to extract opinion words and targets iteratively given only a seed opinion lexicon of small size. The extraction is performed using identified relations between opinion words and targets, and also opinion words/targets them. The relations are described syntactically based on the dependency grammar. The author also proposes novel methods for new opinion word polarity assignment and noisy target pruning.

Brody [2] present an unsupervised system for extracting aspects and determining sentiment in review text. The method is simple and flexible with regard to domain and language, and takes into account the influence of aspect on sentiment polarity. They introduce a local topic model, which works at the sentence level and employs a small number of topics that automatically infer the aspects.

An approach to extract product features and to identify the opinions associated with these features from reviews through syntactic information based on dependency analysis is described in [3].

In [4] Tao and Yi is proposed a novel approach to learn from lexical prior knowledge in the form of domain-independent sentiment-Laden terms, in conjunction with domain-dependent unlabeled data and a few labeled documents. This model is based on a constrained non-negative tri-factorization of the term-document matrix which can be implemented using simple update rules.

The process of assessing the helpfulness of the review comments by assessing the reviewer characteristics, add strength to the review analysis process [5]. The review comments could come from chat rooms or online discussion forums. In many scenarios, it is advisable to use an automated consumer review agent for collecting and creating review models [6]. Various researchers use different machine-learning techniques for performing analysis such as classification, clustering, summarization etc.

Mei *et al.* [7] utilized a probabilistic topic model to capture the mixture of aspects and sentiments simultaneously. Su *et al.* [8] designed a mutual reinforcement strategy to simultaneously cluster product aspects and opinion words by iteratively fusing both content and sentiment link information

All the techniques discussed in this section have some advantages and limitations. Hence a comprehensive technique is still needed to overcome their limitations.

III. METHODOLOGY

The purpose of this analysis is to extract, organize, and classify the information contained in the reviews. This section presents the architecture and functional details of our proposed sentiment classification. Figure 1 shows the architecture of our proposed system, which consists of different functional components.

Review Extraction

In this section the online customer reviews are extracted from Web. The http://www.testfreaks.co.in site is used to extract the reviews. We extract the review of different types of digital cameras like Nikon, Sony, Kodak, Canon, FujiFilm etc.

Sample reviews are

- Amazing camera with great quality pictures; Comes with a perfect combo of stand and all other essentials to start your photo shoot sessions
- Good camera and his picture quality is too good. And his range is too good and I hope that is very well in this range
- I am not impressed with this camera.....I have a very hard time getting clear shots that aren't blurry. Not impressed at all.

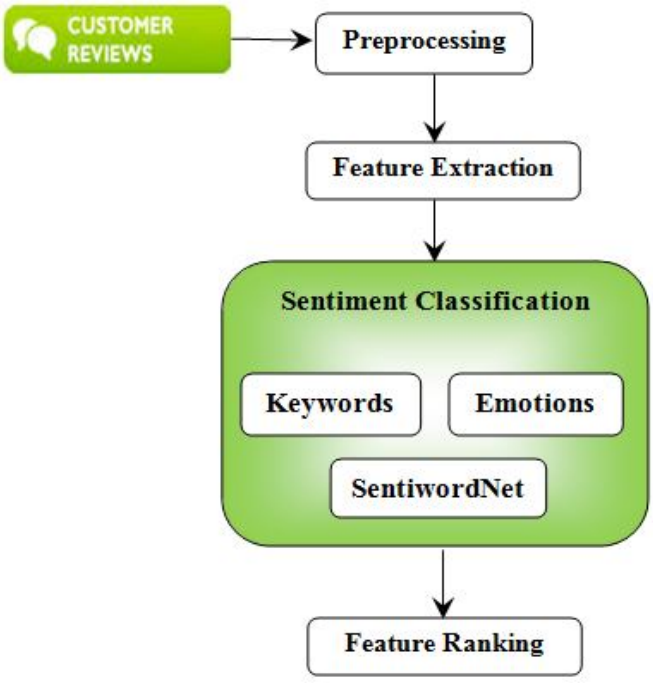

Figure 1 Architecture of Proposed System

Review Preprocessing

In this section the extracted reviews are preprocessed. The following steps are used in preprocessing: Stop word Removal (Remove unwanted words - a, an, the, are, it, was), Stemming Process (impressive- impress, worked-work), POS Tagging (canon/NN, good/JJ, worked/VBD), Feature Extraction (meaningful words).

Keywords based Classification

It uses 'bag of words' approach. Words are domain independent. Each word in the list has been classified as positive/negative. We have to provide words in correct spelling to be classified. Every word has the same weight. There may be a combination of positive/negative words in a review which may result in incorrect classification of review as neutral. Table 1 shows the sample positive and negative keywords. Based on these keyword the review is classified.

Table 1 Sample Positive and Negative Keywords

Positive	Negative
Accurate	**Abnormal**
Gain	**Bad**
Neat	**Fake**
Joy	**Sad**
Valuable	**Upset**
Fast	**Cheat**
Faith	**Zombie**

Emotions Based Classification

This classification is done on the basis of emoticons. It uses regular expressions to detect presence of emoticons which are then classified into positive or negative using a rich set of emoticons which are manually tagged as positive or negative. It uses a list of positive and negative emotions which are actually two text files that include positive and negative emotion symbols respectively.

Table 2 Sample Positive and Negative Emotions

Positive	Negative
:)	**:(**
:-)	**:-(**
=)	**=(**
:D	**:'(**
:-D	**:'s**
B^D	**:[**
xD	**:c**

SentiWordNet

The SentiWordNet is used to classify the reviews. SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, neutrality. each synsets is associated to three numerical scores Pos(s), Neg(s), and Obj(s) which indicate how positive, negative, and "objective" (i.e., neutral) the terms contained in the synset are.

Review Ranking

The reviews are ranked based on the extracted features. Each review has overall ratings. The following formula is used to rank the review

$$O_r + \sum_{i=1}^{n} w_i$$

Here O_r is the overall rating and W_i is the weight for each features in the review. After computing these values sort the rank.

A. EXPERIMENTAL RESULT

This section presents the experimental results on the performance of our proposed techniques. The approaches are implemented using JAVA. All the experiments were run on a Windows 7 with an Intel Pentium(R) CPU P6200 (@2.13GHz) and 2GB RAM.

The customer review of digital camera is extracted from ecommerce site. Table3 shows the review information

Table 3 Summary of Customer Reviews

REVIEW	COUNT
TOTAL REVIEW	5576
POSITIVE	2291
NEGATIVE	315
NEUTRAL	170
UNDEFINED	2800

Table 4 Classification Result of 3 Methods

Method	Pos	Neg	Neu	Undef
Keyword	2550	720	806	1500
Emotion	1456	530	1700	1890
SentiWord	2110	513	203	2750

Figure 4 shows the experimental results of our proposed methods.

Table 5 Weka Classification Result

Alg	Pos	Neg	Both	Undef	Acc
J48	1781	1595	0	2200	53.3
NB	2115	1199	103	2159	96.6
SMO	1987	1254	127	2208	96.6
IBk	1952	1059	288	2277	100
Random Forest	1419	1210	186	2761	96.6
RandomTree	1204	1213	34	3125	100

Figure 5 shows the classification result of weka.

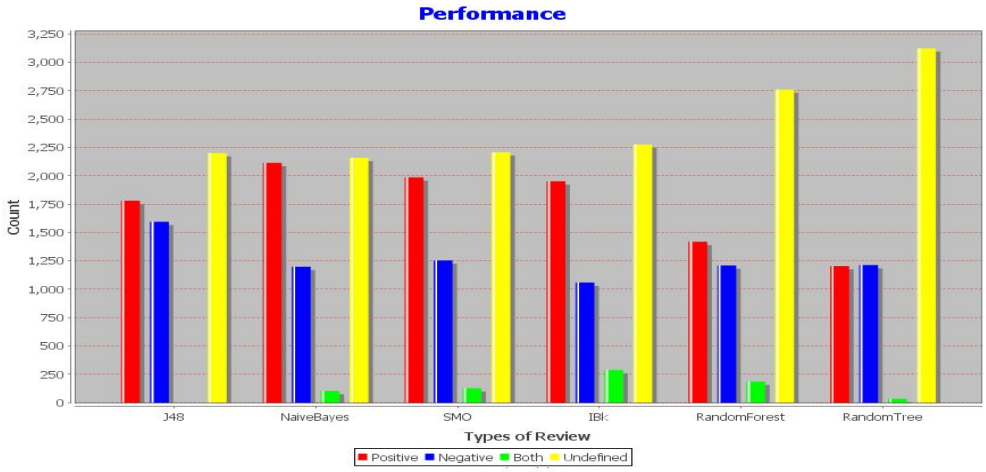

Figure 2 shows the performance of weka result

IV. CONCLUSIONS

In this work, a review analyzer system has been proposed based on performing the sentimental words' analysis for sentiment classification. This paper compares three types of sentiment classification methods. Based on the experimental results the keyword and sentiwordnet gives more accuracy then the emotions based method.

REFERENCES

[1] G. Qiu, B. Liu, J. Bu, C. Chen. (2011). Opinion word expansion and target extraction through double propagation, Computational linguistics, 37(1), 9-27
[2] S. Brody, N. Elhadad, (2010) An unsupervised aspect-sentiment model for online reviews, in: Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics. Publishing, Association for Computational Linguistics, pp. 804-812.
[3] Pimwadee Chaovalit, Lina Zhou, "Extracting Product Features and Opinions from Product Reviews Using Dependency Analysis", Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010, Yantai, Shandong, pp. 2358-2362
[4] L. Tao, Z. Yi, and V. Sindhwani, "A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge," in *Proc. ACL/AFNLP*, Singapore, 2009, pp. 244–252
[5] Aciar, S., Zhang, D., Simoff, S., Debenham J. , "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics", IEEE Transactions on Knowledge and Data Engineering, 2011, Vol. 23, No.10, pp. 1498-1512.
[6] Pollach, I., "Automating user reviews using ontologies: an agent-based approach", Springer Journal on World Wide Web, 2012, Vol. 15, No.3, pp. 285-323.
[7] Q. Mei, X. Ling, M. Wondra, H. Su, and C. X. Zhai, "Topic sentiment mixture: Modeling facets and opinions in weblogs," in *Proc. 16th Int. Conf. WWW*, Banff, AB, Canada, 2007, pp. 171–180
[8] Q. Su *et al*., "Hidden sentiment association in Chinese web opinion mining," in *Proc. 17th Int. Conf. WWW*, Beijing, China, 2008, pp. 959–968

