IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

Sentiment Analysis using Feature Based Opinion Mining Algorithms

T.M. Saravanan[1], Dr. A. Tamilarasi[2], A. Raja[3]

Assistant Professor, Department of Computer Applications, Kongu Engineering College, Perundurai, Erode, India[1, 3]

Professor and Head, Department of Computer Applications, Kongu Engineering College, Perundurai, Erode, India[2]

Abstract: The Explosion of blogs, forums and social networks presents a new set of challenges and opportunities in the way information is searched and retrieved. This huge quantity of information on web platforms put together feasible for exercise as data sources, in applications based on opinion mining and classification. Interest in Opinion Mining has been growing steadily in the last years, mainly because of its great number of applications and the scientific challenge it poses. An effective sentiment analysis process proposes in this research for mining and classifying the opinions. Even though facts still play a very important role when information is sought on a topic, opinions have become increasingly important as well. Opinions expressed in blogs and social networks are playing an important role influencing everything from the products people buy to the product they support. Thus, there is a need for the retrieval of opinions. This paper presents an algorithm which not only analyzes the overall sentiment of a document/review, but also identifies the semantic orientation of specific components of the review that lead to a particular sentiment. The algorithm effectively optimizes the scores of the nouns to extract the potential features. The implementation is carried out on Customer Review Datasets and Additional Review Datasets and also the experimentation results are analyzed.

Keywords: Sentiment Analysis, Opinions.

I. INTRODUCTION

The propagation of blogs and social networks presents a new set of challenges and opportunities in the way information is searched and retrieved. According to comScore, a marketing research company that provides marketing data and services to many of the Internet's largest businesses, out of the 1.1 billion people who actively use the Internet around the globe, 738 million are regular users of social networking sites – about 67% [1]. It further states that if the regular users of other social computing activities such as blogging are added the figure rises to 76%. It is thus clear that there exists vast amount of information in social networking sites such as blogs, review sites, social networking applications, etc.

This information can be leveraged for many purposes, including re-ranking and presenting the results of a search engine. A typical search engine works on the basis of keyword similarity: a user submits a keyword-based query and the search engine returns a list of items that are relevant to this query, as well as user ratings/reviews, if available. An important aspect of this type of search is related to the features of the product, which play a crucial role in the decision making process of the potential customer. It is these features that distinguish one product from other similar products from different brands. Most of the companies focus on a specific feature as their selling point. With the expanse of the e-commerce and the social networking sites, most of the people are using the Internet to check the reviews of products before buying them. They also want to keep themselves updated about any social issue in the neighborhood, in the state, in the country and then across the globe. As the number of reviews has been increasing in a rapid pace, it becomes difficult for the end user to sort the helpful reviews from the ones that do not contain valuable information.

For example, the user may seek to buy a camera, or find a dentist at his/her area. Even though a search engine will return some results, the user needs to first filter out the ones relevant to his/her search, and then iterate through the numerous reviews/ratings for further details, including feedback on the items/services, useful features, etc. This online experience differs significantly from how a similar real-world search would take place. People have always depended on other people's opinion and experiences while buying products or selecting service providers. It is human nature to learn from others' experiences and an ideal search engine should reflect and satisfy this need.

In this work we leverage the information found online in various social networks and use it to create a friendlier search experience for the end user. We focus on analyzing opinions expressed by people in customer reviews, blogs, and social networking applications. We propose an opinion mining and ranking algorithm that first classifies a review as positive, negative or neutral but also identifies the product's more representative features and assigns overall "impression" weights to each one of them. In other words, it also classifies each feature as positive, negative, or neutral in various levels of importance and presents the most important ones to the end user. For example, if the user's search is for an "iPad", the opinions regarding the iPad are retrieved. In addition important features of iPad like “screen size”, “applications”, “touch interface”, etc. are identified and extracted from such reviews. The algorithm summarizes and ranks the opinions about these product features by giving them scores.

II. ASSOCIATED WORK

Many interesting works exist that focus on extracting the opinions from the customer reviews. Some works focus on performing opinion mining to identify the semantic orientation of a review overall, whereas others focus on identifying and extracting the opinion words that will determine the semantic orientation. This line of work further divides into those who focus on the opinion word identification and semantic orientation, and those who also employ features as an additional tool in representing the semantic orientation of a review. Our work is mostly related to the latter category, but we do provide an overview of related work in the other research areas as well.

In [2] the authors analyze and propose the semantic orientation of conjoined adjectives using a log-linear regression model to predict if the two adjectives joined by conjunction are of same or different orientation. In [3] the authors calculate the semantic orientation of words based on their semantic association with pre-determined positive and negative words. The work presented in [4] proposes an unsupervised method to extract syntactic structures that specify the orientation of clauses for domain oriented semantic analysis.

Both [2] and [4] use conjunction rules to extract context-dependent opinion words from large corpora. In [5] the authors propose methods to determine the term subjectivity and term orientation using semi-supervised learning process while in [6] the orientation of the subjective terms is determined by utilizing the term definitions contained in the glossaries and dictionaries. Most of the aforementioned approaches, however, only lay stress on opinion words and do not consider the features. Moreover, they do not incorporate the proposed methodology in a broader opinion mining framework.

A few works exist that perform sentence-level sentiment analysis (i.e. sentiment analysis that is using words but is not extracting representative features) [7–9]. In [7] the opinion words are classified individually and then the polarity of the opinion sentence is calculated by combining the individual opinion word polarity while in [9] the sentiment of each sentence is analyzed by identifying the sentiment expressions and subject terms. Sometimes the opinions regarding the products may not be explicitly mentioned on the customer review sites but they exist in web blogs. Techniques to extract opinions contained in the blogs are proposed in [10]. Finally there exist a few product-ranking techniques based on opinion mining of product reviews for specific languages, such as Chinese [11]. This line of work performs sentence-level sentiment analysis, without focusing on the determination of representative features of the review.

Feature-based opinion summarization [12] allows the customer to drill down the chain of reviews pertaining to a specific feature. Various data mining techniques to summarize the opinions of the existing customers by predicting semantic orientation of the words are proposed in [13]. People often use different words or phrases to describe the same feature. Grouping these features is crucial for effective opinion summary. The words that describe a feature of the product are referred to as “feature expression”. Grouping the feature expression used for a particular feature of a product is addressed in [14] by semi-supervised learning, namely Expectation Maximization (EM) and Naive Bayesian Expectation Maximization. The work of [15] presents multilevel latent semantic association for categorizing the product features. The Opinion Observer, proposed in [16], provides visualizations that can help the potential customer to compare products by a mere glance at these visualizations. In [17] the authors propose a way to automatically mine the product features and the opinions by integrating the semi-structured and unstructured review sources. In this approach the mining results of the semi-structured reviews are treated as prior knowledge and used as a base to mine opinion and product features from the unstructured source using clustering based approach. It then finally integrates product features and opinions to form feature-opinion pairs using the Point wise Mutual Information (PMI) statistics.

Finally, the most similar approach to ours is that of [18], where the semantic orientation of a review sentence is determined by using the a summation as a function of opinion word, set of opinion words containing idioms, distance between feature and semantic orientation of each opinion word. Contrary to our work, however, the algorithm presented in [18] assigns the same weight to all the opinion words and do not follow the same approach as ours in identifying the features (they focus on idioms, whereas we focus on adjectives). Moreover, in our work, we have experimentally shown that assigning more weight to the opinion words expressed in the title of the review yields better results.

III. SYSTEM ARCHITECTURE

Our motivation in this work is to not only mine the opinions but will also extract useful information related to the item’s features and use it to rate them as positive, neutral, or negative. This feature based opinion mining will help the user focus on the features of the opinion/product he/she is interested in. This will help the user spend less time going through reviews that do not add any value in decision-making. In what follows we briefly outline the main components of our system, illustrated in Fig. 1.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

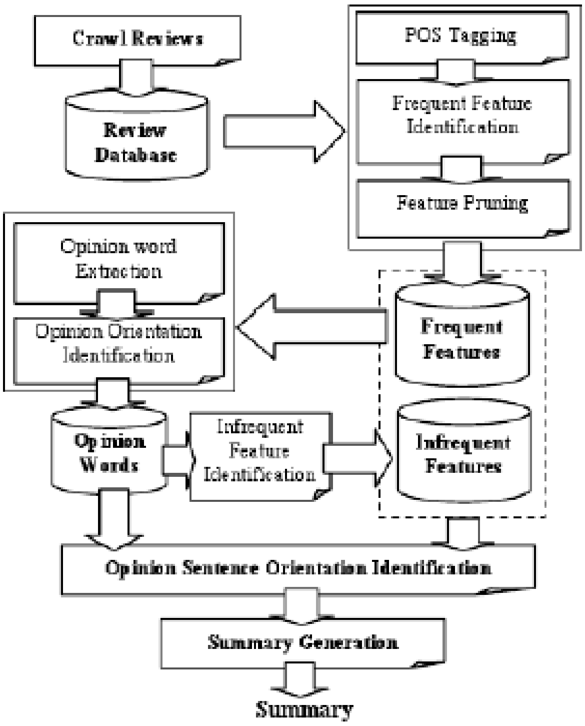

Fig.1 System Architecture

Data Preprocessing: The data from the data set is preprocessed so as to set the data in the format which is acceptable to the data processing algorithms. For example, the tag [t] is inserted at the beginning of the title to indicate that the sentence following [t] is the title of the review. Moreover, the reviews' file which corresponds to a particular product is split into text files containing one review each.

Opinion Mining Engine: The opinion mining engine comprises a POS (parts-of-speech) tagger module [19] and other utility modules used to process the text, such as, identifying the title of the review, and calculating the distance between a noun and its closest adjective.

Opinion Ranking Algorithm: The opinion ranking algorithm ranks the users' opinions based on the scores assigned to the derived features. These scores are used to decide the orientation of the opinion. The details of the ranking algorithm are described later in this paper.

Indexing: The extracted features and opinions are indexed using the indexer module to enable efficient retrieval and presentation from the user interface of the opinion search engine.

Query Engine: The query engine takes the query string as input, and preprocesses it. The preprocessing involves stop words' removal and stemming. We have defined our own list of stop words' and used Porter's algorithm [20] for stemming.

User Interface: The opinion search engine has a web user interface that enables the user to issue a query for the search engine. The user interface has a simple text box for the query input and a search button to submit the query to the backend engine. It displays the results in the form of positive, negative, and neutral opinions. It also displays the specifications or feature ratings of the queried input. Moreover, a search summary is displayed on the left side of the screen that gives the use a quick overview of the search in terms of most important features of the product as mined from the reviews.

IV. OPINION MINING AND RANKING ALGORITHMS

Here we use two algorithms, one for identifying and extracting the features that are deemed as the most important and characteristic of each review, and one that takes as input these features, assigns ranks to them and decides the final classification of the review as positive, neutral, or negative.

4.1. The High Adjective Count Algorithm

The algorithm we propose to identify potential features is called the High Adjective Count (HAC) algorithm. In a nutshell, the main idea behind the algorithm is that the nouns for which reviewers express a lot of opinions are most likely to be the important and distinguishing features than those for which users don't express such opinions. Instead of merely using the term frequency of the keywords, our algorithm starts by identifying the adjectives and nouns in the document collection. The scores of the nouns are initialized to zero. Each adjective is associated with the noun to which it is the closest. This is the noun which the adjective is most likely to describe. For each such adjective, the score of the noun is increased by one. After processing all reviews in the document collection, the algorithm will have assigned scores for each of the nouns. In what follows, we'll refer to these as opinion scores. The opinion scores are used to rank the nouns such that, the higher ranked nouns will be the ones having more adjectives used to describe them. The scores are assigned such that there is one score per noun per item that is reviewed. The ranking can then used to filter the nouns and identify potential features by selecting the nouns which have a score above a particular threshold.

```
HighAdjectiveScores(reviews)
nouns_score_map<- { }
foreach review in reviews do
   Assign part of speech tags to the review
   Apply stemming
foreach line in the review do
foreach adjective in the line find the closest noun
noun_score_map[noun]++
potential_features_map<- { }
foreach noun in  noun_score_map
if noun_score_map[noun] > threshold
potential_features_map[noun] = threshold
return potential_features
```

Fig.2. High Adjective Count Algorithm

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

This threshold is a parameter of the algorithm and can be chosen based on experiments and human evaluation on different review data sets. The pseudocode of the aforementioned algorithm is shown in Fig. 2.

4.2. The Max Opinion Score Algorithm

The second algorithm proposed is the one that ranks the extracted features using the opinion scores assigned in the previous step. This algorithm takes three inputs, each of which is described in the paragraphs below.

This first input is the list of adjectives which are used to express opinions. We refer to those as opinion words. In addition, for each of the adjectives in the list, we need to assign a score which indicates how positive or negative the opinion is. For example, "awesome" indicates a very strong positive opinion whereas "satisfactory" indicates an opinion which, while positive, is not as strong as that of "awesome". We have chosen to manually assign scores in the range [−4, 4] to each of the opinion words. A negative score indicates a negative opinion, whereas a positive score indicates a positive opinion. A higher score indicates a stronger positive opinion than a lower score.

The second input to the algorithm is a list of inversion words. These are words like "not" which invert the sense of the opinion word. When these words occur in the left context of opinion words, they can invert the opinion sense. For example "not good" is a negative opinion. For this reason, when we are assigning scores to opinion words, we also maintain the left context, and if an inversion word appears in the context, we multiply the original score of the opinion word by −1.

The third input to this algorithm is the list of potential features. This can be identified using algorithms like the proposed HAC, or simpler, state-of-the-art ones, such as those that use word count (e.g. TF and TF-IDF).

We process each review sentence by sentence. For each sentence, we look at the opinion words and identify the feature closest to each one in the sentence. The score of the feature is the summation of the scores of the opinion words associated with that feature. The score of the identified features are further summed up to calculate the score of the review. For each feature we compute the average score per opinion word. This score is used to rank the features, based on the intuition that for positive reviews this will identify the features which reviewers like the most, and for negative reviews this will identify the ones which they are most unhappy with. The pseudocode illustrating the Max Opinion Score algorithm is shown in Fig. 3.

Inputs: Reviews
Potential features
Adjectives
Adjective counts
Opinion scores

Outputs: Average scores

Assumptions: R_N →Review Noun scores
R_{NA} →Review Noun Adjective counts
LC →Left Context
LS →Line Scores

Pseudocode

Begin

For each Review
  Initialize $R_N = 0$ *and* $R_{NA} = 0;$
  For each Line in a Review
    Initialize $LC = 0$ *and* $LS = 0;$
    For each Word in Line
      If Word in Adjective scores
        Then Score = Adjective score [Word];
      If Inversion Words in LC
        Then Score = -1 * Score;
      Check for Closest Noun [Word]
      $R_N\,[Closest\ \ Noun\,] += Score\ ;$
      $R_{NA}\,[Closest\ \ Noun\,] ++\ ;$
      $LS\ += Score\ ;$
      Update LC ;
    End
  End
  $Total\ Score = \sum R_N\ ;$
  $Total\ Adjectives = \sum R_{NA}\ ;$
  $Average\ \ Score = \dfrac{Total\ Score}{Total\ Adjectives}\ ;$
End

Stop

Fig.3. Maximum Opinion Score Algorithm

We should note that, for each review we score the features extracted from the title and body separately, as follows:

$$\text{Review Score} = \frac{\alpha \cdot \text{Title Score} + \text{Body Score}}{\alpha + 1} \tag{1}$$

where α is the title weight coefficient, Title Score = $\frac{\sum asc_t}{|a_t|}$, Body Score = $\frac{\sum asc_b}{|a_b|}$, asc_t and asc_b represent the adjective scores in the title and body respectively, and $|a_t|$ and $|a_b|$ represent the number of adjectives in the title and body respectively. The reasoning is that the title is most often a good summary that captures the overall mood of the reviewer and thus should be given a larger weight, as

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

expressed by the title weight coefficient α. Nevertheless, α is a parameter of our algorithm and can be tuned appropriately depending on the data set on which the algorithm is applied [21].

V. EXTENSION FOR OTHER SOCIAL MEDIA

The proposed algorithm can be easily adapted to work with social media other than review sites, which provide us with a very specific structure of title/body of review. In the absence of the title signal the proposed algorithm will continue to function, at the expense of accuracy. However, this signal can be replaced by other indicative signals, depending on the specific application on which it is being applied. For instance, consider a social networking application like Facebook. Even though the title is missing from the users' comments in such sites, we can exploit other signals that clearly indicate whether an opinion is positive or negative, such as the "Like" or the "Share" actions. Similarly a "retweet" is a signal showing a positive opinion on a specific tweet on the micro-blogging application Twitter.5 Such signals can replace the title signal of Title Score in Eq. (1), whereas the weight coefficient α can be appropriately adjusted through experimentation.

VI. EXPERIMENTAL EVALUATION

In this section, the results for our proposed implementation is shown and explained in detail with experimentation results. Initially, we take the dataset Canon G3 of Customer Review Dataset and do preprocessing in the first phase. Followed by the pre-processing phase, our proposed Improved High Adjective Count Algorithm employs on the Noun words, which are considered as the features of opinion mining work.

The sample results of third phase, opinion mining and extraction are tabulated for the whole Canon G3 data. Nouns from IHAC algorithm are considered as features and its relevant opinion words are obtained from the adjectives. Based on these opinion words, we classify the whole review datasets and can identify the particular product is good or bad by the review classification results. The whole datasets are processed by our proposed method and the final sample of review classification. Here, the results of four camera types Canon G3, Canon S100, Nikon coolpix 4300 and Canon Powershot SD500 are given in table 1.

Table 1: Classification results of reviews

Descript-ion	Canon G3	Canon S100	Nikon coolpix 4300	Canon PowerShot SD500
Positive	25.00	49.00	25.00	1.0
Neutral	20.00	2.00	9.00	0.0
Negative	0.00	0.00	0.00	0.0
Result	Positive	Positive	Positive	Positive

6.1. Evaluation Metrics

An evaluation metric is used to evaluate the effectiveness of opinion mining systems and to justify theoretical and practical developments of these systems. It consists of a set of measures that follow a common underlying evaluation methodology. Some of the metrics that we have chosen for our evaluation purpose are Recall, Precision and the F-measure.

In order to employ our proposed method for the effective classification of reviews with mining, we require these evaluation metric values to be computed. The metric values are found based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) with the possibility of classification of reviews. The following table 2 shows how the positive and negative values are described.

Table 2: Description of TP, TN, FP and FN values

Descriptions		OUTPUT: Classified as Positive	OUTPUT: Classified as not Positive
INPUT	Actually Positive	TP	FN
INPUT	Actually not Positive	FP	TN

Using these four basic values, the metrics of Precision, Recall, F-Measure and Accuracy are calculated in our proposed method. The representation of these evaluation metrics are given below in equations.

Precision

The precision estimates how many of the reviews classified to be Positive (Negative or Neutral) are actually Positive (Negative or Neutral) by means of the equation

$$\mathrm{Precision} = \frac{TP}{FP + TP}$$

Recall

The recall indicates how many of the reviews of Positives (Negatives or Neutrals) classes actually are classified. The percentage of Positives (Negatives or Neutrals) correctly classified is represented using recall. It is also equal to Sensitivity.

$$\mathrm{Recall} = \frac{TP}{FN + TP}$$

F-Measure

F-Measure combines precision and recall is the harmonic mean of precision and recall.

$$F - Measure = \frac{2\,(\mathrm{Precision} \times \mathrm{Recall})}{\mathrm{Precision} + \mathrm{Recall}}$$

DOI 10.17148/IJARCCE.2016.51068

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

Accuracy

weighted percentage of Positive, Negative and Neutral reviews that are correctly classified is measured by accuracy.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100$$

6.2. Performance Analysis

The evaluation results for our proposed Improved High Adjective Count based opinion mining work is illustrated in the following table 3 and Fig. 3.

Table 3: Performance evaluation of our proposed work

Product Names	Evaluation Metrics (in %)			
	Precisi-on	Recall	F-Mea sure	Accuracy
Canon G3	94.56	76.35	76.32	93.67
Canon S100	93.26	75.54	85.14	91.73
Nikon CoolPix 4300	95.35	86.32	84.26	94.24
Canon PowerShot SD500	94.63	84.52	84.35	92.56

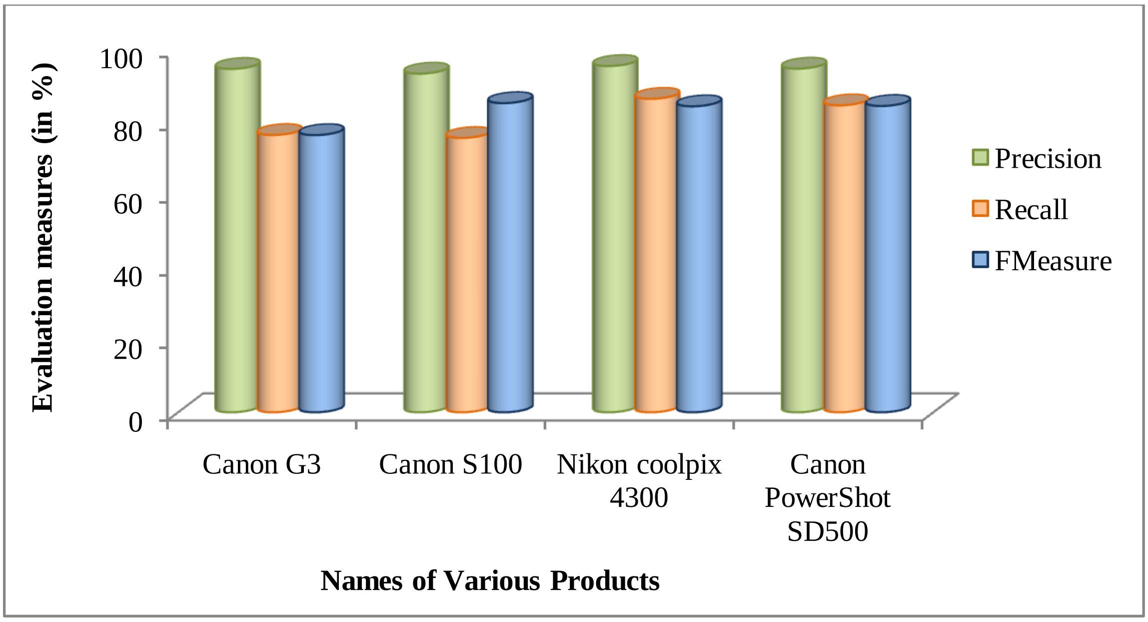

Fig. 3: Evaluation measures precision, recall and f-measure of various products for our proposed work

With the assist of our proposed Improved High Adjective Count method, the assessments for the same product of various companies are brought together in our work. The goods that are reviewed for examining the measure accuracy are Canon G3, Canon S100, Nikon CoolPix 4300 and Canon PowerShot SD500.

Accuracy measure values are also acquired by the description of TP, TN, FP and FN values. Comparing with all these types of camera products, Nikon CoolPIx 4300 model acquires higher accuracy value with 94.24% and the products Canon G3, Canon PowerShot SD 500 and Canon S100 have 93.67%, 92.56% and 91.73% of accuracy, respectively.

The average value for the accuracy is 93.05% of accuracy on average value for our proposed work. Thus, we achieve incredibly excellent accuracy values for the opinion mining and classification of reviews by the results and we can prove that our proposed Improved High Adjective Count algorithm successfully mines the opinions and classifies the online reviews.

VII. CONCLUSION

With the propagation of social networking and e-commerce the information contained in the opinions/reviews expressed by the people has grown by leaps and bounds. In this work we incorporated two novel opinion mining algorithms. The proposed effective opinion mining and classification algorithm was carried out on Customer Review Datasets and Additional Review Datasets. The opinions are based on features and the orientation of these opinions is also largely based on the features rather than a product as a whole.

The proposed framework not only classifies a review as positive or negative, but also extracts the most representative features of each reviewed item, and assigns opinion scores on them. These products were reviewed the various kinds of the product camera, by which, the performance of the proposed work was found. According to the TP, TN, FP and FN values, the reviews were analyzed using the evaluation metrics precision, recall, f-measure and accuracy values. An initial experimental

evaluation on several customer review data sets has shown that our algorithm achieves very high levels of accuracy.
Our plans for future work include experimenting with datasets from other social media, as discussed earlier in this paper. We also plan to further explore the idea of focusing on particular parts in a user's expressed opinion and extract features from there instead of the whole text.

REFERENCES

[1] D. Winder, "The importance of social mobility", Gemalto Review Magazine (2010) 9.
[2] V. Hatzivassiloglou, K.R. McKeown, "Predicting the semantic orientation of adjective"s, Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, ACL'98, 1997, pp. 174–181.
[3] P.D. Turney, M.L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association", ACM Trans. Inf. Syst. 21 (2003) 315–346.
[4] H. Kanayama, T. Nasukawa, "Fully automatic lexicon expansion for domain-oriented sentiment analysis", Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP'06, 2006, pp. 355–363.
[5] B. Pang, L. Lee, "Opinion mining and sentiment analysis", Found. Trends Inf. Retr. 2 (2008) 1–135.
[6] A. Esuli, F. Sebastiani, "Determining the semantic orientation of terms through gloss classification", Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM'05, 2005, pp. 617–624.
[7] S.-M. Kim, E. Hovy, "Determining the sentiment of opinions", Proceedings of the 20th International Conference on Computational Linguistics, COLING'04, 2004.
[8] A. Meena, T.V. Prabhakar, "Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis", Proceedings of the 29th European Conference on IR Research, ECIR'07, 2007.
[9] T. Nasukawa, J. Yi, "Sentiment analysis: Capturing favorability using natural language processing", Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP'03, 2003, pp. 70–77.
[10] W. Zhang, C. Yu, W. Meng, "Opinion retrieval from blogs", Proceedings of the 16th ACM Conference on Conference on Information and Knowledge Management, CIKM'07, 2007, pp. 831–840.
[11] P. Tian, Y. Liu, M. Liu, S. Zhu, "Research of product ranking technology based on opinion mining, Intelligent Computation Technology and Automation", vol. 4, International Conference, 2009, pp. 239–243..
[12] M. Hu, B. Liu, "Mining opinion features in customer reviews", Proceedings of the 19th National Conference on Artificial Intelligence, AAAI'04, 2004, pp. 755–760.
[13] M. Hu, B. Liu, "Mining and summarizing customer reviews", Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'04, 2004, pp. 168–177.
[14] Z. Zhai, B. Liu, H. Xu, P. Jia, "Grouping product features using semi-supervised learning with soft-constraints", Proceedings of the 23rd International Conference on Computational Linguistics, COLING'10, 2010, pp. 1272–1280.
[15] H. Guo, H. Zhu, Z. Guo, X. Zhang, Z. Su, "Address standardization with latent semantic association", Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'09, 2009, pp. 1155–1164.
[16] B. Liu, M. Hu, J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web", Proceedings of the 14th International Conference on World Wide Web, WWW'05, 2005, pp. 342–351.
[17] Q. Miao, Q. Li, R. Dai, "An integration strategy for mining product features and opinions", Proceeding of the 17th ACM Conference on Information and Knowledge Management, CIKM'08, 2008, pp. 1369–1370.
[18] X. Ding, B. Liu, P.S. Yu, "A holistic lexicon-based approach to opinion mining", Proceedings of the International Conference on Web Search and Web Data Mining, WSDM'08, 2008, pp. 231–240.
[19] Stanford log-linear part-of-speech tagger, 2010, http://nlp.stanford.edu/software/tagger.shtml.
[20] M. Porter, "An algorithm for suffix stripping", Program 14 (3) (1980) 130–137.
[21] Magdalini Erinaki, Shamitha Piscal, Jabinder Singh, "Feature Based Opinion Mining", Journal of Computer and System Sciences 78(2012), 1178-1184.