Journal of Web Engineering, Vol. 11, No. 1 (2012) 051-063

A FEATURE–OPINION EXTRACTION APPROACH TO OPINION MINING

BOLANLE A. OJOKOH and OLUMIDE KAYODE

Department of Computer Science, Federal University of Technology, P.M.B. 704, Akure, Nigeria
{bolanleojokoh, imd_kyd}@yahoo.com

Received September 2, 2011
Revised November 24, 2011

With the rapid expansion of the web and e-commerce in recent times, increasingly numerous products are bought and sold on the Web. A lot of product reviews which would be very useful for potential customers to make better decisions are generated by web users. It is highly essential to produce a correct and quick summary of these reviews. In this paper, we propose a method that extracts feature and opinion pairs from online reviews, determines the polarity and strength of these opinions with the aim of summarizing and determining the recommendation status of the customers' reviews. The evaluation results on opinion extraction from the reviews of digital camera demonstrate the effectiveness of the proposed technique.

Key words: Opinions, electronic commerce, text mining, features, digital camera
Communicated by: M. Gaedke & T. Tokuda

1 Introduction

Opinion Mining or Sentiment Analysis is a broad area of natural language processing, computational linguistics and text mining. It is the task of identifying positive and negative opinions, emotions, and evaluations about an object (products, service, article or event) [18]. Textual information in the world can be broadly classified into two main categories, *facts and opinions*. Facts are objective statements about entities and events in the world. Opinions are subjective statements that reflect people's sentiments or perceptions about the entities and events. Many of the existing researches on text information processing have been (almost exclusively) focused on mining and retrieval of factual information, as in information retrieval or web search. Little work has been done on the processing of opinions until only recently [11]. With the rapid growth of the World Wide Web, product reviews generated by users on e-Commerce websites are becoming more and more useful for potential customers to make purchase decision. It has become a common practice for online merchants to ask their customers to share their opinions and hands-on experience on products they have purchased. Such information is highly valuable to manufacturers, online advertisers and potential customers. But it is always stressful and time consuming to read all the reviews because of the huge number. Given a set of evaluative text documents that contain opinions (or sentiments) about an object, opinion mining aims at extracting attributes and components of the object that have been commented on in each document and determining whether the comments are positive, negative or neutral [10].

Opinion Mining has two main research directions, document level opinion mining and feature level opinion mining [8]. Document level mining involves determining the document's polarity by calculating the average semantic orientation of extracted phrases. In feature level opinion mining, reviews are summarized and classified by extracting high frequency feature and opinion keywords. Feature-opinion pairs are identified by using a statistical approach or labeling approach and dependency grammar rules to handle different kinds of sentence structures. Generally, feature level opinion mining has greater precision over document level and uses basically product features in analysis and evaluation.

Online customer reviews have become a cognitive source of information which is very useful for both potential customers and product manufacturers. Customers can now utilize this piece of information to support their decision on whether or not to purchase the product. For product manufacturers, understanding the preferences of customers is highly valuable for product development, marketing and consumer relationship management. In many cases, opinions are hidden in long forum posts and blogs. It is very difficult for human reader to find relevant sources, extract pertinent sentences, read, summarize and organize them into usable forms. An automated opinion mining and summarization system is thus needed. This paper focuses on sentiment analysis at the feature level. It describes the results of an unsupervised learning technique aimed at basically summarizing and classifying customer reviews based on the features of the product on which the customers have expressed their opinion to aid the decision of potential buyers. It will help to generate unbiased judgment based on product features. It can also help to track the shifting attitudes of the general public toward a particular product and simulate human behaviour in suggesting recommendation and decision making. Section two reviews related works. Section three describes the framework of the proposed technique. Section four contains the experiments and evaluation, while section five concludes the work and presents proposed future research work.

2 Related Work

Opinion analysis has been studied by many researchers in recent years. Two main research areas are explored. They are document level opinion mining and feature level opinion mining. Document level opinion mining is concerned with the opinion expressed in the document. Within it, several subtasks can be identified, all of them having to do with tagging a given document according to an expressed opinion. It includes determining document subjectivity, determining document orientation (or polarity), and determining the strength of document orientation [11]. Turney [16] presented an approach for determining document's polarity by calculating the average semantic orientation (SO) of extracted phrases using Pointwise Mutual Information (PMI). Turney and Littman [17] further expanded the work by using cosine distance in latent semantic analysis as the distance measure. Dave et al. [2] describes another technique of document level opinion mining by training a classifier using a corpus of self-tagged reviews available from major websites. It then refines the classifier using this same corpus before applying it to sentences mined from broad web searches. Esuli and Sebastiani [4] also tackled the issue of identifying the orientation of subjective terms contained in text, that is, determining whether a term that carries opinionated content has a positive or a negative connotation. This is believed to be of key importance for identifying the orientation of documents, since it is by considering the combined contribution of these terms that one may hope to solve the problem. At the feature level, some researchers have participated in solving a number of problems. Hu and Liu [6] mined product features that have been commented on by customers, identified opinion sentences, determining their polarity and summarizing the result. They did not consider the strength of the opinions. Ding et al. [3] proposed a lexicon-based approach to opinion mining. Jianshu et al. [7] mined reviews for product comparison and recommendation. Kessler [9] also carried out a related research. Popescu and Etzioni [12] worked on extracting product features and opinions from reviews. Their work focused on product reviews, though their methods apply to a broader range of opinions. They proposed a method called OPINE to achieve their objectives. Their work differs from ours because they used relaxation-labeling in order to find the semantic orientation of potential opinion words. Jin and Ho [8] proposed a lexicalized Hidden Markov Model based learning framework for opinion mining. Their method is different in that it is a supervised learning method. Feng et al. [5] also proposed another related method to solve the problem of opinion mining and applied this for some electronic products from amazon. There are a few differences in the approach used in this work. The tools used in this work are different and our work proposes a threshold value for recommendation of a product, hence a difference in the aggregation of the sentiment values.

3 The Proposed Technique

Figure 1 gives the architectural overview of our proposed system. The several steps involved are described as follows.

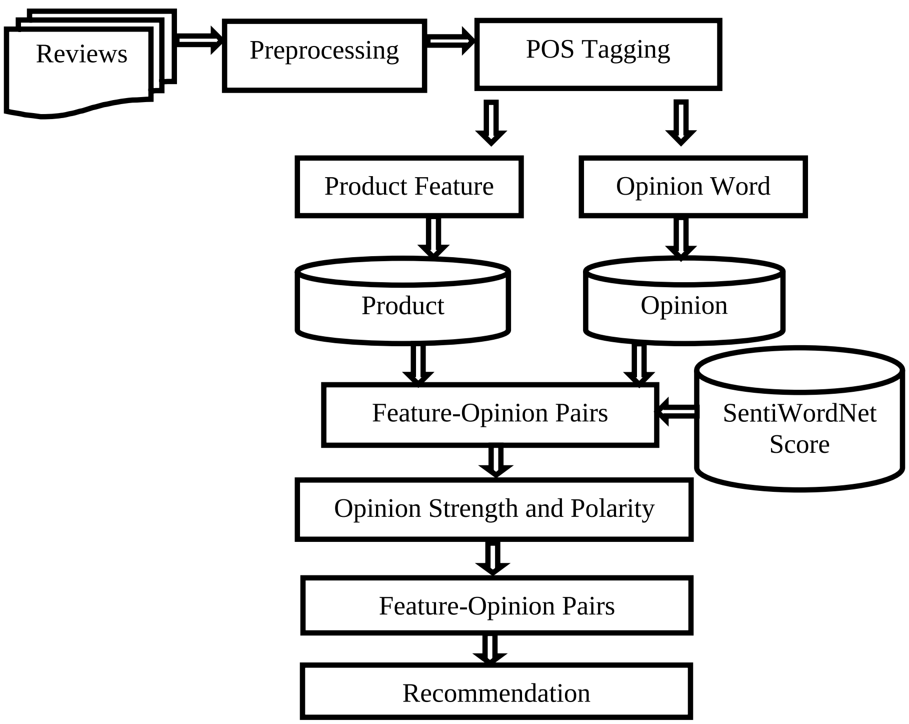


Figure 1 Proposed system architecture

3.1 Part-of-Speech Tagging

Product features are usually nouns or noun phrases and opinion words are usually adjectives in the review sentences. This makes part-of-speech tagging very vital. The OpenNLP Library Part-of-Speech (POS) Tagger was used in this work to produce a part-of-speech for each word in the reviews. A POS Tagger is a piece of software that reads text in some language and assigns parts of speech to each word (such as noun, verb, adjective) of a sentence. As an example, the tagged output for the POS tagger for the sentence "The battery is awesome" is shown below:

The/DT battery/NN is/VBP awesome/JJ

Where the abbreviation DT means determinant, NN means noun, VBP means verb and JJ means adjective. The POS Tagger coded abbreviations are listed in table 1.

3.2 Product Features Mining

Product features often appear frequently in user reviews in the form of nouns or compound nouns. Users often use single nouns (like screen) and two-word compound nouns (like battery life) to represent product features. The method adopted here is therefore, the mining of frequent single nouns and noun bigrams [6].

3.3 Opinion Words Extraction

The next phase after features extraction is the extraction of opinionated words used on the product features in the reviews. These are words that are primarily used to express subjective opinions. The method adopted for extracting the opinion words in this work is built on some established facts [6] about distinguishing sentences used to express subjective opinions from sentences used to objectively describe some factual information. As a result, we used adjectives tagged by the POS tagger as opinion words and limit the opinion words extraction to those sentences that contain one or more product features, as the interest is in customers' opinions on product features. Where there are polysemous words (words with multiple meanings), a threshold of 0.125 was set in the process of selecting the sentiment scores for such words. The relevant opinion words whose positive and negative score for a particular meaning is above this threshold is selected. In a situation where a word still exists more than once with a threshold above 0.125 for diverse meanings, the word sense or meaning that seems most appropriate for quantifying inanimate object or feature qualities is selected, keeping in mind that it will be used to get the sentiment score of opinion words used on the features of a digital camera.

3.4 Feature-Opinion pairs generation

This entails the identification of the product features and the associated opinion words as a pair and its generation for further processing. After the application of POS tagger, each sentence in the review containing a product feature (noun or compound noun) and the corresponding adjective words (the opinion words) are identified. Thus, for each feature in each sentence the nearby adjective is identified as its effective opinion. The typed dependencies obtained, is used to generate the grammatical relationships of a sentence. Here, the generation of feature-opinion pairs is achieved through the application of some rules based on the typed dependencies (as used in Stanford Parser [14]) An illustration is shown below:

"amod" (adjectival modifier) Rule

"This is a great phone"

amod(phone-5, great-4)

In the example above, the feature-opinion pair (phone, great) is extracted.

Table 1 The POS tagger coded abbreviations

Tag	**Tag Description**
CC	Coordinating Conjunction
DT	Determiner
IN	Preposition/subordinate conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, Superlative
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper Noun, singular
NNPS	Proper Noun, plural
PRP	Personal Pronoun
PRPS	Possessive Pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
SYM	Symbol
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, present participle
VBN	Verb, past participle
UH	Interjection
FW	Foreign word

3.5 Opinion strength and Polarity determination

The next phase involves the determination of the polarity and opinion strength of the opinionated words used in the online reviews. In order to indicate both the polarity and strength, a sentiment score ranging from -1.0 to 1.0 is assigned to each feature-opinion pair. The SentiWordNet [13] sentiment score for each opinionated word was used. Given an opinion word, if its positive score is larger than the negative score, its sentiment score is the positive score; thus having positive polarity and the absolute value of the positive score is its strength. Otherwise, if its negative score is greater than the positive score, its sentiment score is the negative score multiplied by -1; thus having negative polarity and the absolute value of the negative score is its strength.

3.6 Feature-Opinion pair aggregation

This involves the separate aggregation of all the positive feature-opinion pairs and the negative feature-opinion pairs in a review. Thus, given a review, r and its positive opinion pair set, pFO, and negative opinion set, nFO, its aggregated positive sentiment score, r.pSS and negative sentiment score, r.nSS are obtained as follows:

$$r.pSS = \sum_{(f_i,o_i)\in pFO} (f_i, o_i).SentimentScore$$

$$r.nSS = \sum_{(f_i,o_i)\in nFO} |(f_i, o_i).SentimentScore|$$

3.7 Opinion Summary generation and Recommendation Status

The final stage involves the generation of the final feature-based review summary. Given the aggregated positive sentiment score (r.pSS) and aggregated negative sentiment score (r.nSS), the procedure involved in opinion summary generation is described below:

1. Begin
2. Double Total;
3. Total = r.pSS + r.nSS;
4. Double positivePercent = (r.pSS / Total) * 100;
5. Double negativePercent = (r.nSS / Total) * 100;
6. If (positivePercent > 69.0) then
7. Status = “Recommended”;
8. Else
9. Status = “Not Recommended”;
10. End

The recommendation status is derived, based on the threshold (70%) set for recommendation. If the positive sentiment score is greater or equal to the threshold, it means the authors recommend the product. Otherwise, it means the authors do not recommend this product. For instance, if the recommendation summary generated is:

Positive: 90%

Negative: 10%

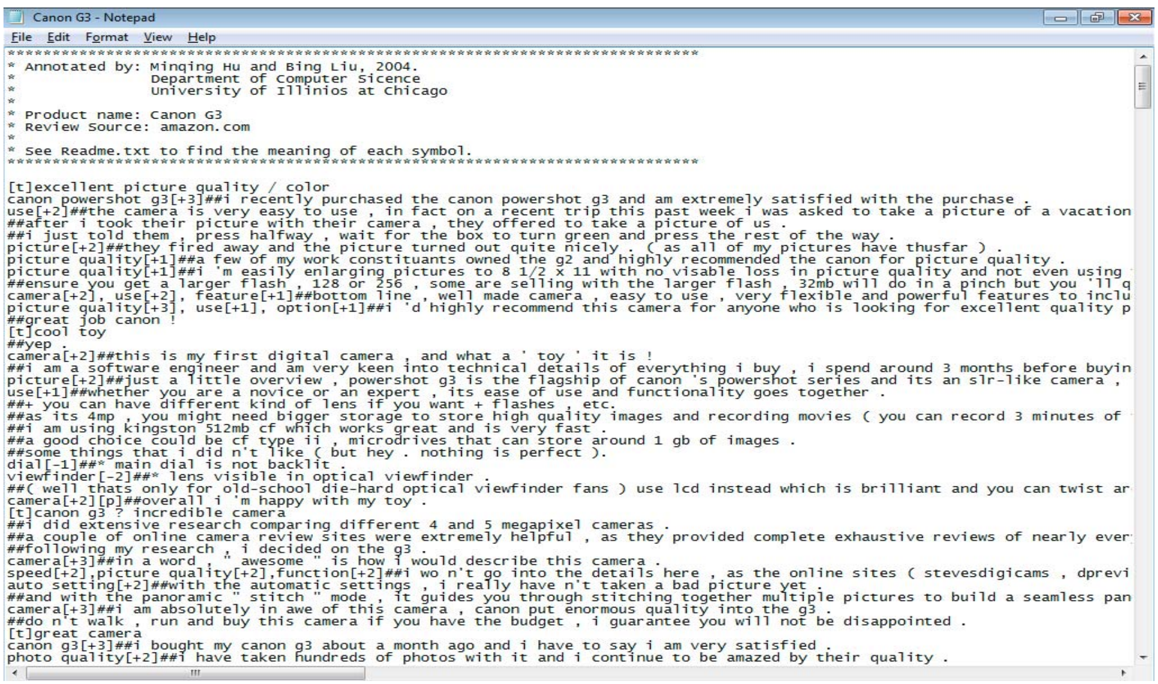
The Recommendation Status is “**Recommended”**.

The threshold of 70% was arrived at after performing experiments with some rated downloaded reviews from amazon. It is reasonable to assume that the 4 star and 5 star reviews should lead to a positive recommendation, while 1 and 2 star should lead to negative recommendation (not recommended). A 3 star review may mean an indecisive recommendation. For all the downloaded rated 4 and 5 star reviews, the least score obtained, after testing with our proposed method was 70.2%, hence the decision to make 70% the threshold for recommendation.

4 Experiments and Evaluation

4.1 Data and tools

The user reviews used in this work were downloaded from the amazon web service [1]. A sample user review for Canon g3 is shown in figure 2 below:


```
Canon G3 - Notepad
File Edit Format View Help
******************************************************************************
* Annotated by: Minqing Hu and Bing Liu, 2004.
*               Department of Computer Sicence
*               University of Illinios at Chicago
*
* Product name: Canon G3
* Review Source: amazon.com
*
* See Readme.txt to find the meaning of each symbol.
******************************************************************************

[t]excellent picture quality / color
canon powershot g3[+3]##i recently purchased the canon powershot g3 and am extremely satisfied with the purchase .
use[+2]##the camera is very easy to use , in fact on a recent trip this past week i was asked to take a picture of a vacation
##after i took their picture with their camera , they offered to take a picture of us .
##i just told them , press halfway , wait for the box to turn green and press the rest of the way .
picture[+2]##they fired away and the picture turned out quite nicely . ( as all of my pictures have thusfar ) .
picture quality[+1]##a few of my work constituants owned the g2 and highly recommended the canon for picture quality .
picture quality[+1]##i 'm easily enlarging pictures to 8 1/2 x 11 with no visable loss in picture quality and not even using
##ensure you get a larger flash , 128 or 256 , some are selling with the larger flash , 32mb will do in a pinch but you 'll q
camera[+2], use[+2], feature[+1]##bottom line , well made camera , easy to use , very flexible and powerful features to inclu
picture quality[+3], use[+1], option[+1]##i 'd highly recommend this camera for anyone who is looking for excellent quality p
##great job canon !
[t]cool toy
##yep .
camera[+2]##this is my first digital camera , and what a ' toy ' it is !
##i am a software engineer and am very keen into technical details of everything i buy , i spend around 3 months before buyin
picture[+2]##just a little overview , powershot g3 is the flagship of canon 's powershot series and its an slr-like camera ,
use[+1]##whether you are a novice or an expert , its ease of use and functionality goes together .
##+ you can have different kind of lens if you want + flashes , etc.
##as its 4mp , you might need bigger storage to store high quality images and recording movies ( you can record 3 minutes of
##i am using kingston 512mb cf which works great and is very fast .
##a good choice could be cf type ii , microdrives that can store around 1 gb of images .
##some things that i did n't like ( but hey . nothing is perfect ).
dial[-1]##* main dial is not backlit .
viewfinder[-2]##* lens visible in optical viewfinder .
##( well thats only for old-school die-hard optical viewfinder fans ) use lcd instead which is brilliant and you can twist ar
camera[+2][p]##overall i 'm happy with my toy .
[t]canon g3 ? incredible camera
##i did extensive research comparing different 4 and 5 megapixel cameras .
##a couple of online camera review sites were extremely helpful , as they provided complete exhaustive reviews of nearly ever
##following my research , i decided on the g3 .
camera[+3]##in a word , " awesome " is how i would describe this camera .
speed[+2],picture quality[+2],function[+2]##i wo n't go into the details here , as the online sites ( stevesdigicams , dprevi
auto setting[+2]##with the automatic settings , i really have n't taken a bad picture yet .
##and with the panoramic " stitch " mode , it guides you through stitching together multiple pictures to build a seamless pan
camera[+3]##i am absolutely in awe of this camera , canon put enormous quality into the g3 .
##do n't walk , run and buy this camera if you have the budget , i guarantee you will not be disappointed .
[t]great camera
canon g3[+3]##i bought my canon g3 about a month ago and i have to say i am very satisfied .
photo quality[+2]##i have taken hundreds of photos with it and i continue to be amazed by their quality .
```

Figure 2 A sample of digital camera online reviews

The SharpNLP (http://sharpnlp.codeplex.com/) used in this work is a C# port of the Java OpenNLP tools, plus additional code to facilitate natural language processing. It is a collection of natural language processing tools written in C# which provides the following NLP tools: a sentence splitter, a tokenizer, a part-of-speech tagger, a chunker (used to "find non-recursive syntactic annotations such as noun phrase chunks"), a parser, and a name finder. To use this suite, the runtime binaries (SharpNLP 1.0.2529) and

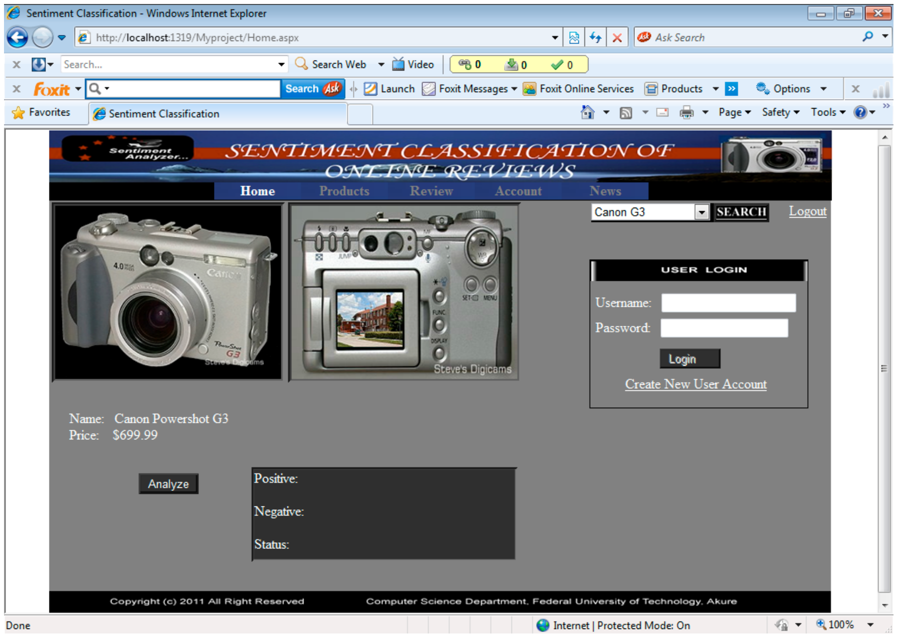


Figure 3 Home page with Canon G3 selected

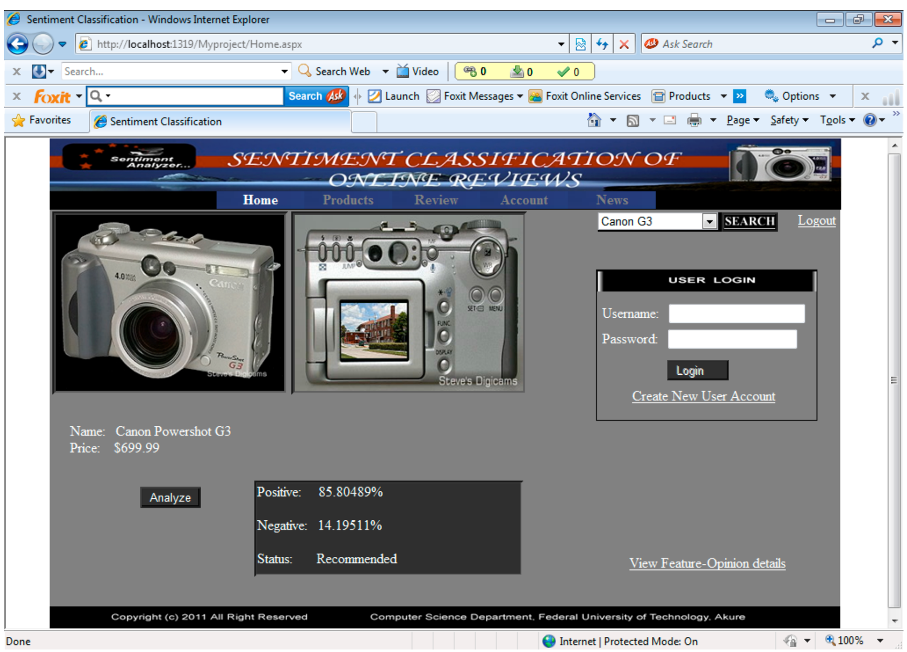


Figure 4 Sentiment classification of Canon G3 reviews

the model files needed was downloaded from [15]. Other tools employed during the implementation phase include Microsoft SQL Server 2005 (Standard Edition) – Database Management System, Microsoft Visual Studio 2008 Integrated Development Environment, SharpNLP Library (C# port of the OpenNLP Library) including the EnglishPOS.bin and EnglishTok.bin binaries model files, and SentiWordNet [13] – a lexical resource of opinion terms in text document. SentiWordNet is a dictionary of opinionated terms containing opinion information on terms extracted from the WordNet database and made publicly available for research purposes. It is built via a semi supervised method and provides a valuable resource for performing opinion mining tasks: it provides a readily available database of sentiment information term for the English language, and could be used as a replacement to the process of manually deriving ad-hoc opinion lexicons. It is an opinion lexicon where each term is associated with numerical scores indicating positive and negative sentiment information.

4.2 Experiments

Experiments were carried out with online reviews of four different digital cameras (Canon PowerShot G3, Canon PowerShot SD500, Canon PowerShot S100 and Nikon Coolpix 4300). The home page is shown in figure 3.

Figure 4 shows the sentiment classification of canon g3 digital camera. Clicking on View Feature-Opinion details redirect the user to the Feature-Opinion pair page in figure 5.

Figure 5 Canon G3 feature-opinion details page

4.3 Evaluation

In the evaluation of the product features extraction, the features identified in the data set were manually tagged and compared with those analyzed by the system. For each preprocessed file, the product-feature pair and score generated by the system was uploaded from the database into a text file. Encarta Dictionary was used to manually label the adjectives in each sentence of the reviews. A comparison of the two files was carried out to generate true positive and negative, and false positive and negative values as defined below in order to generate precision, recall, accuracy and f-score measures of the system.

The precision (*p*), recall (*r*), accuracy (*a*) and F-score (*F*) were calculated as follows:

$$p = \frac{A}{A+C}, \quad r = \frac{A}{A+B}$$

$$F = \frac{2*p*r}{p+r}, \quad a = \frac{A+D}{A+B+C+D}$$

where A is the number of correctly extracted features (true positives), B is the number of features in the reviews but not extracted as features (false negatives), C is the number of features extracted but associated with wrong label (false positives), while D is the number of features in labels that did not exist and were extracted (true negatives).

The results of this evaluation are shown in Table 2.

Table 2 The evaluation result of product features extraction by the system

Product Name	Number of Product Features	Precision	Recall	F-score	Accuracy
Canon g3	159	0.914	0.914	0.914	0.849
Canon sd500	152	0.898	0.870	0.884	0.803
Canon s100	127	0.953	0.904	0.928	0.874
Nikon 4300	140	0.930	0.883	0.906	0.843
Average	145	0.924	0.893	0.908	0.842

Table 3 The evaluation result of opinions extraction

Product Name	Number of opinions	Precision	Recall	F-score	Accuracy
Canon g3	203	0.963	0.940	0.951	0.909
Canon sd500	218	0.930	0.975	0.952	0.910
Canon s100	161	0.924	0.980	0.951	0.910
Nikon 4300	192	0.944	0.955	0.950	0.907
Average	194	0.940	0.963	0.951	0.909

Table 3 shows the results of opinion words extraction, the opinion words were tagged manually and compared with those mined by the system as in the case with features extraction. This was done using the standard evaluation measures of precision (*p*), recall (*r*), F-score (*F*) and accuracy (*a*).

The feature-opinion pair generation of the system at the sentence-level was also evaluated. Table 4 shows these results.

Table 4 The evaluation result of opinion sentence extraction by the system

Product Name	Number of Opinion sentences	Precision	Recall	F-score	Accuracy
Canon g3	170	0.914	0.848	0.880	0.794
Canon sd500	163	0.898	0.803	0.848	0.748
Canon s100	133	0.953	0.858	0.903	0.835
Nikon 4300	147	0.930	0.835	0.880	0.803
Average	153	0.924	0.836	0.878	0.795

Although there is a difference in the products on which our method was tested and many of the existing methods for opinion mining, the method can still be compared relatively with some related systems. It performs averagely better in recall (0.89) than OPINE (0.77) in features extraction and better in both Precision (0.92) and Recall (0.84) for Opinion sentence extraction as compared to 0.79 and 0.76 respectively reported in [12] for Opinion phrase extraction.

Table 5 Comparison of opinion sentence extraction

Product Name	Method	Precision	Recall	F-score
Digital Camera	FOE	0.924	0.836	0.878
Digital Camera	Lexicalized HMM	0.859	0.855	0.857

Table 6 Results of recommendation

Digital Camera	Accuracy
5 star reviews	1.00
4 star reviews	1.00
2 star reviews	0.75
1 star reviews	0.80

According to the report of [5] of the evaluation of product features extraction for their Opinion Based Classifier (OBC) System, our proposed system performs relatively better (0.93, 0.89 and 0.91 precision, recall and f-score respectively) on the average than OBC when compared with theirs on digital cameras, which is the same sample product' opinions on which our system is tested. OBC reported 0.80, 0.55 and 0.63 precision, recall and F-score respectively on features extraction for digital cameras and average of 0.81, 0.52 and 0.63 for all the products.

Jin and Ho [8] reported the results of opinion sentence extraction for three categories of digital cameras (related to this work) using Lexicalized HMM. Table 5 shows the results of a comparative analysis of their reported evaluation results with our Feature-Opinion Extraction (FOE) method.

25 rated reviews were downloaded from amazon to test the accuracy of the recommendation decision made by the system. The result of any 4 or 5 star review that receives a recommendation decision is termed correct, while any 1 or 2 star review that receives a recommendation decision is termed incorrect. Table 6 shows these results.

5 Conclusion and Future Work

In this paper, we proposed a set of techniques for mining and classifying opinions expressed on the features of products as recommended or not recommended. The experimental results indicate that the proposed techniques are effective at performing this task. In the future, we intend to extend the work to improving on the methods adopted in this work to handle more complex sentences and also include product comparison.

References

1. Amazon Web Service (http://www.aws.amazon.com)
2. Dave, K., Lawrence, S. , and Pennock, D.(2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In Proceedings of the 12th International Conference on World Wide Web, WWW'03, pp. 519-528.
3. Ding, X., Liu, B., and Yu, P.S. (2008). A Holistic Lexicon-Based Approach to Opinion Mining. In Proceedings of WSDM '08, California, USA, pp. 231-239.
4. Esuli A. and Sebastiani E. (2006). SentiWordNet: A publicly available lexical resource for opinion mining, In Proceedings of Language Resources and Evaluation (LREC)
5. Feng, S., Zhang, M. and Zhang, Y. (2010). Recommended or Not Recommended? Review Classification through Opinion Extraction. In Proceedings of APWEB Conference.
6. Hu, M and Liu, B (2004). Mining and summarizing customer reviews, In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
7. Jianshu, S. Long, C., Zhu, X., and Minlie, H. (2009). Mining Reviews for Product Comparison and Recommendation. Available at http:///www.gelbukh.com/polibits/2009_39/39_05.pdf
8. Jin, W. and Ho, H.H. (2009). A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining. In Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada,
9. Kessler, W. (2011). Introduction to Sentiment Analysis. Available at http://www.ims.uni-stuttgart.de/~kesslewd/lehre. Accessed 20th July, 2011.

10. Liu, B. (2010). Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing.
11. Liu, B., Hu, M., and Cheng, J. (2005). Opinion Observer: Analyzing and comparing opinions on the Web. In Proceedings of WWW-05.
12. Popescu M. and Etzioni O. (2005). Extracting product features and opinions from reviews. In Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP '05, pp. 339-346.
13. SentiWordNet (http://sentiwordnet.isti.cnr.it/)
14. Stanford Parser (http://nlp.stanford.edu/software/lex-parser.shtml)
15. The SharpNLP (http://sharpnlp.codeplex.com/)
16. Turney, P. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the Conference of the Association of Computational Lingustics, ACL'02, pp. 417-424.
17. Turney, P.D. and Littman, M.L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS). 21(4):315-346.
18. Wilson, T., Wiebe, J., and Hoffman, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pp. 347-354.