AN IMPROVED FRAMEWORK FOR ASPECT-BASED OPINION EXTRACTION FROM CUSTOMER REVIEWS

M. Lovelin Ponn Felciah, R.Anbuselvi, Ph.D.,
Asst. Prof in Dept of Computer Applications
Bishop Heber College (Autonomous)
Tiruchirappalli, India

ABSTRACT

Text is the key method of communicating information in the digital age. Blogs, messages, articles, reviews and opinionated information overflows on the Internet. Many people purchase products online and post their opinions about purchased items. The posted reviews helps others customers to make better decision to purchase the product .Extracting the needful data and summarize it as a useful information still enhance the decision making process ease and efficiently. The work focus on the accuracy of extraction by combining different techniques from three major areas, namely Data Mining, Natural Language Processing techniques and Ontologies. The enhanced framework sequentially mines product's aspects from users' opinions, groups with similar aspects, and generates useful information. This paper focuses on the task of extracting product aspects from reviews by extracting all possible aspects and finds the polarities for each aspect from reviews using natural language, ontology, and frequent "tag" sets. The proposed framework, when compared with an existing baseline model, produced better results.

Keywords *Data Mining, Opinion Mining, Sentiment Analysis, Aspect Extraction, Customer Reviews.*

1. INTRODUCTION

Now a day, the Internet contains huge amounts of textual information on people's expressed opinions, making the Internet an excellent source from which to gather data about a specific object within a specific domain [1]. The feedback posted by customers' has prompted the urgent need for systems that can automatically summarize documents [4]. Searches for information about items available for purchase return huge quantities of information, making it difficult to find useful data easily. Useful online information needs to be presented in a summarized form that includes the relevant data in easy-to-read and easy-to-understand format [2].

Discussion groups, reviews, forums, and blogs available on the Internet contain opinions and information. If mined and summarized, those opinions could provide useful data for decision makers. The process of summarizing opinions be dependent on primarily on identifying and extracting dynamic opinionated information from text. Efficiency of the process and quality of the resulting summary depends on the extraction of key information and exclusion of redundant details [7]. Both individuals and businesses seek opinion summaries to enhance their decisionmaking processes. The customer feedback about purchased items can be independent and accurate [9], [10]. One customer's opinions may not fully represent the opinions of all customers, underscoring the importance of collecting and analyzing opinions from many different opinion holders to evaluate the object under study [3]. The need to understand customers' subjective feedback has made opinion extraction and summarization a hot subject in recent years. In sentiment mining the opinions are extracted, analyzed, summarized, and then presented along with the corresponding opinionated information [5].

Many researchers have studied different types of extraction and summarization, as well as methods to create and evaluate the final summary [8]. This proposed work reviews recent work and covers some techniques on extracting and summarizing opinions. The work focus to achieve by improving the accuracy of the aspect-based opinion summarization model.To enhances the quality of opinion summarization from customers' reviews. The flow of this model is to identify the product entities like components, functions, opinions etc., and associate them to the corresponding opinions along with its subjective sentence. .

2. PROPOSED FRAMEWORK

The enhanced framework was designed to summarize customer reviews and produce "aspect - based opinion summary". To produce a result, some essential information must be extracted. The framework is divided into four major tasks to use text files containing customer reviews as input and then perform the four tasks to produce the final output. Step one involve to mine entities of the product under study and identify the associated opinion orientation of each aspect. The second task is to group aspects based on similarities. The third task is to select the most popular aspects from the reviewed sentences. The last step is to generate an opinionated summary that is based on product aspect as a proposed framework shown in Figure 1.

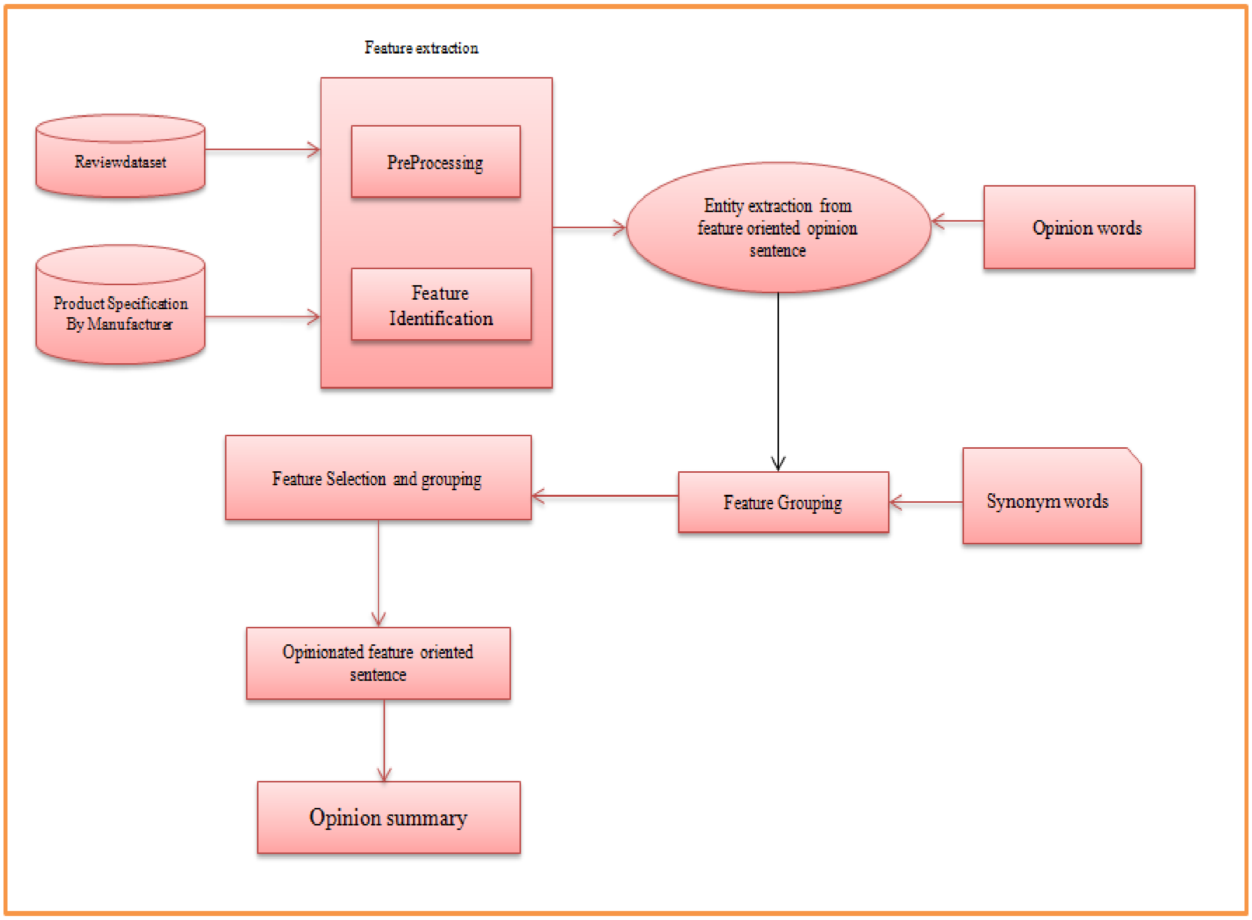

Feature extraction
Reviewdataset
PreProcessing
Product Specification By Manufacturer
Feature Identification
Entity extraction from feature oriented opinion sentence
Opinion words
Feature Selection and grouping
Feature Grouping
Synonym words
Opinionated feature oriented sentence
Opinion summary

Figure 1. Proposed Framework

2.1. Entity Extraction

The first task of the proposed framework is "entity extraction". Entities include aspects/features, components, parts, functions, and opinions of the object being studied. For our work, entity extraction is handled as two extractions: product aspects extraction and opinions extraction. Furthermore, the extraction of aspects is decomposed into two-step process.

2.2. Aspect Grouping

Once entities have been extracted, they are grouped by based on synonyms[9]. Customer may express their opinions about the same aspect using different words and/or phrases. To produce a useful summary, those different words about the same aspect must be grouped. Those words and/or phrases are domain synonym they share the same meaning and so must group them under the same aspect group. In a mobile phone domain, for instance, “capacity” and “memory” are two different expressions referring to the same aspect.

In this paper, aspect grouping is complex due to the numerous possible synonyms. The level of sufficiency is low for two reasons. First, although words may refer to the same aspects, some dictionaries do not consider words to be synonyms. Second, many synonyms are domain synonyms; they are likely to refer to the same aspect in one domain but not in another. We aim to achieve aspect grouping using natural language possessing techniques, shared words and lexicon similarity. Some aspects may share words e.g., (“battery,” “battery life,” “battery usage,” and “battery power”), all of which refer to the same aspect “battery”.

2.3. Aspect Selection

Hence after the aspects have been grouped, the most representative aspect sentences must be selected to form the final opinionated summary. This step can be accomplished by analyzing the strength of each opinionated sentence and then select sentences with the highest weight. The strength of all “adjectives, adverbs and verbs “, within the sentence, will determine the total weight of that sentence. Sentence importance is one of the most critical determinations of this proposed framework. In this paper, we calculate the weights for all “adjectives, adverbs and verbs “for each the sentence. The calculation is done by adding up all weights for each “adjectives, adverbs and verbs “within the sentence, as presented in Table 2. For example, “ear piece is very comfortable”, the sentence has an “adjective = comfortable” and “adverb = very”, therefore, the earned weight for this sentence is “2”. The weights are calculated based on the a method to score a combination of tags (adjective, verb, adverbs) to give weight to each aspect sentence, as indicated in Table 2 for adjectives and adverbs and Table 3 for verbs based on the approach proposed.

Table 2. Adjective and adverb weights

Tags	**Description**	**Weight**
JJ	Adjective	1
JJR	Comparative Adjective	2
JJS	Superlative Adjective	3
RB	Adverb	1
RBR	Comparative Adverb	2
RBS	Superlative Adverb	3

On other hand, verbs are treated differently from adjectives and adverbs. If a sentence contains a verb from positive categories then “+1” will be added to the weight and if the verb is from negative categories then “-1” will be subscribed form the total weight. Based on final sentence’s weights, the selection can be easily made. We will select sentences with the highest weight to be candidatures for the final summary.

2.3. Summary Generation

The final task of the process is summary generation is. It is based on outcomes of the preceding tasks in which the extracted aspects and its corresponding opinion are selected and then weights are given to all the sentences. The summary could be presented in various forms, such as diagram, or text. Our expected output summary takes the form of pros and cons along with a horizontal histogram, where the pros indicate the set of positive product aspects/opinions and the represent the set of negative aspects/opinions. The horizontal histogram included as the percentage of positive opinions compared to negative opinions for all sentences.

3. PROPOSED EXTRACTION TECHNIQUE

As illustrated in previous sections, system input is a list of customers' reviews of a specific product and the output is a summary of all reviews of that product. The initial tasks of this paper rely on part-of-speech (POS) tagging

3.1 Part-of-Speech (POS) Tagging

To extract useful information such as aspects and opinions from reviews, the reviews must be parsed and parts of speech tagged accordingly. Part-of-speech (POS) tagging is the process of parsing each word of the sentence based on identifying linguistic tags. Table 4shows a list of linguistic POS tags.To illustrate the use of POS tagging, we offer the example of a customer's review of an iPhone5s. The original sentence is, "I love my new Iphone5s, it is the best Smartphone ever, and it has a great camera that captures the best photos." The tagged sentence is "I/PRP love/VBP my/PRP$ new/JJ IPhone/NN 5s/NNS, /, it/PRP is/VBZ the/DT best/JJS smartphone/NN ever/RB, /, it/PRP has/VBZ a/DT great/JJ camera/NN that/WDT captures/VBZ the/DT best/JJS photos/NNS /" where every word is tagged using the categories shown in Table 4.

Table 4.Part-of-speech (POS) tagging

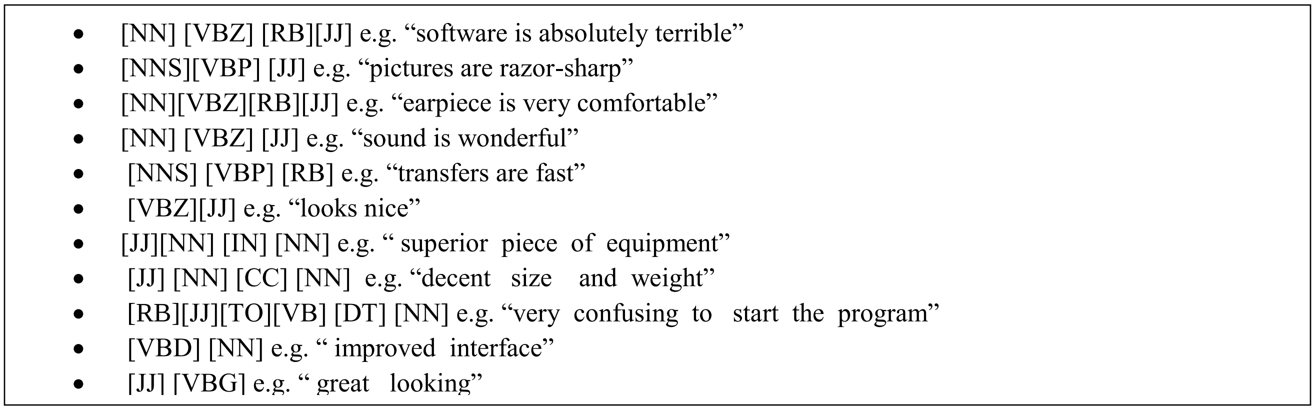
Tag	Description	Tag	Description
JJ	Adjective	RBR	Comparative adverb
JJR	Comparative adjective	RBS	Superlative adverb
JJS	Superlative adjective	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
NN	Noun, singular or mass	VBG	Verb, gerund, or present participle

NNS/NNP	Noun, plural noun, singular	VBN	Verb, past participle
NNPS	Proper noun, plural	VBP	Verb, non-3rd-person singular/p
RB	Adverb	VBZ	Verb, 3rd-person singular present

Earlier research demonstrated that product aspects tend to be nouns or/and noun phrases and opinions tend to be adjectives or/and adjective phrases. In sentiment analysis research showed that some combination of tags contribute to aspects and opinion extraction. Unlike these previous studies, the current research made more use of the sentence parsing process by considering more parts of the sentence to be aspects or/and opinions. The proposed framework is designed to determine what people like and dislike about a given product. Identifying the aspects of the product is the first task, followed by finding the corresponding opinions. Understanding natural language is not easy, so the extraction process is not easy as well. The major difficulty is to understand the implicit meaning of a specific sentence. For example, "using Iphone5 is a piece of cake," the phrase "piece of cake"means it is easy to use. However, there is no explicit word to show that hidden meaning. To solve such issues, semantic understanding is needed.

3.2. Product Aspects Extraction

Aspect extraction involves extraction of aspects of the product being studied about which customers have expressed their opinions on. Aspects are usually nouns or/and noun phrases, for example, "face recognition", "zoom", and "touch screen" are aspects of the product "camera". We must examine all review sentences to know which POS items presented as aspects and which presented as opinions. In natural language, people tend to write almost similar sentence structure. From here, we choose to use frequent sets based on its success in analyzing and understanding customer purchasing behavior. Mining frequent sets plays a great role in data mining, it aims to find interesting patterns form large amount of data. Frequent sets were introduced by to analyze customer behavior and how customers tend to purchase sets of items together. The main motivation to search frequent "tag" sets, came from the need to analyze how people tend to express their feelings in natural language. In other words, how people tend to write opinionated reviews.The result is the frequent sets, which consisted of frequent tags that define the product aspects, the opinion words and the relationship between those two tags.For instance, the tag of aspect appears first, therefore, the sequent of tags looks like [NN][VBZ][RB][JJ] which correspond to the sentence "software is absolutely terrible" . Figure 3 and Figure 4show tags that are more frequent, whereas Figure 5 shows how those tags are extracted.



- [NN] [VBZ] [RB][JJ] e.g. “software is absolutely terrible”
- [NNS][VBP] [JJ] e.g. “pictures are razor-sharp”
- [NN][VBZ][RB][JJ] e.g. “earpiece is very comfortable”
- [NN] [VBZ] [JJ] e.g. “sound is wonderful”
- [NNS] [VBP] [RB] e.g. “transfers are fast”
- [VBZ][JJ] e.g. “looks nice”
- [JJ][NN] [IN] [NN] e.g. “ superior piece of equipment”
- [JJ] [NN] [CC] [NN] e.g. “decent size and weight”
- [RB][JJ][TO][VB] [DT] [NN] e.g. “very confusing to start the program”
- [VBD] [NN] e.g. “ improved interface”
- [JJ] [VBG] e.g. “ great looking”

Figure 3. Frequent tags" Aspect appears first"

Pseudo code for Aspect Tag Extraction

```
Input  :  Review sentences from the dataset
Output: Opinionated sentences with polarity classification

Step1   For each review Identify the subjective sentence S
Step 2   Pre-Processing each sentence with tokenize
Step 3   Apply the tag extraction by its frequency
Step 4  Extract the feature from each sentence and assign weights for tag
Step 5   Counts the opinion by the total weight for each aspect from sentence S
Step6   Summarize the opinion of the aspects with its polarity
```

3.3. Opinion Words Extraction

The second task of the extraction process is opinion extraction. This task involves extracting corresponding opinion words that customers used for every product aspects. Opinion words are usually adjectives that describe or express what customers think about product aspects. Usually, opinion words are located near aspects in the sentence. Some researches located opinion words as the closest adjective to the aspects.Nevertheless, we first locate the opinions words in the sentence and from there we determine the corresponding aspects by searching the sentence backwards first for the closest aspect, if we did not find, then we search forwards. The extraction algorithm is shown below.

4. EXPERIMENTAL RESULTS AND DISCUSSION

To experiment the proposed frame work, we chose the domain data set which is provided in amazons and epinions web sites, which provides a review platform for a wide variety of products and services. The total corpus consist of about 2000 sentence extracted from about 5000 reviews .The sentences from the reviews for this study was restricted to be lower than 200 character to avoid mistakes done by the tokenizes. For both the positive and negative expressions of opinion we calculate the accuracy, precision and recall

5. CONCLUSION

In this paper, the proposed framework to produce an opinionated summary from customer reviews. The main achievement involved the task of aspect and opinion extraction. The extraction was based on the frequent set of opinionate aspects and the weights assigned to the tags for those opinionated aspects and finally summarize with the polarity of the aspects of the product from the given review dataset. The main objective of this study is to provide "aspect-based opinionated summary" from customer reviews of online sold products. The experimental results showed great promise for the technique. This work achieved very high precision and a normal recall performance compared to the baseline model in extracting aspect and opinion.

REFERENCES

[1] Wogenstein, F., Drescher, J., Reinel, D., Rill, S. and Scheidt, J. (2013): Evaluation of an algorithm for aspect-based opinion mining using a lexicon-based approach. Proc. 2nd International Workshop on Issues of Sentiment Discovery and Opinion Mining, 5, ACM.

[2] Zhang, L. and Liu, B. (2014): Aspect and entity extraction for opinion mining, Data Mining and Knowledge Discovery for Big Data, 1–40. Chu, W.W. (ed), Springer, Berlin Heidelberg.

[3] Taboada, M., Brooke J., Tofiloski, M., Voll, K. and Stede, M. (2011): Lexicon-based methods for sentiment analysis. Computational Linguistics 37(2): 267–307.

[4] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver,Colorado, 2015, pp. 486–495.

[5] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, and M. AL-Ayyoub, "Semeval-2016 task 5:Aspect based sentiment analysis," in Proceedings of the 10th International Workshop on Semantic Evaluation, Association for Computational Linguistics, San Diego, California, 2016.

[6] A. El-Halees, "Arabic opinion mining using combined classification approach," 2011.

[7] Vivek Narayanan, Ishan Arora and Arjun Bhatia, "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model", 14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013, pp 194- 201.

[8] Heui Lim, "Improving kNN Based Text Classification with Well Estimated Parameters", LNCS, Vol. 3316, Oct 2004, Pages 516-523.

[9] Esuli, A, Sebastiani, F, "SentiWordNet: A publicly available resource for opinion mining", In Proceedings of the 6th international conference on Language Resources and Evaluation (LREC'06), 2006, pp.417–422.

[10] Stefano Baccianella, Andrea Esuli, and FabrizioSebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining", In Proceedings of the 10th International conference on Language Resources and Evaluation (LREC'10),2010.

[11] Qi, L. and Chen, L. (2010): A linear-chain CRF-based learning approach for web opinion mining. Web Information Systems Engineering–WISE 2010, 128– 141, Springer.