Extracting Patterns for Feature-based Opinion Words from Customer Reviews

Su Su Htay
University of Computer Studies, Mandalay
susuhtay009@gmial.com

Abstract

Most of online Merchants sites allow customers to write comments about particular product that they have purchased. Customer reviews express opinion about products or service which are collectively referred to as customer feedback data. Opinion extraction about products from customer reviews is becoming an interesting area of research. Therefore efficient method and techniques are needed to extract opinions from reviews. In this paper we proposed a novel idea to find opinion words or phrases for each feature from customer reviews in an efficient way. Our focus in this paper is to get the patterns of opinion words/phrases about the feature of product from the review text through adjective, adverb, verb and noun. The extracted features and opinions are useful for generating a meaningful summary that can provide significant informative resource to help the user as well merchants to track the most suitable choice of product.

1. Introduction

Much of the existing research on textual information processing has been focused on mining and retrieval of factual information. Little work had been done on the process of mining opinions until only recently. Automatic extraction of customers' opinions can better benefit for both customers and manufacturers. Product review mining can provide effective information that are classified customer reviews as "recommended" or "not recommended" based on customers' opinions for each product feature. Customer reviews highlight opinion about product features from various Merchant sites. But, many reviews are so long and only a few sentences containing opinions for product features. For a popular product, the number of reviews can be in hundreds or even in thousands, which is difficult to be read them one by one. Therefore, automatic extraction and summarization of opinion is required for each feature. Actually when a user expresses opinion for a product then he/she states about the product as a whole or about its features one by one. Feature identification in product is the first step of opinion mining application. To produce a useful summary, we have to extract opinion for each feature of a product.

In this paper, we take a written review as input and produce a summary review as output. Given a set of customer reviews of a particular product, we need to perform the tasks: (1) Identify product feature that customer commented on. (2) Extracting opinion words or phrases through adjective, adverb, verb and noun and determining the orientation. We use a part-of-speech tagger to identify phrases in the input text that contain adjective or adverb or verb or nouns as opinion phrases. A phrase has a positive semantic orientation when it has good associations (e.g., "awesome camera") and a

negative semantic orientation when it has bad associations (e.g., “low battery”). (3) Generating the summary.

The rest of the paper is organized as follows. Section 2 describes the related work of this paper. Section 3 elaborates motivation for generating opinion summary. Section 4 express system overview and section 5 describes conclusions.

2. Related Work

There are several techniques to perform opinion mining tasks. Hu et al [1] proposed a technique based on association rule mining to extract product features. Then frequent itemsets of nouns in reviews are likely to be product features while the infrequent ones are less likely to be product features. This work also introduced the idea of using opinion words to find additional (often infrequent) features.

Liu et al [4] presented to extract product features from “Pros” and “Cons” as type of review format 2. They proposed a supervised pattern mining method to find language patterns to identify product features. They do not need to determine opinion orientations because of using review format 2 indicated by “Pros” and “Cons”.

Hu et al [2] proposed a number of techniques based on data mining and natural language processing methods to mine opinion/product features. It is mainly related to text summarization and terminology identification. Their system does not mine product features and their work does not need a training corpus to build a summary.

Su et al [5] proposed a novel mutual reinforcement approach to deal with the feature-level opinion mining problem. Their approach predicted opinions relating to different product features without the explicit appearance of product feature words in reviews. They aim to mine the hidden sentiment link between product features and opinion words and then build the association set.

An approach for mining product feature and opinion based on consideration of syntactic information and semantic information in [6]. The methods acquire relations based on fixed position of words. However, the approaches are not effective for many cases.

Turney, P. D. [7] presented a simple unsupervised learning algorithm for classifying reviews as *recommended* (thumbs up) or *not recommended* (thumbs down). The classification of a review is predicted by the average *semantic orientation* of the phrases in the review that contain adjectives or adverbs.

Wu et al [8] implemented for extracting relations between product features and expressions of opinions. The relation extraction is an important subtask of opinion mining for the relations between more than one product features and different opinion words on each of them.

Zhang et al [9] proposed a method to deal with the problem in which sentences involved objective but not subjective and imply nouns and noun phrases.

3. Background Theory

The Web has dramatically changed the way that people express their opinions. They can now post reviews of products at merchant sites and express their views on almost anything in Internet forums, discussion groups, blogs, etc. These online customer reviews, thereafter, become a cognitive source of information which is very useful for both potential customers and product manufacturers. Customers have utilized this piece of this information to support their decision on whether to purchase the product. For product manufacturer perspective, understanding the preferences of customers is highly valuable

for product development, marketing and consumer relationship management.

3.1. Opinion Mining

In this paper, we only focus on mining of opinions on the Web. The task is not only technically challenging because of the need for natural language processing, but also very useful in practice. For example, businesses always want to find public or consumer opinions about their products and services from the commercial web sites. Potential customers also want to know the opinions of existing users before they use a service or purchase a product. Moreover, opinion mining can also provide valuable information for placing advertisements in Commercial Web pages. If in a page people express positive opinions or sentiments on a product, it may be a good idea to place an ad of the product. However, if people express negative opinions about the product, it is probably not wise to place an ad of the product. A better idea may be to place an ad of a competitor's product.

There are three main review formats on the Web. Different review formats may need different techniques to perform the opinion extraction task.

Format (1) - Pros and Cons: The reviewer is asked to describe Pros and Cons separately.

Format (2) - Pros, Cons and detailed review: The reviewer is asked to describe Pros and Cons separately and also write a detailed review.

Format (3) - free format: The reviewer can write freely, i.e., no separation of Pros and Cons.

For the review formats 1 and 2, opinion (or semantic) orientations (positive or negative) of the features are known because pros and cons are separated. Only product features need to be identified. We concentrate on reviews format (3) and we need to identify and extract both product features and opinions.

3.1.1. Feature-based Opinion Mining

This task goes to the sentence level to discover details, i.e., what aspects of an object that people liked or disliked. The object could be a product, a service, a topic, an individual, an organization, etc. For example, in a product review, this task identifies product features that have been commented on by reviewers and determines whether the comments are positive or negative. To obtain such detailed aspects, we need to go to the sentence level. Two tasks are apparent [245]:

1. Identifying and extracting features of the product that the reviewers have expressed their opinions on, called **product features**. For instance, in the sentence "the picture quality of this camera is amazing," the product feature is "picture quality".
2. Determining whether the opinions on the features are positive, negative or neutral. In the above sentence, the opinion on the feature "picture quality" is positive.

In the sentence, "the battery life of this camera is too short," the comment is on the "battery life" and the opinion is negative. A structured summary will also be produced from the mining results.

3.2. Dataset of the system

We used annotated customer reviews data set of 5 products [1] for testing. All the reviews are from commercial web sites such as amazon.com, epinion.com, etc. Only the 2 products of camera reviews data are used such as Canon G3 and Nikon coolpix 4300. Each review consists of review title and detail of review text. The reviews are re-tagged manually based on our own feature list. Each camera review sentence is attached with the mentioned features and their associated opinion words. Therefore, we only

focus on the review sentence containing product features and opinion. ”The pictures are absolutely amazing - the camera captures the minutest of details” will receive the tag: picture [+3]. Words in the brackets are those we found to be associated with the corresponding opinion orientation of feature whether positive or negative.

4. Extracting Patterns of Opinion Phrases

The goal of OM is to extract customer feedback or opinions on products and present the information in the most effective way that serves the chosen objectives. Customers express their opinion in review sentences with single words or phrases. We need to extract these opinion words or phrases in efficient way. Pattern extraction approach is useful for commercial web pages in which customers can able to write comments about products or services. Let us use an example of the following review sentence:
“The battery life is long.”

In this sentence, the feature is “battery life” and opinion word is “long”. Therefore, we first need to identify the feature and opinion from the sentence.

Figure 1 shows the process for generating the results of feature-based opinion summarization. The system input is customer reviews datasets. We first need to perform POS tagging to parse the sentence and then identify product features and opinion words. The extracted opinion words/phrases are used to determine the opinion orientation which is positive or negative. Finally, we summarize the opinion for each product feature based on their orientations.

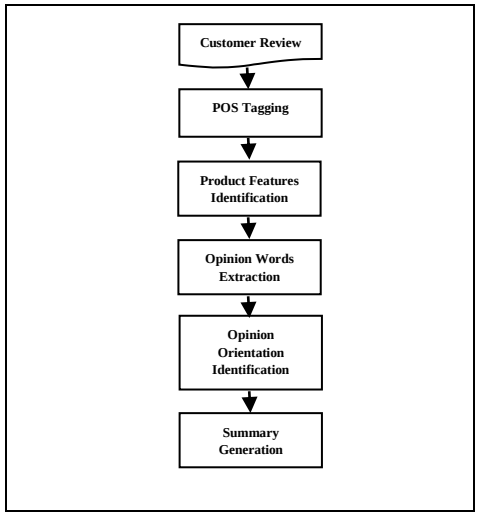

Figure 1. Processing steps for generating feature-based opinion summary

4.1. Identify Product Features

In feature extraction phase, we need to perform part-of-speech tagging to identify nouns/ noun phrases from the reviews that can be product features. POS tagging is important as it allow us to generate general language patterns. We use NLProcessor Linguistic parser (NLProcessor 2000) to parse each sentences and yields the part-of-speech tag of each word (whether the word is a noun, adjective, verb, adverb, etc) and identifies simple noun and verb groups (syntactic chunking). For instance:
<S> <NG><W C=’PRP’ L=’SS’ T=’w’ S=’Y’> ***I*** </W></NG> <VG><W C=’VBP’> ***am*** </W><W C=’RB’> ***absolutely*** </W></VG> <W C=’IN’> ***in*** </W> <NG><W C=’NN’> ***awe*** </W></NG> <W C=’IN’> **of** </W> <NG><W C=’DT’> ***this*** </W> <W C=’NN’> ***camera*** </W></NG> <W C=’.’> **.** </W></S>

We focus on identifying candidate product features that are talked about by many customers. Nouns and noun phrases are the most likely to be product features. After POS tagging is done, we need to identify frequent features that are nouns

or noun phrases using association rule mining in Hu and Liu [1] because frequent features are likely to be product features. The frequent features are then used to find nearest opinion words with adjective/adverb.

4.2. Determining Opinion and its Orientation

To decide the opinion orientation of each sentence, we need to perform three sub-tasks. First, a set of opinion words (adjectives, as they are normally used to express opinions) is identified. If an adjective appears near a product feature in a sentence, then it is regarded as an opinion word. Opinion words are words which contain positive or negative opinion of potential customers. We can extract opinion words from the review using all the remaining frequent features (after pruning). For instance:

"The strap is horrible and gets in the way of parts the camera you need access to."

"After nearly 800 pictures I have found that this camera takes incredible pictures."

"It comes with a rechargeable battery that does not seem to last all that long, especially if you use the flash a lot."

For the first sentence, the feature, *strap*, is near the opinion word horrible. And the second example, feature "picture" is close to the opinion word incredible. In this case, we can extract the nearby adjective as opinion word if the sentences contain any frequent features.

However, for the third sentence, the feature, battery, can not be able to extract nearby adjective to meet the opinion word "long". The nearby adjective "rechargable" dose not bear opinion for the feature "battery".

Moreover, adjectives and adverbs are good indicators of subjectivity and opinions. Therefore, we need to extract phrases containing adjective, adverb, verb and noun that imply opinion. We also consider some verbs (like, recommend, prefer, appreciate, dislike, love) as opinion words. Some adverbs like (not, always, really, never, overall, absolutely, highly, well) are also consider. We extract two or three consecutive words from the POS tagged review if their tag conform to any of the pattern in table 1.

Table 1. Extracted Phrases Patterns

Pattern	First word	Second word	Third word
Pattern 1	JJ	NN/NNS	--
Pattern 2	JJ	NN/NNS	NN/NNS
Pattern 3	RB/RBR/RBS	JJ	--
Pattern 4	RB/RBR/RBS	JJ/RB/RBR/RBS	NN/NNS
Pattern 5	RB/RBR/RBS	VBN/VBD	--
Pattern 6	RB/RBR/RBS	RB/RBR/RBS	JJ
Pattern 7	VBN/VBD	NN/NNS	--
Pattern 8	VBN/VBD	RB/RBR/RBS	--

We found that opinion words/phrases are mainly adjective/adverb that are used to qualify product features with nouns/noun phrases. Therefore, we extract two/three consecutive words containing adjectives and/or adverbs from the POS tagged reviews that are adjacent with nouns/noun phrases if their tags conform to any of the patterns in Table 1.

We collect all opinionated phrases of mostly 2/3 words like (Adjective, Noun), (Adjective, Noun, Noun), (Adverb, Adjective), (Adverb, Adjective, Noun), (Verb, Noun) etc. from the processed POS Tagged Review.

The resulting patterns are used to match and identify opinion phrases for new reviews after the POS tagging. However, there is more likely opinion words/phrases in the sentence but it is not extracted by any pattern. From these

extracted patterns, most of adjectives or adverbs imply opinion for nearest nouns/noun phrases. We further need to prune useless data from the extracted patterns that are not mention as mainly opinionated words like adjectives/adverbs. Table 2 described some example of opinion phrases.

Table 2. A few example of extracted opinionated phrases

(Adjective, Noun)	(low battery), (good memories), (awesome camera) etc.
(Adjective, Noun, Noun)	(high quality pictures)
(Adverb, Adjective)	(extremely pleased), (very easy), (really annoying), (absolutely amazing) etc.
(Adverb, Adjective, Noun)	(very compact camera), (very good pictures) etc.
(Adverb, verb)	(personally recommend)
(Adverb, Adverb, Adjective)	(not so bad) etc.
(Verb, Noun)	(recommend camera), (appreciate picture) etc.
(Verb, Adverb)	(perform well)

Second, for each opinion word, we determine its semantic orientation, i.e., positive or negative. We found that the present of adjectives/adverbs is useful for predicting whether a sentecne is subjective, i.e, expressing an opinion. From these patterns, we extract nearby adjectives/adverbs as opinion words that modifies nouns and noun phrases. In this way, we build up the opinion list.

The semantic orientation of a phrase is calculated as the mutual information between the given phrase and the word "excellence" minus the mutual information between the given phrase and the word "poor". In addition we need to determine the direction of phrase's semantic orientation.

The pointwise mutual information (PMI) between two words defined as follow.

$$PMI(WORD_1, WORD_2) = LOG_2\left[\frac{P(WORD_1 \& WORD_2)}{P(WORD_1)P(WORD_2)}\right] \quad (1)$$

This ratio is a measure of statistical dependence between the two words. The log of this ratio is the amount of information that we acquire about the presence of the words when we observe the others.

The semantic orientation of a phrase is calculated as follow.

$$SO(phrase) = PMI(phrase, "excellent") - PMI(phrase, "poor") \quad (2)$$

SO is positive when phrase is more strongly associated with "excellent" and negative when phrase is more strongly associated with "poor". Finally, we compute the average SO of all phrases in the given review and classify the review as recommended if the average SO is positive and otherwise not recommended. The final output is feature and SO its semantic orientation.

4.3 Summary Generation

After the previous steps, we generate the opinion summary which consists of the following steps:

- For each discovered features, we classify the opinionated sentences into positive and negative categories according to the opinion orientation.
- We compute the count to show the numbers of positive/negative opinionated sentences for each feature.

The following sentences show the example summary of customer reviews classify into positive/negative category of opinion for the product feature in figure 2. The figure shows the product feature "picture" of digital camera and

the numbers of positive/negative opinion for that feature and individual sentences in detail.

Feature: **picture**
Positive: 12
• Overall this is a good camera with really good picture clarity.
• The pictures are absolutely amazing - the camera captures the minutest of details.
• After nearly 800 pictures I have found that this camera takes incredible pictures.
…

Negative: 2
• The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
• Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

Figure 2. An example summary

5. Conclusion

Most of Opinion Mining researches use a number of techniques for mining opinion features and summarizing product reviews based on data mining and natural language processing methods. Review text is unstructured and only a portion or some sentences include opinion-oriented words. Therefore opinion mining system needs only the required sentences to be processed to get knowledge efficiently and effectively. Actually when a user expresses an opinion about a product then he/she states about the product as a whole or about its features one by one. In product reviews, users write comments about features of products to describe their views according to their experience and observations. The first step of opinion mining in classifying reviews documents is extracting features and opinion words. We proposed the ideas to extract patterns of features and/or opinions phrases. We expected to achieve good results by extracting features and opinion-oriented words from review text with help of adjective, adverbs, noun and verb. We believe that there is rich potential for future research. For identifying feature, we need to extend both explicit and implicit feature that constitute in determining the polarity of product/feature.

References

[1] Hu, M, and Liu, B. "Mining and Summarizing Customer Reviews", *in Proceedings of KDD*, 2004, 168-177.

[2] Hu, M., and Liu, B. 2004. "Mining Opinion Features in Customer Reviews." To appear in AAAI'04, 2004.

[3] Liu., B. "Sentiment Analysis and Subjectivity", A chapter in Handbook of Natural language Processing, Second edition.

[4] Liu, B., Hu, M., Cheng, J. "Opinion Observer: Analyzing and Comparing Opinions on the Web", IW3C2, Chiba, Japan, May 10-14, 2005

[5] Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., Swen, B. and Su, Z. "Hidden Sentiment Association in Chinese Web Opinion Mining." i*n the International World Wide Web Conference Committee (IW3C2)*, 2008.

[6] Somprasertsri, G., and Lalitrojwong, P. "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization." *Journal of Universal Computer Science*, 2010.

[7] Turney, P. D., 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*.

[8] Wu, Y., Zhang, Q., Huang, X. and Wu, L. "Phrase Dependency Parsing for Opinion Mining." i*n Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*.

[9] Zhang, L., Liu, B. "Identifying Noun Product Features that Imply Opinions", ACL, 2011.