

Biologically Inspired Opinion Mining: Features Extraction based on Linguistic Patterns for Customer Reviews

Su Su Htay

*University of Computer Studies, Myanmar
susuhtay009@gmail.com*

Abstract

Today, most of online Merchants sites are available to write comments for their customers about particular products/services that they have purchased. Customer reviews express opinion about products/services which are collectively referred to as customers' feedback data. The task of feature extraction for customer reviews is becoming an interesting part of research area in Opinion Mining application. Therefore, efficient method and techniques are needed to extract features from customer reviews. In this paper, we proposed an idea which is based on the knowledge of linguistic filtering patterns to find product features from customer reviews in an efficient way. Our focus in this paper is to get the patterns of words/phrases about the product features from the review text through nouns, verbs, and adjectives. Experiments will carry out to show the promising results using five electronic product reviews datasets from amazon.com.

Keyword: *Opinion Mining, Feature Extraction, Customer Reviews, sentiment orientation.*

1. Introduction

Much of the existing research on textual information processing has been focused on mining and retrieval of factual information. Little works had been done on the processes of mining opinions until only recently. Automatic extraction of customers' opinions can be 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 on a product then he/she states about the product as a whole or about its features one by one. Feature extraction is the critical task for feature-level opinion mining. To produce a useful summary, we have to extract opinions for each feature of a product. In this paper, we take a written review as input for the domain of consumer products. Given a set of customer reviews of a particular product, we need to perform the tasks:

- (1) Extract product feature that customers commented on.
- (2) Extract opinion word for each feature.
- (3) Determine the orientation of opinion for that feature.
- (4) Generate the summary of opinion based on extracted features.

In this paper, we focus on extracting product features which is based on linguistic patterns knowledge.

The rest of the paper is organized as follows. Section 2 describes the related work of this paper. Section 3 elaborates background theory for opinion mining application. Section 4 express proposed techniques for feature extraction task 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.

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.

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.

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.

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.

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. Therefore, mining others’ opinions from online reviews is becoming an interesting area of research.

3.1 Opinion Mining

In this paper, we only focus on mining customers’ opinions for online product reviews. This 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 features are known because pros and cons are separated. Only product features need to be identified.

We concentrate on reviews format (3). Therefore, we need to extract both feature and opinion to determine positive or negative orientation of opinion for that feature.

3.2 Feature-based Opinion Mining

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.

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.

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. 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 feature level. Two tasks are apparent:

- i. Identifying and extracting features of the product that the reviewers have expressed their opinions on, called **product features**.
- ii. Determining whether the opinions on the features are positive, negative or neutral.
- iii. Generating a structured summary from the mining results.

For instance;

The picture quality of this camera is amazing.

The battery life of this camera is too short.

In the above sentence, "picture quality" is the product feature and the reviewer expressed *positive* opinion for that feature. For the next sentence, the comment is on the product feature "battery life" and the opinion for that feature is negative.

3.3 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. Each review consists of review title and detail of review text. The reviews are re-tagged manually based on our own feature list. Each 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 for Product Features

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. In this paper, we focus on product feature identification as the first step of feature-based opinion mining and summarization.

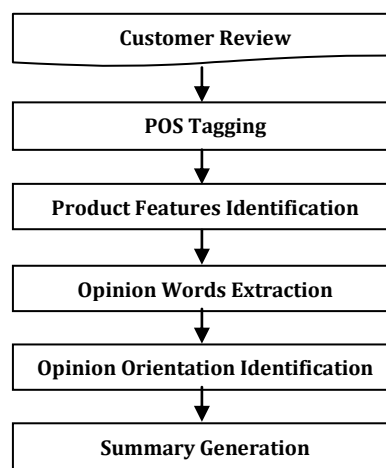


Figure 1: Processing steps for generating feature-based opinion summary

4.1 Identify Product Features

In feature extraction phase, we need to perform part-of-speech tagging to extract nouns/ noun phrases from the reviews that can be product features. POS tagging is important as it allow us to generate the general language patterns. We use Stanford-postagger to parse each sentences and yields the part-of-speech tags of each word (whether the word is a noun, adjective, verb, adverb, etc). For instance:

The_DT photo_JJ quality_NN is_VBZ amazing_JJ and_CC i_FW know_VBP i_FW 'm_VBP going_VBG to_TO have_VB fun_NN with_IN all_PDT the_DT features_NNS ._.

We focus on extracting product features that are talked about by many customers. After POS tagging is done, we need to extract product features using predefined patterns of nouns/ noun phrases based on linguistic knowledge because most of them are likely to be product features.

Feature patterns	Example Features
NN	picture, disc, size, zoom, screensaver, warranty, ringtone, weight, lcd, vibration, earpiece, keypad, headphone, playlist, etc
VB	remove, load, read, delete, rewind, recognize, etc
VB RP	set up, look up
DT NN	no disc
NN NN	battery life, movie mode, print quality, lens cap, sunset feature, memory card, scroll wheel, mp3 player, volume range, etc
JJ NN	audio devices, optical mode, indoor image, online support, rechargeable battery, mobile service, onscreen display, etc
JJ VB NN	progressive scan player
NN VB NN	autofocus assist light
JJ NN NN	continuous shot mode, dual layer dvd, hot shoe flash, etc
NN IN NN	ease of use, line out jack, etc
NN NN NN	menu dial knob, customer support websites, panel button layout, battery charging system, etc
NN IN DT NN	bang for the buck, quality of the picture, size of this camera, etc
NN TO NN NN	signal to noise ratio

Table 1: Extracted feature patterns and example of product features for each pattern

Input:	S= Set of tagged sentences; $s=s_1, s_2, \dots, s_n$ P= Set of noun phrase patterns FC=Feature training corpus
Output:	FS= Set of product feature candidates
<pre> FS= \emptyset For each tagged sentence $s_n \in S$ PC= \emptyset For i=0 to end of sentence s_n If $i < \text{Length}(s_n) - 2$ then $x=3$ Else If $i = \text{Length}(s_n) - 2$ then $x=2$ Else If $i = \text{Length}(s_n) - 1$ then $x=1$ Else $x=0$ End End For j= x to 0 GT=T_i to T_{i+j} /*POS Tag of word$_i$ to word$_{i+j}$ of s_n GW= word$_i$ to word$_{i+j}$ If $GT \in P$ and $GW \notin FC$ then i= i+j PC = PC \cup GW Break End End End FS= FC \cap S \cup PC End </pre>	

Figure 2: Algorithm for finding features consistent with patterns

However, in some case, not all these patterns are likely to be product features. Therefore, we further need to consider all of the extracted features are term feature or not. To solve this problem, we used annotated training corpus which are manually labeled by human tagger to identify whether the extracted features are the term features or not for a specific product.

4.2 Experimental results

For evaluation, we manually read all the reviews. All the results generated by our system are compared with the manually tagged results. Tagging is fairly straightforward for both product features and opinions.

Product Name	No. of tagged features	No. of review sentences	Accuracy of Feature Extraction%
Apex	110	739	69.5
Canon	100	597	71.3
Creative	180	1716	75.6
Nikon	74	346	68.5
Nokia	109	546	73.3

Table 2: Accuracy of feature extraction for five different products

A minor complication regarding feature tagging is that features can be explicit or implicit in a sentence. Most features appear explicitly in opinion sentences, e.g., *pictures* in “*the pictures are absolutely amazing*”.

Some features may not appear in sentences. We call such features as implicit features, e.g., *size* in “*it fits in a pocket nicely*”. Both explicit and implicit features are easy to identify by the human tagger.

5. Conclusion

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 product features extraction. In this paper, we proposed an idea to mining product features based on linguistic patterns which is effectively support for generating review summary of opinions for those features from customer reviews. We used five electronic products available from amazon.com for feature extraction. We expected to achieve good results by extracting these patterns from review text with the help of adjectives, nouns and verbs that are most likely to be product features. For identifying feature, we need to use an annotated training corpus that is manually tagged by human tagger for explicit features only.

References

- [1] Hu, M & Liu, B 2004a. "Mining and Summarizing Customer Reviews", in *Proceedings of KDD*, 168-177.
- [2] Hu, M & Liu, B 2004b. "Mining Opinion Features in Customer Reviews." To appear in AAAI'04.
- [3] Liu, B. "Sentiment Analysis and Subjectivity", A chapter in *Handbook of Natural language Processing*, Second edition.
- [4] Liu, B, Hu, M & Cheng, J 2005. "Opinion Observer: Analyzing and Comparing Opinions on the Web", IW3C2, Chiba, Japan.
- [5] Su, Q, Xu, X, Guo, H, Guo, Z, Wu, X, Zhang, X, Swen, B & Su, Z 2008. "Hidden Sentiment Association in Chinese Web Opinion Mining." in *the International World Wide Web Conference Committee (IW3C2)*.
- [6] Somprasertsri, G & Lalitrojwong, P 2010. "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization." *Journal of Universal Computer Science*.
- [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 & Wu, L 2009. "Phrase Dependency Parsing for Opinion Mining." in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*.
- [9] Zhang, L & Liu, B 2011. "Identifying Noun Product Features that Imply Opinions", ACL.