Mining Product Features from Online Reviews

by

Weishu Hu

Master of Software Engineering

2010

Faculty of Science and Technology
University of Macau
Mining Product Features from Online Reviews

by

Weishu Hu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Software Engineering

Faculty of Science and Technology
University of Macau

2010

Approved by ______________________________________________________________

Supervisor

Date ______________________________________________________________________
In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

Address:  
Telephone:  
Fax:  
E-mail:  

Signature _____________________  
Date _________________________
MINING PRODUCT FEATURES FROM ONLINE REVIEWS

by Weishu Hu

Thesis Supervisor:
Associate Professor, Zhiguo Gong
Master of Software Engineering

With the advance of the internet, e-commerce systems have become extremely important and convenient to human being. More and more products are sold on the Web, and more and more people are purchasing products online. As a result, an increasing number of customers post product reviews at merchant websites and express their opinions and experiences in any network space such as internet forums, discussion groups, and blogs. So there is a large amount of data records related to products on the Web, which are useful for both manufacturers and customers. Mining product reviews becomes a hot research topic, and existing researches mostly base on product features to analyze the opinions. So mining product features is the number one step to further reviews processing. In this thesis, we present how to mine product features efficiently and accurately. The proposed extraction approach is different from the previous methods because we only mine the features of the product from opinion sentences in which the customers have expressed their positive or negative sentiment. In order to find opinion sentences, a SentiWordNet-based algorithm is proposed. There are three steps to perform our task: (1) Identifying opinion sentences in each review which is positive or negative via SentiWordNet; (2) Mining product features that have been commented on by customers from opinion sentences; (3) Pruning feature to remove these incorrect features. Compared to previous work, our experimental result achieves higher precision and recall. It executes fast enough for practical use.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Figure Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1: An Example of Summary</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Figure 2: A System Architecture</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>Figure 3: The Architect of Product Features Extraction System</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Figure 4: An Example of Output</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Figure 5: Head Part of SentiWordNet</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>Figure 6: An Example of Input Data</td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>Figure 7: One Part Result of the System</td>
<td></td>
<td>42</td>
</tr>
</tbody>
</table>


LIST OF TABLES

<table>
<thead>
<tr>
<th>Number</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1: Illustration of the terms for Precision and Recall</td>
<td>39</td>
</tr>
<tr>
<td>Table 2: Precision and Recall at Opinion Sentences Identification</td>
<td>40</td>
</tr>
<tr>
<td>Table 3: Comparison of Hu’s, Opine’s and our results</td>
<td>42</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>AMA</td>
<td>Association Mining Algorithm</td>
</tr>
<tr>
<td>CBA</td>
<td>Classification Based on Associations</td>
</tr>
<tr>
<td>JWI</td>
<td>Java Wordnet Interface</td>
</tr>
<tr>
<td>NB</td>
<td>Naïve Bayes</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>OM</td>
<td>Opinion Mining</td>
</tr>
<tr>
<td>PMI</td>
<td>Point-wise Mutual Information</td>
</tr>
<tr>
<td>PN</td>
<td>Positive &amp; Negative</td>
</tr>
<tr>
<td>POS</td>
<td>Part of Speech</td>
</tr>
<tr>
<td>SLSC</td>
<td>Sentence-Level Semantic Classification</td>
</tr>
<tr>
<td>SO</td>
<td>Subjective &amp; Objective</td>
</tr>
<tr>
<td>SGML</td>
<td>Standard Generalized Markup Language</td>
</tr>
<tr>
<td>URL</td>
<td>Universal Resource Locators</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to convey my sincere appreciation to my supervisor Prof. Zhiguo Gong for his professional guidance, great patience and valuable suggestions throughout the whole process of thesis writing. Prof. Gong always encouraged me when the research could not move forward. He provided much useful idea to me, and also gave instructive tutorial to improve my research capability. Prof. Gong was also an important mentor like a friend. He provided a critical but supportive perspective on my research. He would gently remind me of the importance of conciseness, clarity, readability and proper use of technical terms in writing both papers and thesis.

I acknowledge with gratitude all my research group members, Guangyi Xiao, Yiyang Yang, Haitao Zou, whose cooperation helped me to improve my research. We worked in a cooperative and collegial environment.

I am grateful to other professors in University of Macau, Prof. Enhua Wu, Dr. Jingzhi Guo, Dr. Qiwen Xu, Dr. Si Yain Whar, Dr. Tham Yiu Kwok, Dr. Vong Chi Man, Dr. Xiaoshan LI, Dr. Yan Zhuang, and Dr. Fong Chi Chiu. Their impressed teaching inspired me to make further research during my graduate study.

Last but not least, I wish to give my thanks to my family and my girl friend Ziwei Yang for their support and encouragement all these years.
CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

As internet has become one part of people’s daily lives, e-commerce has been developed into one of the most important methods in business deal. More and more products are sold on the Web, and more and more people are shopping on the Web. In order to share their shopping experiences and feedbacks, an increasing number of customers post product reviews at merchant websites or express their opinions and experiences in various network space such as internet forums, discussion groups and blogs, which is a great wealth of opinions about products. As a consequence, mining opinion has become a perspective research topic since [8, 21, 27]. It is quite different from traditional text summarization [11, 28] in some ways: (1) Opinion mining is a structured summarization rather than a free document which is generated by most text summarization systems; (2) Users are mainly interested in the product features that customers have positive or negative opinions on. A systematic procedure of opinion mining is formed in many researchers’ effort, and Hu and Liu in [21] propose one useful method which is feature-based opinion summarization of reviews. It is performed in three steps:

1. Mining the features of the product that customers have expressed opinions on, and then ranking the features according to their frequencies that they appear in the reviews.

2. For each feature, identifying opinion sentences in each review and determining each opinion sentence’s sentiment orientation. The specific reviews that express opinions are attached to the feature, which facilitates browsing the reviews by potential customers.

3. Making summary using the discovered information.
Hu and Liu summarize the reviews of the digital_camera_1, which looks like the following:

Figure 1: An Example of Summary

<table>
<thead>
<tr>
<th>Digital_camera_1:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature: photo quality</td>
</tr>
<tr>
<td>Positive: 223 &lt;individual reviews&gt;</td>
</tr>
<tr>
<td>Negative: 7 &lt;individual reviews&gt;</td>
</tr>
<tr>
<td>Feature: weight</td>
</tr>
<tr>
<td>Positive: 164 &lt;individual reviews&gt;</td>
</tr>
<tr>
<td>Negative: 15 &lt;individual reviews&gt;</td>
</tr>
</tbody>
</table>

In Figure 1, there are two product features: photo quality and weight. About the photo quality, 223 customer reviews express positive opinions while only 7 customer reviews express negative opinions. The link of <individual reviews> is used to point to the specific reviews that praise or criticize the product feature.

A feature-based opinion summary is helpful for a potential customer to find out the feeling of the existing customers about the digital camera. If potential customers are very concerned with a certain feature, they can click the <individual reviews> link to find out what existing customers like or why they make a complaint.
1.2 MOTIVATION

“How other people think and feel”, is always an important piece of information for most of us to make decisions. In [7, 15], a survey of more than 2000 American adults:

- Between 73% and 87% consumers who have read online reviews of restaurants, hotels, and various services such as travel agencies or doctors, report that reviews made a significant influence on their purchase.

- Consumers report that they are willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item (the variance stems from what type of item or service is considered).

- 32% have provided a rating on a product, service via an online ratings system, and 30% (including 18% of online senior citizens) have posted an online comment or review regarding a product or service.

With the explosion of Web 2.0 platforms (e.g., blogs, discussion forums, peer-to-peer networks, and various other types of social media), Zabin and Jefferries [16] report that consumers like to share their brand experiences and opinions, positive or negative, regarding any product or service. For a popular product, there are hundreds or even thousands reviews on the Web. It is difficult for a potential customer to read them to decide to purchase the product or not. It is also difficult for the manufacturer of the product to survey and manage the opinions. Mining product reviews is very useful for both consumers and manufacturers. And Users mostly concern the product feature which is positive or negative. The feature-based opinion processing is the main approach to mine and summarize opinions. So the first step of the review summarization is mining the product features.

In [21], the merchant or the manufacturer of the product cannot provide a list of features. The reasons are:

- The merchant sells a large number of products so that it is very difficult for him to provide the product features.
• The words used by common users of the product may not be the same as those used by merchants or the manufacturer, although they may refer to the same features. It makes a problem in identifying what the customers are concerned with. In addition, customers may make some comments on the lack of certain features of the product.

• Customers may discuss some features that the manufacturer has never thought about, for example, unexpected features.

• The manufacturer may wish users of its product are not aware of certain weak features.
1.3 OBJECTIVE

*Issue 1*: How to identify the features of the product that the customers have expressed opinions on?

*Issue 2*: How to identify the number of customer reviews express positive or negative opinions?

The objective of this thesis is to solve the above *Issue 1*, *Issue 2*, they are:

- Our first objective is to accurately and reliably extract product features as many as possible, especially for the infrequent features and implicit features.

- Our second objective is to provide a new technique based on natural language pattern and SentiWordNet to extract product features and classify reviews’ sentiment.

- Our last objective is to get best evaluation in efficiency.
1.4 SOLUTION: SENIWORDNET-BASED APPROACH

Extracting product features is the foundation of opinion mining. In order to mine product features, we propose a different method to extract the product features from online reviews. In this thesis, we aim to mine product features that the reviewers have commented on. This task mainly involves three subtasks: (1) Identifying opinion sentences in each review which are positive or negative via SentiWordNet-based algorithm [3]; (2) Mining product features that have been commented on by customers from opinion sentences; (3) Pruning feature to remove those incorrect features.

The essence of product features mining is: many of the previous researchers concentrate on mining product features to identify opinion sentences. Product features are likely to appear in the opinion sentences. This thesis proposes using SentiWordNet-based algorithm to distinguish the opinion sentences which is a new attempt to utilize a number of techniques based on data mining and natural language processing methods to mine product features. Our experimental results show that these techniques are highly effective.
1.5 CONTRIBUTIONS OF THE THESIS

The main contribution of this thesis is to propose a novel method to mine opinion features from opinion sentences, which includes:

- Different from previous method which is mining product features to extract opinion sentences, our method starts with identifying the opinion sentences, and then mining the product features from opinion sentences. It is helpful for accurately and reliably extracting product features as many as possible. First of all, product features are tend to appear in the opinion sentences. If the product features are not in the opinion sentences, they are useless for opinion processing. People pay little attention to product feature without sentiment. So we just focus on the product features in the opinion sentences. Second, the method of product features mining in [2, 6, 21, 22] cannot extract the infrequent features efficiently from the review sentences. Not all the features are modified by the fixed pattern. Opinion sentences are one part of the review sentences, so the frequency of infrequent features in opinion sentences is much higher than in the review sentences. In opinion sentences, infrequent features can be identified more easily.

- A SentiWordNet-based algorithm, which is a theory contribution for identifying the opinion sentences. The sentiment score of each sentence is decided by the sentiment words which are positive or negative in the sentence. The sentiment score of each word is calculated by the SentiWordNet, a public lexical resource in which each WordNet synset $s$ is associated to three numerical scores $\text{Obj}(s)$, $\text{Pos}(s)$ and $\text{Neg}(s)$, describing how objective, positive, and negative the terms contained in the synset are. If the sentiment score of one sentence satisfies certain requirement, it is an opinion sentence.
1.6 ORGANIZATION OF THIS THESIS

The remainder of this thesis is organized as follows:

**Chapter 2** investigates some useful related works for opinion mining algorithms. It is divided into four parts; the first part is about the existing methods to mine product features: (1) features extraction using Association Mining Algorithm (AMA) [21]; (2) OPINE [2] which is a review-mining system whose novel components including the use of relaxation labeling; (3) an application of a probability-based algorithm to extract product features. The second part is about the existing researches on opinion sentences search including (1) how to extract opinion sentences relevant to an open-domain query; (2) how to apply OpinionFinder [25] to identify subjective sentences. The third part is about how to determine the sentiment of opinions and the introduction of SentiWordNet. The last part of chapter 2 is a short conclusion about these related works, and why we should integrate all these three technologies together.

**Chapter 3** gives an overview of our system, describes and evaluates the main components. We present our methodology for using SentiWordNet to calculate opinion words’ score for each sentence. Firstly, we pre-process the review text includes the deletion of stopwords, stemming and fuzzy matching, and then using openNLP to POS tagging. Secondly, calculate the sentiment score of each opinion word with SentiWordNet to generate the sentiment score of each sentence which is the average of all the opinion words in the sentence. And opinion sentences are identified by their sentiment score. Finally, product features including explicit and implicit features are extracted from the opinion sentences. Explicit features can be divided into frequent and infrequent features. Explicit features can be extracted directly, and implicit features are processed by associative spread (Feature Mapping) which is an application of WordNet. After extracting the product features, the feature pruning is employed to improve the accuracy of the extraction.

**Chapter 4** discusses the experiment and evaluation of the implemented application. The evaluation compares the SentiWordNet-based algorithm to other existing product
features mining including AMA and OPINE. The contents of the comparison include the recall, the precision and F-measure of opinion sentences and product features.

**Chapter 5** makes a conclusion with a summary of our contributions, points out the limitation of this research, and gives a brief account of some important future work.
CHAPTER 2: RELATED WORK

This Chapter investigates some useful works relevant to the thesis topic. Section 2.1 presents the existing research on mining product features. Section 2.2 reviews the existing researches on opinion sentences identification and the application of the OpinionFinder. Section 2.3 introduces the existing researches on determining the sentiment of opinions and SentiWordNet. Section 2.4 makes a short conclusion about the significant of our topic about mining product features by SentiWordNet-based algorithm.
2.1 EXISTING RESEARCH ON PRODUCT FEATURES MINING

Many researchers have proposed different methods to solve the problem of mining product features. Our work is closely related to [2, 6, 21, 22] on mining opinion features in customer reviews. The problem is generally decomposed into three main subtasks: (1) Identifying specific product features according to the topic; (2) Identifying opinion expression about the features; (3) Classifying the sentiment orientation of the opinions [18].
2.1.1 **MINING PRODUCT FEATURES BY ASSOCIATION RULE**

Hu and Liu in [21] design a system to perform the summarization in two main steps: feature extraction and opinion direction identification. The inputs to the system are a product name and an entry page for all the reviews of the product. The output is the summary of the reviews.

The system performs five tasks: (1) POS tagging [20], which parses each sentence and assigns the part-of-speech tag of each word, including identifying whether the word is a noun, verb, adjective, etc, and simple noun and verb groups (syntactic chunking). (2) Frequent feature generation. Hu and Liu apply association rule mining [1] to detect all frequent itemsets, which is a set of words or a phrase that occurs together. The association rule miner CBA [19] based on the Apriori algorithm in [1], finds all frequent itemsets in the transaction set with the user-specified minimum support 1%. (3) Feature Pruning aims to remove those incorrect features. Two types of pruning are presented: (a) Compactness pruning checks features that contain at least two words, which are named feature phrases, and removes those that are likely to be meaningless. (b) Redundancy pruning removes redundant features that contain single words. (4) Opinion words extraction with all the remaining frequent features after pruning. (5) Infrequent feature identification. Hu and Liu suppose that people like to use the same opinion word to describe different features. So they can use the opinion words to look for features that cannot be found in (2). If one sentence contains no frequent feature but one or more opinion words, think the nearest noun or noun phrase of the opinion word as an infrequent feature.

The proposal in [21] can produce a number of features, but only explicit features could be found and implicit feature cannot be extracted. The irrelevant sentences may be thought as opinion sentences, and the nouns in irrelevant sentence would be extracted as features. By addition, the premise (people like to use the same opinion word to describe different features) of finding infrequent features is not so reasonable. Different features trend to be described by different opinion words.
2.1.2 UNSUPERVISED INFORMATION-EXTRACTION SYSTEM

In [2], Popescu and Etzioni introduce an unsupervised information-extraction system (OPINE), which mines reviews in order to build a model of important product features, the evaluation by reviewers and the relative quality across products. OPINE performs four main tasks: (1) Identifying product features; (2) Identifying opinions regarding product features; (3) Determining the polarity of opinions; (4) Ranking opinions based on their strength.

OPINE is built on top of KnowItAll, a domain-independent information extraction system based on Web [23], which instantiates relation-specific generic extraction patterns into extraction rules to find candidate facts. A form of Point-wise Mutual Information (PMI) is assigned by KnowItAll’s Assessor between phrases that is estimated from Web search engine hit counts [26]. The PMI is computed between each fact and automatically generated discriminator phrases. Given fact $f$ and discriminator $d$, the PMI score is computed as (1):

$$ PMI(f, d) = \frac{\text{Hits}(d + f)}{\text{Hits}(d) + \text{Hits}(f)} $$

(1)

In order to output a probability associated with each fact, the PMI scores are converted to binary features for a Naive Bayes Classifier.

Given a product class, OPINE extracts explicit feature from parsed review data. (1) The system recursively identifies both the parts and the properties of the given product class and their parts and properties, in turn, continuing until no candidates are found. (2) The system finds related concepts and extracts their parts and properties. Opine achieves 22% higher precision than Hu’s, but has 3% lower recall.
2.1.3 **Apply a Probability-Based Algorithm to Extract Product Features**

Scaffidi presents a probability-based algorithm and compares it to an existing support-based approach in [6]. Specially, Scaffidi uses two algorithms (probability-based and support-based) to extract features from 7 Amazon.com product categories to ask end users to rate the features in terms of helpfulness for choosing products. The features identified by the probability-based algorithm are preferred by the end users. The probability-based algorithm can identify features that comprise a single noun or two successive nouns (which end users rated as more helpful than features comprising only one noun), yet even for collection of tens of thousands of reviews. The system is implemented by six steps: (1) Configuring with baseline frequency Data; (2) Calculating single-lemma statistics; (3) Calculating the probability of observed lemma occurrence counts; (4) Calculating the probability of observed bigram occurrence counts; (5) Filtering words inherent to the category; (6) Selecting the final features.

There are mainly two limitations in Scaffidi’s system. The system needs some supervised machine learning to offer quality feature-extraction, and cannot extract implicit features from reviews.
2.2 OPINION SENTENCE IDENTIFICATION

How to automatically extract and analyze reviews about a specific subject on the Web has become a hot research topic. Some existing related works are provided in this section. They are developed to analyze sentiments about open-domain queries in blog space and text document. Researchers focus on positive or negative measurement. Sentiments and different kinds of subjective information such as neutral opinions, requests, and judgments provide useful information.
2.2.1 Extracting Opinion Sentences Relevant to an Open-Domain Blog

Furuse etc. introduce a search engine used to extract opinion sentences relevant to an open-domain query from Japanese blog pages [10]. Based on not only positive or negative measurements but also neutral opinions, request, advice, and thoughts, the engine could identify opinions efficiently. To retrieve a number of opinion sentences that is reasonable and that a user could be expected to read, researchers attempted to extract only explicitly stated writer's opinions at the sentence-level without quoted or implicational opinions.

In this search engine, opinion sentences are identified by features such as opinion clue expressions, and the relevance is checked to the query of each identified opinion sentence. Opinion clue expressions, which are more restrictive than sentence-level subjectivity in conventional methods, are thought as a criterion for judging opinion sentences. By searching opinion sentences from web pages using these clue expressions, authors created a prototype opinion sentence search system in blog space.

The search engine implemented two tasks: (1) Extracting opinion sentences relevant to a user's query phrase about open-domain topics on products, persons, events, and social phenomena; (2) Identifying opinion sentences based on sentiments, neutral opinions, requests, advice, and thoughts. Comparing the output of the proposed opinion search engine with that of human judgments to whether the sentences were opinions, the experimental results show that the proposed engine satisfied the requirement as a practical application.
2.2.2 INTRODUCTION OF OPINIONFINDER

OpinionFinder [25] is a system that analyses documents and automatically identifies subjective sentences as well as various aspects of subjectivity within sentences, including agents who are sources of opinion, direct subjective expressions and speech events, and sentiment expressions. The goal of OpinionFinder is to develop a system capable of supporting other Natural Language Processing (NLP) applications by providing them with information about the subjectivity in documents. OpinionFinder operates as one large pipeline, which can be divided into two parts. The first part performs mostly general purpose document processing (e.g., tokenization and part-of-speech tagging). The second part performs the subjectivity analysis. The results of the subjectivity analysis are returned to the user in the form of SGML/XML markup of the original documents.

For general document processing, OpinionFinder firstly executes the Sundance partial parser to provide semantic class tags, identify Named Entities, and match extraction patterns that correspond to subjective language. Next, openNLP 1.1.0 is used to tokenize, sentence split, and part-of-speech tag the data, and the Abney stemmer2 was used to stem.

For subjectivity analysis, four components are combined: (1) Subjective Sentence Classification. (2) Speech Events and Direct Subjective Expression Classification. (3) Opinion Source Identification. (4) Sentiment Expression Classification.

OpinionFinder is not suitable for indentifying the opinion sentences in reviews, because identifying opinion sentences does not care about the sources of the opinion and the sentiment expressions are free styles to process in reviews. So we propose a new method to find opinion sentences using SentiWordNet.
2.3 CLASSIFYING THE SENTIMENT OF OPINIONS

Sentiment classification is an embranchment of the natural language processing. In general, sentiment classification means to analyze whether the author holds positive or negative sentiment to a specific subject [1].

Sentiment classification can be used in many applications. It can be used in business to summarize the feedback of their customers. For example, they can use it to classify the favorable or unfavorable reviews towards their products. It can also be used in text filtering. One can use it to recognize and discard unfavorable messages spreading through the internet.

Sentiment classification or semantic orientation classification [1, 2] is a way to automatically classify product reviews into two classes: recommended and not recommended, thus helping customers read them. This classification approach is usually used to classify a customer’s review in a whole to determine its class. However, in reality, a customer often expresses mixed feelings in one review by pointing out some aspects are excellent but others are not so satisfactory. In this case, it is not reasonable to make an overall classification on it.
2.3.1 DETERMINING THE SENTIMENT OF OPINIONS

It is a challenging problem to identify sentiments (the affective parts of opinions). Figure 2 presents a system architecture, which automatically extracted the people who hold opinions about given topic and the sentiment of each opinion [29]. There is a module for determining word sentiment and another for combining sentiments within a sentence in the system.

![Figure 2: A System Architecture](image)

Given a topic and a set of texts, the system operates in four steps as Figure 2: (1) Selecting sentences that contain both the topic phrase and holder candidates. (2) Delimiting the holder-based regions of opinion. (3) Calculating the polarity of all sentiment-bearing words by the sentence sentiment classifier individually. (4) Combining them to produce the holder’s sentiment for the whole sentence.

In [9], authors analyze the previous work in sentiment classification and propose sentiment orientation obtained through phrase patterns. To classify sentiment
orientation, researchers must get enough information from the text in order to infer the author’s sentiment. In [2], Tetsuya Nasukawa states the main tasks in sentiment classification as follows: (1) Finding the sentiment expressions in a text. (2) Distinguishing the polarity and strength of the expressions. (3) Finding the relationship between the expressions and the subject.

Wang etc. [5] proposed an approach to determine customers’ semantic orientations in product reviews at a smaller granularity level (i.e. sentence level). This sentence-level semantic classification (SLSC) approach employs a naïve bayes (NB) classifier, which is used widely in text classification tasks ([3, 4]), as its base classification model. It performs part-of-speech tagging to review sentences and uses certain types of words (adjectives, adverbs) as its features, which is different from common feature selection methods used in building naïve bayes classification models. During classification, heuristic rules are combined with the naïve byes classifier to predict semantic orientations of review sentences.
2.3.2 INTRODUCTION OF SentiWOrdNet

Opinion mining (OM), which is also known as “sentiment classification”, is concerned not with the topic a text is about, but with the opinion it expresses using information retrieval and computational linguistics. Opinion-driven content management has several important applications including determining critics’ opinions about a given product by classifying online product reviews, or tracking the shifting attitudes of the general public towards a political candidate by mining online forums or blogs. Within OM, several subtasks can be identified, all of them having to do with tagging a given text according to expressed opinion:

1. Determining text Subjective & Objective (SO) polarity such as deciding whether a given text has a factual nature (i.e. describes a given situation or event, without expressing a positive or a negative opinion on it) or expresses an opinion on its subject matter.

2. Determining text Positive & Negative (PN) polarity such as deciding if a given Subjective text expresses a Positive or a Negative opinion on its subject matter.

3. Determining the strength of text PN-polarity such as deciding e.g. whether the Positive opinion expressed by a text on its subject matter is Weakly Positive, Mildly Positive, or Strongly Positive.

The task of determining whether a term is indeed a marker of opinionated content has instead received much less attention. Note that in these works no distinction between different senses of a word is attempted, so that the term, and not its senses, is classified (although some such works distinguish between different POSs of a word).

The method used to develop SentiWordNet is an adaptation to synset classification for deciding the PN-polarity and SO-polarity of terms. The method relies on training a set of ternary classifiers, each of them capable of deciding whether a synset is Positive, or Negative, or Objective. Each ternary classifier differs from the other in the training set used to train it and in the learning device used to train it, thus producing different classification results of the WordNet synsets. Opinion-related scores for a synset are
determined by the (normalized) proportion of ternary classifiers that have assigned the corresponding label to it. If all the ternary classifiers agree in assigning the same label to a synset, that label will have the maximum score for that synset, otherwise each label will have a score proportional to the number of classifiers that have assigned it.
2.4 CONCLUSION OF RELATED WORK

Lots of people research on product features extraction, opinion search and sentiment classification. However there are few researches on integrate these three parts together, and no researcher mines product features from opinion sentences. Therefore, this thesis focuses on integrating these three parts to mine product feature by SentiWordNet-based algorithm. The motivation of this topic is very significant. Besides, our object is that we should accurately and efficiently extract the most product features which are making use of future work. The biggest difference between our approach and the previous researches is that we think about the three parts together to make a simple and practical implementation. The relation of opinion sentence and sentiment is obvious, so using SentiWordNet which is a tool of sentiment classification to search opinion sentences is reasonable and acceptable. At the same time, mining product features from opinion sentences has high performance and ideal result.
CHAPTER 3: THE PROPOSED METHOD

In this section, we present how to search product features from opinion sentences which are identified by SentiWordNet-based algorithm. In 3.1, we give the overview of our system and the output. Our system is mainly divided into four components: (1) Pre-processing and POS tagging; (2) Opinion sentence identification; (3) Feature Generation; (4) Feature Pruning. In 3.2, we describe the detail of POS tagging, which is common procedure of nature language processing. In this system, we apply the openNLP 1.4.3 to do the POS tagging. Next section 3.3, SentiWordNet-based algorithm is proposed to search opinion sentences. Product features are extracted from the opinion sentences in 3.4. After generating product features, the pruning process is used to find product feature accurately.
3.1 SYSTEM OVERVIEW

Figure 3 gives an architectural overview for our product feature extraction system. The system performs the extraction in two main steps: opinion sentence identification and product feature extraction. The inputs to the system are all the reviews of the product in the database. The output is the feature set of the reviews as Figure 4.

We download (or crawl) all the reviews, and put them in the review database. The feature extraction function, which is the focus in this thesis, firstly identify the opinion sentence that a lot of people have expressed their positive or negative opinions on, and then extract the product feature including explicit feature and implicit feature. Finally we filter the result via pruning irrelevant feature. Below, we discuss each of the functions in feature extraction in turn.
3.2 PART-OF-SPEECH TAGGING (POS)

POS tagging is the part-of-speech tagging [20] from natural language processing. Firstly, we give some sample sentences from some reviews to describe what kinds of opinions are handled, and the application of part-of-speech tagging from natural language processing will be discussed later.

For a given product, the aim of our system is to find what people like and dislike. Therefore it is an important step to find out the product features that people talk about. However, it is difficult to understand natural language and deal with some types of sentences [21]. The following are some easy and hard sentences from the reviews of a digital camera:

“The images are very vague.”

“Totally an undoubted very smart camera.”

In the first sentence, the user is satisfied with the picture quality of the camera, image is the feature that the user talks about. Similarly, the second sentence shows that camera is the feature that the user expresses one’s opinion. While the features of these two sentences are explicitly mentioned in the sentences, some features are implicit and hard to find. For example,

“While light, it will not easily fit in pockets.”

This customer is talking about the size of the camera, but the word “size” is not explicitly mentioned in the sentence. To find such implicit features, semantic understanding is needed, which requires more sophisticated techniques. Thus in this paper, we propose a simple method to find the implicit feature, called Feature Mapping using associative spread. For example, it is too heavy to carry. And we can imply the weight of this subject from the adjective word. The detail will be shown in Feature Generation section.
However, implicit features occur much less frequent than explicit ones. So we focus on finding features that appear explicitly as nouns or noun phrases in the reviews. To identify nouns/noun phrases from the reviews, we use the part-of-speech tagging.

In this work, we use the openNLP [24] linguistic parser, which to parse each sentence and yield the part-of-speech tag of each word (whether the word is a noun, verb, adjective, etc) and identifies simple noun and verb groups (syntactic chunking). The following shows a sentence with the POS tags.

```
I/PRP am/VBP absolutely/RB in/IN awe/NN of/IN this/DT camera/NN ./.
```

The openNLP system generates XML output. For instance, camera/NN indicates a noun. Each sentence is saved in the review database along with the POS tag information of each word in the sentence.

A transaction file is then created for the identification of frequent features in the next step. In this file, each line contains words from a sentence, which includes only preprocessed noun, verb, adjective, adverb phrases of the sentence. The reason is that other components of a sentence are unlikely to be used in expressing opinions. Here, pre-processing includes the deletion of stopwords, stemming (Stanford POS Tagging [17] and JWI [14] for WordNet [30]) and fuzzy matching. Fuzzy matching [13] is used to deal with word variants or misspellings. For example, “redeye” and “red-eye” actually refer to the same feature. All the occurrences of “redeye” are replaced with “red-eye”.

3.3 OPINION SENTENCES IDENTIFICATION

In opinion mining, users primarily care about what the customers like and dislike. So we only need to extract these sentences called Opinion sentences that people use to express a positive or negative opinion. Observing that people often express their opinions of a product feature in opinion sentences, we can extract opinion sentences from the review database using all the opinion words. In a review, there are many sentences. Some are opinion sentences, others are irrelevant sentences. For instance, let us look at the following two sentences:

“The performance of this camera is perfect.”

“I bought this camera yesterday.”

In the first sentence, the feature, performance, is in the opinion sentence. But in the second example, it’s not an opinion sentence, and no feature exists.

In order to find the opinion sentences, we use opinion words which are sentiment words that people used to express their positive or negative attitudes. Only four kinds of words can express the sentiment, they are nouns, adjectives, adverbs and verbs. Because we use nouns as product features, and few opinion sentences use nouns to express the sentiment. So in this thesis, we only think about adjectives, adverbs and verbs as opinion words. Sentiment includes three types: positivity, negativity and neutrality. Neutrality is usually used to describe a fact, and users pay much attention to the positivity and negativity. In review extraction, we take more care of the positivity and negativity of the opinion words. For evaluating the sentiment of each sentence, we make the quantization of opinion words to calculate the sentiment score of each sentence. The sentiment score of each opinion word is acquired from SentiWordNet.

SentiWordNet is a lexical resource for opinion mining [3]. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity. The assumption that underlies SentiWordNet’s switch from terms to synsets is that different senses of the same term may have different opinion-related properties. Each
of the three scores ranges from 0.0 to 1.0, and their sum is 1.0 for each synset. This means that a synset may have nonzero scores for all the three categories, which would indicate that the corresponding terms have, in the sense indicated by the synset, each of the three opinion-related properties only to a certain degree. For example, the synset [estimable(3)], corresponding to the sense “may be computed or estimated” of the adjective estimable, has an Obj score of 1.0 (and Pos and Neg scores of 0.0), while the synset [estimable(1)] corresponding to the sense “deserving of respect or high regard” has a Pos score of 0.75, a Neg score of 0.0 and an Obj score of 0.25. Following is the head part of SentiWordNet database:

<table>
<thead>
<tr>
<th>POS</th>
<th>offset</th>
<th>PosScore</th>
<th>NegScore</th>
<th>SynsetTerms</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1000003</td>
<td>0.0</td>
<td>0.125</td>
<td>form-only#a#1</td>
</tr>
<tr>
<td>a</td>
<td>1000159</td>
<td>0.25</td>
<td>0.0</td>
<td>dress#a#1</td>
</tr>
<tr>
<td>a</td>
<td>1000307</td>
<td>0.0</td>
<td>0.0</td>
<td>titular#a#5</td>
</tr>
<tr>
<td>a</td>
<td>1000440</td>
<td>0.0</td>
<td>0.0</td>
<td>prescribed#a#4</td>
</tr>
</tbody>
</table>

From this observation, we can extract opinion sentences in the following way:

For each sentence in the review database, if it satisfies that the positive or negative score is greater than certain score, to classify this sentence as opinion sentence. The positive and negative score can be calculated in below formulas.

Firstly, calculate the opinion word’s sentiment. In SentiWordNet, each opinion word has different sentiment scores in different scenes. It’s difficult to classify the opinion
word into the right scene. Here, we use the sentiment average of the opinion word in each scene as its final sentiment score in review.

\[
Score_{pos}(Word_{(v.,\text{adj.},\text{adv.})}) = \frac{1}{n} \sum_{i=1}^{n} Score_{pos}(i)
\]

(2)

\[
Score_{neg}(Word_{(v.,\text{adj.},\text{adv.})}) = \frac{1}{n} \sum_{i=1}^{n} Score_{neg}(i)
\]

(3)

After acquiring the sentiment of each opinion word, we need to get the adjectives, adverbs and verbs’ sentiment score in the sentence respectively. The opinion word in the same part of speech has the same sentiment weight in the sentence, so we think the sentiment average in each part of speech as the sentiment score of adjectives, adverbs and verbs.

\[
Score_{pos}(v.,\text{adj.},\text{adv.}) = \frac{1}{n} \sum_{i=1}^{n} Score_{pos}(Word(v.,\text{adj.},\text{adv.}))
\]

(4)

\[
Score_{neg}(v.,\text{adj.},\text{adv.}) = \frac{1}{n} \sum_{i=1}^{n} Score_{neg}(Word(v.,\text{adj.},\text{adv.}))
\]

(5)

There are three types of words, including verb, adjective and adverb. Calculating the positivity and negativity of each sentence respectively as follow:

\[
Score_{pos} = \frac{Score_{pos}(v.) + Score_{pos}(adj.) + Score_{pos}(adv.)}{3}
\]

(6)

\[
Score_{neg} = \frac{Score_{neg}(v.) + Score_{neg}(adj.) + Score_{neg}(adv.)}{3}
\]

(7)

It’s an opinion sentence, if \(\max(Score_{pos}, Score_{neg}) \geq \alpha\); It’s not an opinion sentence, if \(\max(Score_{pos}, Score_{neg}) < \alpha\). In our experiment, we find \(\alpha\) equals 0.08.

As shown in the previous example, perfect is an adjective that can be considered as the opinion word. So we can use opinion sentence identification to decide the sentence is an opinion or not. We explain the process in the follow example:

\[
\text{The/DT pictures/NNS are/VBP razor-sharp/JJ and/CC clear/JJ ./}
\]
In this sentence, we have three opinion words, including are, razor-sharp and clear. First we use stemmer to get each opinion word’s base form, and then calculate its sentiment score. Verb be is a stopword, so we omit it. Firstly, using formula (2) and (3), razor-sharp is 0.313 in positivity and 0.0 in negativity, and clear is 0.410 in positivity and 0.049 in negativity. Secondly, using formula (4) and (5), we get that the sentiment score of adjective is 0.362 in positivity and 0.025 in negativity. At last, under formula (6) and (7), the sentence sentiment score is 0.362 in positivity and 0.025 in negativity. Max(0.362, 0.025) > 0.08, so it is a opinion sentence.
3.4 FEATURES GENERATION

This step is to find features that people are most interested in. The features are classified into explicit and implicit. Explicit features are more easily extract than the implicit features, so we apply the associative spread (Feature Mapping) with WordNet to find the features. In order to obtain better features, we filter the features generated by above algorithm. Feature Pruning is presented to make a better result.
3.4.1 **Explicit Feature**

Explicit feature is the product features that appear in noun form in the reviews. Most of the features are expressed explicitly in the reviews. So we focus on extracting the majority of the explicit features which means finding most of the feature in the reviews. Instants below are the explicit features in the opinion sentence. In order to extract the explicit features, we do not use association rule mining [1]. We use probability-based algorithm to extract the product feature. Explicit features are divided into frequent and infrequent features. In the previous step, we have filtered the irrelevant sentences, so the infrequent feature also can have a high probability appear in the result of probability-based algorithm. We don’t need to extract the infrequent features independently, because the amount of the infrequent features which can not be found in our algorithm is very small.

In [21], Hu and Liu use Association rule mining to find all frequent itemsets, which are a set of words or a phrase that occur in one sentence. However, its computation is too large, and the result still needs compact pruning. We use a much rigorous method to finish this step. Each noun in the same sentence should appear nearby, and then the noun phrases can be considered as the candidates of product features.

*The picture is great, but the flash is terrible.*

*The battery life is not so good.*

*The quality of the picture is perfect.*

Here ‘picture’ and ‘flash’ will be misunderstood as one feature in [21]. We will distinguish the two nouns into two features, because they are not nearby. In sentence 2, ‘battery’ and ‘life’ is appearing nearby, so they are considered to organize one feature. In sentence 3, ‘quality’ and ‘picture’ separated by the preposition are also considered nearby and describe one feature.

For acquiring the better nouns used for product feature, we employ the nearby principle to extract the nouns are described by the adjective or verb. One adjective or verb only can be used for one object.
The picture is beautiful in this camera.

The lens and flash are very good.

In sentence 1, picture has the ‘beautiful’ to describe, but camera does not have any description. So we only extract picture as candidate feature. In sentence 2, it is a conjunction sentence, so lens and flash can be described by one adjective or verb. Lens and flash are two features.
3.4.2 Implicit Feature

The amount of implicit features is few in the reviews, and most of them use some specific adjectives to express implicit opinions. The adjective almost has clear meaning, which is one adjective mapped to one noun. So we can create a mapping database to obtain the implicit feature.

*It is too heavy to carry out.*

*It is too expensive.*

The feature ‘weight’ can be understood in Sentence 1, and the price also can be gotten in Sentence 2. So we map heavy to weight, and expensive to price. In order to understand more adjectives, we use WordNet 3.0 to extend the adjective set. The synonymset and antonymset of one adjective are used to describe the same feature. So if one adjective belongs to one synonymset or antonymset, the mapping noun is the feature expressed implicitly. If there are no explicit features in the opinion sentence, we use this method to find the implicit feature. In order to get implicit feature, we first stem the adjective, and then use JAWS [12] to get the synonymset and antonymset from WordNet. If the adjective belongs to one synonymset or antonymset, the corresponding feature will be extract as candidate feature.

After getting the candidate features, we can calculate their probability of occurring in the reviews to identify whether they are product features or not.

\[
\text{Probability} = \frac{\text{Occurrence}}{\text{Sentence}} \tag{8}
\]

*Occurrence: the occurrence number of the candidate feature in the reviews;*

*Sentence: the number of opinion sentences.*

The experiment shows the occurrence probability of the product feature is bigger than 0.6%, which is much smaller than 1% in [21].
3.4.3 Feature Pruning

The probability-based algorithm still can generate useless or fake features, which are some uninteresting and redundant ones. So feature pruning aims to remove these incorrect features. We use the pruning method proposed in [21].

Redundancy pruning focuses on removing redundant features that contain single words. Hu and Liu in [21] define that p-support (pure support) of feature ftr is the number of sentences that ftr appears in as a noun or noun phrase, and these sentences must contain no feature phrase that is a superset of ftr. For example, feature manual has the support of 10 sentences, which is a subset of feature phrases manual mode and manual setting in the review database. The support of the two feature phrases are 4 and 3 respectively, and the two phrases do not appear together in any sentence. Then the p-support of manual would be 3.

If a feature has a p-support lower than the minimum p-support and the feature is a subset of another feature phrase, it is pruned. Here, the minimum p-support is 3. In the previous example of manual with a p-support of 3, it is not pruned. Thus all the three features, manual, manual mode, manual setting, could be meaningful.
CHAPTER 4: EXPERIMENT

A set of experiments has been conducted to evaluate the effectiveness of our mining model based on SentiWordNet algorithm. The main objective of the experiments was to compare our method with Hu’s and OPINE, and test the performance of our system. So we adopt the same data as Hu’s and OPINE. In 4.1, we describe the test data, which is a classical review set. And then we provide the criterion to evaluate your system in 4.2. At last section, we discuss about the experiment result.

Our experiments are conducted in two stages: (a) Evaluate the result of opinion sentences identification. (b) Focus on the output of product features.
4.1 DATA

In order to do our experiments, we use the review data from http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip, which is provided by Liu [21, 22]. There are 5 products in the data package: (1) Digital camera - Canon G3; (2) Digital camera - Nikon coolpix 4300; (3) Cellular phone - Nokia 6610; (4) Mp3 player - Creative Labs Nomad Jukebox Zen Xtra 40GB; (5) Dvd player - Apex AD2600 Progressive-scan DVD player. All the reviews were from amazon.com. This data is used in [2, 21, 22]. All of these products also are pre-processed by delete the stopwords, stem and fuzzy match. Then we use the POS tagging to mark the reviews. After tagging the documents, we employ the SetiWordNet-based algorithm to identify opinion sentences. Finally, product features are generated with pruning.

Figure 6: An Example of Input Data

[51] poor quality = problems with dual-layer dvd's .
player has a problem with dual-layer dvd 's such as alien season 1 and season 2 .
for the money , get a better quality player .
player works and looks great - if you can get the dvd 's to play .
I know the saying is " you get what you pay for " but at this stage of game dvd players must have better quality than this - there is no excuse .
I will never purchase again .
customer service and technical support are overloaded and unresponsive - tells you about the quality of their products and their willingness to stand behind them .
I sent this player up , and never believe the reviews on a product right before christmas .
I have destroyed several of my dvs , andser .
for the first few weeks , this player was everything i expected it to be , an affordable multi-format dvd player with a stylish slim case as advertised .
when my dvs would stop playing in the middle , or not even be read at all ,
new dvs would allways begin skipping after a few plays .
i thought it was just the player , but then i started checking the discs to find that the apex 2600 is actually ruining my media .
there are funny little ridges on the dvd that look like the edge of a tree trunk !
this player is not worth my price and i recommend that you do n't purchase it .
i intend to contact the company , but from what i 've read ,
i just got screwed out of fifty bucks .
apex does n't answer the phone .
4.2 EVALUATION APPROACH

4.2.1 PRECISION AND RECALL

Precision and recall are two widely used statistical classifications.

In the context of classification tasks, the terms true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn) are used to compare the given classification of an item (the class label assigned to the item by a classifier) with the desired correct classification (the class the item actually belongs to). This is illustrated by the table below:

<table>
<thead>
<tr>
<th>Obtained Result</th>
<th>Obtained result / classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>tp (true positive)</td>
</tr>
<tr>
<td>E2</td>
<td>fn (false negative)</td>
</tr>
<tr>
<td>E1</td>
<td>fp (false positive)</td>
</tr>
<tr>
<td>E2</td>
<td>tn (true negative)</td>
</tr>
</tbody>
</table>

\[
\text{Precision} = \frac{tp}{tp + fp} \quad (9)
\]

\[
\text{Recall} = \frac{tp}{tp + fn} \quad (10)
\]

4.2.2 F-MEASURE

A measure that combines Precision and Recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

\[
F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)
\]
4.3 RESULT AND DISCUSSION

In our experiments we use sets of reviews for 5 product classes from Hu and Liu [21, 22]. There are five electronics products: 2 digital cameras, 1 DVD player, 1 mp3 player, and 1 cellular phone. Hu’s system and OPINE are the review mining system most relevant to our work. Hu’s system uses association rule mining to extract frequent review noun phrases as features. Frequent features are used to find potential opinion words (only adjectives) and the system uses WordNet synonyms/antonyms in conjunction with a set of seed words in order to find actual opinion words. Finally, opinion words are used to extract associated infrequent features. The system in [21] only extracts explicit features. OPINE is built on top of KnowItAll, a Web-based, domain-independent information extraction system. OPINE’s Feature Assessor uses PMI assessment to evaluate each candidate feature and incorporates Web PMI statistics in addition to review data in its assessment. Our system uses pre-processing in identifying the opinion sentences which can reduce the interference of irrelevant sentence. By addition, we use Linguistic pattern and probability-based algorithm to extract the feature instead of Associated Rule and PMI, this simple method gains a good result. In the following, we analyse the experiment’s result.

Table 2: Result of Opinion Sentences Identification

<table>
<thead>
<tr>
<th>Product Name</th>
<th>No. of manual Opinion Sentences</th>
<th>SentiWordNet Approach</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td>F-measure</td>
<td></td>
</tr>
<tr>
<td>Digital camera 1</td>
<td>466</td>
<td>0.903</td>
<td>0.931</td>
<td>0.917</td>
<td></td>
</tr>
<tr>
<td>Digital camera 2</td>
<td>274</td>
<td>0.923</td>
<td>0.966</td>
<td>0.944</td>
<td></td>
</tr>
<tr>
<td>Cellular Phone</td>
<td>398</td>
<td>0.915</td>
<td>0.922</td>
<td>0.918</td>
<td></td>
</tr>
<tr>
<td>Mp3 player</td>
<td>1219</td>
<td>0.882</td>
<td>0.904</td>
<td>0.893</td>
<td></td>
</tr>
<tr>
<td>DVD player</td>
<td>488</td>
<td>0.891</td>
<td>0.912</td>
<td>0.901</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>569</td>
<td>0.908</td>
<td>0.927</td>
<td>0.916</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows the recall and precision of SentiWordNet Approach are very high, the average of recall is 0.908 and the average of precision is 0.927, which means this approach is effective to identify the opinion sentences. The errors happen in the following cases:
This camera keeps on autofocussing in auto mode with a buzzing sound which cannot be stopped.

Get a "system error" problem 30 days after purchase.

Awesome!

In sentence 1, the opinion words (buzzing sound) which are nouns can not be marked as an opinion sentence in our method, but these situations are very limited. So we can neglect using nouns as opinion words. In sentence 2, the sentence does not have any opinion words, but it still expresses the opinion implicitly, and the feature also can be gotten from this sentence. This expression style is seldom in reviews. In sentence 3, it is considered as an opinion sentence, but there are no features. These situations do not interfere with the feature extraction.

From Table 3, we know that our method is better than Hu's either in precision or in recall. Compared with OPINE, the precision of our system is not so good, but the recall is much better. Overall our proposal will have better performance than OPINE.

In order to show that our system's performance is robust across multiple product classes, we used two sets of reviews downloaded from Epinions.com for Movie and Laptop. In Movie's reviews, we get 0.78 in precision and 0.89 in recall. In Laptop, we get 0.83 in precision and 0.91 in recall. So the average of the two products is 0.805 in precision and 0.90 in recall.
Table 3: Comparison of Hu’s, Opine’s and Our Result

<table>
<thead>
<tr>
<th>Product Name</th>
<th>No. of manual Features</th>
<th>Hu’s</th>
<th>OPINE</th>
<th>Opinion Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td>F-m</td>
</tr>
<tr>
<td>Digital camera 1</td>
<td>79</td>
<td>0.822</td>
<td>0.747</td>
<td>0.783</td>
</tr>
<tr>
<td>Digital camera 2</td>
<td>96</td>
<td>0.792</td>
<td>0.710</td>
<td>0.749</td>
</tr>
<tr>
<td>Cellular phone</td>
<td>67</td>
<td>0.761</td>
<td>0.718</td>
<td>0.739</td>
</tr>
<tr>
<td>Mp3 player</td>
<td>57</td>
<td>0.818</td>
<td>0.692</td>
<td>0.750</td>
</tr>
<tr>
<td>DVD player</td>
<td>49</td>
<td>0.797</td>
<td>0.743</td>
<td>0.769</td>
</tr>
<tr>
<td>Average</td>
<td>69</td>
<td>0.80</td>
<td>0.72</td>
<td>0.758</td>
</tr>
</tbody>
</table>

In summary, with the average of recall of 90% and the average precision of 79%, we believe that our techniques are quite promising, and can be used in practical application.

Finally, we give a demo of the system.

Figure 7: One Part Result of the System

<table>
<thead>
<tr>
<th>Paragraph Number: 839</th>
<th>Opinion Number: 477</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Number: 200</td>
<td>Sentence Number: 339</td>
</tr>
<tr>
<td>video frequency: 12</td>
<td>Ratio: 0.03539923</td>
</tr>
<tr>
<td>audio frequency: 3</td>
<td>Ratio: 0.000495575</td>
</tr>
<tr>
<td>player Frequency: 91</td>
<td>Ratio: 0.25643658</td>
</tr>
<tr>
<td>control Frequency: 9</td>
<td>Ratio: 0.016546672</td>
</tr>
<tr>
<td>button Frequency: 21</td>
<td>Ratio: 0.051946902</td>
</tr>
<tr>
<td>picture Frequency: 15</td>
<td>Ratio: 0.044247767</td>
</tr>
<tr>
<td>screen Frequency: 6</td>
<td>Ratio: 0.017699115</td>
</tr>
<tr>
<td>price Frequency: 18</td>
<td>Ratio: 0.053097345</td>
</tr>
<tr>
<td>feature Frequency: 25</td>
<td>Ratio: 0.073746316</td>
</tr>
<tr>
<td>model Frequency: 8</td>
<td>Ratio: 0.02359882</td>
</tr>
<tr>
<td>amazon.com Frequency: 3</td>
<td>Ratio: 0.0088495575</td>
</tr>
<tr>
<td>unit Frequency: 17</td>
<td>Ratio: 0.050147492</td>
</tr>
<tr>
<td>output Frequency: 9</td>
<td>Ratio: 0.023548672</td>
</tr>
<tr>
<td>cable Frequency: 4</td>
<td>Ratio: 0.01179941</td>
</tr>
<tr>
<td>sound Frequency: 4</td>
<td>Ratio: 0.01179941</td>
</tr>
<tr>
<td>movie Frequency: 12</td>
<td>Ratio: 0.0339823</td>
</tr>
<tr>
<td>size Frequency: 4</td>
<td>Ratio: 0.01179941</td>
</tr>
<tr>
<td>format Frequency: 4</td>
<td>Ratio: 0.01179941</td>
</tr>
</tbody>
</table>

In Figure 7, there are 839 paragraphs, 477 opinion sentences, 288 product features and 339 opinion sentences satisfying fixed pattern.
In this thesis, we propose a different method to extract the product features, which applies data mining and natural language processing methods. Comparing to previous works, we propose a new pre-process to identify the opinion sentence using SentiWordNet-based algorithm, which can get 0.927 in precision and 0.908 in recall. The simple extraction based on probability also can get a better result than Hu’s. Although OPINE has the best precision, our method improves the precision and recall at the same time. So our proposed techniques are effective in performing the feature extraction.

This thesis leaves many avenues for future work, and we plan to further improve these techniques. Some Limitations of this framework as following:

1. The product features which are appeared in non-opinion sentences cannot be extracted. However, the frequency of this situation is very small and the product features are little value for evaluating the product in current use. These features may be a hot topic in the future.

2. The sentiment of each word is identified by the average of each scene score in SentiWordNet. So the sentiment score is not so accurate to classify the opinion sentences.

3. In this thesis, opinion words are limited to adjective, adverb and verb. But other words such as nouns, word groups which also can express the sentiment are neglected.

4. Our algorithm cannot respond very well in case of the implicit product features mining by the associative spread (Feature Mapping).

In order to solve these limitations in the next stage, we plan to use following methods:
1. We should improve our algorithm to extract product features from non-opinion sentences. We can think it as a special case to propose a new method to solve this limitation.

2. We also collect the context of each word to analyze what’s the sentiment score of the word in the sentence. In order to identify more opinion words, we should provide a new pattern to calculate the sentiment score from all opinion words. So more natural language processing technologies should be employed in the next step.

3. Implicit product features is a difficult problem as everyone knows. And our main task is accurately and efficiently mining the product features as many as possible. We will attempt to create a better dictionary from WordNet to enhance the capability of associative spread (Feature Mapping).

In addition, we classify features according to the opinions’ strength that have been expressed, e.g., to sort the features from the strongest to the weakest. Feature extraction is the first step, and the summarization will be done in the future work.
BIBLIOGRAPHY


[12] JAWS, the Java API for WordNet Searching. 
http://lyle.smu.edu/cse/dbgroup/sw/jaws.htm


VITA

Weishu Hu, Weiss

University of Macau

2010

Academic Qualification

Bachelor of Computer Science & Technology in College of Information Science and Technology, Jinan University, China (2007)

Research Interest

Opinion mining and summarization
User Trust and Opinion
E-commerce
Web Development

Activity

2008 - 2009  Head of Propaganda Department of Postgraduate Association, University of Macau