CHAPTER 1

INTRODUCTION

This chapter introduces the purpose and details of the research. It begins with the discussion of aspect-based opinion mining background. Next, the chapter continues with the discussion of motivation, problem statement and objectives that triggers such work to be conducted. Then, the significance of research, scope as well as contribution of this study is described. This is followed by the research methodology and also the organization of the thesis.

1.1 Introduction

What others think always has been an important piece of information in everyday life. People need opinions to let them know about the advantages and disadvantages associated with products or services for making decisions. For example, if an individual wanted to buy a car, he or she would like to find out people's opinions about the features of the car such as color, price and seat type before purchasing it.

There are many ways to seek for opinions. Traditionally, individuals refer to friends, family, experts, catalogues and consumer reports to acquire opinions. On the other hand, organizations used to know consumer opinions expressed towards their product or services by conducting surveys or interviews. These manual ways of getting opinions are quite burdensome as it involves spending of huge amount of money and time.

With the proliferation of World Wide Web, the way people share and seek for opinions have changed. Both customers and organizations can get freely available opinions expressed by the people all around the world. Sites such as blogs, social

networks, review sites, forums and e-commerce websites provides better understanding of customer's perceptions towards products and services.

Unfortunately, the opinionated texts which are meant to be helpful, become overwhelming with the passage of time due to the frequency of posts by users. Hence, an automatic way to mine the unstructured contents in text is needed to get a relevant piece of information. As a solution to this problem, opinion mining is introduced by researchers.

Opinion mining or sentiment analysis is a field of study that analyzes people's opinions towards an item (e.g. product) or issue (e.g. service) (Liu, 2012). Aspect based opinion mining relatively give more detailed insights towards a product or service. This is because; a typical customer review will have the description about aspect of product or service as well as general feedback about a product or service. The emotion or view expressed in the reviews towards aspect is known as sentiment.

Diverse techniques have been utilized by researchers since the last decade to implement aspect based opinion mining such as syntactic techniques (Hu & Liu, 2004; Popescu & Etzioni, 2007; Zhao & Li, 2009) and semantic techniques (Zhao & Li, 2009; Penalver-martinez et al., 2014) in extraction of aspect and sentiments in customer reviews.

In this thesis, a syntactic oriented aspect-based opinion mining is used to jointly extract aspect and its corresponding sentiment by utilizing aspect-sentiment pair extraction rules and compound noun lexicons.

1.2 Motivation

Opinion helps customers and organizations gain some insight about the quality of product or service. However, if review sentences are present in a vast amount, they

can be burdensome for humans to analyze opinion in every sentence on the review site to make the final decision upon purchase of a product or service.

Even though the reviews comprised of vast volumes of sentences, but not all the sentences express opinion towards products or services. Therefore, extracting terms that represent opinion in sentences would be useful. The important terms that present in opinionated sentences are aspect and its corresponding sentiment.

Hence, to represent the opinion of the product or service from review sentence, aspect-sentiment pair extraction is crucial. The aspect-sentiment pairs extracted can be utilized to represent a decomposed view of opinion in review sentences as shown in Figure 1.1.
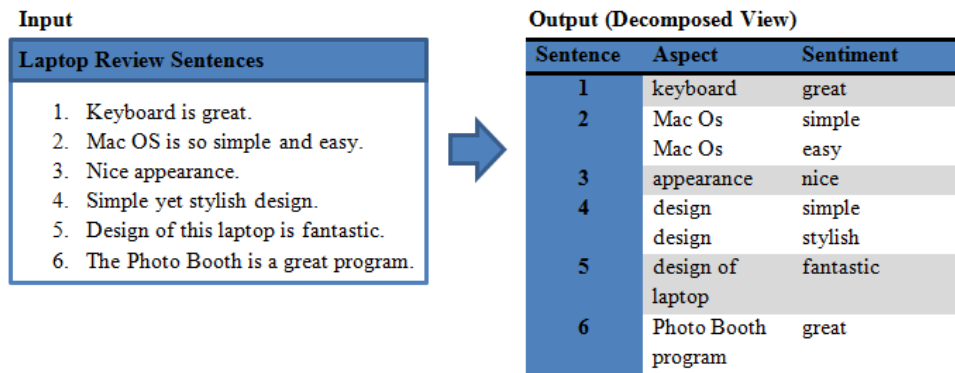
Input

Laptop Review Sentences

1. Keyboard is great.
2. Mac OS is so simple and easy.
3. Nice appearance.
4. Simple yet stylish design.
5. Design of this laptop is fantastic.
6. The Photo Booth is a great program.

Output (Decomposed View)

Sentence	Aspect	Sentiment
1	keyboard	great
2	Mac Os	simple
	Mac Os	easy
3	appearance	nice
4	design	simple
	design	stylish
5	design of laptop	fantastic
6	Photo Booth program	great

Figure 1.1: Decomposed view of opinion in review sentences

1.3 Problem Statement

Let $S = \{S_1, S_2, ..., S_k\}$ be lists of sentences in product or service reviews, where k, is the order of sentences in the review. In each sentence, S_k, there are aspect-sentiment pairs, $P = \{P_1, P_2, ..., P_n\}$, where, n represents the order of aspect-sentiment pairs corresponding to its sentence. For example, in S_1 = "The screen is good", the P_1 = (screen, good). The goal of this work is to improve the extraction of the aspect-sentiment pairs, P_n in product or service review sentences, S_k. The key issue in the

aspect-sentiment pair extraction is to identify aspect-sentiment pairs without resulting in a noticeable loss in aspect and sentiment terms in a sentence.

Aspect-sentiment pair extraction is quite a hard task as it involves processing the sentences written by reviewers in different ways. Hence, there is no predefined solution for this problem. Most of the previous works have extracted aspect-sentiment pair using syntactic and semantic based approaches. Recently, syntactic based approaches are used more frequently, such as the dependency based (Moghaddam & Ester, 2012; Ravi & Raghuveer, 2012; Chinsha & Joseph, 2013; Bancken et al., 2014) and rules based methods (Bross & Ehrig, 2013; Veselovská & Tamchyna, 2014) are able to identify aspect-sentiment pairs. However, those works still have limitation in identifying multi-word aspects that consists of parent-child relationship and non-noun-noun compound noun.

Most of the important aspects are missed in the existing works, due to parsing error in the identification of non-noun-noun compound noun aspects. The parsing error causes misidentification of the aspects as sentiments. For example, in the compound noun "hard disk", "hard" is identified as sentiment and "disk" is identified as an aspect. The parent-child aspects are also missed because of inadequate dependency relations and generalized rules are unable to relate the parent-child aspect to their corresponding sentiment. For example, in the sentence, "The screen of laptop is big", the aspect, "Screen of laptop" is not extracted as whole, only screen (child) or laptop (parent) is extracted.

Besides that, some of the works (Bancken et al., 2014; Veselovská & Tamchyna, 2014) do not consider different orders of aspect-sentiment arrangement in the extraction, such as one aspect and one sentiment, one aspect and many sentiments,

many aspects and one sentiment, one sentiment and one aspect, one sentiment and many aspects as well as many sentiments and one aspect. The mapping of aspect to its sentiment is important to produce correct aspect-sentiment pairs. For instance, in many sentiments and one aspect arrangement, one aspect should be assigned to both the sentiments. For example, in the sentence "Excellent and quick service", the aspect-sentiment pairs are (service, excellent) and (service, quick).

To overcome these drawbacks, the aspect-sentiment pair extraction model with rules and lexicons is proposed to improve aspect-sentiment pair extraction. Different types of aspects such as compound nouns, non-noun-noun compound nouns and parent-child aspects are identified using rules and compound noun lexicons. Besides that, the order of aspect-sentiment pairs arrangement in sentences are captured in the rules. Then the templates associated with the rules are used to map the aspects to their sentiments. With regard to his, more aspects can be discovered. Identification of more aspects eventually results in discovery of more sentiments which finally results in more aspect-sentiment pairs.

1.4 Objectives

In this thesis, we propose aspect-sentiment pair extraction model for extraction of aspect-sentiment pairs using rules and compound noun lexicons. With respect to this, our objectives are:

1) To determine the optimized number of review sentences to train aspect-sentiment pair extraction model.

2) To improve the aspect-sentiment pairs extraction via enhancement of rules and compound noun lexicons.

1. 5 Significance of Research

The decomposed view that consists of aspect-sentiment pairs of customer's review sentences can be a beneficial utilization for consumers, organizations and also for opinion mining systems. Consumer are no more bound to read a lot of reviews and can save time when making decisions before proceeding to purchase products or services. Similarly, product manufacturers can avoid trivial job of processing and analyzing review sentences to obtain customer's feedback. In addition, this decomposed view can also be applied as an input to opinion mining systems such as decision making and recommendation system.

1.6 Scope

The scope of the proposed approach is to apply aspect-based opinion mining for customer reviews such as product or service reviews. The key task that is considered is to extract aspect and its corresponding sentiment in English language product reviews. The aspect-sentiment pair extraction needs both aspect and sentiment to be present in the sentence. Aspect without sentiment or sentiment without aspect extractions is beyond the scope of this work. Besides that, the approach is limited to extract explicit aspect, not the implicit aspect. For example, "Laptop was very expensive". Here, the purpose is not meant to identify the aspect "price".

The review sentences dataset used consists of minor grammatical mistakes in terms of punctuation (",", "!", "." and etc.) and spelling errors ("bateery"). The sentences with sarcasm, improper sentence structure and informal way of writing reviews (colloquial language) are not considered in this work. Other than that, only the task of extracting aspect and its sentiment is considered, assigning scores or calculation of aspect and sentiment polarity or intensity is not included. The compound noun lexicons generation is focused on two-worded compound noun

word. Furthermore, this work only considers short sentences within the range of two to fifteen.

1.7 Research Contributions

A lot of researches have been done to this date for the extraction of aspect and sentiment pair from reviews. In the existing work, the task is focused on aspect and sentiment extraction without considering aspect detection improvement. Many solutions have been studied to extract aspect sentiment pairs, however, less consideration is given to the detection of aspect sentiment pair types such as parent-child aspect. The identification of non-noun-noun compound noun as aspects are also not taken into account.

Thus, the contribution of this work is the aspect-sentiment pair extraction with rules and compound noun lexicons (ASPERC) model for customer reviews e.g. products and services reviews to discover different types of aspects such as single-word aspect (noun) and multi-word aspect (noun phrase). With enhanced rules for the detection of compound noun and parent-child aspects, more aspects can be discovered. To cover the non-noun-noun compound noun, the compound noun lexicons are generated in accordance with related domain. These lexicons can be referred if the POS tagger cannot identify the non-noun-noun compound noun.

Besides that, all order of arrangement of aspects and the sentiment in sentences for aspect-sentiment pair extraction, such as one aspect and one sentiment, one aspect and many sentiments, many aspects and one sentiment, one sentiment and one aspect, one sentiment and many aspects as well as many sentiments and one aspect are taken into account. Additionally, the parent-child aspect sentiment pair in many

aspects and one sentiment type also can be identified by using the rules. The aspects are mapped to sentiments with the aid of templates created by the rules.

Hence, more aspect-sentiment pairs are discovered by enhancing the aspect-sentiment rules as well as utilizing compound noun lexicons. Our work recall is higher than the baseline, indicating that the coverage to identify the aspect-sentiment pair is better.

1.8 Research Methodology

This section discusses the flow when conducting this research. Overall, five main phases such as planning, analysis, design, implementation and evaluation as shown in Figure 1.2 are followed. The discussion of the phases will be elaborated in the next section.
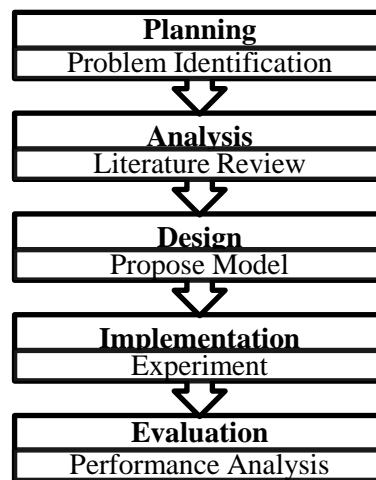
```
┌─────────────────────────────┐
│         Planning            │
│    Problem Identification    │
└─────────────────────────────┘
              ⇩
┌─────────────────────────────┐
│         Analysis            │
│      Literature Review      │
└─────────────────────────────┘
              ⇩
┌─────────────────────────────┐
│          Design             │
│       Propose Model         │
└─────────────────────────────┘
              ⇩
┌─────────────────────────────┐
│       Implementation        │
│         Experiment          │
└─────────────────────────────┘
              ⇩
┌─────────────────────────────┐
│         Evaluation          │
│    Performance Analysis      │
└─────────────────────────────┘
```

Figure 1.2: Research workflow

Planning

Aspect-sentiment pair extraction is important to represent the opinion from review sentence. The opinion can be leveraged towards customers and manufacturers to get feedback about a product or services for decision making. The problem in aspect-sentiment pair extraction that needs a solution is in terms of aspect identification. By identifying different types of aspects, eventually, the sentiment associated with the

8

aspects also can be identified. Hence, aspect-sentiment pair extraction can be improved by discovering more aspect-sentiment pairs. With more aspect-sentiment pair extraction, better information regarding a product or service can be acquired.

Analysis

The literature review on aspect-based opinion mining techniques, terminologies, experiments as well as the performance metrics is studied. The task in aspect-based opinion mining such as aspect-sentiment pair extraction is selected as the focus of this work as it provides useful information to society. This study is also a demanding research in since last decade as many companies are utilizing the opinions from reviews to get free opinions about their product or services. The gap of aspect-sentiment pair extraction in existing work is found as a motivation of doing our work.

Design

An aspect-sentiment pair extraction using rules and compound noun lexicons (ASPERC) model is proposed to extract aspect-sentiment pairs from the customer reviews such as product or service review sentences. This model is enhanced with rules and compound noun lexicons to identify more aspect-sentiment pairs.

Implementation

The model is implemented using benchmarked dataset that is widely used in the previous aspect-sentiment pair extraction works. The implementation involves two main findings in our model such as aspect-sentiment pair extraction rules and compound noun lexicons. Several steps are carried out to produce compound noun lexicons, generic rules, aspect-sentiment pair rules, aspect-sentiment pair templates as well as aspect-sentiment pair extractions.

Evaluation

The proposed model is evaluated against baseline with the golden standard aspect-sentiment pair annotations. The performance measure such as precision and recall are used to evaluate the ability of the model in extraction of aspect-sentiment pair.

1.9 Organization of Thesis

This thesis organization is as follows: Chapter 2 describes the literature review and research gap. Chapter 3 presents the aspect-sentiment pair extraction based on rules and compound noun lexicons (ASPERC) model. Chapter 4 gives insights of evaluation of proposed work with baseline. Chapter 5 discusses conclusion and future work.