19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

Mining Associative and Comparative Patterns for Thai Sentiment Analysis

Choochart Haruechaiyasak and Alisa Kongthon
Speech and Audio Technology Laboratory (SPT)
National Electronics and Computer Technology Center (NECTEC)
National Science and Technology Development Agency (NSTDA)
Thailand Science Park, Klong Luang, Pathumthani 12120, Thailand
Email: choochart.har@nectec.or.th, alisa.kon@nectec.or.th

***Abstract*—The aspect-based sentiment analysis has been popularly applied for analyzing product reviews. The results from the analysis could help summarize the customer satisfaction towards the products. Previous aspect-based sentiment analysis only focuses on relating sentiment polarity with product aspect. In a competitive market, it is more important to gain some insight of the sentiment and aspect towards a product brand and in comparison with other brands. To enhance the existing aspect-based sentiment analysis, we propose the following extensions, (1) association of sentiment and aspect with the product brand, and (2) comparison between two product brands. Our proposed approach is based on the pattern analysis of simplified patterns with dynamic filler terms. The simplified patterns help increase the effectiveness of pattern extraction. To evaluate the proposed approach, we performed experiments using product reviews in the smartphone domain. The results show that the performance of product brand and aspect extraction is significantly improved with the simplified patterns for both associative and comparative pattern minings.**

Keywords: Sentiment analysis, pattern mining, product reviews, Thai language.

I. INTRODUCTION

Today with a large population on social networking websites, the amount of social media contents has increased exponentially over the years. A majority of social media contents includes posts and comments which usually contain sentiments or opinions from the users. In many businesses, social media has been adopted as an active communication channel between companies and customers. Many companies regularly use social networking websites to promote new products and services, and post announcements to their customers. On the other hand, customers often post comments to express their sentiments towards products and services. Due to the real-time nature of the social media, monitoring customers' comments has become a critical task in customer relation management (CRM). Sentiment analysis has received much attention among market research community as an effective approach for analyzing social media contents.

The research in opinion mining and sentiment has gained a lot of interest in text mining and NLP communities [15]. Previous works in this area focused on analyzing reviews as being positive or negative either at the document level [3], [12] or sentence level [16], [17]. For instance, given some reviews of a product, the system classifies them into positive or negative reviews. No specific details or features are identified about what customers like or dislike. To obtain such details, the feature-based opinion mining or aspect-based sentiment analysis approach has been proposed [5], [8], [13], [14]. This approach consists of two steps as follows.

1) Identifying and extracting aspects (or features) of an object from a sentence upon which a reviewer expressed his/her opinion.
2) Determining whether the sentiment regarding the extracted features are either positive or negative.

The aspect-based sentiment analysis could provide users with summarized sentiment on a product. For example, for smartphone reviews, the aspect-based sentiment analysis allows users to view positive or negative sentiments on smartphone aspects such as *price*, *design*, *battery life*, *camera* and *screen size*. Breaking down reviews into aspect level is very essential for making a purchase decision. Different customers have different preferences on smartphone specifications. For example, some might prefer smartphones with high quality photo taking, however, some might prefer nice design.

Previously proposed aspect-based sentiment analysis approaches only focus on relating sentiment polarity with aspect. To enhance the existing aspect-based sentiment analysis, we propose the following extensions.

1) Association of sentiment and aspect with the product brand, and
2) Comparison between two product brands.

The analysis results from the associative mining could help relate a product brand with sentiment and aspect in a sentence. For example, a sentence "iPhone has many great applications", the associative mining could identify "iPhone" as having a positive sentiment on the "application" aspect. The analysis results from the comparative mining could help compare a product brand with others in terms of sentiment and aspect. For example, a sentence "iPhone has better camera than Samsung", previous aspect-based sentiment analysis could only identify positive sentiment on the "camera" aspect without relation to product brands. Using comparative mining, the results could identify "iPhone" as containing a positive sentiment and "Samsung" as receiving a negative sentiment.

The above two analysis extensions are based on the technique of mining associative and comparative patterns. To increase the effectiveness of the pattern analysis, we propose an approach of simplified patterns with dynamic filler terms. The simplified patterns help increase the effectiveness of pattern extraction. To construct the patterns, we design two lexicon types: *domain-dependent* and *domain-independent*. The domain-dependent lexicons include entities (i.e., product brands and models), aspects (i.e., product features) and polar words (both positive and negative sentiments). The domain-independent lexicons are negators (e.g., "not", "never", "unlikely") and comparative prepositions (i.e., "than", "more than", "less than"). Using these lexicons, we could construct a set of associative and comparative patterns from the tagged corpus. We evaluated the proposed framework on the domain of smartphone reviews. The experimental results showed that our proposed approach is very effective in extracting entities and aspects from unseen input texts.

The remainder of this paper is organized as follows. In next section, we review some related works on different approaches for sentiment analysis. In Section 3, we present the proposed approach for mining associative and comparative patterns in aspect-based sentiment analysis. In Section 4, we perform experiments by using a corpus of smartphone product reviews. The evaluation results will be presented with some discussion. Section 5 concludes the paper with the future work.

II. RELATED WORK

The research in opinion mining and sentiment analysis has gained a lot of interest in text mining and NLP communities [15]. Much work in this area focused on evaluating reviews as being positive or negative either at the document level [1], [12] or sentence level [17]. For instance, given some reviews of a product, the system classifies them into positive or negative reviews. No specific details or features are identified about what customers like or dislike. To obtain such details, a *feature-based* opinion mining approach has been proposed [5].

The problem of developing subjectivity lexicons for training and testing sentiment classifiers has recently attracted some attention. Although most of the reference corpora has been focused on English language, work on other languages is growing as well. Ku and Chen [9] proposed the bag-of-characters approach to determine sentiment words in Chinese. This approach calculates the observation probabilities of characters from a set of seed sentiment words first, then dynamically expands the set and adjusts their probabilities. Later in 2009, Ku et al. [10], extended their bag-of-characters approach by including morphological structures and syntactic structures between sentence segment. Their experiments showed better performance of word polarity detection and opinion sentence extraction. Haruechaiyasak et al. [4], proposed a framework for constructing Thai language resource for feature-based opinion mining. The proposed approach for extracting features and polar words is based on syntactic pattern analysis.

Our work is related to previous works in mining comparative sentences. Most previous works focus on English language. Jindal and Liu [6], [7] proposed an approach for categorizing comparative sentences into different types. Their approach integrated pattern discovery and supervised learning approach to identifying comparative sentences from text documents. For other languages, Liu et. al. [11] proposed a method for analyzing comparative sentences from Chinese text documents by combining rule-based methods and statistical methods. The method performed the broad extraction results by using comparative words, sentence structure templates and dependency relation analysis. Yang and Ko [18] proposed an approach for extracting comparative sentences from Korean text documents. Their proposed approach is based on a collection of comparative keywords for searching comparative-sentence candidates.

Our main contribution in this paper is to enhance the existing aspect-based sentiment analysis in Thai language. The proposed approach can analyze associative and comparative patterns of sentiment and aspect with the product brands. The results from the analysis help increase the effectiveness of the aspect-based sentiment analysis.

III. THE PROPOSED APPROACH

The performance of the aspect-based sentiment analysis relies on the design and completeness of related lexicons. Previous works in aspect-based sentiment analysis considered only aspect and polar terms. However, to perform an in-depth associative and comparative analysis of the product brands, our proposed approach includes the product brands and models as entities. Our designed lexicon consists of two types, domain-dependent and domain-independent.

In this paper, we illustrate the proposed approach through an example of smartphone domain. The domain-dependent lexicon consists of four categories as follows.

- **Entity (ENT)**: The entities are terms describing brands and/or models of the products.
- **Aspect (ASP)**: Aspect refers to the features of the product in the specific domain. Examples of aspects in smartphone domain are "price", "screen size" and "camera".
- **Positive polar words (POS)**: Positive polar words are sentiment words which represent *positive* views on features.
- **Negative polar words (NEG)**: Negative polar words are sentiment words which represent *negative* views on features.

It is worth noting that although some polar words are domain-independent and have explicit meanings such as "beautiful", "modern", "expensive" and "terrible". Some polar terms are domain-dependent and have implicit meanings depending on the contexts. For example, the word "large" is generally considered positive for the *screen size* aspect of smartphone domain. For other domains, however, the word "large" could be considered as negative.

The domain-independent lexicon consists of general words which have different parts of speech (POS) in a sentence. For sentiment analysis task, we design six different domain-independent lexicons as follows (some examples are shown in Table I).

- **Negator (NGT)**: Like English, these words are used to invert the sentiment polarity. Examples are "not", "unlikely" and "never".
- **"than" (THA)**: The word "than" acts as a preposition for comparison sentence. In Thai language, the word "than" is followed by either positive and negative polar words when a product brand is compared to another.
- **"more than" (MTH)**: The word "more than" is a special case of "than". When a positive polar word is followed by "more than", it yields positive sentiment for the preceding entity. For example, a phrase "*A* beautiful more than *B*" has a positive sentiment on *A* and a negative sentiment on *B*. The same logic applies for a negative polar word.
- **"less than" (LTH)**: The word "less than" is a special case of "than". When a positive polar word is followed by "less than", it yields negative sentiment for the preceding entity. For example, a phrase "*A* beautiful less than *B*" has a negative sentiment on *A* and a positive sentiment on *B*. However, when a negative polar word is followed by "less than", it yields positive sentiment for the preceding entity.

TABLE I. DOMAIN-INDEPENDENT LEXICONS

Tag	Example
Negator (NGT)	ไม่ (not), ไม่ค่อย (unlikely), ไม่เคย (never), ไม่ค่อย (not really), ไม่ควร (should not), ไม่น่าจะ (probably not)
"than" (THA)	กว่า (than)
"more than" (MTH)	มากกว่า (more than)
"less than" (LTH)	น้อยกว่า (less than)

Figure 1 shows the processes and work flow under the proposed approach. The process starts with a tagged corpus (C_T) which is annotated according to the domain-dependent and domain-independent tag sets. From the tagged corpus, we construct associative and comparative patterns (P_T) and lexicons (L_T). The lexicon construction is performed by simply collecting words which are already tagged with the lexicon types. The pattern construction is performed by generalizing the repeated patterns. The pattern construction process consists of the following steps.

1) **Text tokenization**: The first step is to tokenize a given input text into a series of word tokens. Thai language is considered as an unsegmented language in which words are written continuously without the use of word delimiters. In this paper, we apply a dictionary-based word segmentation program to tokenize texts.
2) **Tag assignment**: For each word token, we assign tags based on the designed domain-independent and domain-dependent lexicons. Words which are not found in these lexicons are assigned with the tag <term>, which is considered as a term filler.
3) **Pattern simplification**: The last step is to simplify the patterns. The patterns with fixed number of <term> tags are very strict, i.e., it would be difficult to match other unseen texts. To increase the effectiveness of the pattern matching, we propose the concept of dynamic filter terms. Any <term> could be expanded based on the regular expression technique and set by filter threshold parameter, *Filler_TH*.

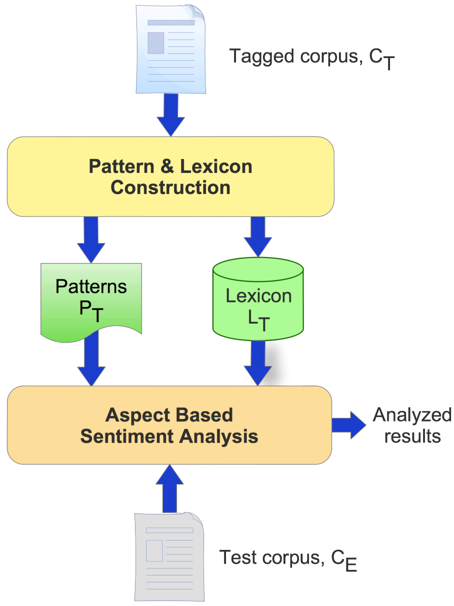

Fig. 1. The process for associative and comparative pattern mining for sentiment analysis

Table II gives an example of associative pattern mining process. An input text is tokenized into word tokens. It can be observed from the example that in Thai written language, many times the writers omit the subject (such as "I") in the sentence. Next, the word token is assigned an appropriate tag. To assign a tag, the word is looked up in the lexicons, L_T, which were previously constructed. Words not found in the constructed lexicons are assigned with general <term> tags. For example, the associative pattern "<ENT><term>*<ASP><POS>", would match any sentences which have the tag sequence as shown. The tag <term>* represents zero or more word tokens, therefore, allowing flexible pattern matching on many unseen sentences. In the experiment section, we will perform an empirical analysis to optimize this parameter. Table III gives an example of comparative pattern mining process. The process is similar to the associative pattern mining.

IV. EXPERIMENTS AND RESULTS

To evaluate our proposed approach, we performed experiments in the domain of smartphones. The data set was collected from many smartphone review websites in which users provide comments and reviews on various smartphone's brands and models. One of the major source of reviews is webboards, e.g., *Pantip* [19]. The training corpus consisting of approximately one thousand reviews are randomly selected from the collected data set. The corpus was manually tagged by seven annotators using the tag set as described in previous section.

TABLE II. EXAMPLE OF ASSOCIATIVE PATTERN MINING PROCESS

Original text:	ซื้อ iPhone เพราะดีไซน์สวยหรู (I) bought iPhone because design (is) elegant.				
Tokenized text:	ซื้อ bought	iPhone iPhone	เพราะ because	ดีไซน์ design	สวยหรู elegant
Assigned tags:	term	**ENT-P**	term	**ASP**	**POS**
Simplified pattern:	**<ENT-P>** <term>* **<ASP> <POS>**				

TABLE III. EXAMPLE OF COMPARATIVE PATTERN MINING PROCESS

Original text:	S4 แบตน่าจะอึดกว่า iPhone 5 S4 battery probably lasts longer than iPhone 5.					
Tokenized text:	S4 S4	แบต battery	น่าจะ probably	อึด lasts longer	กว่า than	iPhone 5 iPhone 5
Assigned tags:	**ENT-P**	**ASP**	term	**POS**	**THA**	**ENT-N**
Simplified pattern:	**<ENT-P> <ASP>** <term>* **<POS> <THA> <ENT-N>**					

From the corpus, the total number of tagged entities (ENT), i.e., brands and models, is equal to *545*. The total number of tagged aspects (ASP), including their synonyms, is equal to *278*. The number of tagged positive (POS) and negative (NEG) polar words are *609* and *837*, respectively. Table IV gives some examples of lexicon obtained from the tagged corpus. It can be observed that entities contain variation of terms describing brands or brands with their models, in both Thai and English. Sometimes abbreviations are used to denote a brand and its model, for example, "ip4s" refers to "iPhone 4S". Aspect terms describe features or specifications of smartphones in general, such as, *design*, *screen* and *camera*. Polar terms describe positive and negative sentiments of the smartphone such as *strong*, *elegant*, *expensive*, and *too small*.

Table V and VI give some examples of most frequently occurred associative and comparative patterns constructed from the corpus. The total number of associative patterns is equal to *25*. The total number of comparative patterns is equal to *31*. The tag set symbols are as previously described in Table I and IV with the tag <term> denoting any other general terms. For associative mining, patterns which occur frequently are an entity followed by a polar term, e.g., <ENT-P><POS>, or an entity followed by an aspect term and a polar term, e.g., <ENT-P><ASP><POS>. These patterns show that Thai opinionated texts are mostly very simple. However, Thai written language have high flexibility in its grammars. It is interesting to observe that an aspect term can appear before or after the entity (e.g., pattern no. 2 and 4 in Table V). Also, the role of negators (NGT) is similar to many other language including English. When a negator appears in front of a polar term, it will inverse the polarity of the term, i.e., from positive to negative and vice versa. In our proposed approach, the appearance of a negator in from of a polar term will have an effect of inverting the entity as well, i.e., from ENT-P to ENT-N and vice versa.

TABLE IV. EXAMPLES OF LEXICON IN SMARTPHONE DOMAIN

Lexicon	Example
Entity (ENT)	lumia, zenfone, iphone, xiaomi Mi3, z1, LG G3, oppo, ip4s, ipad, โซนี่ (Sony), Huawei, Galaxy S5, lenovo, xperia, ไอโฟน (iphone), ออปโป (Oppo), ซัมซุง (Samsung)
Aspect (ASP)	ถ่ายรูป (photo taking), เสียง (sound), ดีไซน์ (design), วัสดุ (material), กล้อง (camera), สเปิ๊ค (specification), ฟังก์ชั่น (function), บอดี้ (body), หน้าจอ (screen), น้ำหนัก (weight), ราคาขายต่อ (second hand price), แอพ (application)
Positive Polar words (POS)	ละเอียด (fine-grained), รวดเร็ว (fast), แข็งแรง (strong), ดูคลาสสิค (classic looking), สวยงาม (beautiful), ชัด (vivid), แรง (powerful), สว่าง (bright), หรูหรา (elegant), เบา (light)
Negative polar words (NEG)	ค้าง (hanged), หน่วง (delayed), ห่วย (terrible), มีปัญหา (having problem), แพง (expensive), ใหญ่ไป (too large), เล็กไป (too small), ช้า (slow), ซด (consuming power)

For comparative mining, the most patterns are then an entity is followed by a polar term, the comparison preposition "than", and another entity in opposite sentiment, e.g., <ENT-P><POS><THA><ENT-N>. Sometimes, writers also include an aspect term in the comparative pattern. Similar to the associative pattern, an aspect term can either appear in front of a polar term (e.g., pattern no. 1 in Table VI) or in front of the first entity (e.g., pattern no. 4 in Table VI). We also observed a special case of comparative pattern in which the writers do not mention two entities, ENT-P and ENT-N in the same sentence. Instead of a complete comparative sentence, e.g., "A is better than B", this sentence is simplified to "A is better."

TABLE V. EXAMPLES OF ASSOCIATIVE PATTERNS

No.	Example
1	**<ENT-P><ASP><POS>** <ไอโฟน><หน้าจอ><แจ่ม> <iPhone><screen><vivid>
2	**<ENT-P><POS>** <ซัมซุง><ดีเลิศ> <Samsung><excellent>
3	**<ENT-P><ASP><NGT><NEG>** <zenphone><ราคา><ไม่><แพง> <Zenphone><price><not><expensive>
4	**<ASP><ENT-N><NEG>** <หน้าจอ><iphone><เป็นรอยง่าย> <screen><iphone><easily scratched>
5	**<ENT-P><ASP><term><ASP><POS>** <S5><แบต><กับ><หน้าจอ><ชนะเลิศ> <S5><battery><and><screen><be the winner>

TABLE VI. EXAMPLES OF COMPARATIVE PATTERNS

No.	Example
1	**<ENT-P><ASP><POS><THA><ENT-N>** <ไอโฟน><หน้าจอ><แจ่ม><กว่า><ซัมซุง> <iPhone><screen><vivid><than><Samsung>
2	**<ENT-P><POS><THA><ENT-N>** <ซัมซุง><สวย><กว่า><ไอโฟน> <Samsung><beautiful><than><iPhone>
3	**<ENT-P><POS><THA>** <zenphone><คุ้ม><มากกว่า> <Zenphone><worth the money><than>
4	**<ASP><ENT-N><NEG><THA><ENT-P>** <แบต><iphone><ห่วย><กว่า><samsung> <battery><iPhone><terrible><than><Samsung>
5	**<ENT-N><ASP><POS><LTH><ENT-P>** <S5><เคส><ทนทาน><น้อยกว่า><ip6> <S5><case><sturdy><less than><iPhone 6>

Another interesting observation is the comparative patterns has direction when the term “more than” or “less than” appear in the sentence. If a positive polar term is followed by “more than”, it will have a positive sentiment on the entity (e.g., pattern no. 2 of Table VI). However, if a positive polar term is followed by “less than”, it will have a negative sentiment on the entity (e.g., pattern no. 5 of Table VI). The same logic applies for the case of negative polar terms.

Using the constructed patterns, we performed experiments to evaluate how well the patterns could analyze the sentiment by identifying entities, aspect terms and polar terms in a sentence. From the tagged corpus, we randomly selected a set of test corpus. The number of associative test sentences is *233*. The number of comparative test sentences is *251*. We use the extraction accuracy as the evaluation metric. The accuracy is defined as the number of correctly identified terms (i.e., entities, aspect terms, polar terms) divided by the total number of test sentences.

In the first experiment, we performed empirical analysis on the filter threshold parameter, *Filler_TH*. For both associative and comparative patterns, we varied the number of *Filler_TH* from 0 to 9. The results of empirical analysis for associative and comparative pattern mining are shown in Fig. 2 and Fig. 3, respectively. From both figures, it can be observed that as we relaxed the number of filler terms in the patterns, the extraction accuracies of entities, aspect and polar terms gradually increases until reaching a stable fixing point. For associative mining, the stable fixing point is when the number of filler terms is equal to 7. For comparative mining, the stable fixing point is when the number of filler terms is equal to 4. Therefore, by allowing dynamic filter terms could help increase the effectiveness of both pattern mining processes. The findings are not surprising. From our previous observations have shown that in Thai language, a written associative or comparative sentence could be very flexible and does not always follow perfect grammatical rules.

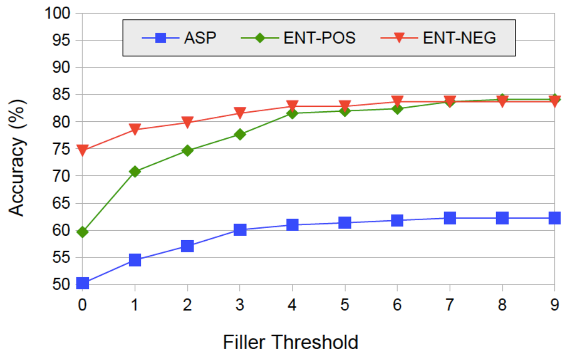

Fig. 2. An empirical analysis of *Filler_TH* for associative pattern mining

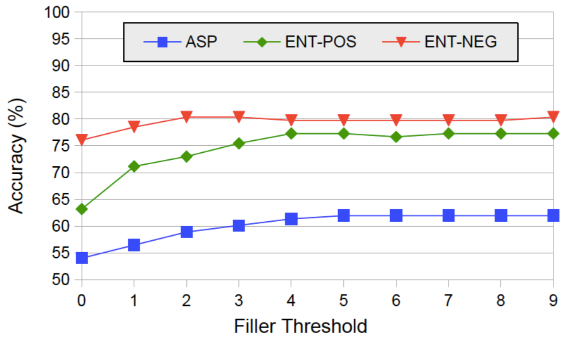

Fig. 3. An empirical analysis of *Filler_TH* for comparative pattern mining

Table VII shows the summarized evaluation results of associative and comparative pattern mining for sentiment analysis. For every test sentences in the corpus, the proposed mining approach was able to detect polar terms (either positive or negative) in every sentences. It can be observed that for associative mining, the accuracies of ENT-P and ENT-N extraction are both higher that those of comparative mining. This could be due to the associative patterns are mostly more simple than those of comparative patterns. In associative mining, only one entity must be identified compared to one or two entities for comparative mining. The main cause of entity identification errors are due to the lexicon does not contain the newly unseen brands or models appear in the test corpus. Other cause is the writers use variations or abbreviations when referring to brands or models of smartphones. Examples include the term "5" to refer to "iPhone 5" or "z2" to refer to "Sony Xperia Z2".

Another interesting observation is the extraction of aspect terms are not very high, only around 61-62% for both associative and comparative minings. The main cause of low detection rate is due to the aspect terms in smartphone domain are quite varied. From our error analysis, we found that some aspect terms appear in the test corpus are very fine grained or not directly related to the smartphone hardware specifications such as *front camera*, *shutter*, *games*, *OS*, *after sale service* and *Internet surfing*. The most effective and simplest approach to increase the extraction accuracies of entities, aspect and polar terms is to increase the number of tagged sentences in the corpus. Another practical solution is to keep the list of brand and model names in the smartphones updated regularly. The process of maintaining the lexicon is also applicable for other business domains, especially the ones which are very competitive and regularly launching new products and models such as automobile and consumer goods.

TABLE VII. EVALUATION RESULTS OF ASSOCIATIVE AND COMPARATIVE SENTIMENT ANALYSIS

Analysis	**Accuracy (%)**		
	ASP	**ENT-P**	**ENT-N**
Associative	62.23	84.12	83.69
Comparative	61.96	77.30	80.36

V. CONCLUSION AND FUTURE WORK

We proposed an approach for mining associative and comparative patterns for aspect-based sentiment analysis in Thai language. To increase the performance of pattern extraction, our proposed approach is based on the analysis of relaxed patterns with dynamic filler terms. A set of domain independent lexicon was designed to support the pattern construction process. The proposed approach first constructs associative and comparative patterns from a tagged corpus. The constructed patterns are then used to automatically analyze the sentiment by extracting entities, aspect and polar terms from an input sentence. The performance evaluation was done using a corpus of smartphone reviews obtained from various webboards and websites. From the experimental results, the entity extraction for associative mining yielded better performance (in terms of accuracy) than comparative mining, approximately 83-84% compared to 77-80%. The extraction of aspect terms yielded an accuracy of approximately 61-62%, for both associative and comparative minings. The main cause of extraction errors are due to the lexicon does not contain the newly unseen brands or models appear in the test corpus. Also, we found that some aspect terms appear in the test corpus are very fine grained or not directly related to the smartphone hardware specifications

For future work, we plan to apply the algorithm, DIPRE (Dual Iterative Pattern Relation Expansion) [2], to automatically collect entities and aspect terms by using the constructed patterns. Another future work includes applying the constructed associative and comparative patterns for cross business domains. We would like to study how patterns in one business domain can be used to perform the aspect-based sentiment analysis in different domains.

REFERENCES

[1] Philip Beineke, Trevor Hastie and Shivakumar Vaithyanathan, "The sentimental factor: improving review classification via human-provided information," *Proc. of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 263–270, 2004.

[2] Sergey Brin, "Extracting Patterns and Relations from the World Wide Web," *Proc. of the Int. Workshop on The World Wide Web and Databases*, pp. 172–183, 1998.

[3] Kushal Dave, Steve Lawrence and David M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," *Proc. of the 12th int. conf. on World Wide Web*, pp. 519–528, 2003.

[4] Choochart Haruechaiyasak, Alisa Kongthon, Pornpimon Palingoon, Chatchawal Sangkeettrakarn, "Constructing Thai Opinion Mining Resource: A Case Study on Hotel Reviews," *Proc. of the Eighth Workshop on Asian Language Resources*, pp. 64–71, 2010.

[5] Minqing Hu and Bing Liu, "Mining and summarizing customer reviews," *Proc. of the 10th ACM SIGKDD int. conf. on Knowledge discovery and data mining*, pp. 168–177, 2004.

[6] Nitin Jindal and Bing Liu, "Identifying comparative sentences in text documents," *Proc. of the 29th annual int. ACM SIGIR conf.*, pp. 244–251, 2006.

[7] Nitin Jindal and Bing Liu, "Mining comparative sentences and relations," *Proc. of the 21st int. conf. on artificial intelligence*, pp. 1331–1336, 2006.

[8] Yohan Jo and Alice H. Oh, "Aspect and sentiment unification model for online review analysis," *Proc. of the fourth ACM int. conf. on Web search and data mining*, pp. 815–824, 2011.

[9] Lun-Wei Ku and Hsin-Hsi Chen, "Mining opinions from the Web: Beyond relevance retrieval," *Journal of American Society for Information Science and Technology*, 58(12):1838–1850, 2007.

[10] Lun-Wei Ku, Ting-Hao Huang and Hsin-Hsi Chen, "Using morphological and syntactic structures for Chinese opinion analysis," *Proc. of the 2009 empirical methods in natural language processing*, pp. 1260–1269, 2009.

[11] Quanchao Liu, Heyan Huang, Chen Zhang, Zhenzhao Chen, and Jiajun Chen, "Chinese Comparative Sentence Identification Based on the Combination of Rules and Statistics," *Proc. of the 9th int. conf. on Advanced Data Mining and Applications* pp. 300–310, 2013.

[12] Pang, Bo, Lillian Lee and Shivakumar Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," *Proc. of the ACL-02 conf. on empirical methods in natural language processing*, pp. 79–86, 2002.

[13] Popescu, Ana-Maria and Oren Etzioni, "Extracting product features and opinions from reviews," *Proc. of the conf. on human language technology and empirical methods in natural language processing*, pp. 339–346, 2005.

[14] Tun Thura Thet, Jin-Cheon Na, and Christopher S.G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *J. of Inf. Sci.*, 36(6): 823–848, 2010.

[15] Mikalai Tsytsarau and Themis Palpanas, "Survey on mining subjective data on the web," *Data Minining and Knowledge Discovery*, 24(3): 478–514, 2012.

[16] Wiebe, Janyce and Ellen Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," *Proc. of conf. on Intelligent Text Processing and Computational Linguistics*, pp. 486–497, 2005.

[17] Wilson, Theresa, Janyce Wiebe and Paul Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," *Comput. Linguist.*, 35(3):399–433, 2009.

[18] Seon Yang and Youngjoong Ko, "Finding relevant features for Korean comparative sentence extraction," *Pattern Recogn. Lett.* 32(2): 293–296, 2011.

[19] Pantip Webboard. Available: *http://www.pantip.com*.