

USING SENTIMENT ANALYSIS TECHNIQUE FOR ANALYZING THAI CUSTOMER SATISFACTION FROM SOCIAL MEDIA

Todsanai Chumwatana

*Faculty of Information Technology, Rangsit University, Pathumthani 12000, Thailand
Email: Todsanai.c@rsu.ac.th*

ABSTRACT. With the rapidly increasing number of Thai online customer reviews available in social media and websites, sentiment analysis technique, also called opinion mining, has become an important task in the past few years. This technique aims to analyze people's emotions, opinion, attitudes and sentiments. The classical approaches for opinion mining represents the reviews as bag-of-words as many words can be used to identify positive or negative feedbacks. This makes these methods work well with European language reviews which are segmented texts. However, these bag-of-words based methods face problem with Thai customer's review which is non-segmented text, since Thai texts are formed as a long sequence of characters without word boundaries. Up to now, not much research conducted on sentiment analysis for Thai customer reviews. This paper proposes a sentiment analysis technique for Thai customer's reviews. The proposed technique is based on the integration of Thai word extraction and sentiment analysis techniques for mining Thai customer's opinion. To demonstrate the proposed technique, experimental studies on analyzing Thai customer's reviews from social media are presented in this paper. The results show that the proposed method provides significant benefits for mining Thai customer's opinion from social media.

Keywords: sentiment analysis, Thai word segmentation, opinion mining

INTRODUCTION

Recently, large amount of customer reviews are available on the internet. A significant number of website, blogs, forums and social media allow people to provide their opinion about the services and products (Bagheri, 2013). This online reviews open the new era of business intelligence and online marketing in today world. The words contained in customer's review can be regarded as the main factor to judge the customer satisfaction. Due to this rapid increase of online customer reviews in digital form, sentiment analysis technique has become an important task in recent year. This technique is used to analyze people's emotions, opinion, attitudes and sentiments. Sentiment Analysis, also called Opinion Mining, is one of the recent research field in the area of information extraction and text processing. Sentiment analysis provides many opportunities to develop the performance of business work by using the extracted reviews on the internet. For instance, the comments about products and services proposed by customers can be computed into account aspects such as positive and negative opinions about these products or services. These comments will then be used to improve the service and product quality. Most of the researches in sentiment analysis of website, blog, forum and social media have been performed in European languages where words are clearly

defined by word delimiter such as white space or other special symbols. This is because European texts are explicitly segmented into word tokens, then word tokens are used as a bag-of-words to be parameters for the sentiment analysis process. However, when apply the sentiment analysis to Thai customer's reviews, Thai texts need to be parsed and tokenized into individual words first before analysis. In Thai, there have been very few attempts to work on sentiment analysis of social media. This is because the syntax of Thai language is highly ambiguous and Thai language is non-segmented (i.e. a text document is written continuously as a sequence of characters without explicit word boundary delimiters). Figure 1 shows an example of Thai language (Jaruskulchai, 2003;Chumwatana, 2013).

กำลังช่างใจคิดว่า ถอย Nokia Lumia 520 มาเพื่อใช้งานเปิดมัล
เปิดไฟล์เอกสาร ราคาไม่แพงด้วย แต่ปัญหาคือเราต้องพกมือถือ 3

Figure 1. Example of the Thai language

Because of this problem, this paper proposes a sentiment analysis technique for Thai customer's reviews. The proposed technique is based on the integration of Thai word extraction and sentiment analysis techniques for mining Thai customer's opinion.

RELATED WORKS

Today, the sentiment analysis has become an important task and also spread widely used in any country and any language. This makes many approaches have been proposed in different type of languages. To study the sentiment analysis from web service, Jesus Serrano-Guerrero presented Sentiment analysis: A review and comparative analysis of web services (Serrano-Guerrero, 2015). This work provided review and compare some free access web services, analyzing theirin capabilities to classify and score different pieces of text with respect to the sentiments contained. In addition, A survey on sentiment detection of reviews was revealed by Cheng in 2009 (Cheng, 2009). This survey discussed Document sentiment classification and opinion extraction and some main approaches in this area.

Furthermore, as the sentiment technique is also employed to non-English language, some researches that work for the other languages have also been proposed. In 2008 (Tan, 2008), the empirical study of sentiment analysis for Chinese documents were proposed. This technique presents an empirical study of sentiment categorization on Chinese documents. The experimental results indicate that IG performs the best for sentimental terms selection. In Czech social media, supervised sentiment analysis was presented by Ivan Habernal (Habernal, 2014). This article describes in-depth research on machine learning methods for sentiment analysis of Czech social media. The author also established a common ground for further research by providing a large human-annotated Czech social media corpus.

USING SENTIMENT ANALYSIS FOR ANALYSING THAI CUSTOMER SATISFACATION FROM SOCIAL MEDIA

In this section, the proposed method that combines Thai word extraction technique and sentiment analysis technique is described. In this paper, the proposed method aims to analysis Thai customer's opinion from the reviews on social media using Thai word segmentation to extract Thai words and sentiment analysis to check customer's opinion. Therefore, the proposed method consists of two main steps: **1)** a process of extracting Thai words and text processing, and **2)** a process of sentiment analyzing the extracted Thai words as shown in Figure 2.

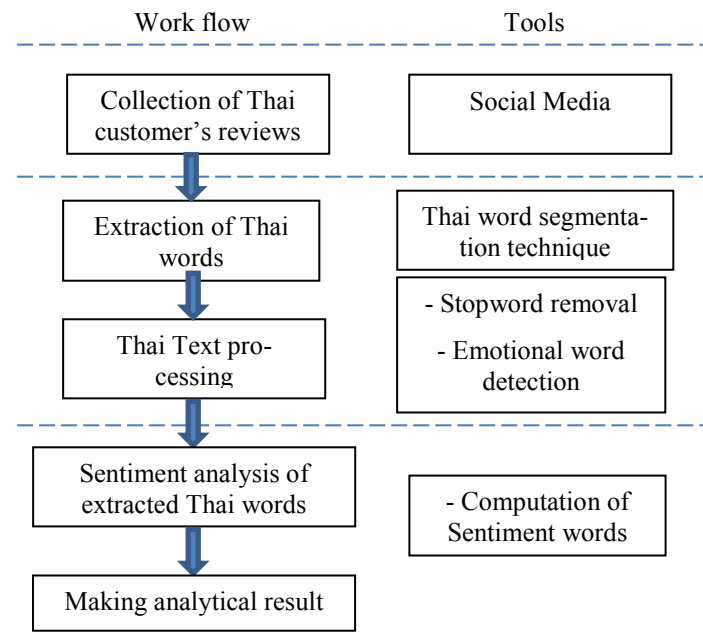


Figure 2. System Architecture

From Figure 2, Thai customer's reviews crawled from social media are used as an input to extract Thai words using Thai word segmentation technique. Thai word segmentation technique is the way to pass human knowledge to computers, so that computers can understand the Thai language like humans do. As a result, the computer can use this technique to extract Thai words before the next process. Thai word segmentation technique can be mainly categorized into three approaches: Rule base, Dictionary base and Machine learning based approaches (Chumwatana, 2013). In this paper, the machine learning technique is employed to segment Thai customer's reviews as this technique is well known and most efficient (Haruechaiyasak, 2004; Kruengkrai, 2006). This technique uses the machine learning technique to learn from a Thai text corpus. As learning is the essential part of this approach, it is necessary to have an appropriate word-segmented Thai text corpus, called the ORCHID corpus. The ORCHID corpus in 2008 consisted of 113,404 manually tagged words since 1996 (Haruechaiyasak, 2008; Sornlertlamvanich, 1997). This makes the machine learning technique work well in term of Thai word segmentation. The example of segmenting customer's reviews by using the machine learning technique is shown in the following.

| | |
|-----------------------------|--|
| Customer's review | มือถือยี่ห้อนี้ค้างบ่อยและก็ไม่ดีด้วย |
| Translate Customer's review | This mobile phone brand is always hang and it is not good as well. |
| Thai word segmentation | มือ-ถือ-ยี่ห้อ-นี้-ค้าง-บ่อย-และ-ก็-ไม่-ดี-ด้วย |

Figure 3. Segmentation of: “มือถือยี่ห้อนี้ค้างบ่อยและก็ไม่ดีด้วย” by using Machine learning technique

After these Thai words are segmented and extracted from reviews, the next process is Thai text processing which consist of Thai stopword removal and emotional word detection.

Thai stopwords are frequently occurring insignificant words in the Thai language (Chuleerat, 1998). These Thai words do not represent the content of the texts. Therefore, such words should be removed from the set of extracted words. Then, a further process is to perform emotional word tagging, which assigns each extracted word. In this process, the set of emotional words in database will be matched to the rest of Thai extracted words from the first step in order to check whether it is the emotional word as shown in figure 4.

| | |
|--------------------------|--|
| Customer's review | มือถือที่หือนี้ค้างบ่อยและก็ไม่ดีด้วย |
| Thai word segmentation | มือถือ ที่หือ นี้ ค้าง บ่อย และ ก็ ไม่ดี ด้วย |
| Stopword removal | มือถือ ที่หือ นี้ ค้าง บ่อย และ ก็ ไม่ดี ด้วย |
| Emotional word detection | มือถือ ที่หือ นี้ ค้าง บ่อย และ ก็ ไม่ดี ด้วย |

Figure 4. Example of Thai word extraction and text processing

Finally, the sentiment analysis technique will be applied to verify the extracted word which is the emotional word. This is because the emotional words can reveal the customer's emotions, opinion, attitudes and sentiments as well as their satisfaction. In this process, emotional words in database are represented in 3 levels: 1). Positive opinion represented by +1, 2). No opinion represented by 0, and 3). Negative opinion represented by -1 as shown in the table below.

Table 1. Emotional word database

| Emotional word | Sentiment level |
|-------------------|-----------------|
| เฉย ๆ (so so) | 0 |
| เยี่ยม (great) | +1 |
| แย่มาก (terrible) | -1 |

After all process has been done, the customer's reviews are finally assigned by sentiment level: +1, 0 and -1 as depicted in the table 2.

Table 2. Sentiment level values for reviews

| Reviews | Sentiment level | Category | Source |
|---|-----------------|----------------|----------|
| จะว่าไปพักหลังๆ โนเกียทำตัวเครื่องไม่สวยแล้วอะ สู้สมัซ N8 N9 Lumia 800 900 ไม่ได้เลย | -1 | Nokia Thailand | Facebook |
| กำลังช่างใจ คิดว่า ออช Nokia Lumia 520 มาเพื่อใช้งานเปิดเมมส์ เปิดไฟล์เอกสาร ราคาไม่แพงด้วย แต่ปัญหาคือ เราต้องพก มือถือ 3 เครื่องกันเลย !! | 0 | Nokia Thailand | Twitter |
| Nokia Lumia เล่นมันดีแฮะ | +1 | Nokia Thailand | Twitter |

Consequently, the results from the proposed technique can be used to educate the business's products or services. The numerous feedbacks from customer together with sentiment level can help the business to improve their product quality and services in order to be able to compete with the business candidate.

EXPERIMENTAL STUDIES

In this section, an experiment for sentiment analyzing customer’s review based on the proposed method is presented. 2348 reviews were used as an input in the experimental study. All reviews used are crawled from Facebook and Twitter websites that consist of 5 different categories: Nokia Thailand, Nokia X, Overall mobile, Nescafe red cup machine and Shabushi restaurant as shown in table 3.

Table 3. Experimental categories

| Category | No. of reviews | Channel |
|-------------------------|----------------|----------|
| Nokia Thailand | 1939 | Twitter |
| Nokia X | 91 | Twitter |
| Overall Mobile | 7 | Twitter |
| Nescafe Red Cup Machine | 112 | Twitter |
| Shabushi restaurant | 199 | Facebook |

In order to determine whether the proposed method is appropriate for sentiment analyzing Thai customer’s reviews, the experimental study is needed. In this experiment, Thai words are first extracted from the input reviews by using Thai word segmentation technique. The Thai text processing is then applied to the resultant set of extracted words. Finally, the sentiment analysis technique is hired to analysis customer’s opinion from the reviews. From the experimental study, the customer’s opinion in negative and positive ways is reveal in the following figure 5.

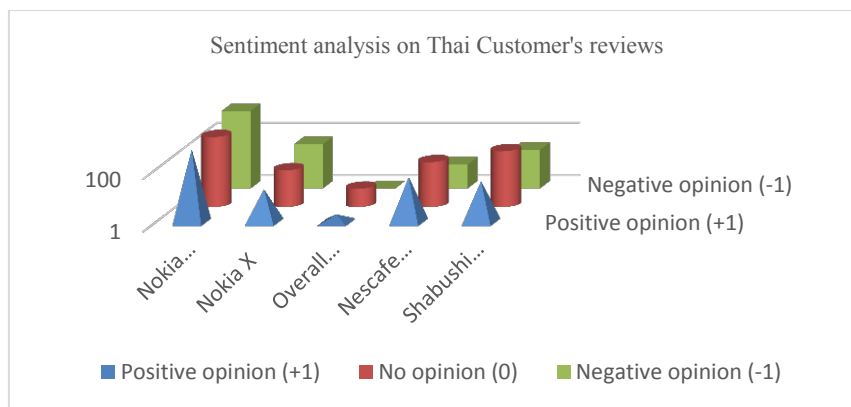


Figure 5. Graph of sentiment analysis of Thai customer’s reviews

From the graph in Figure 9, the experiment result shows that each category provides different results depend on their product quality. Shown in Figure 5, more number of Thai customer have negative opinion with Nokia Thailand and Nokia X when compared to the number of customer who have the positive opinion while most of Thai people just talk about the general thing for Overall mobile category. In contrast, many people provide good opinion reviews for 2 categories: Nescafe Red Cup Machine and Shabushi restaurant. These results would help the company to know about the weakness of their products and services and improve the quality of the products and services.conclusion

Today, numerous number of customer reviews are available on the internet. Website, blogs, forums and social media are the main channel which people use to provide their opin-

ion about the services and products. This makes the sentiment analysis has become an important task to analysis the customer's satisfaction. However, it is a challenge task to process sentiment analysis for Thai review as its nature of being non-segmented texts. Due to this problem, this paper proposes a sentiment analysis technique for Thai customer's reviews. The proposed technique is based on the integration of Thai word extraction and sentiment analysis techniques for mining Thai customer's opinion. The proposed method consists of two main steps: (1) a process of extracting Thai words and Thai text processing, and (2) a process of sentiment analyzing the extracted words. The experiment results show that different categories provide different results depending on the brand and quality. This reveals the customer's opinion about the products and services. Consequently, these such results would be the benefit for business to open the way of improvement.

REFERENCES

- Bagheri, A., Sarae, M & Jong, F D. (2013). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems*. 52, 201-213. (ISI Journal)
- Jaruskulchai, C. & Kruengkrai, C. (2003). A practical text summarizer by paragraph extraction for Thai. In *Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages*. Sapporo, Japan. pp. 9-16.
- Chumwatana, T. (2013). A Survey of Automatic Indexing Techniques for Thai Text Documents. In *Information Technology Journal*. King Mongkut's University of Technology North Bangkok, Thailand. Volume 17.
- Serrano-Guerrero, J., Olivás, J.A., Romero, F.P., & Herrera-Viedma, E. (2015). Sentiment Analysis: A Review and Comparative Analysis of Web Service. *Information Science*, 311, 18-38.
- Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications*. 36(7).
- Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for Chinese documents. *Expert Systems with Applications*. 34(4), 2622-2629.
- Habernal, I., Ptáček, T., & Steinberger, J. (2014). Supervised sentiment analysis in Czech social media. *Information Processing & Management*, 50693-707. Online publication date: 1-Sep-2014
- Chumwatana, T. (2013). A Survey of Automatic Word Segmentation Techniques for Thai Text Documents. In *Rangsit Journal of Information Technology (RJIT)*. Rangsit University. Thailand. 1(2), 36 – 43. July-December 2013.
- Haruechaiyasak, S., Kongyoung, S. & Damrongrat, C. (2004). LearnLexTo: A Machine-Learning Based Word Segmentation for Indexing Thai Texts. In *ACM 17th Conference on Information and Knowledge Management*.
- Kruengkrai, C., & Isahara, H. (2006). A Conditional Random Field Framework for Thai Morphological Analysis. In *Proceeding of the Fifth International Conference on Language Resources and Evaluation (LREC-2006)*.
- Haruechaiyasak, C., Kongyoung, S., & Dailey, M. N. (2008). A Comparative Study on Thai Word Segmentation Approaches. In *Proceedings of Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*.
- Sornlertlamvanich, V., Charoenporn, T., & Isahara, H. (1997). ORCHID: Thai Part-Of-Speech Tagged Corpus. In *Technical Report TR-NECTEC-1997-001*: Thai National Electronics and Computer Technology Center.
- Chuleerat, J. (1998). An Automatic Indexing for Thai Text Retrieval. vol. *Ph.D. Thesis U.S.A: George Washington University*.