

Thai Sentiment Analysis for Consumer's Review in Multiple Dimensions Using Sentiment Compensation Technique (SenseComp)

Paitoon Porntrakoon, Chayapol Moemeng
Vincent Mary School of Science and Technology
Assumption University Samuthprakarn, Thailand
{paitoon, mchayapol}@scitech.au.edu

Abstract—Trustworthiness of an e-vendor in the e-marketplace can be determined in multiple dimensions: product, price, and shipping. An e-vendor who has high trust level in more dimensions is more likely to have the competitive advantage than others. A consumer's review is analyzed to find its polarity in different dimensions. Positive sentiment in consumers' reviews helps increase the trustworthiness of e-vendors which in turn influences consumer's purchase intention. In this paper, we propose the method to automatically analyze Thai sentiment of consumer's review in product, price, and shipping dimensions by using multi-dimensional lexicon and sentiment compensation technique. A consumer's review in Thai language is tokenized using the longest matching algorithm. Then, it is analyzed to find its sentiment. Sentiment compensation technique is used to automatically compensate the sentiment to a dimension where consumer's review mentions the sentiment without a dimension. The results show that our proposed method outperform sentiment to dimension (S2D) and dimension to sentiment (D2S) methods with the overall accuracy 93.60%.

Index Terms—multi-dimensional trust, multi-dimensional lexicon, Thai sentiment analysis, sentiment compensation, e-marketplace

I. INTRODUCTION

Web 2.0 technologies provide the ability for Internet users to generate their own contents and share them to other people in the community. Ratings and reviews of a product are the examples of user-generated contents which play a very important role in e-marketplace websites. Consumers can compare the same products from different e-vendors before making the purchase decision. Positive ratings and reviews of a product help increase the purchase intention of consumers. Trust is very important factor in making the purchase decision. Trustworthiness of an e-vendor in e-marketplace websites should not be measured in single dimension but rather in

multiple dimensions: product, price, and shipping. An e-vendor may have high trust score in one dimension but low trust score in the others. Therefore, multi-dimensional trust provides the opportunity to consumers to decide which dimension of e-vendor is the most important that can help them making the good purchase decision [1], [2], [3].

Sentiment analysis or opinion mining can be used to identify the polarity of consumer's review in different dimensions which will become the trust score. The polarity of sentiment in a dimension can be positive, negative, or neutral which is equivalent to the trust score 1, -1, and 0 respectively. A number of frameworks used to automatically analyze the sentiment in Thai language were proposed such as SenseTag [4], S-Sense [5], and HotelOpinion [6]. These frameworks extract a consumer's review into feature and polar words in order to determine the positive and negative opinion of a consumer [6], [7]. Thai word segmentation is required for this extraction. There are three major types of word segmentation approaches: rule based, dictionary based, and machine learning based approaches. Rule based approach compares the consumer's review with Thai language linguistic rules [8], [9]. Dictionary based approach compares the consumer's review with the words in dictionary by using the longest matching, maximum matching, or decision tree algorithm [8], [9], [10]. Machine learning based approach compares the consumer's review by using the mathematical and statistical models such as Markov Chain [8] and Hidden Markov [10] models. The precision of word segmentation helps increase the accuracy of sentiment analysis.

In this paper, we propose a method to automatically analyze Thai sentiment of consumer's review in product, price, and shipping dimensions by using multi-dimensional lexicon and sentiment compensation technique (SenseComp). It analyzes

Thai sentiment of consumer's reviews in product, price, and shipping dimensions by applying the longest word matching algorithm in the dictionary based approach. Moreover, the sentiment compensation technique is used to automatically compensate the sentiment to missing dimension in the consumer's review. For instance, if a consumer's review contains the positive polarity without a specific dimension then such positive polarity will be automatically assigned to the pre-defined dimension. We did the experiments to compare the accuracy of SenseComp to manual, sentiment to dimension (S2D), and dimension to sentiment analyses (D2S).

The remainder of this paper is organized as follows. Section 2 provides the overview of Thai sentiment analysis and Thai word segmentation. Section 3 explains the details of proposed method (SenseComp). Section 4 describes the experiments and discusses the results. Section 5 concludes the paper with the future work.

II. RELATED WORK

A. Multi-dimension Trust Measurement

In an e-marketplace website in Thailand, such as Lazada, the five-star rating system is used in consumers' reviews in which the reviews may not be consistent with the given stars. A positive review may obtain only 3 stars while a negative review may obtain 5 stars. Therefore, trustworthiness of an e-vendor should be measured in product, price, and shipping dimensions in order to improve the accuracy of consumers' reviews [1]. Fig. 1 shows the multi-dimensional trust model for Thailand e-marketplace where the sentiment of a consumer's review is analyzed to find its polarity in multiple dimensions.

Score in each dimension is given based on its sentiment. Sentimental score 1 is given to positive, -1 is given to negative, and 0 is given to neutral sentiments respectively. Multi-dimensional trust score is calculated to show the strength and weakness of an e-vendor in different dimensions. Overall trust score of an e-vendor is calculated by using the weight to balance the importance of all dimensions. To facilitate and support this model, a method to automatically extract the consumers' reviews and analyze the sentiments should be used. Table I shows the examples of manual sentiment analysis and trust scores in multi-dimensions.

B. Thai Sentiment Analysis

Product review from a consumer is useful for the others to make a good purchase decision. Positive review can increase the trustworthiness of e-vendor which in turn influences the consumers' purchase intention [1], [2]. Thai sentiment analysis is a challenge because Thai is one of unsegmented languages

Figure 1. Multi-dimensional trust model for Thailand e-marketplace.

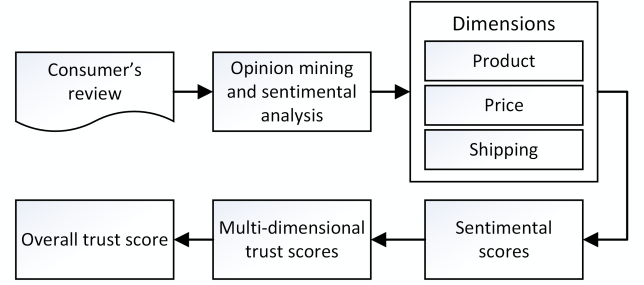


Table I
EXAMPLES OF MANUAL SENTIMENT ANALYSIS IN MULTI-DIMENSIONS

Review	Product	Price	Shipping
ได้สินค้าเร็ว แฟนชอบมาก ราคาถูกครับ (Fast delivery. My girlfriend likes it very much. Cheap price.)	1	1	1
สินค้าดี ใช้การได้ดี ได้รับสินค้าเร็ว ไม่มีปัญหาครับ (Good product. Good conditions. Fast delivery. No problem.)	1	0	1
เดี๋ยวนี้อะไรลดเยอะ แพงกว่าเดิมเยอะมาก (No more discount. Much more expensive than before.)	0	-1	0
การส่งสินค้าไม่ดีอย่างมาก ไม่มีการแพคเกจเลย (Shipping is not good. No packing at all.)	0	0	-1

[4]. A number of frameworks were proposed to analyze the sentiments of Thai opinions such as SenseTag [4], S-Sense [5], and HotelOpinion [6]. These frameworks extract Thai opinions into sentences. Each sentence can be tokenized into words by using the longest word matching algorithm which is the popular algorithm in Thai word segmentation [8], [9], [10]. Then, each word is analyzed using the domain-dependent lexicons to identify whether it is the feature or polar word [4], [5], [6]. Feature word shows the specific feature of product in the consumer's review where the polar word shows the positive or negative polarity of the sentiment in such review. Table II shows the examples of feature and polar words in consumers' reviews.

III. PROPOSED METHOD

In this paper, we propose a method to automatically analyze Thai sentiment of consumer's review in product, price, and shipping dimensions by using multi-dimensional lexicon and sentiment compensation technique (SenseComp). The sequence of processes in SenseComp is shown in Fig. 2. First, text in a consumer's review is extracted into sentences by using the blank space. Second, each sentence is cleaned by removing English characters and special symbols to ensure

Figure 2. Proposed method for SenseComp.

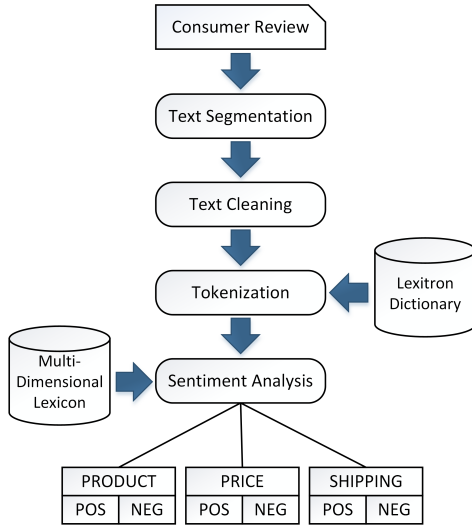


Table II
EXAMPLES OF FEATURE AND POLAR WORDS

Review	Feature Words	Polar Words	
		Positive	Negative
กล้องไอโฟนดี (iPhone has good camera.)	กล้อง (camera)	ดี (good)	-
หน้าจอไม่สวยเลย (Screen is not beautiful.)	หน้าจอ (screen)	-	ไม่สวย (not beautiful)

that it contains only numbers and Thai characters. Third, each sentence is tokenized using the longest matching algorithm. The LEXiTRON¹ dictionary developed by NECTEC is used in the tokenization. Fourth, each word is analyzed to identify whether it refers to product dimension (PRD), price dimension (PRC), shipping dimension (SHP), positive polar (POS), negative polar (NEG), or others (OTH) by using predefined multi-dimensional lexicon. Finally, scores 1, -1, and 0 are given to the positive, negative, and neutral polarities of the dimensions respectively.

In multi-dimensional lexicon, the words referring to dimensions and sentiments are manually defined based on the observation of real consumers' reviews. A sentiment word can be the polarity in one or more dimension. For instance, แพง (expensive) is the negative polarity in price dimension where ไม่ดี (not good) can be the negative polarity in product and shipping dimensions. Therefore, a dimension in each sentence of the review is important to determine a reference for the sentiment. In some reviews, consumers mention the sentiment but do not mention the specific dimension which leads to the difficulty in sentiment analysis. For instance, สวยมาก (very beautiful) is the review which has sentiment without a specific

¹LEXiTRON, <http://lexitron.nectec.or.th>

Table III

SENTIMENT ANALYSIS USING SENTIMENT COMPENSATION TECHNIQUE

Consumer's review

ได้สินค้าเร็ว แฟนชอบมาก ราคาถูกครับ

Fast delivery. My girlfriend likes it very much. Cheap price sir.

SHP	POS	PRD	POS	PRC	POS	OTH
ได้สินค้า	เร็ว	สินค้า	ชอบมาก	ราคา	ถูก	แฟน, ครับ
delivery	fast	product	like very much	price	cheap	my girlfriend, sir

dimension.

In SenseComp, the sentiment compensation technique is used to automatically assign a dimension to the sentiment when the sentence does not mention a specific dimension and such sentiment can refer to one and only one dimension. This technique helps increase the accuracy of Thai sentiment analysis in multiple dimensions. Table III shows the sentiment analysis of a consumer's review “ได้สินค้าเร็ว แฟนชอบมาก ราคาถูกครับ (Fast delivery. My girlfriend likes it very much. Cheap price sir.)” where สินค้า (product) is the product dimension compensated by the sentiment ชอบมาก (like very much). From this example, it is shown that the positive polar is given to product dimension although the sentence does not mention such dimension. This is because such positive polar is defined for the product dimension only.

IV. EXPERIMENTS AND RESULTS

To evaluate the accuracy of SenseComp, we conducted the experiments by comparing the accuracy of SenseComp to manual, S2D, and D2S methods. In manual method, each review is manually analyzed to identify the polarity in product, price, and shipping dimensions. In each dimension, the scores 1, -1, and 0 are given to the positive, negative, and neutral polarities respectively. For S2D method, the sentiment of each sentence in a consumer's review is identified and analyzed to find its polarity. Then, the dimension of such polarity is determined. D2S method is opposite to S2D. In D2S method, the dimension in each sentence is identified before finding out its polarity.

A total of 2,500 consumers' reviews were collected from Lazada Thailand website by a crawler program. Each of them was manually analyzed to identify the polarity in product, price, and shipping dimensions as shown in Table I. One thousand from these reviews were randomly selected and used to create and update the multi-dimensional lexicon using MySQL database. Multi-dimensional lexicon contains the feature and polar words in product, price, and shipping dimensions. A web-based application was developed using PHP in order

Table IV
NUMBER OF REVIEWS CLASSIFIED BY DIMENSIONS

Dimension	Number of Reviews	Percentage
None	180	7.2
Product Only	945	37.8
Price Only	130	5.2
Shipping Only	255	10.2
Product and Price	345	13.8
Product and Shipping	435	17.4
Price and Shipping	70	2.8
All	140	5.6
Total	2,500	100

Table V
ACCURACY COMPARISON BETWEEN SENSECOMP, S2D, AND D2S METHODS

Dimension	Accuracy (%)		
	SenseComp	S2D	D2S
None	100.00	100.00	50.00
Product Only	78.84	50.79	35.98
Price Only	76.92	50.00	23.08
Shipping Only	62.75	41.18	13.73
Product and Price	82.61	46.38	37.68
Product and Shipping	81.61	51.72	34.48
Price and Shipping	71.43	64.29	42.86
All	78.57	32.14	28.57
Total	93.60	62.20	42.40

to support the experiments. Table IV shows the number of consumers' reviews classified by the dimensions. Among these 2,500 reviews, 180 (7.2%) reviews did not mention any dimension, 1,330 reviews (53.2%) mentioned only one dimension, 850 reviews (34.0%) mentioned two dimensions, and 140 reviews (5.6%) mentioned all of three dimensions.

Table V shows the results of experiments. SenseComp outperforms the S2D and D2S methods because it uses the sentiment compensation technique. In consumers' reviews that mention only one dimension, product dimension has the highest accuracy (78.84%) when compared to S2D (50.79%) and D2S (35.98%). In two dimensions, product and price has the highest accuracy (82.61%) when compared to S2D (46.38%), D2S (37.68%), and other combinations of dimensions such as product and shipping (81.61%) and price and shipping (71.43%). It shows that the accuracy of SenseComp is highest when there is the product dimension in the consumers' reviews. In three dimensions, SenseComp has the highest accuracy (78.57%) when compared to S2D (32.14%) and D2S (28.57%). When counted the number of correct analyses, SenseComp has highest accuracy (93.60%) when compared to S2D (62.20%) and D2S (42.40%).

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed SenseComp – a method to automatically analyze Thai sentiment of consumer's review in product, price, and shipping dimensions by using multi-dimensional lexicon and sentiment compensation technique. We discussed multi-dimensional trust measurement and Thai sentiment analyses where the sentiments of consumer's review become the trust score of product and e-vendor. Moreover, we discussed the techniques used in the segmentation and tokenization of consumers' reviews. The results from our experiments show that the sentiment compensation technique increases the accuracy of SenseComp in all dimensions when compared to S2D and D2S methods. In the future, this proposed method should be improved to support more complex consumer's reviews which may have multiple dimensions and multiple sentiments in a single sentence.

REFERENCES

- [1] P. Porntrakoon and C. Moemeng, "A model for multi-dimensional trust measurement in thailand e-marketplace," in *2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Phuket, 2017, pp. 73–76.
- [2] J. Chen and X.-L. Shen, "Consumers' decisions in social commerce context: An empirical investigation," *Decision Support Systems*, vol. 79, pp. 55–64, 2015.
- [3] Z. Wu and L. Wang, "Trustworthiness Measurement of E-commerce Systems Using Fuzzy Hybrid Multi-criteria Analysis," in *2015 IEEE Trustcom/BigDataSE/ISPA*, 2015, pp. 668–675.
- [4] K. Trakultaweekoon and S. Klaitin, "Sensetags: A tagging tool for constructing thai sentiment lexicon," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Khon Kaen, 2016, pp. 1–4.
- [5] C. Haruechaiyasak, A. Kongthon, P. Palingoon, and K. Trakultaweekoon, "S-sense: A sentiment analysis framework for social media sensing," in *Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP)*, 2013, pp. 6–13.
- [6] A. Kongthon, C. Haruechaiyasak, and C. Sangkeetrakarn, "Hotelopin-ion: An opinion mining system on hotel reviews in thailand," in *2011 Proceedings of PICMET '11: Technology Management in the Energy Smart World (PICMET)*, Portland, OR, 2011, pp. 1–6.
- [7] C. Haruechaiyasak, A. Kongthon, P. Palingoon, and C. Sangkeetrakarn, "Constructing thai opinion mining resource: A case study on hotel reviews," in *2010 Proceedings of the 8th Workshop on Asian Language Resources*, Beijing, China, 2010, pp. 64–71.
- [8] K. Chaonithi and S. Prom-on, "A hybrid approach for thai word segmentation with crowdsourcing feedback system," in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Chiang Mai, 2016, pp. 1–6.
- [9] C. Mahatthanachai, K. Malaivongs, N. Tantranont, and E. Boonchieng, "Development of thai word segmentation technique for solving problems with unknown words," in *2015 International Computer Science and Engineering Conference (ICSEC)*, Chiang Mai, 2015, pp. 1–6.
- [10] W. Thanadechteemapat and C.-C. Fung, "Thai word segmentation for visualization of thai web sites," in *2011 International Conference on Machine Learning and Cybernetics*, Guilin, 2011, pp. 1544–1549.