



A Hierarchical Emotion Classification Technique for Thai Reviews

Jirawan Charoensuk* & Ohm Sornil

Graduate School of Applied Statistics, National Institute of Development Administration, 118 Seri-Thai Road, Bangkok, Bangkok, 10240, Thailand

*E-mail: jirawan.cha@stu.nida.ac.th

Abstract. Emotion classification is an interesting problem in affective computing that can be applied in various tasks, such as speech synthesis, image processing and text processing. With the increasing amount of textual data on the Internet, especially reviews of customers that express opinions and emotions about products. These reviews are important feedback for companies. Emotion classification aims to identify an emotion label for each review. This research investigated three approaches for emotion classification of opinions in the Thai language, written in unstructured format, free form or informal style. Different sets of features were studied in detail and analyzed. The experimental results showed that a hierarchical approach, where the subjectivity of the review is determined first, then the polarity of opinion is identified and finally the emotional label is calculated, yielded the highest performance, with precision, recall and F-measure at 0.691, 0.743 and 0.709, respectively.

Keywords: *hierarchical emotion classification; speech synthesis; opinion classification; text processing; Thai language.*

1 Introduction

With the growth of Internet communication, emotion classification on text-based resources, e.g. blogs, online newspapers and social networks, is becoming a more interesting and challenging task. Emotion classification can be applied in researches on subjects such as speech synthesis [1], image processing [2-4] and especially text processing [5-8]. Many people regularly buy products from websites or mobile applications. These websites or applications want to have product feedback after customers have used their products. Customers' reviews are usually written in unstructured format, free form or informal style. The content of a review usually expresses opinions and emotions about the quality and quantity of the product that the reviewer ordered and used. These emotions can affect the profit and image of a company. If reviewers express positive opinions or positive emotions toward a product on a website, there is a tendency to attract more customers to it. Due to the large amount of reviews, it is difficult for readers to identify emotions manually. An automatic emotion classification method is needed to solve this problem.

Most emotion classification methods are proposed for English and Western languages. However, these cannot be directly applied to the Thai language. Thai texts are written in continuous form, without spaces, punctuation marks, full stops (‘.’), commas (‘,’) or semicolons (‘;’) to identify boundaries of words and sentences [9]. Haruechaiyasak, *et al.* [10] proposed the S-Sense framework for three Thai social media sources to identify intention and sentiment labels from text. They used three lexicon resources, i.e. LEXiTRON (electronic Thai-English dictionary), the Thai Twitter corpus and clue words, and applied the Multinomial Naive Bayes algorithm as the classification model. Chirawichitchai [6] proposed emotion classification on social networks in the Thai language by using a corpus-based approach. This research used Boolean, term frequency and Tf-Idf weights as feature sets and applied the Support Vector Machine, Naïve Bayes, Decision Tree and K-nearest Neighbor algorithms to detect six emotions. The experimental results showed that Boolean weighting with Support Vector Machine yielded the best performance. Lastly, Chumwatana [11] proposed sentiment classification on social media and websites. This method extracts emotional words from text and assigned each word with a sentiment score (‘+1’ for a positive word, ‘0’ for a neutral word and ‘-1’ for a negative word). The sentiment score of an opinion is calculated by summing the word scores together. Previous researches on Thai opinion classification have suggested that an effective feature set can be constructed from corpus-based and lexicon approaches. This research proposes a hierarchical framework to identify emotions (i.e. objective, anger, disgust, fear, sadness, happiness and surprise) for actual customer reviews written in the Thai language.

The hierarchical classification framework consists of 3 levels: the opinion level, the sentiment level and the emotion level. The opinion level separates customers’ reviews into two types, i.e. objective and subjective reviews [12]. An objective or neutral emotion expresses factual information or no opinion. A subjective opinion expresses a reviewer’s opinion, which can be classified as a positive or a negative opinion. The second level is the sentiment level, which categorizes a subjective opinion as a positive or a negative sentiment [13]. The emotion level then assigns an emotion label to an opinion.

Based on Ekman [14], six basic emotions can be used to describe facial expressions in all human traditions. These emotions are: anger, disgust, fear, sadness, happiness and surprise. Breazeal in 2003 [15] proposed polarity in arousal-valence (A-V) graph space based on Ekman’s emotions, where the x-axis represents valence by mapping a scale of pleasant versus unpleasant or positive versus negative sentiment, while the y-axis represents arousal by mapping a scale of being relaxed vs aroused. Positive emotions are happiness and surprise, while negative emotions are anger, disgust, fear, and sadness. The emotion classification used in this research was organized accordingly.

2 Proposed Emotion Classification Techniques

The proposed method consists of three main processes: (1) text preprocessing (2) feature extraction, and (3) emotion classification. Text preprocessing provides necessary information and normalization of the words that occur in the reviews. Feature extraction generates a set of features by using corpuses and lexicons. Then, the emotion classification applies a classification algorithm to the extracted features to identify the emotion labels. Three approaches for emotion classification were studied, as shown in Figure 1.

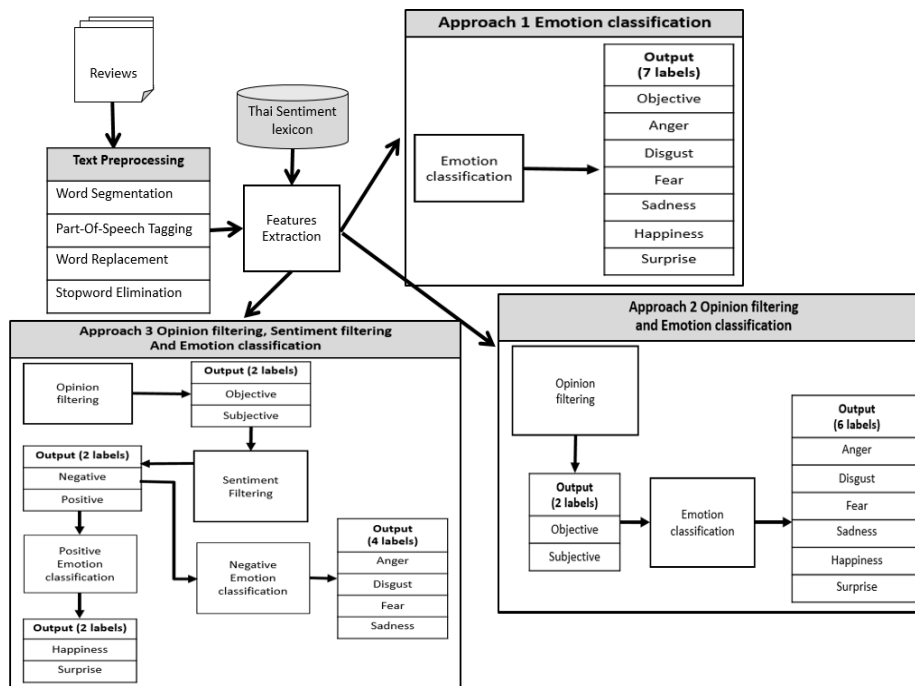


Figure 1 Three approaches of emotion classification framework.

The first approach uses a classifier directly to identify seven labels of opinion analysis, i.e. objective opinion and Ekman's six human emotions. The second approach is a two-level structure of opinion filtering and emotion classification. Opinion filtering determines whether a review contains a subjective opinion of the reviewer. Emotion classification then classifies the emotion of the opinion identified in the previous step into six emotion labels, i.e. anger, disgust, fear, sadness, happiness and surprise. The last approach contains three levels of opinion filtering, sentiment filtering and positive and negative emotion classification. The difference with the second approach is that after opinion filtering instead of classifying emotion directly, each opinion is separated first

into positive and negative opinions and then each positive opinion is classified into a positive emotion and each negative opinion is classified into a negative emotion.

2.1 Text Preprocessing

Preprocessing includes word segmentation, part-of-speech (POS) tagging, word replacement and stopword elimination. Examples of the preprocessing steps are shown in Table 3.

2.1.1 Word Segmentation

Unlike English, the Thai language has distinctive syntactical and semantic characteristics. The language has no specific symbols (e.g. ‘.’, ‘?’, ‘;’) to identify the end of a sentence or clause. Furthermore, there is no space between words. Accordingly, identifying the boundaries of each word is a nontrivial problem for Thai. This research specifically focused on customer reviews usually written in unstructured format, free form or informal style. Word segmentation was performed by KuCut [16], which is based on global and local unsupervised learning to segment unknown words.

2.1.2 Part-Of-Speech Tagging

Part-of-speech (POS) is considered an important element at the morphology level to represent the role of token words such as ‘verb’, ‘noun’ and ‘conjunction’. POS has essential information at the word level for identifying opinion categories. Hence, the Jitar tagging tool [17] was applied to assign a POS label to each word. Jitar is based on a trigram hidden Markov model (HMM) and the Naist corpus [18]. The Naist corpus consists of 60,511,974 words that were collected from Thai magazines and has 49 part-of-speech tag sets in 17 groups.

2.1.3 Word Replacement

Word replacement reduces typographical errors and words with repeated characters. Typographical errors are caused by mistyping, for example, ‘แอลกอฮอล์’ (alcohol) can be mistyped as ‘แอลกอฮอล์’, ‘แอลกอฮอล์’ or ‘แอลกอฮอล์’, while words with repeated characters are caused by a reviewer repeating characters on the keyboard to express a strong opinion. For example, in ‘ดีมากกกกก’ (very good) the character ‘ก’ is repeated 5 times. The Thai language has the symbol ‘ๆ’ to signify the repetition of the previous word. Therefore, ‘ดีมากกกกก’ is becomes ‘ดีมากๆๆๆ’. Word replacement was implemented by regular expression rules. In addition, this research defined five

new POS tags for punctuation that indicates opinion labels. The new POS tag set is shown in Table 1.

Table 1 New POS tag set.

Naist tag set	New tag set
?/punc	?/Qmark
(/punc or /(punc	‘/Qparent
๑/punc	๑/Qrepeat
!/punc	!/Qexclamation
‘/punc or ’/punc or “/punc or ”/punc	”/Qquote

2.1.4 Stopword Elimination

Some extremely common words in text have little value to identify types of human emotions. These words are called stopwords and consist of a set of common words (e.g. a, the, for, at), punctuations (e.g. (,], ?, ‘), numbers (e.g. 1, 2, 3) and symbols (e.g. %, \$, @). These were eliminated.

The remaining words are the main words that the reviewer uses to express his or her opinion. This research used information from POS tagging to ignore words that do not express opinions or emotions of reviewers. There are three types of filters, i.e. eleven POS tags that are words expressing no opinion, blank and English words for brands or ingredients of products, or reviewers’ names. The list of eleven POS tags is shown in Table 2.

Table 2 Eleven POS tags for stopword elimination.

No	POS Tag	Description	Example	No	POS Tag	Description	Example
Participle				Symbol			
1	aff	Affirmative	ค่ะ, ครับ	7	sym	Symbol	๑๗%,%
2	part	Particle	นะ, นั่นเอง	Noun			
Classifier				8	ntit	Title noun	นาย, นางสาว
3	cl	Classifier	ขึ้น, กล้อง	9	nlab	Label noun	2, ก
Prefix				10	nnum	Cardinal number	หมื่น, 1000
4	pref1	Prefix1	การ, ความ	Punctuation			
5	pref2	Prefix2	ผู้, นัก	11	punc	Punctuation	.,- ,_
6	pref3	Prefix3	ชาว				

Table 3 Examples of each text preprocessing step.

Process	Example	Remark
Input	“ivory_caps”_ไม่เห็นได้ผลเลยใช้มาจะ_6_ขวดแล้วอะ [“ivory caps”, there are not any results, although I have used it for 6 bottles]	underscore (_) represents a space
Word Segmentation	“ ivory _caps ” ไม่ เห็น ได้ ผล เลย ใช้ มา จะ _ 6 _ ขวด แล้ว อะ 	vertical bar () represents a segmented sign
POS Tagging	“/punc ivory /nnp _/punc caps /nnp ”/punc _/punc ไม่ neg เห็น vt ได้ vpost ผล ncn เลย part _/punc ใช้ vt มา vpost จะ prev _/punc 6 /nnum _/punc ขวด cl แล้ว vpost อะ aff ”/Qquote ivory /nnp _/punc caps /nnp ”/Qquote	
Word Replacement	_/punc ไม่ neg เห็น vt ได้ vpost ผล ncn เลย part _/punc ใช้ vt มา /vpost จะ prev _/punc 6 /nnum _/punc ขวด cl แล้ว vpost อะ aff	replace “/punc and ”/punc with ”/Qquote
Stopword Elimination	”/Qquote ”/Qquote ไม่ neg เห็น vt ได้ vpost ผล ncn เลย part ใช้ vt มา vpost จะ prev แล้ว vpost	remove 3 tokens; 6/nnum , ขวด/cl , อะ/aff and remove 5 spaces and 2 English words

2.2 Feature Extraction

Feature extraction constructs a vector representation for each review. This research used five corpus-based and lexicon feature subsets.

2.2.1 Term Weighting

The term frequency (tf) and inverse document frequency (idf) weighting technique [19] was used, where tf_{ij} is the number of times a term t_i appears in document d_j and f_i is the raw frequency count of term t_i in document d_j . The normalized tf_{ij} formula is shown in Eq. (1), where the maximum is computed over all terms that appear in the document and $|V|$ is the vocabulary size. idf_i is the inverse document frequency of term t_i , where N is the total number of documents and df_i is the number of documents in which term t_i appears. The formula for idf_i is shown in Eq. (2). The weight $Tf - Idf_{ij}$ can be calculated with Eq. (3):

$$tf_{ij} = \frac{f_{ij}}{\max(\{f_{1j}, f_{2j}, \dots, f_{|V|j}\})} \quad (1)$$

$$idf_i = \log \frac{N}{df_i} \quad (2)$$

$$Tf - Idf_{ij} = tf_{ij} * idf_i \quad (3)$$

The Tf-Idf weight of a unigram word (TUW) is calculated for each term while that of a bigram word (TBW) is calculated for each word pair.

2.2.2 Part-of-Speech Weighting

The part-of-speech (POS) weight feature is the Tf-Idf weight for a POS tag, calculated from the training data. There are weights for both unigram part-of-speech (TUP) and bigram part-of-speech (TBP).

2.2.3 Thai Sentiment Lexicon

The last subset of features is a Thai sentiment lexicon [20] that creates two attributes for identifying positive and negative words from customer reviews. These attributes were derived from a Thai sentiment lexicon that is available online and consists of 1,031 words, where 321 words constitute the positive lexicon (PL) and 710 words constitute the negative lexicon (NL). The lexicon includes nouns, verbs and adjectives. This feature is represented as two integer attributes: *positive_score* and *negative_score*. The value of each attribute is calculated according to the pseudocode in Figure 2.

<p>Input: a sequential token word ($w_{1..n}$) in each review</p> <p>Output: summary of positive score (PL) and summary of negative score (NL)</p> <p>Initial value: positive_score = 0, negative_score = 0</p> <p>if token $w_{1..n} \in PL$ then positive_score++</p> <p>else token $w_{1..n} \in NL$ then negative_score++</p>

Figure 2 Pseudocode for calculating lexicon score.

2.3 Opinion Classification

Three classification algorithms were used in combination with the proposed method to identify emotion labels, i.e. Decision Tree, Multinomial Naïve Bayes and Support Vector Machine.

2.3.1 Decision Tree

A decision tree consists of internal nodes that represent attribute tests and leaf nodes that contain output classes. Information gain is computed as the decrease in entropy after a data set is split on an attribute and the attribute with the highest information gain is selected for the current split. Information gain and entropy can be calculated according to Eqs. (4) and (5), respectively [21].

$$Gain(S, A) = Entropy(S) - \sum_i \frac{|S_i|}{S} \cdot Entropy(S_i) \quad (4)$$

$$Entropy(S) = \sum_{i=0}^c -p_i \log_2 p_i \quad (5)$$

where S is the training data set, A is the attribute set, p is the proportion of instances belonging to class i , and c is the total number of classes.

2.3.2 Multinomial Naive Bayes

Multinomial Naive Bayes [22] is a popular classification technique in the context of text analytics. It calculates the conditional probability of observing features x_1 through x_n given some class c , where $p(x_i|c)$ is shown in Eq. (6).

$$P(x_1, x_2, \dots, x_n | c) \quad (6)$$

With the independence assumption, Multinomial Naive Bayes can be expressed as Eq. (7). Its application to text classification considers the positions of words in a document as shown in Eq. (8).

$$C_{NB} = \operatorname{argmax}_{c \in C} P(c_j) \prod_{x \in X} P(x|c) \quad (7)$$

$$C_{NB} = \operatorname{argmax}_{c \in C} P(c_j) \prod_{i \in \text{positions}} P(x_i | c_j) \quad (8)$$

2.3.3 Support Vector Machine

Support Vector Machine (SVM) [19,23] is a supervised learning method relying on a linear separation of input data with high dimensions. SVM represents the training data with different categories as points in a vector space and uses a margin to define the distance between the separating hyperplane and the training data that are closest to this hyperplane. A kernel function $K(x,y)$ represents our desired notion of similarity between data x and y , allowing the learning of a non-linear model. The polynomial kernel, shown in Eq. (9), is a commonly used function and was applied in this research.

$$K(x, y) = (\langle x, y \rangle + 1)^d \quad (9)$$

3 Data Set

There is no standard test collection for free-text reviews in the Thai language specifically for opinion classification. The data set used in this research consisted of customer reviews of cosmetics collected from three popular beauty websites, i.e. www.lazada.co.th, www.kony.com and www.vanilla.in.th with a total of 2,770 reviews. Each review was annotated by five readers who were familiar with the subject matter and the label with the majority votes was selected as the result. Accordingly, each annotation had three levels: the opinion label, consisting of two types (objective and subjective); the sentiment label, consisting of two types (positive and negative); and the emotion label,

consisting of six labels (anger, disgust, fear, sadness, happiness and surprise). Table 4 shows the characteristics of the data set.

Table 4 Characteristics of the data set.

Label	Number of reviews	Example	Remark
Objective	138	สูตรของ Bio-Oil เป็นการผสมกันของสารสกัดจากพืชและวิตามินที่อยู่ในรูปของน้ำมัน โดยมีสารประกอบ PurCellin Oil ซึ่งทำให้สูตรของ Bio-Oil มีความบางเบาดูดซึมสูผิวได้ง่าย [The Bio-Oil formulation is a combination of plant extracts and vitamins, suspended in an oil base. It contains PurCellin Oil, which makes it light and not greasy.]	
Subjective	2,632	เนื้อครีมมันมาก วันไหนอากาศร้อนตกป้ายหน้ามันเยิ้มเลย [It's very oily. It turns greasy when the weather is hot.]	
Positive	994	หลอดนี้ใช้ได้หลายเดือนคุ้มค่ามาก [A tube lasts for many months. Worth the money!]	positive and negative labels are identified from subjective labels
Negative	1,638	เนื้อครีมมันมาก วันไหนอากาศร้อนตกป้ายหน้ามันเยิ้มเลย [It's very oily. It turns greasy when the weather is hot.]	
Anger	679	แย่มาก ของใกล้หมดอายุ ไม่แจ้งให้ทราบ ขวดใหญ่มาก แล้วจัดไปร 1 แกรม 1 ไปอีก ใครจะไปใช้ทัน [It is very bad. Product was almost expired, but they did not tell the customer. The bottle is big with buy 1 get 1. Who can use it all?]	anger, disgust, fear and sadness are identified from negative labels
Disgust	287	เนื้อครีมมันมาก วันไหนอากาศร้อนตกป้ายหน้ามันเยิ้มเลย greasy.]	
Fear	183	เราใช้แล้วแพ้ะ แสบตามากๆ มันร้อนบวมไม่ถูก [I am allergic to the product. My eyes are very irritated.]	
Sadness	559	ตัวนี้ซื้อมาแอบหวังเล็กๆ วาจะดี แต่กลับเฉยๆ ใช้ไปได้ครึ่งขวดก็เปลี่ยนไปลองยี่ห้ออื่นแล้วค่ะ ไม่เห็นผลใดๆ เลย [I hoped that this would a good product. But after a half of the bottle, I changed to another brand. No good.]	
Happiness	489	เนื้อครีมนุ่มมากผิวสัมผัสดี ซึมไว เวลาทาแล้วทำให้หน้าดูนุ่มๆ ขึ้นผิวสุขภาพดีขึ้น ไม่แพ้ ไม่มีกลิ่นแรงด้วย [The cream is very soft and absorbed into the skin quickly. It makes my face look soft and healthy skin. No strong smell.]	happiness and surprise are identified from positive labels
Surprise	435	หลอดนี้ใช้ได้หลายเดือนคุ้มค่ามาก [A tube lasts for many months.]	

4 Experimental Evaluations

The feature extraction process generates five feature subsets for emotion classification. They are: Tf-Idf of unigram words (TUW), Tf-Idf of bigram words (TBW), Tf-Idf of unigram POS (TUP), Tf-Idf of bigram POS (TBP) and Thai sentiment lexicon (TL). This research used Decision Tree, Multinomial

Naïve Bayes and Support Vector Machine as classifiers. Eighty percent of the data were used to construct the model and the remaining twenty percent were used as test data. The first experiment studied the performance of Approach 1 with six patterns of feature sets and three classification algorithms. The results of the first approach are shown in Table 5. We can see that the TBW feature had the best performance, especially with Multinomial Naïve Bayes. The highest precision, recall and F-measure values of this approach were 0.689, 0.652 and 0.657, respectively.

Table 5 Effectiveness of the first approach.

Classifier	Feature set	Effectiveness of Approach 1		
		Emotion Classification (7 labels)		
		Precision	Recall	F-measure
Decision Tree	TUW	0.505	0.492	0.493
	TUW+TL	0.491	0.483	0.483
	TUW+TL+TUP	0.499	0.482	0.486
	TBW	0.505	0.494	0.494
	TBW+TL	0.482	0.468	0.469
	TBW+TL+TBP	0.492	0.473	0.477
Multinomial Naïve Bayes	TUW	0.648	0.619	0.625
	TUW+TL	0.614	0.570	0.570
	TUW+TL+TUP	0.623	0.590	0.591
	TBW	0.689	0.652	0.657
	TBW+TL	0.632	0.125	0.134
	TBW+TL+TBP	0.668	0.190	0.227
Support Vector Machine	TUW	0.605	0.595	0.596
	TUW+TL	0.609	0.600	0.600
	TUW+TL+TUP	0.611	0.604	0.604
	TBW	0.640	0.633	0.632
	TBW+TL	0.649	0.640	0.639
	TBW+TL+TBP	0.627	0.619	0.619

The results of the second approach (two-level hierarchy of opinion filtering and emotion classification) are shown in Table 6. They show that the TBW+TL+TBP feature set with Support Vector Machine had the best performance in filtering opinions, with 0.985, 0.986 and 0.984 for precision, recall and F-measure, respectively. Next, the emotion of each opinion was classified and the results show that Support Vector Machine with the TBW feature performed best with 0.688, 0.684 and 0.681 for precision, recall and F-measure, respectively.

Table 6 Effectiveness of the second approach.

Classifier	Feature set	Effectiveness of Approach 2					
		Opinion Filtering and Emotion Classification					
		Level 1 Opinion Filtering (2 labels)			Level 2 Emotion Classification (6 labels)		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Decision Tree	TUW	0.963	0.960	0.960	0.515	0.513	0.511
	TUW+TL	0.966	0.962	0.962	0.488	0.490	0.488
	TUW+TL+TUP	0.963	0.962	0.962	0.504	0.498	0.499
	TBW	0.962	0.957	0.959	0.496	0.487	0.486
	TBW+TL	0.965	0.960	0.962	0.501	0.494	0.492
	TBW+TL+TBP	0.959	0.957	0.958	0.501	0.494	0.488
Multinomial Naïve Bayes	TUW	0.979	0.969	0.972	0.652	0.624	0.615
	TUW+TL	0.979	0.968	0.972	0.659	0.635	0.629
	TUW+TL+TUP	0.982	0.972	0.975	0.651	0.635	0.629
	TBW	0.959	0.697	0.783	0.693	0.426	0.377
	TBW+TL	0.960	0.734	0.810	0.682	0.449	0.394
	TBW+TL+TBP	0.962	0.820	0.869	0.648	0.452	0.419
Support Vector Machine	TUW	0.979	0.980	0.978	0.646	0.643	0.643
	TUW+TL	0.978	0.979	0.978	0.640	0.635	0.636
	TUW+TL+TUP	0.982	0.982	0.982	0.635	0.631	0.630
	TBW	0.983	0.983	0.967	0.688	0.684	0.681
	TBW+TL	0.983	0.983	0.981	0.678	0.673	0.670
	TBW+TL+TBP	0.985	0.986	0.984	0.668	0.665	0.662

The third approach is a three-level hierarchy, where an opinion is first filtered, then the polarity of the opinion is identified and finally the emotion of an opinion with positive or negative polarity is classified accordingly. Its effectiveness in opinion filtering and sentiment classification is shown in Table 7 and the result of classifying positive and negative emotion classification is shown in Table 8. The results of opinion filtering using the third approach were in the same direction as those of the second approach. Table 7 shows that the best sentiment-filtering configuration was TBW+TL+TBP with Support Vector Machine. The precision, recall and F-measure of this type were 0.947, 0.947 and 0.946, respectively. The positive and negative opinions were sent to the positive and the negative emotion classifiers, respectively. The results of positive emotion classification show that the TUW+TL+TUP feature set with Multinomial Naïve Bayes had the best performance with 0.768, 0.765 and 0.764 for precision, recall and F-measure. Lastly, the results of negative emotion classification show that the TUW+TL feature set with Support Vector Machine had the best performance with 0.719, 0.709 and 0.705 for the three measures, respectively.

Table 7 Effectiveness of opinion filtering and sentiment filtering in the third approach.

		Effectiveness of Approach 3					
		Opinion Filtering, Sentiment Filtering and Emotion Classification					
Classifier	Feature set	Level 1 Opinion Filtering (2 labels)			Level 2 Sentiment Filtering (2 labels)		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Decision Tree	TUW	0.963	0.960	0.960	0.812	0.802	0.804
	TUW+TL	0.966	0.962	0.962	0.823	0.806	0.808
	TUW+TL+TUP	0.963	0.962	0.962	0.770	0.764	0.766
	TBW	0.962	0.957	0.959	0.807	0.791	0.793
	TBW+TL	0.965	0.960	0.962	0.805	0.787	0.789
Multinomial Naïve Bayes	TBW+TL+TBP	0.959	0.957	0.958	0.808	0.795	0.797
	TUW	0.979	0.969	0.972	0.878	0.878	0.878
	TUW+TL	0.979	0.968	0.972	0.865	0.863	0.864
	TUW+TL+TUP	0.982	0.972	0.975	0.864	0.863	0.864
	TBW	0.959	0.697	0.783	0.902	0.897	0.895
Support Vector Machine	TBW+TL	0.960	0.734	0.810	0.906	0.905	0.904
	TBW+TL+TBP	0.962	0.820	0.869	0.868	0.867	0.865
	TUW	0.979	0.980	0.978	0.878	0.875	0.875
	TUW+TL	0.978	0.979	0.978	0.885	0.882	0.883
	TUW+TL+TUP	0.982	0.982	0.982	0.881	0.878	0.879
	TBW	0.983	0.983	0.967	0.932	0.932	0.932
	TBW+TL	0.983	0.983	0.981	0.939	0.939	0.939
	TBW+TL+TBP	0.985	0.986	0.984	0.947	0.947	0.946

Table 8 Effectiveness of positive and negative classification in the third approach.

		Effectiveness of Approach 3					
		Opinion Filtering, Sentiment Filtering and Emotion Classification					
Classifier	Feature set	Level 3-1 Positive Emotion Classification (2 labels)			Level 3-2 Negative Emotion Classification (4 labels)		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Decision Tree	TUW	0.675	0.673	0.672	0.624	0.618	0.617
	TUW+TL	0.613	0.612	0.610	0.627	0.618	0.618
	TUW+TL+TUP	0.584	0.582	0.575	0.631	0.618	0.615
	TBW	0.635	0.633	0.630	0.625	0.612	0.608
	TBW+TL	0.637	0.633	0.628	0.636	0.612	0.608
	TBW+TL+TBP	0.655	0.653	0.651	0.625	0.618	0.619
Multinomial Naïve Bayes	TUW	0.757	0.745	0.741	0.723	0.703	0.699
	TUW+TL	0.765	0.755	0.752	0.707	0.691	0.685
	TUW+TL+TUP	0.768	0.765	0.764	0.718	0.697	0.688
	TBW	0.733	0.602	0.531	0.680	0.558	0.516
	TBW+TL	0.733	0.602	0.531	0.671	0.570	0.534
	TBW+TL+TBP	0.689	0.684	0.680	0.658	0.564	0.524
Support Vector Machine	TUW	0.695	0.694	0.693	0.701	0.703	0.700
	TUW+TL	0.695	0.694	0.693	0.702	0.703	0.700
	TUW+TL+TUP	0.717	0.714	0.713	0.694	0.697	0.693
	TBW	0.664	0.663	0.662	0.712	0.703	0.698
	TBW+TL	0.654	0.653	0.652	0.719	0.709	0.705
	TBW+TL+TBP	0.663	0.663	0.663	0.711	0.703	0.699

Table 9 compares the performance between the two-level hierarchical classification (Approach 2) and the three-level hierarchical classification (Approach 3) using the feature sets and algorithms that yielded the highest accuracy according to Tables 6, 7 and 8. The results show that the third approach achieved the best performance with accuracy at 69.60%. According to the precision and F-measure measurement, Approach 3 achieved better performance in classifying sentiments and emotions than Approach 2. For some emotions, the recall of Approach 2 was higher than that of Approach 3.

Table 10 shows a comparison between Approach 1 and Approach 3. For all the negative emotions (anger, disgust, fear, sadness) and one positive emotion (happiness), Approach 3 achieved higher precision than Approach 1. For objective, sadness and happiness, Approach 1 achieved higher recall than Approach 3.

Overall, the Tf-Idf of bigram word feature is the most effective feature subset to be used for filtering opinions, determining polarity and classifying negative emotions. Lexicon resources such as a Thai sentiment lexicon and a POS tag set at the morphology level can improve accuracy for the opinion filtering in Approaches 2 and 3.

Support Vector Machine achieves high performance in identifying contrasting opinions such as objective versus subjective opinions, and positive versus negative sentiments. Multinomial Naïve Bayes achieves high performance in identifying closely related emotions, such as happiness versus surprise in positive emotion classification.

There are four main reasons for incorrect classification. Firstly, inaccurate Thai word segmentation and POS tagging due to the complexity of the Thai language. Secondly, ambiguities in the Thai sentiment lexicon: (1) The scores of positive and negative emotion do not conform with the answer. For example, an anger emotion getting a positive score. (2) Reviewer can be sarcastic, for example ‘ดีจริงจริงเลย แป้งนี้หือนี้’ [this face powder is good], where the review actually has a negative sentiment. (3) The Thai sentiment lexicon only has two labels (positive and negative), which is not sufficient to classify emotions. (4) Bigrams cannot detect patterns of word pairs whose distance is further than two words, especially negative words. For example, a positive review [ไม่|เหนียว|เหนอะ|หนะ] can generate 3 bigrams, such as ‘ไม่เหนียว’, ‘เหนียวเหนอะ’, ‘เหนอะหนะ’.

The word ‘ไม่’ [not] is a negative word; ‘ไม่เหนียว’ expresses a positive sentiment, but ‘เหนียวเหนอะ’ and ‘เหนอะหนะ’ express a negative sentiment. The resulted sentiment of the entire opinion is thus negative since the weight of a

negative bigram is higher than the weight of a positive one. Thus, bigram is not effective in handling this type of problems.

Table 9 Effectiveness of hierarchical emotion classification.

Label	Approach 2 - Two-level Hierarchy Classification			Approach 3 - Three-level Hierarchy Classification		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Objective	0.769	1.000	0.870	1.000	0.769	0.870
Anger	0.844	0.692	0.761	0.844	0.667	0.745
Disgust	0.724	0.750	0.737	0.808	0.724	0.764
Fear	0.500	0.769	0.606	0.526	0.769	0.625
Sadness	0.652	0.732	0.690	0.547	0.630	0.586
Happiness	0.654	0.607	0.630	0.700	0.673	0.686
Surprise	0.592	0.617	0.604	0.646	0.738	0.689
Average	0.676	0.738	0.699	0.691	0.743	0.709
Accuracy	68.50%			69.60%		

Table 10 Effectiveness of Approach 3 versus Approach 1.

Level	Label	Approach 3 - Three-level Hierarchy Classification			Approach 1 - Emotion Classification		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Level 1	Objective	1.000	0.769	0.870	1.000	0.786	0.880
Opinion label	Subjective	0.989	1.000	0.994			
Level 2	Positive	0.939	0.975	0.957			
Sentiment label (based on result of level 1)	Negative	0.959	0.904	0.931			
	Anger	0.844	0.667	0.745	0.769	0.694	0.730
Level 3	Disgust	0.808	0.724	0.764	0.800	0.556	0.656
Emotion label (based on result of level 2)	Fear	0.526	0.769	0.625	0.500	0.538	0.519
	Sadness	0.547	0.630	0.586	0.530	0.714	0.609
	Happiness	0.700	0.673	0.686	0.633	0.717	0.673
	Surprise	0.646	0.738	0.689	0.656	0.583	0.618
Average		0.691	0.743	0.709	0.688	0.667	0.670
Accuracy		69.60%			66.67%		

5 Conclusion

In this paper, a framework for hierarchical classification of fined-grained emotions of cosmetic reviews written in the Thai language was presented. The proposed framework begins by extracting important words that express the opinion of the reviewer, representing each review with a set of features consisting of characteristics of unigram and bigram words and part-of-speech tags, and a Thai sentiment lexicon.

Three approaches of emotion classification were proposed and studied in detail, i.e. direct emotional classification of review texts; opinion filtering and emotional classification of opinions; and a hierarchical approach of opinion filtering, opinion polarity identification and emotion classification. At each step, Decision Tree, Support Vector Machine and Multinomial Naïve Bayes were tested as classifiers.

A set of experiments was conducted to evaluate the effectiveness of the different approaches and configurations on a collection of actual informal free-text reviews acquired from the Internet. The results showed that the proposed hierarchical approach had the best performance with precision, recall and F-measure at 0.691, 0.743 and 0.709, respectively. In addition, the Tf-Idf of bigram words was found to be the most effective set of features to tackle this problem.

References

- [1] Cowie, R., Douglas-Cowie E., Savvidou, S., McMahon E., Sawey, M. & Schroder M., *Emotion Recognition in Human-computer Interaction*, IEEE Signal Processing Magazine, **18**(1), pp. 32-80, Jan. 2001.
- [2] Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P. & Sejnowski, T.J., *Classifying Facial Actions*, IEEE Trans Pattern Anal Mach Intell, **21**(10), pp. 974-989, Oct. 1999.
- [3] Cohen, I., Sebe, N., Garg, A., Chen, L.S. & Huang, T.S., *Facial Expression Recognition from Video Sequences: Temporal and Static Modeling*, Computer Vision and Image Understanding, **91**(1-2), pp. 160-187, Jul. 2003.
- [4] Mehta, D., Siddiqui, M.F.H. & Javaid, A.Y., *Facial Emotion Recognition: A Survey and Real-world User Experiences in Mixed Reality*, Sensors, **18**(2), pp. 416, Feb. 2018.
- [5] Strapparava, C. & Mihalcea, R., *Learning to Identify Emotions in Text*, in Proceedings of the 2008 ACM Symposium on Applied Computing, New York, NY, USA, pp. 1556-1560, 2008.
- [6] Chirawichitchai, N., *Emotion Classification of Thai Text Based Using Term Weighting and Machine Learning Techniques*, presented at the 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 91-96, 2014.
- [7] Inrak, P. & Sinthupinyo, S., *Applying Latent Semantic Analysis to Classify Emotions in Thai Text*, presented at the 2010 2nd International Conference on Computer Engineering and Technology, **6**, pp. V6-450-454, 2010.
- [8] Burget, R. & Karasek, J., *Recognition of Emotions in Czech Newspaper Headlines*, **20**(1), pp. 39-47, 2011.

- [9] Sriphaew, K., Takamura, H. & Okumura, M., *Sentiment Analysis for Thai Natural Language Processing*, in Proceedings of the 2nd Thailand-Japan International Academic Conference TJIA, pp. 123-124, 2009.
- [10] Haruechaiyasak, C., Kongthon, A., Palingoon, P. & Trakultaweekoon, K., *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), Nagoya, Japan, pp. 6-13, 2013.
- [11] Chumwatana, T., *Using Sentiment Analysis Technique for Analyzing Thai Customer Satisfaction from Social Media*, presented at the Proceedings of the 5th International Conference on Computing and Informatics, Turkey, pp. 659-664, 2015.
- [12] Lui, B., *Sentiment Analysis and Opinion Mining*, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 5(1), pp.1-167, 2012.
- [13] Lee, D., Jeong, O-R. & Lee, S., *Opinion Mining of Customer Feedback Data on the Web*, Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication, Suwon, Korea, pp. 230-235, 2008.
- [14] P. Ekman, *An Argument for Basic Emotions*, Cognition and Emotion, 6 (3-4), pp. 169-200, May 1992.
- [15] Breazeal, C., *Emotion and Sociable Humanoid Robots*, International Journal of Human-Computer Studies, 59(1-2), pp. 119-155, Jul. 2003.
- [16] Sudprasert, S. & Kawtrakul, A., *Thai Word Segmentation Based on Global and Local Unsupervised Learning*, in Proceedings of the 7th National Computer Science and Engineering Conference (NCSEC'2003), 2003, pp. 1-8.
- [17] Kok, D. de, *Jitar HMM Part of Speech Tagger*, <https://github.com/danieldk/jitar>, 2018. (9 July 2018)
- [18] Asanee, K., Supapas, K., Thitima, J. & Chanvit, J., *A Lexibase Model for Writing Production Assistant System*, in Proceeding of the Second Symposium on Natural Language Processing, pp. 226-236, 1995.
- [19] Liu, B., *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, 2nd ed. Berlin Heidelberg: Springer-Verlag, 2011.
- [20] Phatthiyaphaibun, W., *lexicon-thai/sentiment at master · PyThaiNLP/lexicon-thai-GitHub*, 2016. [Online]. Available: <https://github.com/PyThaiNLP/lexicon-thai/tree/master/sentiment>. (9 July 2018).
- [21] Salzberg, S.L., *C4.5: Programs for Machine Learning by J. Ross Quinlan*, Morgan Kaufmann Publishers, Inc., 1993, Mach Learn, 16(3), pp. 235-240, Sep. 1994.
- [22] McCallum, A. & Nigam, K., *A Comparison of Event Models for Naive Bayes Text Classification*, in AAAI-98 Workshop on Learning for Text Categorization, pp. 41-48, 1998.

- [23] Yekkehkhany, B., Safari, A., Hodayouni, S. & Hasanlou, M., *A Comparison Study of Different Kernel Functions for SVM-based Classification of Multi-temporal Polarimetry SAR Data*, in ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **XL-2/W3**, pp. 281-285, 2014.