



Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining

Chakrit Pong-inwong¹ · Wararat Songpan¹

Received: 2 April 2017 / Accepted: 26 February 2018 / Published online: 5 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

This research proposes a new sentiment analysis method called sentiment phrase pattern matching (SPPM). The analysis model extracts the responses and comments from discussions that are posted in a teaching evaluation system in the form of open-ended questions and allows student respondents to provide feedback to their teachers on factors that affect teaching and studying in a classroom. The proposed method consists of three main phases: (1) collect feedback data and perform tokenization via the Teaching Senti-Lexicon; (2) analyze sentiment analysis phrases by SPPM, which is based on the association mining method and integrated with sentiment phrase frequency by using forward bigram traversal, for separating the many phrases from teaching feedback sentences; and (3) sentiment analysis based on sentiment scores from the Teaching Senti-Lexicon. The objective of this research is to obtain feedback from open-ended questions automatically via the proposed method for sentiment classification and to determine the best classification of the responses to the open-ended questions within educational attitude contexts by classifying attitude contexts as positive or negative. Moreover, SPPM is compared to others classifier algorithms. The results indicate that the SPPM method achieves the highest accuracy of 87.94% compared to the other classifier algorithms. In addition, SPPM achieves precision, recall and F-measure values of up to 92.06, 93 and 92.52%, respectively. The main contribution of the proposed model is that it determines the most effective strategy for improving teaching based on students' opinions.

Keywords Sentiment phrase analysis · Sentiment analysis · Sentiment phrase pattern matching · Teaching evaluation

1 Introduction

Recently, due to the growing phenomenon of the exchange of both structured and unstructured data, many transactions have occurred on social media and in feedback review systems. In previous research, it has been claimed that the technology for extracting knowledge from unstructured sentences, which is called natural language processing (NLP), has advanced rapidly [1, 2]. With the rapid growth of online reviews and the explosion of the number of social media reviews, the volume of online customer reviews of products has experienced tremendous growth. The generic data from a variety of sources, such as Facebook, Twitter, and Forum/

discussion web boards, passed the sentiment analysis process into the final sentiment polarity result. The sentiments that are hidden in the reviews can be classified as positive or negative [3]. It is an important task that combines data mining and NLP techniques to extract opinions or sentiments [4–6]. Currently, in educational settings, data mining plays an important role in improving education quality. For example, student success prediction [7] and secondary education placement test score prediction use sensitivity analysis [8]. Sentiment analysis is used for educational benefit improvement. Therefore, teaching evaluation has been one of the most important parts of educational development [9] because it can indicate various aspects of teaching quality. Evaluation for education is the first amongst three major areas in the academic world, and Students' Evaluation of Teaching (SEOT) is the most important and strongest factor for educational management and motivation in teaching which is used to ensure teaching quality at the university level. The results of teaching attitude evaluations represent several teaching properties, and their mean values reflect

✉ Wararat Songpan
wararat@kku.ac.th

Chakrit Pong-inwong
chakrit@kkumail.com

¹ Department of Computer Science, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand

the average quality of teaching. Generally, teaching attitude evaluation is performed at the end of the semester through the use of questionnaires comprising almost all closed-ended questions. The questionnaires that focus on teaching attitude evaluation commonly consist of two phases of the teaching evaluation process: The first phase is the process of obtaining responses from students as respondents, who are provided with multiple choices of answers with different levels of satisfaction, to evaluate their own attitudes. In general, the multiple-choice answers to choose from are strongly agree, agree, uncertain, disagree and strongly disagree, which correspond to numbers from 5 to 1, respectively. The second phase is the analysis process, in which the questionnaires' responses are statistically calculated and analyzed. The problem is that results of the teaching attitude evaluation represent several teaching properties, as the mean values reflect the average of the teaching scores. Moreover, open-ended questions have appeared at the end of the questionnaires in attempts to obtain additional student opinions on their teachers and their teaching, the responses to which are not analyzed. In addition, the information from the students' opinions on the feedback forms in terms of the teacher's teaching performance have been largely ignored.

In this paper, problems in interpreting students' attitudes from open-ended questions were solved. This paper aims at presenting a method for teaching sentiment classification from teaching recommended texts. This paper represents a framework for teaching attitude classification based on a case study of a teaching evaluation system in the Thai context at three Universities, Thailand by using a combination of association rule and sentiment analysis techniques, which is called sentiment phrase pattern matching (SPPM). SPPM utilizes sentiment analysis to extract patterns of language that can be integrated to calculate the polarity of words in open-ended answers from students.

2 Related work

2.1 Sentiment analysis

The growth of the social community via online media has rapidly increased. In other words, the review message appeared in many sources such as Twitter/microblog [10], blogs, Facebook, consumer review forums, movie reviews and other research that is similar to [11, 12]. These reviews caused difficulties in analysing the sentiment polarity, for which sentiment analysis was carried out with respect to the linguistic process. This approach was utilized to identify the sentiment from text reviews in which the phrases or sentences cover both positive and negative opinions. Khan et al. [13] proposed a decision framework for sentiment analysis by sentiment score revision for

SentiWordNet (SWN). Although sentiment analysis that is processed under the SentiWordNet corpus can be highly accurate, when employing the SentiWordNet corpus for languages other than English, the results for both meaning and usage cannot be obtained. This indicates that when transferring and interpreting the meaning in English to other languages, the sentiment polarity values could shift or become neutralized, which makes it difficult to assess the polarity sentiment. The Micro-Blog Sentiment Analysis System (MSAS) was proposed by Chamlerwat et al. [14] for analysing customer opinions on a smart phone with five functions. This includes collecting the customers' opinion posts, filtering posts, detecting polarity, categorizing product features and discussion. For the training set were obtained from Twitter posts for detecting the sentiment polarity with SentiWordNet 3.0 [15] to interpret the customers' reviews. PosScore and NegScore were used to measure positive and negative opinions, respectively, and SVM was exploited to classify their sentiments towards product features.

In addition, attitude explorations have been carried out by many researchers. For instance, the multiple attribute decision making problems for teaching evaluation on Wushu teaching in high schools with 2-tuple linguistic information was presented by Xue [16]. Furthermore, Leong et al. [17] proposed the potential application of sentiment mining for analysing teaching evaluations via trainees' short message system (SMS) texts. They also presented the tree model for classifying student sentiments towards teaching in a training class. However, sentiment analysis was limited to words only and failed to cover sentiment phrases. In practice, texts that are obtained from teaching evaluations are most likely to be opinion-expressive phrases or sentences that are written to convey sentiments. This is similar to the finding of Naradhipa and Purwarianti [18], who proposed sentiment classification for Indonesian messages in social media with SVM and Maximum Entropy algorithms to identify the sentiments of users whose messages were posted on social media. Teaching evaluation using a hybrid of least squares support vector machine (LSSVR) and the chaotic particle swarm optimization (CPSO) (CPSO-LSSVR) was proposed by Jing and Yanqing [9]. Moreover, Yu et al. [19] used the Naïve Bayes (NB) algorithm to analyze part of the impact of social and conventional media on firm equity value. The algorithm facilitated determination of the categories of positive and negative documents. Xianghua et al. [20] proposed the Multi-aspect Sentiment Analysis approach for sentiment classification of Chinese Online Social Reviews, which was called (MSA-COSRs). In this approach, Latent Dirichlet Allocation (LDA) was applied to explore the multi-aspects of social review topics, and the HowNet lexicon method was used to classify the associated sentiments.

2.2 Association rule learning

Association rule mining is an important task in the field of data mining [21]. It is a popular learning association method for discovering interesting relations between elements. Association rule mining consists of two main steps: determining the frequency feature, which generated all features, and rule generation. Furthermore, the method measures the frequency of occurrence within features. For example, let $I = \{i_1, i_2, \dots, i_n\}$ be a set of items and D be a set of transactions, where T is an element transaction of D , and $T \subseteq I$. Let X and Y be sets of items such that $X, Y \subseteq I$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$.

Support: The support value is the statistical probability of the co-occurrence of items in a transaction. The rule $X \Rightarrow Y$ holds with support s if $s\%$ of the transactions in D contains $X \cup Y$. Rules for which s is greater than a user-specified support are said to have minimum support.

$$\text{Support}(X) = P(X \cup Y) \quad (1)$$

Confidence: The confidence of a value, $X \Rightarrow Y$, with respect to a set of transactions T , is the proportion of the transactions that not only contain X but also contain Y .

$$\text{Conf}(X) = \text{Supp}(X \cup Y) / \text{Supp}(X) \quad (2)$$

The probability calculation method is consistent with the method that is applied in this study, which utilizes the probability values of phrases pattern using association rule and phrasal co-occurrences using bi-grams [22] in the texts that students wrote to convey their opinions regarding lecturers' teaching efficiency. This resulted in the discovery of sentiment phrases that benefited the teaching evaluation sentiment analysis.

3 Proposed model

This paper proposed a method for enhancing association rule mining to analyse sentiment phrases, which is applied with the Teaching Senti-Lexicon for sentiment analysis and teaching attitude classification. The framework consists of three main phases: The first phase is the data source mapping phase, in which text data are extracted from open-ended questionnaires in the teaching evaluation systems obtained from three universities. The second phase is the data preparation phase, which consists of tokenization based on Lex-Plus and sentiment phrase pattern matching for finding patterns in any language. The case study uses the Thai language. The third phase expresses the data model for classifying the students' attitudes towards teaching. Lastly, the sentiment polarity is classified, which is compiled as the

result diagram in the model deployment state. The method of Thai sentiment analysis for teaching evaluation is illustrated in Fig. 1.

3.1 Data collection

In the data collection process, the response messages are in forms of the attitude domain for both positive and negative polarities, which are obtained from the Loei Rajabhat University, Buriram Rajabhat University and Roi-Et Rajabhat University teaching evaluation systems, as discussed above. A total of 30,500 messages are obtained. In addition, the open-ended responses from questionnaires regarding the performance of the teacher of the class are collected feedback messages that are stored in the teaching evaluation system. For example, a message is “ควรพูดให้ช้ากว่านี้บางครั้งพูดเร็วทำให้นักศึกษาฟังไม่ทัน” that is interpreted to “*You should speak slowly; sometimes you speak quite fast. Therefore, I did not understand your teachings*”.

3.2 Data preparation

Students were well prepared for the sentiment analysis process. After the students' responses were extracted, as shown in data collection, the messages were added to the appropriate data sets for use in teaching sentiment classification experiments, where set MSG is the set of messages (S), as shown in Eq. (3).

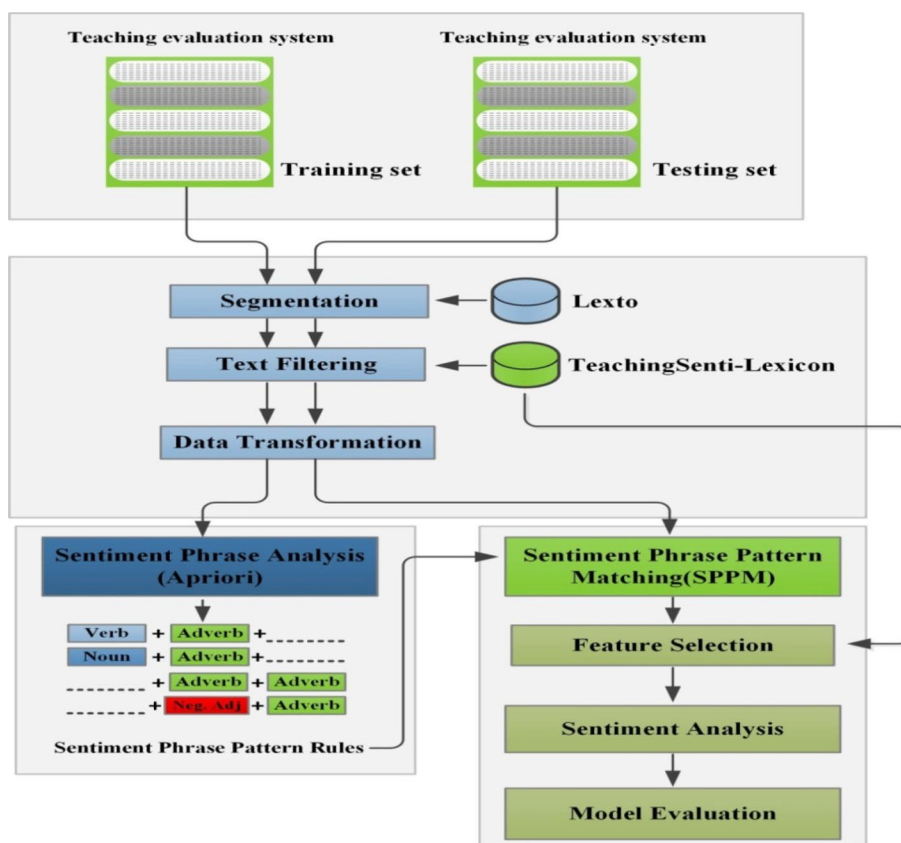
$$MSG = \{S_1, S_2, \dots, S_n\} \quad (3)$$

where MSG is the set of feedback messages; and S_i is a feedback message.

3.2.1 Data cleaning

The teaching evaluation data came from three institutions. In total, 30,500 responses were collected of information that was posted by the students to express their opinions on teaching performances through the teaching evaluation system from each university. Many messages are unavailable for detecting the sentiment phrase that without the teaching evaluation meaning, for example, the word “55555..”. This numerical text cannot be interpreted; thus, these messages are spam and irrelevant to the sentiment texts. The instance filtering handle that compared the text data and Teaching Senti-Lexicon dictionary was removed, and the sentiment terms were deleted. Symbolic texts that do not convey any sentimental meaning, for example, “#%#@”, “@@@)(*”; zero-opinion texts such as “no comment” and “none”) and texts that are irrelevant to the teaching evaluation were filtered out and excluded from the sentiment analysis. Therefore, 12,222 messages remained for training and testing in this teaching evaluation experiment.

Fig. 1 Framework of sentiment phrase pattern matching (SPPM)



3.2.2 Tokenization and filtering

The response texts from teaching evaluations are generally written in Thai, which is quite different from English. Thai writing uses continuous words without blank spaces. Although Thai sentence grammar writing principles explain the sentence structure clearly, most students who write the comments that are stored in the teaching evaluation system post the messages incorrectly in terms of the Thai sentence structure principles, such as emotion, short phrases and continuous words without spaces. In addition, when the texts are parsed completely, they are then assigned to set S_i , which has phrasing elements in each message as follows:

$$S_i = \{P_1, P_2, \dots, P_n\} \quad (4)$$

where S_i is a sentence that contains an element of MSG; and P_i is the phrases that an element in a set of sentence S_i .

To tokenize many feedback texts, we used a tool called A Thai Lexeme Tokenization (LexToPlus), which parses the texts to tokenize the explicit teaching evaluation scope based on the Teaching Senti-Lexicon, which was created specifically for teaching evaluation, to handle text filtering.

3.2.3 Teaching Senti-Lexicon

The texts are tokenized by the previously described process. Separated terms are considered. These terms are unavailable for interpretation without irritant word filtering, and the teaching sentiment evaluation must be limited to the teaching sentiment domain. Therefore, the Teaching Senti-Lexicon is specifically designed for teaching sentiment evaluation. It contains teaching sentiment terms and defines the terms and sentiment weight scores for sentiment polarity computation. In addition, the Teaching Senti-Lexicon is also capable of filtering out the feedback messages that are irrelevant to the teaching assessment. The Teaching Senti-Lexicon consists of five attributes: *TermId*, *Term*, *TypeOfTerm*, *Feature* and *SentimentScore*. Most of comments are not written in correct grammatical structure, and there are several clauses in the text. Therefore, the *TypeOfTerm* attribute is specifically designed, which takes values such as *TV*, *QA*, *Neg. Adv.*, *PPA*, *NPA* and other type of terms shown in Table 1, which represent the terms, term types, features and sentiment weight scores. The sentiment weight scores are assigned values from -1 to -0.1 to represent negative polarity words. In

Table 1 Example of Teaching Senti-Lexicon

Term Id	Term	Type of term	Feature	Sentiment score
T001	ดีมาก (very good)	PPA	Teaching	1
T002	การสอน (teaching)	N	Teaching	0.1
T003	ไม่ค่อย (seldom/rarely)	Neg. Adv.	–	– 0.5
T004	ไม่ (not)	Neg. Adv.	–	– 1
T005	ยิ้ม (smile)	V.	Ethic	0.8

contrast, score values from 0 to 1 are assigned for positive words. The advantage of this methodology lies in the combination of negative and positive words. For example, in the Thai sentiment phrase “ไม่ค่อยยิ้มเลย” (“seldom smile”), the word “ไม่ค่อย” (seldom) has a sentiment score of -0.5 , and the word “ยิ้ม” has a sentiment score of 0.8 . Therefore, the resulting score of this sentiment phrase is -0.4 , which represents negative polarity. The Teaching Senti-Lexicon can be used to determine the score of a compound word by multiplying a negative value and a positive value.

Typically, each phrase is a combination of one or more words. Therefore, the sentence (S) from Eq. (4) requires a set of phrases (P), which each consist of several words. P is defined as a subset of S ($P \subset S$) as in Eq. (5).

$$P_i = \{w_1, w_2, \dots, w_n\} \quad (5)$$

where P_i is a set of weight score values (w), which are calculated from sentiment phrase scores.

3.2.4 Data transformation

In this stage, the data are split into a set of words in set P , to be replaced by word types from the Teaching Senti-Lexicon according to the type of word reference that is obtained. The *TypeOfTerm* attribute used to perform a data transformation, which changes words to types of words. The output dataset S is ready to be processed by SPPM.

3.2.5 SPPM

By clause detection in a sentence, which may be complex or without proper grammatical structure, it is difficult to identify the sentiment phrase in each sentence. In this paper, a solution is proposed, which is called SPPM. The sentiment phrase pattern is analysed via association pattern rules is a high possibility, verified with the phrase frequency to be the most effective. The advantage of this method is that it is flexible in detecting the patterns of language. The process of SPPM consists of two main steps: The first step to find the sentiment pattern via the a priori association rule learning algorithm. The association rules, the pattern with the highest probability, and a sample Thai phrase as presented in Table 3, in which the phrase association rules are extracted from the message. For example, the highest-probability phrase pattern is R01: TV \Rightarrow PPA, which means that the phrase has the same type of term as in Table 2. A Teaching Verb (TV) occurs with a Positive Polarity Adverb (PPA) with a probability of 0.81 , which is calculated as $P(TV \cup PPA)/P(TV)$. Whereas R02: Pos. Adv \Rightarrow Adv in this pattern in a Thai phrase contrasts with the English phrase. For example, Pos. Adv occurs with an Adverb in Thai with a probability as high as 0.75 in the evaluation of a phrase that combines one or more words. In addition, this method is better than calculating the polarity word by word. Second, processing is compared the sentiment phrase pattern rules that were obtained in the previous step to determine the best

Table 2 Examples of sentiment pattern association rules

Rule ID	Phrase pattern rule extraction	Conf.	Thai phrase	English phrase
R01	TV \Rightarrow PPV	0.81	สอน (teach) เข้าใจ—(understand)	Teach-understand
R02	Pos. Adv \Rightarrow Adv	0.75	ประทับใจ (impression) มาก— (very)	Very-impression
R03	Neg. Adv \Rightarrow TV	0.69	ไม่ค่อย (seldom) เข้าใจ— (understand)	Seldom-understand
R04	Pos. Adv \Rightarrow Pos. Adv	0.56	เข้าใจ (understand) มาก—(too much)	Understand-too much
R05	TV \Rightarrow Neg. Adv	0.34	สั่งงาน (assigned) เยอะ—(too much)	Assigned-too much

Table 3 Example of sentiment phrase frequency

Phrase (P)	Terms1 (p1)	Terms2 (p2)	Frequency	Probability
สอนเข้าใจ (teach understand)	สอน (teach)	เข้าใจ (understand)	4192	0.349
สอนดี (good teaching)	สอน (teach)	ดี (good)	2437	0.203
ดีมาก (very good)	ดี (good)	มาก (very)	2336	0.195
เข้าใจง่าย (easy to understand)	เข้าใจ (understand)	ง่าย (easy)	984	0.082
ใจดี (very kind)	ใจดี (kind)	มาก (very)	542	0.045
ไม่เข้าใจ (not understand)	ไม่ (not)	เข้าใจ (understand)	444	0.037
มีเทคนิค (has technique)	มี (have, has)	เทคนิค (technique)	399	0.033
เข้าใจดี (good understand)	เข้าใจ (understand)	ดี (good)	365	0.024

matching pattern with the frequency phrase via selected the phrase where have high probability. The processing steps are shown in Algorithm 2.

The frequent sentiment phrases that were obtained from the students' feedback, based on 12,250 instances of training data, for which the sentiment phrases were separated manually and the probabilities were compute using the forward bi-gram traversal method. The highest-frequency phrases are represented by bold lines, which indicate associations between word occurrences, and the thin lines represent low-frequency phrases, which rarely occur. For example, the words “very” and “good” are connected by a bold line, as they appear together in many comments from students, and the a line connects the words “too much” and “assignment”, which less frequently appear together in the comments from students. According to the training set of frequency sentiment phrases, the phrase probabilities are listed in Table 3. The phrase with the highest accuracy is divided into short phrases (P_i) and completes every phrase in a sentence (S_i). These phrases are calculated independently via Eq. (6) to prepare for the next feature selection step.

Algorithm 1: Sentiment Phrase Pattern Matching

Input: instances of S // *MSG = $\{S_1, S_2, \dots, S_n\}$ *//

The number of message (N);

The number of phrase element dataset (P);

The number of association pattern rules (R) that have an element of term type r_i ;

Output: dataset of feedback instance (S);

for $i = 1; i \leq N$ do

Step 1: searching for the best matching between *sentiment phrase pattern rules* and S items

for $j = 1; j \leq P$ do

for $m = 1; m \leq R$ do

if $r_m = tw_m$ then // * Matching pattern of rule with type of word (tw_m) *//

put w_m to P_j // * Collect word (w_m) into phrase P_j *//

Step 2: comparison of the probability of each frequent phrase (pf_j)

if $pf_j > 0$ then

put P_j to S_i

end if

end if

end for

end for

return S

The result of sentiment phrase pattern matching for the sentiment phrase (P_i) is the weight of the sentiment phrase. For example, the feedback message that is presented by the sentence (S_1) “อาจารย์สอนดีค่ะ |เข้าใจง่าย|” (“Teacher are |good teaching| and |easy to understand|”), in the previous sentence, contains two sentiment phrases: P_1 “สอนดี” (Good teaching), which matches with pattern TV \Rightarrow PPV with probability 0.203, and P_2 “เข้าใจง่าย” (easy to understand), which matches with pattern Pos. Adv \Rightarrow Pos. Adv with probability 0.082. Both probabilities are greater than 0. Therefore, both “good teaching” and “easy

Table 4 Example of types of term attributes

Terms	Type of term	Description
สอน (teach), อธิบาย (explain), สั่งงาน (assign)	TV	Teaching verb
มาก/เยอะ (very much), เกินไป (over), เร็ว (fast), ช้า (slow)	QA	Quantity adverb
ไม่ (not/do not), ไม่มี (haven't), ไม่ค่อย (seldom/rarely)	Neg. Adv.	Negative adverb
ดี (good), ดีมาก (very good), เยี่ยม (excellent), สนุก (funny), ชอบ (like), ประทับใจ (impress), พอใจ-มีความสุข (happy), เก่ง (smart), ง่าย (easy), สุดยอด (awesome-thumb up)	PPA	Positive polarity adverb
เสียใจ (sad), เศร้า (upset), หหมดหวัง (hopeless), เครียด (stress), ซีเรียส (serious), แย่ (bad), ไม่ดี (not good), ยาก (difficult), เหนื่อย (tried)	NPA	Negative polarity adverb

to understand” will be considered in the sentiment polarity score by using word score multiplication, as shown in Eq. (6).

$$WP_i = \prod_{i=1}^n w_i \tag{6}$$

where WP_i is the multiplied weight of phrase i in each sentence (S_i); w_i is the word score value, as indicated by the Teaching Senti-Lexicon; and n is the number of words in a phrase.

3.2.6 Feature selection

The weight of a phrase (WP_i), which is calculated via Eq. (6), has a value between -1 and 1 . The weight of a phrase is determined by various attributes, as stated in the Teaching Senti-Lexicon. The attributes consist of teaching skill, teaching planning, teaching material, teacher ethics and miscellaneous, which are utilized for feature extraction by using the *type of term* attribute in Table 4. Each sentiment phrase represents a feature via the type of term. The sentiment values of sentiment phrases will be replaced by the sentiment weights with both positive and negative values.

3.3 Sentiment polarity classification and evaluation

Defining polarity class is very important for the teaching’s attitude preparation, whereby each instance of student messages should be identified as reflecting positive or negative attitudes. The weighted mean of the polarity scores is used for attitude classification via the proposed automatic sentiment polarity from all of the synsets of the Teaching Senti-Lexicon. There are two numerical values of sentiment measurement: Pos(positive) and Neg(Negative). Some students’ response messages in the teaching evaluation forum are quite similar to the mean result. However, only a small number of the teaching evaluation texts are available to measure the sentiment weight score because the messages are specific to teaching evaluation. As a result, this research used the weight of the sentiment phrase (WP_i) and teaching sentiment

function (T_s) to determine the sentiment polarity, as expressed in Eq. (7). The product of weight T_s represents positive polarity when the score is between 0 and 1 ($0 \leq \text{Positive polarity weight} \leq 1$) and negative polarity when the score is between -1 and 0 ($-1 \leq \text{Negative polarity weight} < 0$).

$$T_s = \frac{\sum_{i=1}^N WP_i}{N} \tag{7}$$

where T_s is the teaching sentiment polarity $\begin{cases} 0 \leq T_s \leq 1; & \text{Positive} \\ -1 \leq T_s < 0; & \text{Negative} \end{cases}$, WP_i is weight of phrase $_i$ ($-1 \leq WP_i \leq 1$); and N is the number of phrases in a sentence.

For example, explicit results on the teaching sentiment class, which is assigned by the teaching attitude function (T_s), are shown in Table 5.

The experimental accuracy is calculated from the confusion matrix, as shown in Table 6.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP is the number of true positives, which is the number of instances that are correctly classified into the positive class; TN is the number of true negatives, which is the number of instances that are correctly classified into the negative class; FP is the number of false positives, which is the number of instances that are incorrectly classified into the positive class; and FN is the number of false negatives, which is the number of instances that are incorrectly classified into the negative class.

Then, the obtained results are tested for efficiency by determining the precision, recall and F-measure. Precision is the proportion of the predictions for the positive class or negative class that were correct. The recall or true positive (TP) rate is the proportion of the predictions for the positive class or negative class that were correct, relative to the actual size of the positive class. The accuracy may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases. Therefore, the F-measure is used. It measures the overall precision and recall, according to Eq. (9).

Table 5 Example of sentiment polarity classification

Teaching feedback message (S)	T_s	Polarity
สนุกตื่นเต้นมากค่ะแต่มีเนื้อหาบางอันที่ไม่ค่อยเข้าใจ <i>Funny and exiting</i> class, but some chapters are <i>not understandable</i> .	0.33	Positive
สอนเร็วมากตามไม่ทันเลยค่ะชอบสั่งงานเยอะ Teach too <i>fast</i> or too quickly and assign <i>too much homework</i>	− 1	Negative

Bolditalic indicates a sentiment phrase that has weight score to measure in the teaching sentiment polarity

Table 6 Confusion matrix

Confusion matrix	Predicted = Positive	Predicted = Negative
Actual = Positive	<i>TP</i>	<i>FP</i>
Actual = Negative	<i>FN</i>	<i>TN</i>

$$F\text{-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

4 Experimental results and discussion

The experiments were performed on data that are collected from three universities in Thailand and, after cleaning, 12,222 instances remained. The teaching sentiment classification approach is compared to 13 other classifier algorithms in terms of effectiveness, as shown in Table 7. The proposed algorithm, which is called SPPM, achieves the highest accuracy of 87.94%, compared with the other classifier algorithms which the accuracy in this case will collect from positive and negative class that are predicted to correct class.

The precision and F-measure of SPPM are given as 92.06 and 92.52% in the predicted positive class. One R is the algorithm with the highest recall of 97.50% in the predicted positive class, followed by SVM with 93.44%. One R achieves high recall because the number of instances that are predicted to be Positive is 9889. However, SPPM is able to predict the negative class with the highest recall and F-measure of 67.40% and 68.82%. The number of instances that are correctly predicted to be in the negative class, compared to actual number of instances in the negative class, is 1627 instances.

In this research, we have implemented the SPPM method to analyze teaching evaluation feedback texts (<http://www.thaisentiminig.info>)

with real data from Thailand. The users are tested, and the results are highly satisfactory in the aspects of accuracy, convenience and processing speed. According to the test results of using the proposed sentiment analysis method on comments from the teaching evaluation system from three universities, namely Loei Rajabhat University, Roi-Et Rajabhat University, and Buriram Rajabhat University, based on a 5-point Likert satisfaction scale, user satisfaction in various aspects can be described as follows: (1) The system provided identify sentiments of text messages: very satisfied. (2) The system performance of sentiment analysis for teaching evaluation in shortly: very satisfied. (3) The processing speed of the sentiment analysis system for teaching evaluation is fast: very satisfied. (4) User interface design is easy to use: satisfied. (5) The dashboard of sentiment polarity showed the resulting interpretation of sentiment analysis: satisfied.

The method of sentiment analysis in teaching evaluation using SPPM is presented and applied to a system in Thailand, in which there are many feature functions, such as Thai word parsing, word segmentation, sentiment phrase analysis, sentiment analysis and the dashboard. In addition, the system can increase or edit the sentiment vocabulary and update the sentiment weight scores in the Teaching Senti-Lexicon. The results of sentiment analysis of teaching evaluation are obtained from students' responses. SPPM could analyze all of the feedback messages from each university. However, there are limitations regarding the disclosure of the list of subjects and teachers. In this case, the privacy policies of each institute do not allow publication of the information on the subjects and teachers. For this reason, SPPM extracted only the sentiment polarity of teaching feedback, without subject attributes. However, SPPM achieves high accuracy for both positive and negative classes. These results are beneficial to the readability of teacher evaluations and in visualize phrase sentiment analysis to address areas of improvement in teaching evaluations.

Table 7 Overall correctness percentage comparison

Algorithm	Class	Confusion matrix	Accuracy	Precision	Recall	F-measure
SPPM (Proposed Model)	Positive	9121 687	87.94	92.06	93.00	92.52
	Negative	787 1627		70.31	67.40	68.82
ID3	Positive	8900 908	80.98	86.27	90.74	88.45
	Negative	1417 997		52.34	41.30	46.17
One R	Positive	9889 254	82.30	83.29	97.50	89.83
	Negative	1984 515		66.97	20.61	31.52
Naïve Bayes	Positive	8323 1323	79.85	88.26	86.28	87.26
	Negative	1107 1307		49.70	54.14	51.82
SVM	Positive	9165 643	86.81	90.44	93.44	91.92
	Negative	969 1445		69.20	59.86	64.19
BFTree	Positive	8805 1003	80.12	86.05	89.77	87.87
	Negative	1427 987		49.60	40.89	44.82
SimpleLogistic	Positive	9292 500	86.70	89.21	94.89	91.96
	Negative	1124 1290		72.07	53.44	61.37
Logistic	Positive	9298 510	86.74	89.33	94.80	91.98
	Negative	1111 1303		71.87	53.98	61.65
BayeNet	Positive	8787 1022	84.78	91.29	89.58	90.43
	Negative	838 1576		60.66	65.29	62.89
Stacking	Positive	9400 408	80.29	82.45	95.84	88.64
	Negative	2001 413		50.30	17.11	25.53
Ada Boost	Positive	8838 970	81.83	87.60	90.11	88.84
	Negative	1251 1163		54.52	48.18	51.15
Attribute selected classifier	Positive	9404 404	84.60	86.42	95.88	90.90
	Negative	1478 936		69.85	38.77	49.87
Zero R	Positive	8838 970	79.44	85.14	90.11	87.55
	Negative	1543 871		47.31	36.08	40.94
Hoeffding tree (VFDT)	Positive	8810 998	81.18	87.12	89.82	88.45
	Negative	1302 1112		52.70	46.06	49.16

Bold value indicates the best results per column

5 Conclusions

This research paper focuses on the investigation of responses to open-ended questions as student feedback from a teaching evaluation system. Generally, it aimed to assess teacher evaluations in order to improve the quality of education. Specifically, the newly designed model is vitally useful for an individual teacher to improve his or her teaching, especially for effectively revising his or her teaching strategies to directly cater to certain needs of students and for more effective classroom management. If considering the evaluation derived from closed-end questionnaires, the proposed model, which is called SPPM, performs teaching evaluation through sentiment analysis from open-ended responses. It determines the students' sentiments that are hidden in the response texts and is beneficial for applications in the field of education. SPPM was able to extract the patterns of language in a case study in the Thai language, which is different from other languages; this is a problem in case studies of sentiment

analysis. The patterns determine and filtering the phrase that appeared a new type of term and can be collected in the system for evaluation of the sentiment score of a word in the Teaching Senti-Lexicon. The results of teaching evaluation by using SPPM show that opinion mining can be effectively applied to identify students' attitudes towards classroom teaching. Moreover, SPPM can be applied to perform sentiment phrase pattern matching to other languages and to automatically create repositories of other languages that may be considered in the future.

References

1. Khan FH, Bashir S, Qamar U (2014) TOM: twitter opinion mining framework using hybrid classification scheme. *Decis Support Syst* 57:245–257
2. Peng Y, Kou G, Li J (2014) A fuzzy PROMETHEE approach for mining customer reviews in Chinese. *Arab J Sci Eng* 39(6):5245–5252

3. Quan C, Ren F (2014) Unsupervised product feature extraction for feature-oriented opinion determination. *Inf Sci* 272:16–28
4. Sun S, Luo C, Chen J (2017) A review of natural language processing techniques for opinion mining systems. *Inf Fusion* 36:10–25
5. Bilici E, Saygin Y (2017) Why do people (not) like me?: mining opinion influencing factors from reviews. *Expert Syst Appl* 68:185–195
6. Martín-Valdivia MT, Martínez-Cámara E, Perea-Ortega JM, AlfonsoUreña-López L (2013) Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Syst Appl* 40(10):3934–3942
7. Natek S, Zwilling M (2014) Student data mining solution–knowledge management system related to higher education institutions. *Expert Syst Appl* 41(14):6400–6407
8. Şen B, Uçar E, Delen D (2012) Predicting and analyzing secondary education placement-test scores: a data mining approach. *Expert Syst Appl* 39(10):9468–9476
9. Jing LV, Yanqing Z (2012) Teaching evaluation method based on least squares support vector machine and chaos particle swarm optimization algorithm. *JDCTA: Int J Digit Content Technol Appl* 6(11):343–351
10. Wu F, Song Y, Huang Y (2016) Microblog sentiment classification with heterogeneous sentiment knowledge. *Inf Sci* 373:149–164
11. Rao Y, Li Q, Mao X, Wenyin L (2014) Sentiment topic models for social emotion mining. *Inf Sci* 66:90–100
12. Ceron A, Curini L, Iacus SM (2016) iSA: a fast, scalable and accurate algorithm for sentiment analysis of social media content. *Inf Sci* 367–368:105–124
13. Khan FH, Qamar U, Bashir S (2016) eSAP: a decision support framework for enhanced sentiment analysis and polarity classification. *Inf Sci* 367–368:862–873
14. Chamlerwat W, Bhattarakosol P, Rungkasiri T, Haruechaiyasak C (2012) Discovering consumer insight from twitter via sentiment analysis. *J Univers Comput Sci* 18(8):973–992
15. Esuli A, Sebastiani F (2006) SENTIWORDNET A publicly available lexical resource for opinion mining. In: *Proceedings of the 5th conference on language resources and evaluation (LREC'06)*, pp 417–422
16. Xue T (2012) Study on the wushu teaching evaluation in high schools with 2-tuple linguistic information. *Adv Inf Sci Serv Sci* 4(9):107–113
17. Leong CK, Lee YH, Mak WK (2012) Mining sentiments in SMS texts for teaching evaluation. *Expert Syst Appl* 39(3):2584–2589
18. Naradhipa AR, Purwarianti A (2012) Sentiment classification for Indonesian message in social media. In: *Proceedings of 2012 international conference on cloud computing and social networking (ICCCSN)*
19. Yu Y, Duan W, Cao Q (2013) The impact of social and conventional media on firm equity value: a sentiment analysis approach. *Decis Support Syst* 55(4):919–926
20. Xianghua F, Guo L, Yanyan G, Zhiqiang W (2013) Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowl Based Syst* 37:186–195
21. Wu J, He Z, Gu F, Liu X, Zhou J, Yang C (2016) Computing exact permutation p-values for association rules. *Inf Sci* 346–347:146–162
22. Ghiassi M, Skinner J, Zimbra D (2013) Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst Appl* 40(16):6266–6282