

Research Article

Lexicon-Based Sentiment Analysis of Teachers' Evaluation

Quratulain Rajput, Sajjad Haider, and Sayeed Ghani

Faculty of Computer Science, Institute of Business Administration (IBA), Garden/Kiyani Shaheed Road, Karachi 74400, Pakistan

Correspondence should be addressed to Quratulain Rajput; qrajput@iba.edu.pk

Received 13 July 2016; Revised 31 August 2016; Accepted 5 September 2016

Academic Editor: Francesco Carlo Morabito

Copyright © 2016 Quratulain Rajput et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The end of the course evaluation has become an integral part of education management in almost every academic institution. The existing automated evaluation method primarily employs the Likert scale based quantitative scores provided by students about the delivery of the course and the knowledge of the instructor. The feedback is subsequently used to improve the quality of the teaching and often for the annual appraisal process. In addition to the Likert scale questions, the evaluation form typically contains open-ended questions where students can write general comments/feedback that might not be covered by the fixed questions. The textual feedback, however, is usually provided to teachers and administration and due to its nonquantitative nature is frequently not processed to gain more insight. This paper aims to address this aspect by applying several text analytics methods on students' feedback. The paper not only presents a sentiment analysis based metric, which is shown to be highly correlated with the aggregated Likert scale scores, but also provides new insight into a teacher's performance with the help of tag clouds, sentiment score, and other frequency-based filters.

1. Introduction

Evaluating performance of faculty members is becoming an essential component of an education management system. It not only helps in improving the course contents and quality but is also often used during the annual appraisal process of faculty members. The evaluation is typically collected at the end of each course on a set of question which are answered using Likert scale. The aggregate of the responses is used as a metric to gauge the teaching quality of the concerned faculty member.

The evaluation form, however, also provides room for open feedback which typically is not included in the performance evaluation/appraisal due to lack of automated text analytics methods [1–3]. The textual data may contain useful insight about subject knowledge of the teacher, regularity, and presentation skills and may also provide suggestions to improve the teaching of quality. Such information may not be readily extracted from the Likert scale based feedback [4]. However, getting sense out of the textual feedback manually is a laborious task and, as a result, the textual feedback is not properly utilized [3].

This paper aims to analyze textual feedback automatically and to develop quantitative and qualitative metrics that can aid in assessing a teacher's performance and highlighting her major areas of appreciations/concerns. The work comes under the emerging area of sentiment analysis which has gained prominence since the revolution of the World Wide Web. A lot of work has been reported recently where researchers have extracted sentiments from comments posted on online forums [5], movies/items review sites [6, 7], social networking sites [8, 9], teachers evaluation [3, 10], and so forth.

The primary focus of the sentiment analysis is to determine a writer's feeling from a given text. The feeling might be his/her attitude, emotion, or opinion. The most important step of this analysis is to classify the polarity of the given text as positive, negative, or neutral [5, 11]. In a similar fashion, the presented work aims to identify the polarity of a student's feedback in terms of positive, neutral, and negative. In addition, the paper also suggests methods to identify the recurring theme in students' feedback by generating word clouds for visualization, sentiment score, and other frequency-based filters.

The rest of the paper is organized as follows. Section 2 provides a brief survey of the field of sentiment analysis. Section 3 discusses the presented approach for analyzing students' feedback while the results are presented in Section 4. Finally, Section 5 concludes the paper and provides future research directions.

2. Related Work

The field of sentiment analysis is an exciting and new research direction to discover people's sentiments. The text on which sentiment analysis is generally performed can be categorized broadly into two types: facts and opinions. Facts are objective expressions about entities, events, and their properties. Opinions are usually subjective expressions that describe sentiments, emotion, appraisal, or feeling [12]. During the last decade, automatic extraction of opinion or sentiment from text has been an active area of research [11]. The existing approaches can be classified into five main points categorized as listed in the following [13].

(i) *Keyword Spotting*. Classify the text based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Keyword spotting has a limitation in two areas: (a) it cannot reliably recognize affect negated words, and (b) it relies on surface features. For instance, both the sentences such as "today was a happy day" and "today was not a happy day at all" are classified as positive due to the presence of positive affect word "happy." Sometimes, a sentence conveys affect through underlying meaning rather than affect adjectives. For example, the text "My husband just filed for divorce and he wants to take custody of my children away from me" evokes strong emotions but uses no affect keywords.

(ii) *Lexicon-Based Approaches*. This approach can be further classified into dictionary based approaches and corpus based approaches. In the *dictionary* based approach, a small set of opinion words is collected manually as a seed. Then well-known dictionaries [14] or thesaurus [15] are used to expand the set of opinion words by adding their synonyms and antonyms. The newly found words are added to the seed list. The process continues until no more words are found in the dictionary. In the end, manual review is carried out to remove errors. One of these approaches is proposed by Kim and Hovy [16]. The major drawback of this approach is that it is unable to find a domain and context specific opinion words. For example, consider the following two sentences:

"Teacher is fine"

"There will be fine on late payment"

The word *fine* is used in both sentences. In the first sentence, it has positive sentiment while in the other it has negative sentiment.

The *corpus based approach* overcomes the limitation of the dictionary based approach. In addition to the seed word list, this approach identifies context specific opinion words. The finding of such words is based on syntactic or

cooccurrence pattern in the text using linguistic constraints [17]. For example, consider the following two sentences:

"The teacher is good and the course is fine"

"The teacher is good but the course is difficult"

In the first sentence, the linguistic connective "*and*" shows next information with the same sentiment. However the connective word "*but*" is following with the contradictory statement. This approach is applied in the presented work and currently only considers individual word sentiments while cooccurrence pattern will be tested in the future work.

(iii) *Lexical Affinity*. This approach trains probability from linguistic corpora [5]. It not only detects obvious affect words but also assigns sentiment to arbitrary words. For example, lexical affinity might assign the word "accident" a 75 percent probability of indicating a negative effect, as in the sentence "hurt by a car accident." There are two main problems with this approach. First, negated sentences (I avoided an accident) and sentences with other meanings (I met my girlfriend by accident) pretend same lexical affinity because they operate solely on the word level. Second, lexical affinity probabilities are often biased toward the text of a particular kind.

(iv) *Statistical Methods*. This method makes use of machine learning methods such as Bayesian inference and Support Vector Machines [18, 19]. By feeding a large training corpus of affectively annotated texts to a machine learning algorithm, the system not only learns the shades of affect keywords (as in the keyword spotting approach) but also takes into account the valence of other arbitrary keywords, punctuation, and word cooccurrence frequencies. It has been mentioned that the performance of these machine learning based approaches highly depends upon the quality and quantity of the training data and feature selection [3, 5, 20, 21].

(v) *Concept Level Techniques*. Unlike purely syntactical techniques, concept level approaches are leveraged on elements from knowledge representation such as ontologies and semantic networks and, hence, can detect semantics that are expressed in a subtle manner [22, 23].

In all the above approaches, the fundamental step of the sentiment analysis is to identify the polarity as positive, negative, or neutral on either a word level, sentence level, or document level. To facilitate this step, various subjectivity corpora are developed that annotate the lexicon of words with positive, negative, and neutral polarity [24, 25]. In applications where a domain is unknown, the use of such corpus may work fine. However, it is argued that better results can be achieved using a domain specific sentiment language. For this reason, several sentiment annotated corpora have been made freely available. The list includes the MPQA newswire corpus [26], movie review corpus [19], and restaurant and laptop review corpus [5]. It must be noted that, as of now, no such sentiment corpus is available exclusively for the educational domain.

Sentiment analysis can be applied to a broad range of real world problems. Many businesses are adopting text and

TABLE 1: Samples of student's textual feedback.

	Sample students feedback
1	A good teacher
2	I felt that the instructor could've done a better job in making the concepts clear
3	Dull course
4	A course is fine
5	<i>Difficult</i> course, <i>great</i> teacher and also <i>able</i> to relate it to the <i>practical</i> knowledge
6	This course increased my knowledge. Teacher is also very helpful I like this course very very much :)
7	Institute should reconsider its policy of fyp's

sentiment analysis and incorporating it into their processes. Applications in businesses include (i) computation of *customer satisfaction* to get an idea of how happy customers are with a company's products from the ratio of positive to negative tweets about them; (ii) *identification of critics and promoters*: it can be used by customer support for spotting dissatisfaction or problems with products. Several research groups around the world are currently focusing on understanding the dynamics of sentiment in electronic communities through sentiment analysis and CYBEREMOTION [27] is one such project funded by the 7th framework initiative of the European Union with participation of nine institutes in Europe. The primary goal of this project is to understand the role of collective emotions in electronic communities.

In the medical domain, physicians and nurses express their judgments and observations on a patient's health status in clinical narratives. Sentiments in clinical documents differ from the sentiments in user generated content or other text types. Analyzing and aggregating this information over time can support the treatment decisions by allowing a physician to quickly get the health status overview of a patient. In this way, labor intensive user studies for treatment or medication evaluation can be facilitated. Goeuriot et al. [28] performed dictionary based sentiment analysis on clinical text.

In the education domain, students express their opinions about a teacher's teaching abilities. Due to the unavailability of automated tools to process such feedback, however, these comments are not properly utilized. Few efforts have been reported in the literature that aim to address this issue. MacKim and Calvo [1] analyze student's feedback to evaluate their learning experience. The study compares categorical model and dimensional model making use of five emotion categories: anger, fear, joy, sadness, and surprise. Joy and surprise are taken as a positive polarity, whereas anger, fear, and sadness belong to negative polarity.

Leong et al. [2] applied sentiment analysis and text mining by collecting student's feedback that is collected through SMS. They also explored the incomplete text and spelling errors. Each feedback has been categorized based the concepts defined for each category. Each feedback can belong to either no category, one category, or several categories. Jagtap and Dhotre [3] used SVM and HMM based hybrid approach for sentiment analysis of teachers' assessment.

The work by Altrabsheh et al. [10] analyzed student's feedback by collecting via social media such as Twitter. They

not only identified student's feeling in terms of positive and negative but also identified some more refined emotions. Emotions can be negative such as confused, bored, and irritated while positive emotions such as confident and enthusiastic are considered. Different techniques have been used in sentiment analysis, and a few have proved to give superior performance such as Naive Bayes (NB), Max Entropy (MaxEnt), and Support Vector Machines (SVM).

3. Sentiment Analysis Process

This section explains the presented approach to analyze the textual faculty evaluation provided by students. The approach identifies the polarity of a comment as positive, negative, or neutral. A metric has been suggested that provides an alternate to the Likert scale score. In addition, the section also explains how keywords are identified that could assist both the instructor and the administration in pinpointing the major areas for improvement.

To better understand the issues in analyzing sentiments of textual faculty evaluation, consider the few students' feedback shown in Table 1. The comments are comparatively short and highlight a number of challenges. There may be issues of spelling mistakes, grammatical errors, and use of abbreviated shortcuts and emoticons. Besides, sentences may be incomplete and may contain filler words (e.g., um and well). Therefore these kinds of text can be categorized as noisy and unstructured due to its informal writing style. The issues highlight the difficulty in accurately extracting the sentiments from such unstructured and incoherent feedbacks.

Since the students' feedback is available in an unstructured (textual) form, several subprocesses are necessary to generate meaningful insight from it. These include data gathering, data preprocessing, stop words filtering, name entity recognition, transformation, sentiment tagging, sentiment score, and word cloud visualization. The overall process is shown in Figure 1 while the major building blocks are explained as follows.

3.1. Preprocessing. The aim of preprocessing is to remove the unwanted and noisy data. In this paper, the preprocessing stage comprises the following tasks:

- (i) *Tokenization.* This process breaks a stream of text into a list of words.

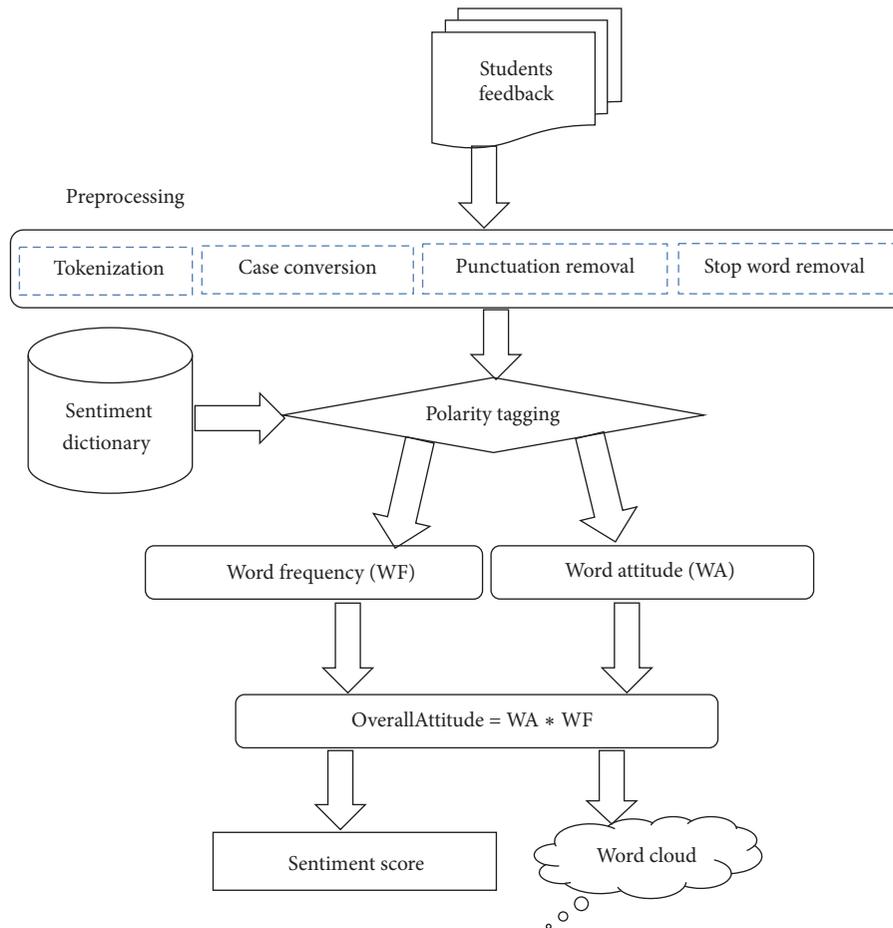


FIGURE 1: Sentiment analysis process.

- (ii) *Stemming*. This step reduces words to their stem or root form as stemming simplifies the sentiment analysis process. The same word can be used in a different flavor for grammatical reasons such as organize, organizes, organizing.
- (iii) *Case Conversion*. This step changes the text into either the lowercase or the uppercase.
- (iv) *Punctuation Removal*. Punctuations in a text generally do not provide any useful information. This step, therefore, erases the punctuation characters from the word.
- (v) *Stop Word Removal*. Stop words consist of prepositions, help verbs, articles, and so forth. They typically do not contribute in analyzing sentiments and are removed from the text.

3.2. *Sentiment Dictionary*. A sentiment dictionary contains a list of words along with their respective polarity. Several such corpora have been developed and are made freely available. In this work, we have used the MPQA corpus [26], consisting of 8221 records (words) where each record consists of six features as listed in the following:

- (i) *type*: representing whether the word has strong or weak subjectivity;
- (ii) *len*: length of the clue in words;
- (iii) *word1*: token/word as the subjective clue;
- (iv) *pos1*: part of speech of the word;
- (v) *stemmed1*: meaning that if the word is a stemmed word, assign “y,” otherwise “n.” If $\text{stemmed1} = y$, this means that the clue should match all unstemmed variants of the word with the corresponding part of speech. For example, “abuse” will match “abuses” (verb), “abused” (verb), and “abusing” (verb), but not “abuse” (noun) or “abuses” (noun);
- (vi) *prior polarity*: which specifies the sentiment of the word as positive, negative, or neutral. Figure 2 shows the words count with respect to their polarity in sentiment dictionary.

3.3. *Polarity Tagging*. This step analyzes each word in a student’s feedback and tags the word as positive, negative, and neutral using its polarity in the sentiment dictionary. The neutral words are removed from the data as they do

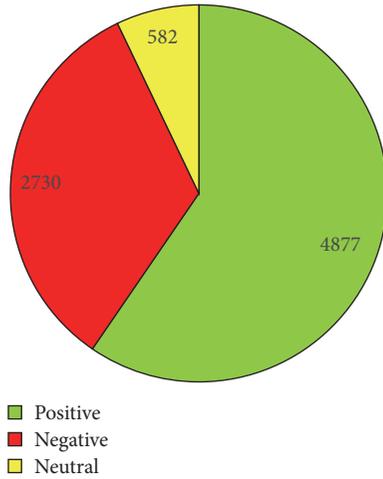


FIGURE 2: Words in sentiment dictionary.

not provide any subjectivity clue. For example, consider a student’s feedback as shown in row 5 of Table 1:

“*difficult* course, *great* teacher and also *able* to relate it to the *practical* knowledge.”

In this example, after the preprocessing step, each word would have been tagged with their polarity as found in the sentiment dictionary. Here the word *difficult* is tagged with negative polarity while *great*, *able*, and *practical* are tagged with positive polarity as shown in the following:

Sentiment words	→	difficult	Great	able	practical
Word polarity	→	negative	Positive	positive	positive

3.4. *Word Frequency.* This step computes the frequency of each word in each comment. In the above example, each word occurs only one time; therefore frequency of each word is one:

Sentiment words	→	difficult	great	able	practical
Word frequency	→	1	1	1	1

3.5. *Word Attitude.* This step converts the polarity of each word into a numeric value to perform further computation. The conversion formula is given as follows:

$$\text{Word Attitude} = \begin{cases} 1, & \text{if priorpolarity} = \text{positive;} \\ -1, & \text{if priorpolarity} = \text{negative.} \end{cases} \quad (1)$$

Sentiment words	→	difficult	Great	able	practical
Word Attitude	→	-1	+1	+1	+1

For the above example, the word *difficult* has negative polarity, so its attitude is -1 while *great*, *able*, and *practical* have positive polarity so their attitude is 1.

3.6. *Overall Attitude.* The overall attitude of a word is obtained by multiplying its attitude with its frequency.

$$\text{OverallAttitude} = \text{Word Attitude} * \text{Word Frequency.} \quad (2)$$

Continuing with the same example of a student’s feedback (row 5 in Table 1), the overallAttitude is computed using (2). Since the frequency of each sentiment word is 1, the overallAttitude is either -1 or +1 as shown in the following:

Sentiment words	→	difficult	Great	able	practical
OverallAttitude	→	-1	+1	+1	+1

3.7. *Word Cloud Visualization.* The overall attitude of each sentiment word from the given list of comments is used to draw a word cloud.

3.8. *Sentiment Score.* In this step, each feedback comment is assigned a sentiment score by adding the overallAttitude of each word in a feedback. This score is then used to evaluate a teacher’s performance.

$$\text{Sentiment Score} = \sum_{i=1}^n \text{overallAttitude} (i), \quad (3)$$

where n is the number of positive and negative words in a feedback and i represents a particular word. For instance, in the above example, the word *difficult* appears one time; therefore its frequency is 1 while its attitude is negative. Thus, the overallAttitude of the word *difficult* is -1. On the other hand, the words *great*, *able*, and *practical* also appear once, so their frequency is 1. Since their attitude is positive, the overallAttitude of all these words is 1. The sentiment score is computed by adding overallAttitude of all positive and negative words while ignoring the neutrals. The sentiment score is the summation of overallAttitude of each sentiment word in a feedback. For the given example, the sentiment score thus becomes

$$\text{Sentiment Score} = (-1) + 1 + 1 + 1 = 2. \quad (4)$$

4. Results

This section applies the presented approach on the textual feedback of students provided at the end of various courses conducted at our institution. The data set comprises 1748 students’ feedback provided at the end of 63 courses conducted during 2010 and 2014. The few samples of students’ feedback are shown in Table 1. The student enrollment in these courses was in the range of 25 to 45.

The sentiment analysis is performed using Knime which provides an open source data analytics platform. It allows users to create workflow and integrate various components such as data mining and text processing. According to the approach discussed in the previous section, a Knime workflow is developed using its text processing component which is shown in Figure 3. The Knime workflow is divided into five main components:

- (i) *Reading Sentiment Dictionary.* This group of nodes reads the sentiment dictionary (subjectivity corpus)

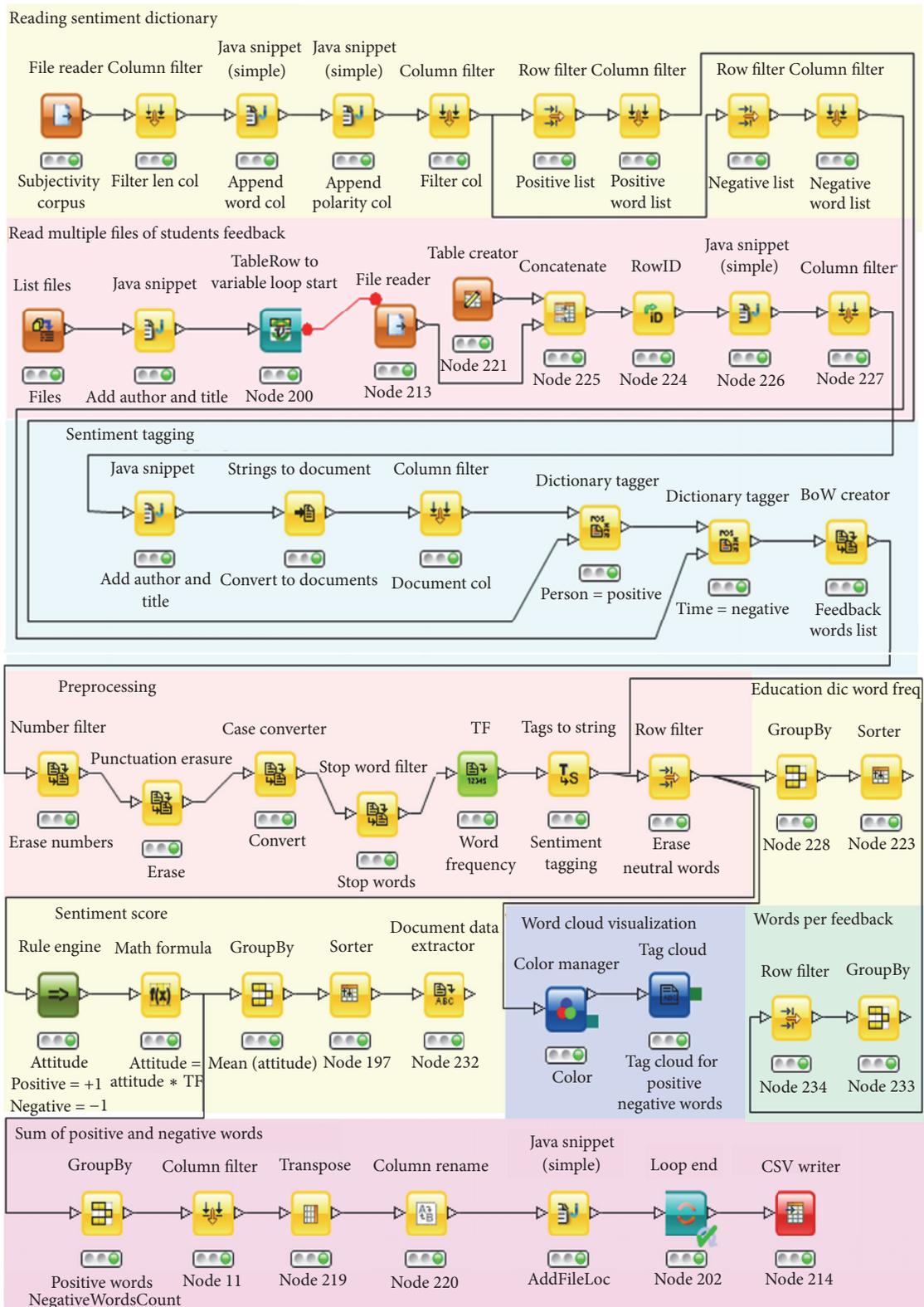


FIGURE 3: Kettle Sentiment analysis workflow.

and separates the words into two sublists: positive words and negative words.

- (ii) *Read Files of Student's Feedback.* A file represents student's textual feedback in a course. The group of nodes reads students' feedback from each file.
- (iii) *Sentiment Tagging.* This group of nodes reads textual feedback and creates a list/bag of words. The words are then tagged as positive or negative using the polarity of words specified in the sentiment dictionary.
- (iv) *Preprocessing.* This group of nodes erases data that has no subjectivity clue such as stop words, punctuations, and also performs additional tasks to reduce noise that include case conversion, stemming, and so forth.
- (v) *Sentiment Analysis.* This group of nodes computes the sentiment score and provides word cloud visualization to be used for teachers' evaluation.

4.1. Sentiment Dictionary Modification. It must be stated that due to an informal way of writing, the extraction of positive and negative words involves a lot of challenges. The polarity of some of the vocabulary used by students in education environment needs to be modified in sentiment subjectivity corpus. Consider the following few student's feedback:

- (i) Good
- (ii) ok
- (iii) Good teacher. I do not have anything negative to say
- (iv) Great teacher while marking is bit strict
- (v) Fun teacher
- (vi) Fine teacher
- (vii) Miss xyz is awesome teacher
- (viii) Awsome teacher, extremely effective learning, and practice the knowledge learned in the course. Thumbs-up

If one follows the standard sentiment dictionary, then some of the above words would be classified as negative. For example, the words *miss*, *lecture*, *fine*, and *fun* are assigned negative polarity by the existing general purpose sentiment dictionary. However, in the context of students' feedback, they are not considered negative. The word *miss* refers to a teacher while *lecture* refers to a class session. The polarity of these words, thus, should be neutral. On the other hand, the words *fine* and *fun* are typically used in a positive sense in an academic environment so their polarity should be positive. In lieu of these differences, the polarity of certain words has been modified in the sentiment dictionary. The list of these words is shown in Table 2.

The distribution of words in terms of polarity is listed in Table 3. The table clearly shows that out of 4877 positive words and 2730 negative words available in the sentiment dictionary, a very small number of unique words have been utilized in the students' feedback. These words could correspond to the educational sentiment vocabulary.

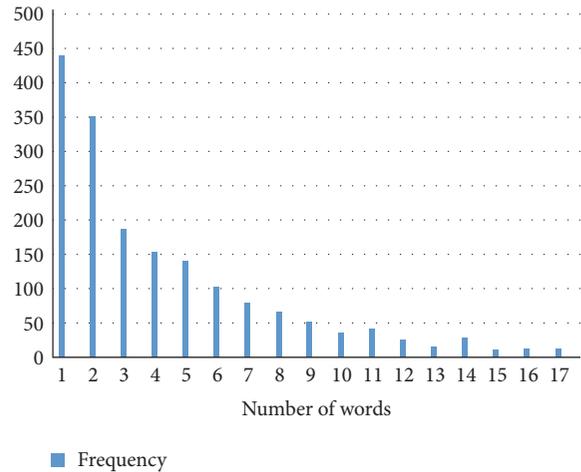


FIGURE 4: Bar chart of number of words in students' comments.

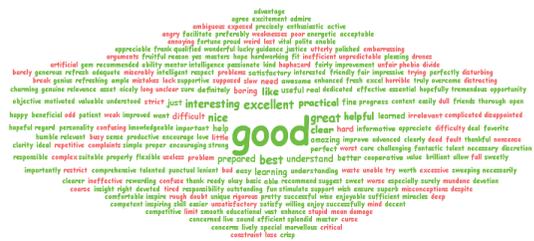


FIGURE 5: Word cloud of student's feedback.

Another observation made from the textual feedback is that comments are typically short as students avoid writing long sentences. Many times, they write only two or three words. Figure 4 shows a bar chart of number of words in students' comments across all 1748 cases. The average across all the cases is nine words per comment.

4.2. Feedback Analysis Using Word Cloud. The word cloud visualization is an excellent way to communicate the findings. For this research, the built-in word cloud visualization feature of Knime has been used. It is called after the preprocessing step as shown in Figure 3. An example word cloud from the complete 1748 comments is shown in Figure 5. The words in red color are those having negative polarity while the words in green have positive polarity. The most frequent words are *good*, *interesting*, *excellent*, *practical*, *great*, *helpful*, and so forth and are shown in bold fonts. This kind of visualization is very useful for the administration to better understand students' point of view regarding a particular course or a teacher. Word clouds can also aid in identifying trends and patterns that otherwise are difficult to identify from reading comments. In addition, the word cloud visualization also helps to track the progress of a teacher over a particular span of time. For example, the word clouds shown in Figures 6 and 7 are of two similar courses taught by the same teacher in two different semesters, Fall 2010 and Spring 2012. It can

TABLE 2: Modified polarity of words in sentiment dictionary.

Word	Original polarity	Modified polarity
Fun	Both positive and negative	Positive
Fine	Both positive and negative	Positive
Miss	Negative	Neutral
Challenge	Negative	Positive
extreme, extremely	Negative	Neutral
Thumb	Negative	Neutral
Object	Negative	Neutral
Overcome	Negative	Positive
Lecture	Negative	Neutral
Negative	Negative	Neutral
Concern, concerned, concerns	Negative	Positive

TABLE 3: Polarity based distribution of words.

	Polarity	
	Positive	Negative
Original sentiment dictionary	4877	2730
Words utilized from original dictionary	213	100
Words utilized from modified dictionary	221	87

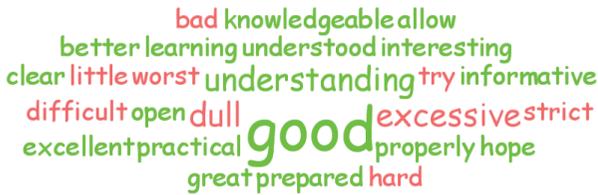


FIGURE 6: Word cloud of a course offered in Fall 2010.

be observed that the teacher had more negative words in Fall 2010 while her teaching style improved in Spring 2012 as highlighted by a lesser number of negative words. This type of temporal word cloud can aid administration in tracking the progress of a teacher over a period of time.

4.3. *Feedback Analysis Using Sentiment Score.* The paper suggests the computation of a sentiment score for each textual feedback as explained in Section 3. The computation is demonstrated in Table 4 which lists few comments, their computed sentiment score, and actual sentiments. A negative sentiment score suggests a negative comment and likewise, a positive score indicates a positive comment. A score of 0 indicates a mixed feedback (equal number of positive and negative words). The last column shows the actual sentiments as identified by reading the feedback manually. For example, the sentiment score of comment 1 is computed as positive and is also identified as such. However, in few cases, the feedbacks are classified incorrectly. For example, the sentiment score of comment 2 is computed as negative whereas actual sentiment

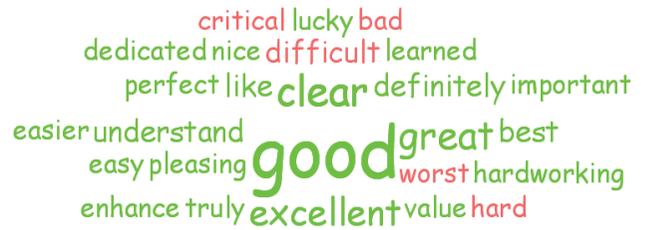


FIGURE 7: Word cloud of a course offered in Spring 2012.

is positive. The reason for this incorrect classification is that currently the paper focuses on unigrams and does not incorporate bigram or higher order n -gram which would have resulted in correct classification. Similarly, few words (such as late and casual) need to be added to the dictionary as this would further increase the accuracy of the presented approach.

Table 5 shows a confusion matrix of sentiment identification as predicted by sentiment scores against the actual sentiments. The confusion matrix helps in understanding the applicability of the presented approach. The table shows that there were 1028 positive feedbacks in our data set and 1002 were classified correctly. Similarly, 94 out of 176 negative comments were identified correctly. In the case of neutral feedback (mix of positive and negative), 29 out of 30 feedbacks were classified correctly, and only 1 was identified as negative. Thus, the proposed approach was able to achieve an accuracy of 91.2%.

To further analyze the performance of the proposed sentiment classification approach, recall, precision, and F -measure were computed for all three classes as reported in Table 6. The best performance was for positive sentiments as it had the highest recall and precision rates. Mixed cases were very few in numbers (only 30) but the approach still managed to get a very good recall although the precision was not very high due to a small number of records. The lower number of records also affects the performance of the negative feedback which has a high recall but a low precision rate.

TABLE 4: Sentiment score of the selected student's feedback.

Student's feedback	Sentiment score	Actual sentiment
Good instructor, a very nice person	2	Positive
The topic is always well explained and the teacher never gets <i>angry</i> for asking too many questions	-1	Positive
A good teacher	1	Positive
Lab was <i>difficult</i> because of size of students. Nobody knows what is going in front portion of the class	-1	Negative
Explains <i>easy</i> questions in class and is giving extreme <i>difficult</i> questions in exam	0	Negative
The number of students in class is high due to which teacher faces <i>problem</i> in paying attention to every student	-1	Negative
Good effort by teacher. Some concepts are <i>unclear</i> , needs to be explained in more detail	0	Mixed
This course is very <i>important</i> for all CS students but the teacher needs to change her casual behavior (most of our classes started late)	1	Negative

TABLE 5: Classification of feedbacks using sentiment score.

		Identified by the presented approach			
		Positive	Negative	Mixed	
Actual	Positive	1002	6	20	1028
	Negative	60	94	22	176
	Mixed	0	1	29	30
		1062	101	71	1234

TABLE 6: Performance measure of sentiment score.

	Recall	Precision	F-measure
Positive	0.97	0.94	0.95
Negative	0.53	0.93	0.67
Mixed	0.97	0.4	0.57

4.4. Comparison between Sentiment Score and Likert Scale Based Teacher Evaluation. This section compares the performance of the sentiment score metric against the Likert-based scores. As shown in the previous subsection, the range of sentiment score for each comment is either negative, positive, or mixed. The Likert scale, on the other hand, has the range between 0 and 5 for a particular course. To bring the sentiment score metric comparable to the Likert scale, a new metric, termed sentiment result, has been suggested and its computation is shown in

$$\text{sentiment result} = \left(\frac{\sum \text{positive}}{\sum \text{positive} + \sum \text{negative}} \right) * 5. \quad (5)$$

The equation computes the sum of positive and negative words found in all feedback of a given course. To avoid zero summation, a prior sampling scheme is utilized which assumes that 2 positive and 2 negative words already exist in each comment. Only those courses are considered that have at least 20 positive and negative words available. By enforcing this threshold, 51 courses were considered out of the initial 63. Table 7 shows the sentiment results of few courses and compares them against the Likert-based scores while Figure 8 shows the values of both metrics across all 51 courses. The correlation of both scores was found to be 0.64. Together with word cloud and sentiment score, the suggested sentiment result metric can provide further insight into a teacher's performance. The Likert scale scores are based on predetermined questions while textual feedbacks are provided as open-ended comments. Due to this reason, students write other positive/negative points which are not specifically asked in the Likert scale based questions. Thus, the sentiment score, in addition to word cloud, gives more insight into a teacher's performance.

5. Conclusion

This paper performed sentiment analysis on faculty evaluation provided by students at the end of a course. A Knime

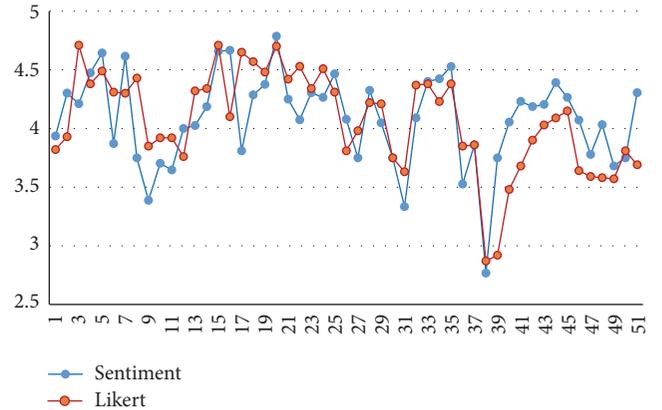


FIGURE 8: Correlation in the sentiment and Likert scale based results.

workflow was developed using its text processing component for sentiment analysis of students' feedback. The presented approach suggests the computation of sentiment score to classify the feedback as either positive, negative, or neutral. To measure its performance, accuracy, recall, precision, and F-measure were computed and the results were found to be very positive. The paper also demonstrated that the sentiment score is comparable to aggregated Likert scale based score. However, the sentiment score, in addition to word clouds, gives more insight that is not possible with the Likert-based score. This is due to the fact that the Likert scores are computed from a predetermined questionnaire that restricts students to comment beyond what is asked in those questions. Textual feedback, on the other hand, is open-ended. The paper also suggested a modified subjectivity corpus to be used in the academic domain to achieve better results. Finally, it showed how word clouds visualization techniques can aid in getting unique insight into a teacher's performance which is typically not available via Likert-based scores. The future work would focus on analyzing bigram and higher order n -gram for computing the sentiment scores. In addition, more words, often used in academic environment, would be added to the sentiment dictionary.

Competing Interests

The authors declare that they have no competing interests.

TABLE 7: Textual feedback result in scale of five.

Courses	Positive words	Negative words	Sentiment result	Likert scale result
CS01	29	5	4.26	4.51
CS02	22	3	4.4	4.3
CS03	66	22	3.75	2.92
CS04	22	5	4.0	4.5

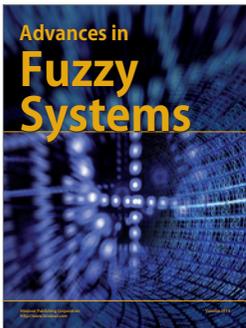
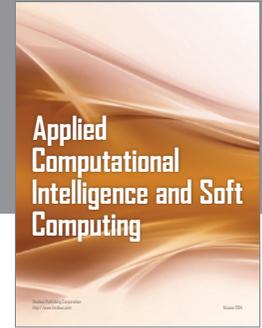
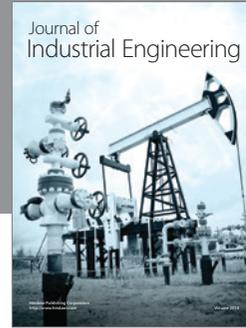
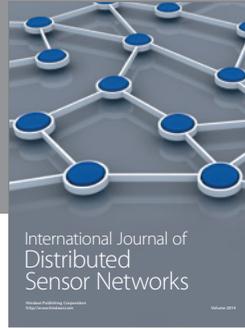
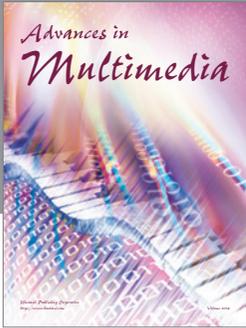
Acknowledgments

The authors are grateful to the members of the Artificial Intelligence Lab for sharing their faculty evaluations that made this work possible.

References

- [1] S. MacKim and R. A. Calvo, "Sentiment analysis in student experiences of learning," in *Proceedings of the 3rd International Conference on Educational Data Mining (EDM '10)*, pp. 111–120, Pittsburgh, Pa, USA, June 2010.
- [2] C. K. Leong, Y. H. Lee, and W. K. Mak, "Mining sentiments in SMS texts for teaching evaluation," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2584–2589, 2012.
- [3] B. Jagtap and V. Dhotre, "SVM & HMM based hybrid approach of sentiment analysis for teacher feedback assessment," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 3, no. 3, pp. 229–232, 2014.
- [4] J. Ogden and J. Lo, "How meaningful are data from Kim, S. M. & Calvo, R. A. (2010). Sentiment analysis in student experiences of learning," in *Proceedings of the 3rd International Conference on Educational Data Mining*, Pittsburgh, Pa, USA, 2012.
- [5] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: a survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [6] A. Minanović, H. Gabelica, and Ž. Krstić, "Big data and sentiment analysis using KNIME: online reviews vs. social media," in *Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO '14)*, pp. 1464–1468, May 2014.
- [7] L. Augustyniak, T. Kajdanowicz, P. Kazienko, M. Kulisiewicz, and W. Tuligłowicz, "An approach to sentiment analysis of movie reviews: lexicon based vs. classification," in *Hybrid Artificial Intelligence Systems: 9th International Conference, HAIS 2014, Salamanca, Spain, June 11–13, 2014. Proceedings*, vol. 8480 of *Lecture Notes in Computer Science*, pp. 168–178, Springer, Berlin, Germany, 2014.
- [8] S. Rosenthal, P. Nakov, A. Ritter, and V. Stoyanov, "Semeval-2014 task 9: sentiment analysis in twitter," in *Proceedings of the 8th International Workshop on Semantic Evaluation*, pp. 73–80, Dublin, Ireland, August 2014.
- [9] H. Saif, M. Fernandez, Y. He, and H. Alani, "Evaluation datasets for twitter sentiment analysis," in *Proceedings of the 1st International Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM) in Conjunction with AI*IA Conference*, Turin, Italy, 2013.
- [10] N. Altrabshah, M. Gaber, and M. Cocea, "SA-E: sentiment analysis for education," in *Proceedings of the 5th KES International Conference on Intelligent Decision Technologies (KES-IDT '13)*, Sesimbra, Portugal, 2013.
- [11] A. Balahur, R. Mihalcea, and A. Montoyo, "Computational approaches to subjectivity and sentiment analysis: present and envisaged methods and applications," *Computer Speech & Language*, vol. 28, no. 1, pp. 1–6, 2014.
- [12] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, Taylor & Francis, Boca Raton, Fla, USA, 2nd edition, 2010.
- [13] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.
- [14] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: an on-line lexical database," *International Journal of Lexicography*, vol. 3, no. 4, pp. 235–244, 1990.
- [15] S. Mohammad, C. Dunne, and B. Dorr, "Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '09)*, vol. 2, pp. 599–608, Singapore, August 2009.
- [16] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*, Association for Computational Linguistics, 2004.
- [17] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL '98)*, pp. 174–181, 1997.
- [18] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, pp. 415–463, Springer US, 2012.
- [19] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP '02)*, vol. 10, pp. 79–86, 2002.
- [20] N. C. Petersen, "A Logical Approach to Sentiment Analysis," Thesis prepared at Department of Informatics and Mathematical Modelling at the Technical University of Denmark, http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/6459/pdf/imm6459.pdf.
- [21] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Mining and Knowledge Discovery*, vol. 24, no. 3, pp. 478–514, 2012.
- [22] E. Cambria, "An introduction to concept-level sentiment analysis," in *Advances in Soft Computing and Its Applications*, Lecture Notes in Computer Science, pp. 478–483, Springer, Berlin, Germany, 2013.
- [23] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 715730, 9 pages, 2015.

- [24] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity in phrase-level sentiment analysis,” in *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP '05)*, pp. 347–354, Stroudsburg, Pa, USA, October 2005.
- [25] E. Riloff and J. Wiebe, “Learning extraction patterns for subjective expressions,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '03)*, pp. 105–112, Stroudsburg, Pa, USA, July 2003.
- [26] J. Wiebe, T. Wilson, and C. Cardie, “Annotating expressions of opinions and emotions in language,” *Language Resources and Evaluation*, vol. 39, no. 2, pp. 165–210, 2005.
- [27] J. Ahn, A. Borowiec, K. Buckley et al., “CYBEREMOTIONS—collective emotions in cyberspace,” *Procedia Computer Science*, vol. 7, pp. 221–222, 2011.
- [28] L. Goeriot, L. Kelly, G. J. F. Jones, H. Müller, and J. Zobel, “Report on the SIGIR 2014 Workshop on Medical Information Retrieval (MedIR),” *ACM SIGIR Forum*, vol. 48, no. 2, pp. 78–82, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

