

Towards a Framework on Sentiment Analysis of Educational Domain for Improving the Teaching and Learning Services

Ke Arah Rangka Kerja Analisis Sentimen di dalam Domain Pendidikan untuk Meningkatkan Servis Pengajaran dan Pembelajaran

Harnani Mat Zin¹, Norwati Mustapha², Masrah Azrifah Azmi Murad²,
Nurfadhlina Mohd Sharef²

¹*Faculty of Arts, Computing, and Creative Industry, Universiti Pendidikan Sultan Idris; harnanimatzin@gmail.com*

²*Faculty of Computer Science and Information Technology, University Putra Malaysia; {norwati, masrah, fadhlina}@fsktm.upm.edu.my*

Abstract

Analysing students' feedback and their expressed emotions toward any subjects could help lecturers to understand their students' learning behaviour. Several platforms are used by students to express their feelings such as through social networking sites, blogs, discussion forums and the university survey systems. However, the feedbacks typically contain thousands of sentences and are from various sources which makes analysing them a cumbersome and tedious work. In this regard, sentiment analysis (SA) has been proposed to automate the process of mining user feedback into valuable information. This paper discusses the principles of SA, its potential benefits, and its application in the educational field based on the synthesis of previous studies. We suggest that SA can help lecturers to easily understand the needs and problems of their students. In particular, a framework and a performance evaluation method were proposed to help guide the implementation of the SA in the education domain.

Keywords: Opinion, sentiment analysis, sentiment analysis in educational domain.

Abstrak

Menganalisis maklum balas pelajar dan emosi mereka yang dinyatakan kepada mana-mana mata pelajaran dapat membantu pensyarah memahami tingkah laku pembelajaran pelajar mereka. Beberapa platform digunakan oleh pelajar untuk menyatakan perasaan mereka seperti melalui laman rangkaian sosial, blog, forum perbincangan dan sistem tinjauan universiti. Walau bagaimanapun, maklumbalas biasanya mengandungi beribu-ribu ayat dan berasal dari pelbagai sumber yang menjadikan analisisnya suatu kerja yang rumit dan membosankan. Dalam hal ini, analisis sentimen (SA) telah dicadangkan untuk mengautomasikan proses perlombongan maklum balas pengguna kepada maklumat yang berharga. Kertas ini membincangkan prinsip SA, manfaat potensinya, dan aplikasinya dalam bidang pendidikan berdasarkan sintesis kajian sebelumnya. Kami mencadangkan agar SA dapat membantu para pensyarah memahami dengan mudah keperluan dan masalah pelajar mereka.

Khususnya, rangka kerja dan kaedah penilaian prestasi dicadangkan untuk membantu memimpin pelaksanaan SA dalam domain pendidikan.

Kata kunci: Pendapat, analisis sentimen, analisis sentimen di dalam domain pendidikan.

INTRODUCTION

The emergence of Internet and Web 2.0 technologies has enabled students to convey their expressions and feelings toward the teaching and learning services. Opinion about products or services, in this case teaching and learning services, is very crucial for both parties, namely the students and the lecturers. For instance, a student may want to see the opinions of other students about a specific subject before they enrol in any courses. Meanwhile, the lecturers may want to analyse the feedback from their students about their teaching for future improvement. Thus, analysing students' opinions and their responses is a very important process. To analyse the reviews written by students, sentiment analysis (SA) is deemed highly appropriate, because it deals with the computational study of sentiments expressed digitally.

SA is defined as a field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions toward entities such as products, services, organizations, individuals, issues, events, topics, and their attributes (Liu, 2012). Essentially, SA is a sub-discipline of a computational linguistics that is concerned with opinions expressed in a document. Specifically, SA is a multi-discipline research that combines approaches from other disciplines such as data mining, statistics, artificial intelligence, and computational linguistics. The term "sentiment analysis" is also called opinion mining. In this regard, an opinion is defined as any subjective statements of a reviewer about any entity, and this opinion can be positive, negative, or neutral statements. For a machine, the opinion used in SA can be described as a quintuple, which is an object consisting of five different things as follows (Pang & Lee, 2008):

$$(O_j, f_{jk}, SO_{ijkl}, h_i, t_i) \quad (1)$$

Where:

- O_j is the opinion target object.
- f_{jk} is the feature of object.
- SO_{ijkl} is the sentiment polarity: positive, negative, or neutral.
- h_i is the source or opinion holder.
- t_i is the time in which the opinion is expressed.

In a previous study of sentiment analysis, a simple function was used to identify the polarity of the reviews as being either positive or negative. If the number of positive reviews posted by the users is higher than their negative reviews, then the function summarizes that the global opinion is positive. And the reverse is true if negative reviews outnumber the positive ones.

$$\begin{aligned} \sum \text{positive (reviews)} > \sum \text{negative (reviews)} &\rightarrow \text{positive} \\ \sum \text{positive (reviews)} < \sum \text{negative (reviews)} &\rightarrow \text{negative} \end{aligned} \quad (2)$$

Applying SA to the education domain is one of the emerging research areas. In particular, the main objective of applying sentiment analysis to education domain is to analyse the educational data contents, models, and also to analyse or summarize learner's perceptions. According to (Muruganathan & ShivaKumar, 2016), sentiment analysis of educational data concentrates on the computing process models that focus on the educational context.

The aim of this paper is to present an overview of sentiment analysis and how this new research area can be utilized in the education domain to analyse educational data. This paper also proposes a basic framework for the sentiment analysis of education domain, together with a proposed performance measurement method. We hope that this field of study will benefit students and lecturers by further improving the teaching and learning process.

The remainder of this paper is organized as follows: Section 2 discusses the sentiment analysis at the three different levels. Section 3 elaborates the approaches of sentiment analysis. Section 4 critically reviews previous studies on sentiment analysis of education domain. Section 5 delves into the discussion regarding the framework of sentiment analysis of education domain. Section 6 presents the conclusions of this paper.

SENTIMENT ANALYSIS AT DIFFERENT LEVELS

Sentiment analysis can be categorized into three different levels, namely the document level sentiment analysis, sentence level sentiment analysis, and word, phrase or aspect level sentiment analysis, with each level having its own strengths and weaknesses (as shown in Table 1). Document level sentiment analysis is concerned with classifying the overall opinion expressed by the opinion holders about a certain object in the entire document as being positive, negative, or neutral (Bing Liu, 2010). Sentence level classification can express a sentiment polarity for each sentence of a review and even for each word of the sentence. However, both the document level analysis and the sentence level sentiment analysis will not be able to discover what exactly that the people like or dislike. Thus, the word, phrase or aspect level sentiment analysis was introduced, which focuses on the recognition of all sentiment expressions within a given document and the aspects to which they refer (Feldman, 2013). In fact, the aim of sentiment analysis at the word, phrase or aspect level is to extract the features of the commented object and subsequently to determine the opinion orientation of the object as being either positive or negative.

Table 1: SA at different levels (Mishra & Jha, 2014).

SA level	Assumptions and tasks involved at different levels
Document level	<ol style="list-style-type: none"> 1. Each document focuses on a single object and contains opinions posted by a single reviewer. 2. It is not pertinent to blogs and forum posts. <p><i>Task 1:</i> Opinion classification of the document. <i>Classes:</i> Positive, negative, and neutral</p>
Sentence level	<ol style="list-style-type: none"> 1. A sentence contains only one opinion posted by a single reviewer. 2. In each document, the sentence boundary is clearly defined. <p><i>Task 1:</i> Identify the subjectivity of the sentence as being either subjective or opinionated. <i>Classes:</i> Objective and subjective</p> <p><i>Task 2:</i> Opinion classification of the sentences. <i>Classes:</i> Positive, negative, and neutral.</p>
Word, phrase, or aspect level	<ol style="list-style-type: none"> 1. The data source focuses on features of a single object posted by a single reviewer. 2. It is not pertinent to blogs and forum posts. <p><i>Task 1:</i> Identify and extract object features expressed by the reviewer. <i>Task 2:</i> Determine whether the opinions on the features are positive, negative, or neutral. <i>Task 3:</i> Group the synonyms for each extracted feature and provide a feature-based opinion summary of multiple reviews.</p>

From table 1, it shows clearly the tasks involved in each level of SA. Among these levels, the word, phrase, or aspect level SA is the most popular SA, because it analyzes text at a finer level of granularity (Zhang & Liu, 2014). Moreover, most of the current works are focused on the feature extraction at the word, phrase, or aspect level SA.

APPROACHES IN SENTIMENT ANALYSIS

There are two most common approaches in sentiment analysis research, namely the machine learning approach and lexicon-based approach. The most popular techniques and approaches in sentiment analysis have been reported by (Pang & Lee, 2008) in their article as shown in Figure 1. While Table 2 summarizes the advantageous and disadvantageous of these two approaches.

The machine learning approach uses several learning algorithms to determine the sentiment polarity by training on a known dataset. Two most common techniques in machine learning approach are supervised learning and unsupervised learning. The supervised method needs to train from examples that are manually classified by humans, while the unsupervised method does not need such training data. As such, the supervised learning method is the most popular method compared with the unsupervised learning method.

In contrast, the lexicon-based approach works by calculating the sentiment polarity of a review based on the semantic orientation of words or reviewed sentences. In essence, the semantic orientation is a

measure of the subjectivity of an opinion of the text. Two most common techniques of lexicon-based approach are corpus-based approach and dictionary-based approach. An opinion word lexicon is constructed in order to identify the words that describe the features of any products or services, and then it will be used again for the classification process. WordNet, SentiWordNet, and SenticNet are among the most popular lexicons used in sentiment analysis research. The dictionary-based approach works as follows: (i) find the opinion seed words, (ii) look up for their synonyms and antonyms in the dictionary. For this approach, we can either develop our own dictionaries or use the existing ones. However, building one’s own dictionary requires more time and (most likely) increases the cost. By contrast, the corpus-based approach starts with a predefined list of opinion words, and then it will find other opinion words existing in a large corpus to help determine the opinion words with context specific orientations. To solve such a problem, the semantic and statistical methods are used accordingly.

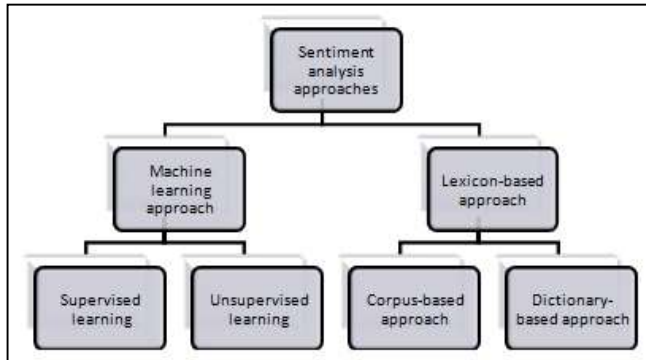


Figure 1: Machine learning approach and lexicon-based approach in sentiment analysis.

Table 2: Advantageous and disadvantageous of machine learning approach and lexicon-based approach.

Approach	Advantageous	Disadvantageous
Machine learning	<ul style="list-style-type: none"> • Perform well in specific domains. • Give high accuracy if the input texts are in simple sentences. • Easy adaptability to change the input. • Able to measure the degree of uncertainty. 	<ul style="list-style-type: none"> • Need pre-defined training data. • For supervised techniques, it does require labelled training data. • Difficult to obtain better result due to the characteristic of natural language.
Lexicon-based	<ul style="list-style-type: none"> • Does not need labelled training data. • Give better result if the domain is less bounded. • Able to mine the meaning behind the dataset. 	<ul style="list-style-type: none"> • Does not perform well in specific domains. • Need to have a corpus or dictionary as pre-defined seed words.

Table 2 can become a guideline in order to choose the sentiment analysis approach. However, the selection of the approaches depends on the research objective and the dataset itself. If the dataset is in

a specific domain and using simple sentences, then the machine learning approach is most suitable. But, if the dataset is in a less bounded domain and contains complex sentences, then we can use the lexicon-based approach. Besides, the lexicon-based approach is able to perform the semantic analysis of the input data since it applied the natural language processing.

PREVIOUS WORK ON SENTIMENT ANALYSIS IN EDUCATION DOMAIN

Research on the application of sentiment analysis in the educational domain has gained traction in the last decade. Of particular interest, many researchers have proposed several methods of such application in their studies. For example, Li and Ma (Li & Ma, 2012) proposed an application of opinion mining system called OMTEM in order to obtain the overall evaluation of each teacher. Likewise, Guitart et al. (Guitart et al., 2013) proposed a method to mine students' opinions of the tutorial classroom at the Open University of Catalonia. Meanwhile, Altrabsheh et al. (Altrabsheh, Cocea, & Fallahkhair, 2014b) (Altrabsheh, Cocea, & Fallahkhair, 2014a) proposed a method for analysing real-time students' feedback, which according to them was the first research of such a nature. Of late, many researchers have begun to examine the application of social networking sites (e.g., Facebook) in educational domain in earnest. For instance, Ortigaso et al. (Ortigosa, Martín, & Carro, 2014) study involved the sentiment analysis of Facebook with respect to its application to e-learning. Similarly, the work of Ravi et al (Ravi, Ravi, Siddeshwar, & Mohan, 2015) dealt with the evaluation of students' feedback on their training programs. Recently, the latest work of Zarra et al. (Zarra, Chiheb, Faizi, & Afia, 2016) was focused on textual similarity and sentiment analysis to analyse learners' comments on discussion forums. Table 3 shows the strengths and weaknesses of the approaches reported in such studies.

Table 3: Summary of previous study.

Reference	Approach	Strength	Weakness
(Li & Ma, 2012)	Lexicon-based	<ul style="list-style-type: none"> • OMTEM overcomes the shortcomings of traditional teacher evaluation method. • Combination of word segmentation and topic extraction is good in extracting students' comments. 	<ul style="list-style-type: none"> • Focus only on Chinese word segmentation. • The use of simple sum function to calculate the overall polarity leads to low quality result.
(Guitart et al., 2013)	Lexicon-based	<ul style="list-style-type: none"> • Able to identify a large number of opinions, because it uses the combination of three language resources (WordNet, FreeLing, and affective dictionary). • Give high-level precision as only the features of words with substantial elements of opinions are listed. 	<ul style="list-style-type: none"> • Limited coverage of effective dictionary, because it is manually compiled by the researcher. • The output produced in the form of tag cloud can only be used in few circumstances.
(Altrabsheh et al., 2014b)	Machine learning	<ul style="list-style-type: none"> • Able to process real-time data. • Use several combinations of n-grams 	<ul style="list-style-type: none"> • Unable to implement the low level pre-processing technique, because

Reference	Approach	Strength	Weakness
(Altrabsheh et al., 2014a)		<p>features.</p> <ul style="list-style-type: none"> • Test the proposed method with several machine learning classifiers (e.g., SVM, NB, ME, and CNB) 	<p>of the relatively clean dataset.</p> <ul style="list-style-type: none"> • Produce good results for the positive and negative classes but not for the neutral class.
(Ortigosa et al., 2014)	Hybrid approach (ML + Lexicon)	<ul style="list-style-type: none"> • Students give more realistic comments using Facebook compared to LMS, thus the dataset will become more meaningful. • The combination of SVM and lexicon-based tagging is better, because it produces higher accuracy. 	<ul style="list-style-type: none"> • Give the same score for each positive word without considering the strength of the sentiment words. • The temporal window to detect the changes is only one week (7 days); however, it can be improved by decreasing the temporal window to lower numbers, such as 3 days.
(Ravi et al., 2015)	Lexicon-based	<ul style="list-style-type: none"> • Apply POS tagging to recognize different parts of a speech in text. • Use SentiWordNet to calculate the sentiment scores of each extracted adjectives. • The measurement is performed based on the correlation coefficient between program sentiment score and program rating. 	<ul style="list-style-type: none"> • Extract only the adjectives word. Nonetheless, improvement can be made by extracting other part-of-speech words, such as nouns and verbs.
(Zarra et al., 2016)	Hybrid approach (ML + Lexicon)	<ul style="list-style-type: none"> • TFIDF approach is used to analyze the behavior of words in a corpus. • Use Latent Semantic Analysis (LSA) to create the relationship between messages and course content. • Apply sentiment analysis techniques to identify students' opinions of educational problems or issues. 	<ul style="list-style-type: none"> • Only focus on document level sentiment analysis. Nonetheless, improvement can be made by working on sentence level and aspect level sentiment analysis.

From Table 3, we can clearly see that the hybrid approach which is the combination of machine learning and lexicon-based approach nowadays become an emerging approach. This hybrid approach is able to tackle the statistic, syntactic and semantic of the review sentences. It also gives better performance results.

PROPOSED FRAMEWORK AND EVALUATION

The main contribution of this paper is the proposed framework of sentiment analysis of the education domain and the performance measurement method that can help validate the educational products or services.

Sentiment Analysis Framework

The proposed sentiment analysis of educational domain comprises five main phases, namely the data collection, pre-processing, feature extraction, feature classification, and output generation. Figure 2 shows all the five phases of the analysis.

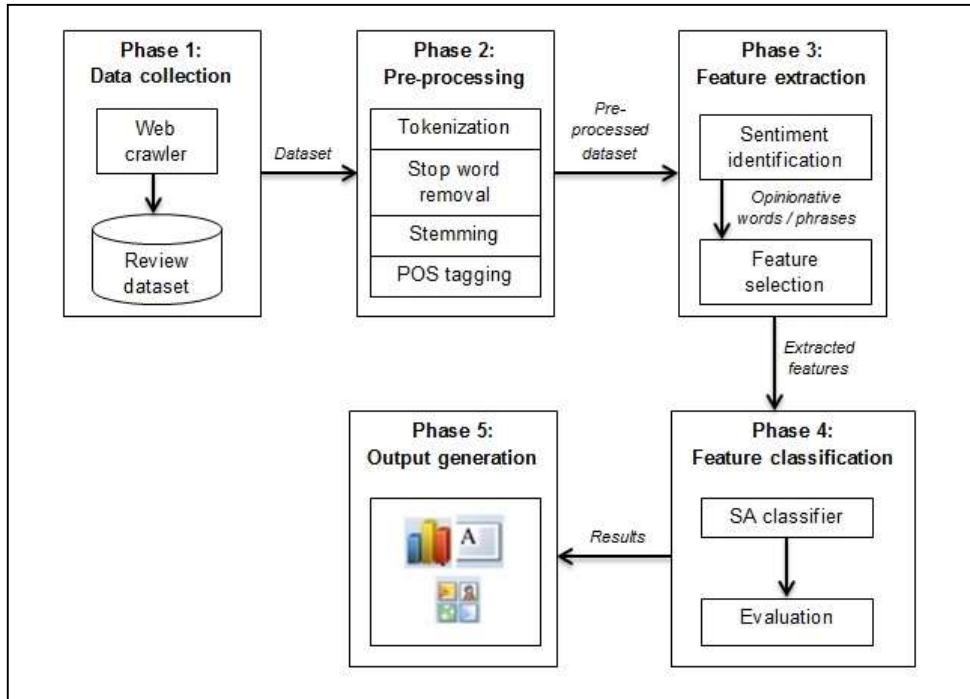


Figure 2: The five phases of the sentiment analysis of educational domain.

a) Data collection

The main object of such research is the dataset itself, entailing the researchers to prepare a quality dataset as an input of this work. Such a dataset can be collected from the Internet with the help of web crawlers or specialized systems that were internally built.

b) Pre-processing

The pre-processing phase is an important phase to be considered in such an analysis as it involves the task of cleaning up the dataset from any noisy data, thus reducing the complexity of a document (Haddi, Liu, & Shi, 2013). Once cleaned, the pre-processed dataset is ready for the following tasks. The pre-processing sub-tasks involved are tokenization, stop word removal, stemming, and part-of-speech (POS) tagging. First, the tokenization task is performed to split up the sentences into terms of tokens (words). Then, the stop word removal task is carried out to remove articles such as *a*, *an*, and

the. Subsequently, the stemming task is performed to transform the tokens into a single type, normally a root word; for example, the word "lecturers" will become "lecturer". As such, the stemming process will reduce redundant words in a document. Finally, the POS tagging task is undertaken to label (tag) each token with its correct part-of-speech.

c) Feature extraction

There are two main sub-tasks involved in this phase, namely the sentiment identification and feature selection. The main objective of sentiment identification task is to distinguish between opinionative and non-opinionative words or sentences. In this regard, only opinionative words or phrases are deemed important, and they will be subjected to the subsequent sub-task, which is the feature selection. The feature selection task will only select relevant features and remove any irrelevant, redundant features. The collection of extracted features will become an input for the next phase.

d) Feature classification

The feature classification is the process of running the sentiment analysis classifier on the extracted features in order to obtain the sentiment polarity (positive or negative) and to evaluate their performance. There are several classifiers that can be used, such as support vector machine (SVM) and naive baye s (NB). In addition, the precision, recall, and f-measure are some of the techniques that can be used for performance measurement.

e) Output generation

Finally, the output generation phase involves generating the output of the results of the last four phases of the SA in various forms, such as a summary table, graphs, star rating, and feature ranking.

Performance Evaluation

A majority of existing studies on the evaluation of the SA results were based on the evaluation metrics that are commonly used in information retrieval and document classification research, such as the precision, recall, and f-measure. Such selection of metrics is hardly surprising as they have been widely and extensively used in the document classification research (Manning, Raghavan, & Schutze, 2009). The precision metric measures the correctness of a classifier, and the recall metric measures the completeness or sensitivity of a classifier. The combination of precision and recall metrics is measured by the f-measure, which is the weighted harmonic mean of precision and recall. The following are the formulas of the three performance criteria:

$$\text{Precision} = \frac{\text{Number of relevant feature retrieved}}{\text{Total number of feature retrieved}} \quad (3)$$

$$\text{Recall} = \frac{\text{Number of relevant feature retrieved}}{\text{Total number of relevant feature in database}} \quad (4)$$

$$\text{Precision} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

CONCLUSION

This paper discusses the background of the sentiment analysis, including the different levels of sentiment analysis, and the most common approaches used in sentiment analysis. In particular, the potential benefits of SA for educational purposes are also highlighted based on emerging studies that are now focusing on applying the analysis in the educational domain. In view of the broad scope of the educational domain (encompassing a wide range of possibilities), the researchers proposed a new framework and a performance evaluation method that can help guide the implementation of such an analysis in the education realm. This study can be extended by implementing and testing the proposed framework using real educational dataset. Such research can help researchers to examine the effectiveness of SA in analyzing educational data, which will be important in helping practitioners to improve their current practices. With more efforts directed toward this goal, research on the use of SA in education domain will become more intense and popular, thus benefitting all parties concerned.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Higher Education of Malaysia for awarding the principal author with the research grant to carry out this research.

REFERENCES

- Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014a). Learning sentiment from student's feedback for real-time interventions in classrooms. *Lecture Notes in Computer Science*, 8779, 40–49. Retrieved from <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6984506>
- Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014b). Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback. In *International Conference on Tools with Artificial Intelligence, ICTAI* (pp. 419–423). <http://doi.org/10.1109/ICTAI.2014.70>
- Bing Liu. (2010). Sentiment Analysis and Subjectivity. In *Handbook of Natural Language Processing* (pp. 1–38). <http://doi.org/10.1145/1772690.1772756>
- Feldman, R. (2013). Techniques and Applications for Sentiment Analysis. *Communications of the ACM*, 54(4), 82–89.
- Guitart, I., Conesa, J., Villarejo, L., Lapedriza, A., Masip, D., Perez, A., & Planas, E. (2013). Opinion Mining on Educational Resources at the Open University of Catalonia. In *2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems* (pp. 385–390). IEEE. <http://doi.org/10.1109/CISIS.2013.70>
- Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. In *Procedia Computer Science* (Vol. 17, pp. 26–32). Elsevier B.V. <http://doi.org/10.1016/j.procs.2013.05.005>
- Li, C., & Ma, J. (2012). Research on online education teacher evaluation model based on opinion mining. In *National Conference on Information Technology and Computer Science (CITCS 2012)* (pp. 1061–1064). <http://doi.org/doi:10.2991/citcs.2012.264>
- Liu, B. (2012). *Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies*. (B. Liu, Ed.), Morgan & Claypool. Morgan & Claypool Publishers. <http://doi.org/10.2200/S00416ED1V01Y201204HLT016>

- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge University Press. <http://doi.org/10.1109/LPT.2009.2020494>
- Mishra, N. ., & Jha, C. K. . (2014). An insight into task of opinion mining. In *Lecture Notes of the Institute for Computer Sciences* (Vol. 117, pp. 185–190). Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-84921384276&partnerID=40&md5=cb39ca050af23697b78393da851e2cc6>
- Murugananthan, V., & ShivaKumar, B. L. (2016). An adaptive educational data mining technique for mining educational data models in elearning systems. *Indian Journal of Science and Technology*, 9(3), 1–5. <http://doi.org/10.17485/ijst/2016/v9i3/86392>
- Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31(1), 527–541. <http://doi.org/10.1016/j.chb.2013.05.024>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations & Trends in Information Retrieval*, 2(1–2), 1–135.
- Ravi, K., Ravi, V., Siddeshwar, V., & Mohan, L. (2015). Sentiment Analysis Applied to Educational Sector. In *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. <http://doi.org/10.1109/ICCIC.2015.7435667>
- Zarra, T., Chiheb, R., Faizi, R., & Afia, A. El. (2016). Using Textual Similarity and Sentiment Analysis in Discussions Forums to Enhance Learning. *International Journal of Software Engineering and Its Applications*, 10(1), 191–200.
- Zhang, L., & Liu, B. (2014). Aspect and Entity Extraction for Opinion Mining. In *Data Mining and Knowledge Discovery for Big Data* (Vol. 1, pp. 1–40). <http://doi.org/10.1007/978-3-642-40837-3>