



Sentiment Analysis of Open-Ended Student Feedback

Timo Hynninen, Antti Knutas and Maija Hujala

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 23, 2021

Sentiment analysis of open-ended student feedback

Timo Hynninen*, Antti Knutas** and Maija Hujala***

*South-Eastern Finland University of Applied Sciences / Department of Information Technology, Mikkeli, Finland

**LUT University / LUT School of Engineering Science, Lappeenranta, Finland

***LUT University / LUT School of Business and Management, Lappeenranta, Finland

timo.hynninen@xamk.fi, antti.knutas@lut.fi, maija.hujala@lut.fi

Abstract—In this paper, we perform a sentiment analysis on a large set of open-ended course feedback from university courses collected between 2016 and 2019. We used the R programming language and environment for statistical computing to categorize feedback texts by their sentiment values (positive, negative). Additionally, we calculate the NRC Emotion values, which categorise the feedback according to eight basic emotions. We present analysis on the trends of how the feedback evolved through the years. Finally we compare the findings from our data to existing literature.

I. INTRODUCTION

Student evaluations of teaching (SET) are commonly used for quality control in higher education. Student evaluations from university courses are used to assess both teaching material and teachers themselves [1], [2]. While the bulk of evaluations usually use quantitative, Likert-type scales, open-ended feedback texts are also common. The use of the text feedback for quality control, however, is less common.

On the other hand, student evaluations are a difficult tool to utilize. For example, the validity of evaluations as a measurement of quality can in some settings be questionable [3], and evaluations do not necessarily reflect about students' learning [4]. In addition to the validity questions of the student evaluations, text feedback is challenging to analyze in large quantities.

This paper presents a study on the sentiment analysis of open-ended student evaluations from university courses. This work is a continuation to our previous study [5] in which we distinguished different themes from open-ended feedback texts using the topic modelling method. In our previous study we established that the feedback students give concerns comments about course arrangements, study motivation, course content, and dissatisfaction in the teaching methods. We found that these topics were in line with types of student feedback established in prior work (see for example [6] or [7]). However, one of the conclusions in our previous study was that in our topic modeling results we could not see a clear disposition towards either positive or negative feedback. This result is somewhat surprising, as previous work maintains that students' comments tend to be more positive than negative. Therefore, further investigation of our feedback data is necessary.

The main research question this paper addresses is, **What emotions can be distinguished from student feedback?** The main research question is further divided into sub-questions, which are listed as follows.

- How does the machine learning algorithm classify the basic emotions in student course feedback texts?
- Is the ratio between positive and negative feedback comments the same as established in SET literature?

Rest of the paper is structured as follows. Section II presents the related work on using open text student feedback and sentiment analysis. Section III presents the research methods, and the procedures for data collection and analysis. The main results are presented in Section IV, and further discussed in Section V. Finally, Section VI concludes.

II. RELATED WORK

In recent years, there has been a growing interest in the analysis and characterization of responses to the open-ended questions of student feedback surveys. A few studies have proposed sentiment analysis approaches for classifying students' written comments as positive, (neutral) or negative [8], [9], [10], [11], [12]. Onan [10] aimed at finding an efficient sentiment classification scheme on feedback comments provided by students. He tested and compared several machine learning-based approaches and deep learning-based approaches to sentiment analysis and found out that the deep learning-based methods outperformed the traditional machine learning classifiers. In turn, Pong-inwong and Songpan [11] proposed a new sentiment analysis method (sentiment phrase pattern matching, SPPM) and tested it with open-ended student feedback. In addition to students' sentiment polarity, Jena [13] and Nimala and Jebukumar [14] used sentiment mining techniques to model and predict students' emotions based on open-ended student feedback. Jena [13] tested various emotion classifiers to identify eight emotions (amused, anxiety, bored, confused, engaged, enthused, excited, and frustrated) from students' written comments whereas Nimala and Jebukumar [14] proposed a sentiment topic emotion mining model which captures eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). In addition to simply classifying students' written comments, Andersson et al. [8] compared

average hours of study outside of class with the average sentiment of student feedback data and found a moderate negative correlation between them. They concluded that the use of sentiment analysis for student feedback data could allow for the simplification of the feedback questionnaires. Furthermore, Ibrahim et al. [15] focused on distinguishing feedback related to specific issues, such as assessment, from student’s comments. Cunningham-Nelson et al. [16], [17] and Pyasi et al. [18], for their part, proposed tools for visually exploring students’ sentiment for analysing written student feedback.

III. METHODS

In this section, we first detail the data collection and then data analysis process.

A. Data collection

The data used in this study comes from the student feedback surveys carried out at a Finnish university between three academic years, beginning in the autumn 2016 and ending in the spring 2019. The questionnaire in 2016-17 had one open-ended question: *“Other feedback about the course (for example, ways to enhance learning during the course)”*. The questionnaire in 2017-18 had five open-ended questions: *“What factors affected my level of motivation?”*, *“What factors affected how much I invested in my learning?”*, *“What factors affected the workload?”*, *“My feedback regarding the teaching methods:”*, and *“What factors promoted my learning and how could learning be supported better?”*. In turn, the questionnaire in 2018-19 had four open-ended questions: *“The course as a whole promoted my learning (1=very little; 5=very much) - If you chose 1-3, please give concrete examples”*, *“The course as a whole promoted my learning (1=very little; 5=very much) - If you chose 4-5, please give concrete examples”*, *“What aspects of the course most need improvement?”*, and *“What was best about this course?”*. In addition, all questionnaires included several 5-point Likert-scale questions about, for example, motivation, workload, and teaching methods.

The survey questionnaires were sent to students via email after they completed the courses. Responses were collected anonymously and voluntarily. This study is restricted to feedback written in English, so we included only those responses that contained answers to open-ended questions written in English. We used the langdetect library [19] with the Python programming language to detect the language, and verified the results by manual inspection.

TABLE I
PERCENTAGE OF POSITIVE AND NEGATIVE COMMENTS EACH YEAR

Year	positive	negative
2016-2017	81%	19%
2017-2018	83%	17%
2018-2019	84%	16%
All 3 years	83%	17%

The sample size for sentiment analysis was in total 4990 feedback responses. 1534 were collected in 2016-17, 1953 in 2017-18, and 1503 in 2018-19.

B. Sentiment analysis process

“Sentiment analysis is defined as the task of finding the opinions of authors about specific entities.“ [20]

In the analysis process, we used the sentiment approach to detect emotions in text. We used the NRC Word-Emotion Association Lexicon [21], which is a list of words and their associations with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The lexicon was manually created using a crowdsourcing platform [22].

We performed the analysis using the R statistical computing environment [23] and its libraries, especially in sentiment analysis [24], text mining [25] and cleaning [26], and visualization [27].

The main part of the analysis was performed using the Syuzhet sentiment analysis library [24] and its included NRC lexicon [21]. The Syuzhet library uses a bag-of-words algorithm, where it compares words from the analyzed text string against the selected lexicon and provides sentiment or emotion data. The exact analysis steps are summarized as follows.

- 1) Import the data:
 - i) Download student feedback data.
 - ii) Sort the feedback by metadata such as course type and language.
 - iii) Select the subset to be analyzed; in this case English language responses and specific range of years.
- 2) Preprocessing:
 - i) Divide each feedback response into an individual string in an array.
 - ii) Remove punctuation, extra whitespaces, and other symbols not relevant to analysis.
 - iii) Remove stopwords using the textmining [25] library’s English stopwords list.
- 3) Sentiment analysis: Use the Syuzhet [24] library and the NRC lexicon [21] to calculate emotion values for each response.
- 4) Analyzing the findings: Export the results as a table and perform descriptive or statistical analysis.

We used the wordcloud [27] library to create a comparative wordcloud that displays the most characteristic words from each emotion and demonstrates the emotions’ comparative prevalence in the dataset. The wordcloud was generated by first assigning each feedback response to their strongest detected emotion and then creating a term-document matrix based on emotions and the associated responses. This means that words co-occurring with the lexicon’s emotion words are also included in the emotion’s wordcloud segment.

TABLE II
PERCENTAGE OF EACH NRC EMOTION IN THE FEEDBACK TEXTS

Year	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
2016-2017	3.4 %	21.7 %	3.6 %	7.4 %	15.1 %	6.8 %	9.3 %	32.6 %
2017-2018	2.9 %	21.0 %	2.8 %	5.9 %	17.8 %	5.5 %	11.3 %	32.7 %
2018-2019	2.5 %	21.7 %	2.6 %	7.0 %	15.7 %	6.1 %	8.8 %	35.7 %
All 3 years	2.9 %	21.6 %	3.0 %	6.9 %	15.9 %	6.2 %	9.5 %	33.9 %

TABLE III
MOST COMMON WORDS ASSOCIATED WITH EACH EMOTION (IN DESCENDING ORDER)

anger	anticipation	disgust	fear	joy	sadness	surprise	trust
students	lectures	lectures	lectures	lectures	lectures	lectures	lectures
lectures	time	students	work	work	work	work	work
time	work	time	students	students	case	students	students
work	students	work	case	time	students	time	time
exam	learning	group	time	learning	time	exam	learning
group	exam	lecture	exam	interesting	group	learning	interesting

IV. RESULTS

After collecting and preprocessing the feedback surveys given to students we were left with 4990 individual open feedback texts. After a manual inspection to ensure the preprocessing had properly cleaned up and formatted the data we proceeded with the sentiment analysis. First, we calculated the overall emotion values for each feedback text. The sentiment analysis yields a sentiment score which determines if the overall emotion in the text is negative (a negative score) or positive (a positive score).

In the analysis, we found an overwhelming number of positive feedback texts. Overall 83% of total feedback was positive and 17% was negative. The number of overall positive and negative comments is presented by year in Table I. The difference between the years is minimal, although the number of positive comments is slowly on the rise (81% in 2016, 83% in 2017, and 84% in 2018), and conversely the number of negative comments is slowly diminishing.

In order to distinguish more detailed emotions from the texts, we also performed an NRC emotion analysis. The analysis uses the NRC lexicon and classifies the text in one or more of the NRC categories (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) by calculating a sentiment score for the text in each category. Table II presents the percentages of each NRC emotion category distinguished by the analysis. The most prevalent emotions emerging from the feedbacks are *trust* (overall 33.9%), *anticipation* (21.6%), and *joy* (15.9%). Less prevalent emotions were *surprise* (9.5%), *fear* (6.9%), *sadness* (6.2%), *disgust* (3.0%), and *anger* (2.9%). Each emotion is present in less than 10% of the texts.

Most common words in the feedbacks related to the emotion categories are listed in Table III. As shown in the table, certain words, for example *lectures* or *time* appear in multiple emotion categories. In order to see how closely associated to the emotions the individual words are, we employed the word cloud approach, which

is visualized in Figure 1. Of the most occurring concepts, other students are associated with anger, exercises and time with anticipation, homework with fear, learning with joy, cases with sadness, exams with surprise, and teaching and lectures with trust.

V. DISCUSSION

By using the Syuzhet [24] library in the R statistical computing environment [23] we were able to extract two dimensions of emotion in our open-text student evaluations of teaching. The overall dominant sentiment value is largely positive, and the NRC emotion analysis provided a breakdown of what those positive emotions likely are: The most common emotions were classified as trust, anticipation and joy.

Our results indicate that, overall, the emotions distinguished from the feedback data were generally mainly positive. During our three-year observation period more than 80% of the feedback was positive and less than was 20% negative. This is an important thing to consider, since usually when faculty evaluates student feedback of teaching, critical feedback is highlighted and negative points are evaluated. This finding is in line with the results of the study by Alhija and Fresko [6], which points out that student feedback is more often positive than negative. Also, according to the results of Sengkey et al. [12] and Cunningham-Nelson et al., [17] student feedback seems to be biased towards positive evaluations of teaching.

What comes to the NRC emotion categories, our results show that many common learning and university related words are the target of many different emotions. Lectures, work, time, and other students are the target of both positive (e.g. joy, anticipation and trust) and negative (e.g. fear, sadness and disgust) emotions. In our previous study on the use of topic modelling for analysing student feedback texts [5], we found that while topic modeling can be used to extract themes from student feedback, its results do not tell much about the overall level of student satis-

- [8] E. Andersson, C. Dryden, and C. Variawa, "Methods of applying machine learning to student feedback through clustering and sentiment analysis," *Proceedings of the Canadian Engineering Education Association (CEEA)*, 2018.
- [9] S. Gottipati, V. Shankararaman, and J. R. Lin, "Latent dirichlet allocation for textual student feedback analysis," in *Proceedings of the 26th International Conference on Computers in Education ICCE 2018*. APSCE, 2018.
- [10] A. Onan, "Mining opinions from instructor evaluation reviews: A deep learning approach," *Computer Applications in Engineering Education*, 2019.
- [11] C. Pong-Inwong and W. Songpan, "Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (sppm) based on association mining," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 8, pp. 2177–2186, 2019.
- [12] D. F. Sengkey, A. Jacobus, and F. J. Manoppo, "Implementing support vector machine sentiment analysis to students' opinion toward lecturer in an Indonesian public university," *Journal of Sustainable Engineering: Proceedings Series*, vol. 1, no. 2, pp. 194–198, 2019.
- [13] R. Jena, "Sentiment mining in a collaborative learning environment: capitalising on big data," *Behaviour & Information Technology*, pp. 1–16, 2019.
- [14] K. Nimala and R. Jebakumar, "Sentiment topic emotion model on students feedback for educational benefits and practices," *Behaviour & Information Technology*, pp. 1–9, 2019.
- [15] Z. M. Ibrahim, M. Bader-El-Den, and M. Cocea, "Mining unit feedback to explore students' learning experiences," in *UK Workshop on Computational Intelligence*. Springer, 2018, pp. 339–350.
- [16] S. Cunningham-Nelson, M. Baktashmotlagh, W. Boles *et al.*, "Visually exploring sentiment and keywords for analysing student satisfaction data," in *29th Australasian Association for Engineering Education Conference 2018 (AAEE 2018)*. Engineers Australia, 2018, p. 132.
- [17] S. Cunningham-Nelson, M. Baktashmotlagh, and W. Boles, "Visualizing student opinion through text analysis," *IEEE Transactions on Education*, vol. 62, no. 4, pp. 305–311, 2019.
- [18] S. Pyasi, S. Gottipati, and V. Shankararaman, "Sufat-an analytics tool for gaining insights from student feedback comments," in *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2018, pp. 1–9.
- [19] *langdetect 1.0.7: Language detection library ported from Google's language-detection*, 2016. [Online]. Available: <https://pypi.org/project/langdetect/>
- [20] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.
- [21] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.
- [22] —, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. Association for Computational Linguistics, 2010, pp. 26–34.
- [23] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019. [Online]. Available: <https://www.R-project.org/>
- [24] M. L. Jockers, *Syuzhet: Extract Sentiment and Plot Arcs from Text*, 2015. [Online]. Available: <https://github.com/mjockers/syuzhet>
- [25] I. Feinerer, K. Hornik, and D. Meyer, "Text mining infrastructure in r," *Journal of Statistical Software*, vol. 25, no. 5, pp. 1–54, March 2008. [Online]. Available: <http://www.jstatsoft.org/v25/i05/>
- [26] T. W. Rinker, *textclean: Text Cleaning Tools*, Buffalo, New York, 2018, version 0.9.3. [Online]. Available: <https://github.com/trinker/textclean>
- [27] I. Fellows, *wordcloud: Word Clouds*, 2018, r package version 2.6. [Online]. Available: <https://CRAN.R-project.org/package=wordcloud>