Multi-Label Emotion Mining From Student Comments

Angelina Tzacheva
University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC, USA, 28223
aatzache@uncc.edu

Jaishree Ranganathan
University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC, USA, 28223
jrangan1@uncc.edu

Rajendra Jadi
University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC, USA, 28223
rjadi@uncc.edu

ABSTRACT
Science, Technology, Engineering, and Mathematics (STEM) education is gaining more attention not today but has been under research, and discussion for the past few decades. Factors that are considered for research include but not limited to the following, culture on campus, teaching and learning models, and student experience in classroom, gender bias, and stereotypes. One of the major factors is the teaching model adopted which have impact on the student learning styles and their experience in the classroom. Teaching models include traditional models, modern flipped class-room models, and active learning approaches. This study focuses on active learning approaches and their impact on students learning and experience. Light-weight team is an active learning approach, in which team members have little direct impact on each other's final grades, with significant long-term socialization. In this work we used data from end of course student evaluation. We propose extend our previous method for assessing the effectiveness of the Light-weight team teaching model, through automatic detection of emotions in student feedback in computer science course by creating multi-label for each text comment. The students are surveyed about their feelings and thoughts about teaching and learning models adopted and student experience in the classroom. Results show that implementation of these methods result in increased positivity in student emotions.

CCS Concepts
Applied computing → Collaborative learning

Keywords
Education; Emotion Mining; Learning Approaches; Light-Weight Team.

1. INTRODUCTION
Light-Weight teams [1] is an Active Learning approach, in which team members have no direct impact on each other's final grades, yet there is a significant component of peer teaching, peer learning and long-term socialization.

ICIEI 2019, July 10–12, 2019, Durham, United Kingdom

ACM ISBN 978-1-4503-7169-8/19/07…$15.00
DOI: https://doi.org/10.1145/3345094.3345112

This innovative pedagogical approach has been studied in Computer Science undergraduate courses and has been reported to have high levels of student engagement [1], [2], [3], [4], and [5]. However, very limited work has been done on assessing Active Learning methods for undergraduate Courses at Department of Computer Science, and in the College of Computing and Informatics, as well as in peer institutions nationwide in terms of understanding if this model has impact retaining female participants in the undergraduate STEM courses.

This study aims to evaluate the effectiveness of teaching model and their impact on student learning styles and experience in classroom and identify factors that help in performance and positive attitude of students towards Computer science course. We propose a novel method for assessing pedagogical innovation through automatic detection of emotions in text, produced by student participants, in computer science courses. Students are surveyed about their sentiment and emotions on the Light-Weight Teams approach, and about their overall experience, and satisfaction with this Active Learning environment. Student Learning Outcomes are measured to determine if the Light-Weight Teams Active Learning approach helps to improve grades and student performance. Student Sentiment and Emotions are assessed to gain deeper understanding of Active Learning interventions compared to classical teaching methods.

2. RELATED WORK
In this section we review studies that have been done in the area of analyzing student evaluations, including text and quantitative data. Sentiment Analysis in education data has been widely applied to Massive Open Online Courses [6], and e-learning [7]. On the other hand we see minimal contribution towards actual classroom student feedbacks and impact of active learning methodologies on student's emotion.

2.1 Sentiment Analysis on MOOC and E-Learning Courses
Massive Open Online Course (MOOC) is wide-spread since it is first introduced in 2006 as part of distant education. It is basically free participation to users from any location and does not bind to any individual university or organization. It is observed that MOOC suffer a high drop-out rate close to 90% [8], [9]. There are lot of factors that influence drop-out, user's perspective or emotions is one of the major concerns. Using survival modelling technique and lexicon based approach, [10] it is shown that student sentiment towards MOOC has impact on the attrition rate over time. On the other hand student performance and learning outcomes are important given the high attrition rates, improving factors that affect performance can help improve drop-out rates. Student generated text data is mined to quantify their impact of

performance and learning outcomes [6]. Authors [6] use a lexicon based approach and study the correlation of student sentiment with quiz and assignment grades.

E-Learning is considered as internet based learning, use of technological and digital tools to deliver educational content [11], [12]. To build an effective e-learning system it is necessary that the instructor gains some insight and knowledge about user's opinion and or sentiment towards the technology used and materials covered. Authors Kechaou et al. [13], use feature selection and hybrid classifier for sentiment classification of e-learning blogs. They suggest that Information Gain (a criteria for measuring goodness) outperforms the other two features Mutual Information and CHI statistics. Similarly [14], [15], [16] proposed lexicon-based approaches and machine learning classifiers including support vector machine, and probabilistic approach based on Latent Dirichlet Allocation (LDA) to identify the student sentiment as either positive, negative or neutral. Authors Binali et al. [17], identify emotional reaction of students towards the e-learning courses. They classify the student response into one of the following emotional state 'confusion', 'happy', 'interest', and 'sad'.

2.2 Sentiment Analysis on Regular Courses

There are number of advantages when student feedback is collected in real-time, it helps the student see their concerns addressed and take benefits during their course of study. Authors Altrabsheh et al. [18], collect real time student feedback and label the data into three sentiment class 'positive', 'neutral', and 'negative' with help of three experts. The learning performance was investigated with the following machine learning techniques: Naive Bayes, Complement Naive Bayes, Maximum Entropy, and Support Vector Machine. They achieve good results with Support Vector Machine and Complement Naive Bayes. In a similar way authors Leong et al. [19] use prompt feedback and propose the use of short message service (SMS) for student evaluation and explore the application of text mining in particular Sentiment Analysis ('positive' and 'negative')on SMS texts. They show the positive and negative aspects of lecture in terms of the conceptual words extracted and text link analysis visualization. Authors Altrabsheh et al. [20] explore approaches for real time feedbacks. This work discusses how feedback is collected via social media such as Twitter and apply Sentiment Analysis to improve teaching called as Sentiment Analysis for Education (SA-E). This system collects data from Twitter where the students provide their feedback. The text data after pre-processing and extracting features including: term presence and frequency, N-gram position, part-of-speech, syntax, and negation. Later the text is analysed via Naive Bayes and/or Support Vector Machine which categorizes the whole post as either 'positive' or 'negative'.

Authors [21], classify student's feedback into 'positive' or 'negative' and suggest that Naive Bayes performs better with good recall. Authors Jagtap et al. [22] classify student feedback data classifying into 'positive' and 'negative' categories by using of hybrid approach combining Hidden Markov Model (HMM) and Support Vector Machine (SVM). Though they have concluded that applying advance feature selection method combined with hybrid approach work well for complex data, their works did not show the results of classification model for validation.

Authors Rajput et al. [23] apply text analytics methods on student's feedback data and obtain insights about teacher's performance with the help of tag clouds, and sentiment score. In this work the authors use sentiment dictionary Multi-Perspective Question Answering (MPQA) [24] to find words with positive and negative polarity. By combing the word frequency and word attitude the overall sentiment score for each feedback is calculated. Finally they have compared the sentiment score with Likert scale based teacher evaluation and conclude that Sentiment score with word cloud provide better insights than Likert scale results.

2.3 Emotion Mining on Student Feedback

Authors Kim et al. [25] compare the performance of categorical model and dimensional model by grouping the fine-grained emotion into more generic classes. They use lexicon-based approach and classification models including Majority Class Baseline (MCB), Keyword Spotting (KWS), and Dimension based estimation, CNMF - NMF based categorical classification. It is observed that NMF based categorical model and dimensional model have better performance.

In this paper we propose analyzing the qualitative end-of-course teacher evaluations for both traditional classroom and online courses with multi-label fine grained emotions such as 'anger-fear', 'trust-anticipation', 'sadness-disgust', 'joy-trust', 'anger-fear', and 'fear-disgust' with the help of National Research Council - NRC Emotion lexicon and combing the word frequency and sentiment score to determine the overall sentiment - emotion associated with student comments.

3. METHODOLOGY

3.1 Data Collection and Extraction

We use the web-based course evaluation system by UNC Charlotte, administered by Campus Labs to collect data for this study. The student feedback (qualitative data) for instructor is collected for the terms 2018 spring and summer sessions. The data is merged with the existing data corpus collected as part of similar study [27] which includes data for the academic years 2013 until 2017. The following fields are extracted from the collected html files using jsoup [28] Java library: academic year, term, course name, question, comments. The data contains 1070 instances. Sample student feedback is shown in Table. 1.

Table 1. Sample Student Feedback

Comments
Easily available to communicate with if needed
The course has a lot of valuable information
Get rid of the group project
There was no enthusiasm in the class. The instructor should make the class more lively and interactive
Best professor

3.2 Pre-Processing

The process of removing noisy data, for instance removing special characters, stop words is called pre-processing. It is the most important step for any text data as removal of noisy data helps improve the accuracy or classification rate. We use python Natural Language Toolkit (NLTK) [29] to perform tokenization, case-folding, stop words removal.

3.2.1 Tokenization

The process of removing certain special characters like punctuation and splitting the sentence into pieces of words called tokens is called tokenization.

3.2.2 Case-Folding

Text written by students on the evaluation forum is not very formal and may contain both upper case and lower case letters. In order to make it better for further processing of the text we perform case-folding where all of the text is converted to lower-case.

3.2.3 Stop Words Removal

Some of the words in English language are frequently used in order to make the sentence more complete in terms of grammar. For instances words like ‘am’, ‘is’, ‘was’, ‘are’ are not much useful in terms of the context of the sentence, in particular to emotions. Python Natural Language Toolkit (NLTK) [29] corpus contains list of stop words which is used as part of this step.

In the pre-processing step, certain comments which are not valid are removed, for instance comments with only ‘n/a’, ‘NA’, etc. The pre-processed dataset contains close to 815 records in the dataset.

3.3 Emotion Labeling

We use the National Research Council - NRC Lexicon [30], [31] for the purpose of Emotion labeling. NRC Emotion lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, disgust, surprise, trust, joy, and sadness) and two sentiments (positive and negative). The Annotations in the lexicon are at WORD-SENSE level. Each line has the format: $<$Term$>$ $<$AffectCategory$>$ $<$AssociationFlag$>$ as shown in Figure. 1.

<term><tab>**<AffectCategory>**<tab>**<AssociationFlag>**

<term>: is a word for which emotion associations are provided;

<AffectCategory>: is one of eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, or disgust) or one of two polarities (negative or positive);

<AssociationFlag>: has one of two possible values: 0 or 1. 0 indicates that the target word has no association with affect category, whereas 1 indicates an association.

Figure 1. NRC Emotion Lexicon – Word Level Annotation.

Each student evaluation comment can contain multiple emotions, as a student can have emotions like ‘trust’ and ‘anticipation’ together in terms of course evaluation. For instance consider the following comment from the dataset “The book which was chosen for this course is an amazing learning tool. There is a lot of very useful and necessary information covered in the textbook. This hardly applies to all classes. I hope that the instructor will continue using this book in the future”. This shows that the student has trust that the book used for the course has good content and also anticipates that it will be used in future semesters. This kind of knowledge extracted from the student evaluation help the instructor gain better understanding the course delivery and student expectations.

Student comments are processed as tokens and calculate score with respect to each of the eight emotions ‘anger’, ‘fear’, ‘sadness’, ‘disgust’, ‘surprise’, ‘anticipation’, ‘trust’. After the entire comment is processed the emotion which has the highest score is assigned as the final label together with the second most frequent emotion with respect to that student comment. As part of Emotion labeling if the final emotion score is zero then those records are omitted from the dataset.

3.4 Visualization

This paper focuses on identifying if the students are feeling better in a way the course is delivered with changes including Light Weight teams, flipped class room, and active learning methodologies. After labeling the students feedback with appropriate multi emotion class, the data is used to visualize the results over the years 2013 to 2018 using temporal properties. For visualization Python and Tableau software [32] is used.

4. RESULTS

We have explored the results using bar graph and word cloud using the temporal property of the data ‘Year’. This helps the instructor better understand how the emotions change over the years and what changes helped students.

Figure. 2. shows the bar graph over the years ‘2013’ until ‘2018’ with the multi-emotions grouped into three sentiment classes of ‘positive’, ‘negative’, and ‘neutral’. The emotions ‘anticipation’, ‘trust’, ‘surprise’, ‘joy’ which are considered to be positive emotions are marked as with a score of +1 and ‘sadness’, ‘disgust’, ‘fear’, ‘anger’ which are considered to be negative emotions are marked with a score of -1. ‘Anticipation’ and ‘Surprise’ can also fall under the negative category, which is handled by the following method. The final emotion class is determined to be positive if the overall score is greater than or equal to 1, negative if the overall score is less than or equal to -1 and neutral if it is 0. After the labeling the data is grouped based on the attribute year.

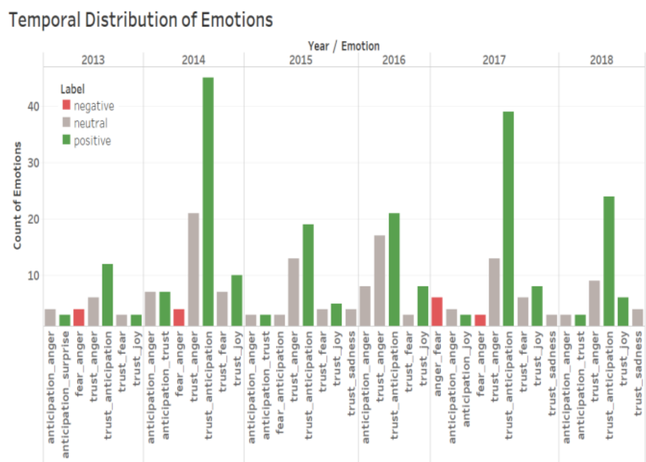

Figure 2. Temporal Distribution of Emotions.

In Figure. 3. we see that 2013, 2014, and 2015 has more negative words compared to the years 2016, 2017, and 2018 where we could observe positive words like ‘helpful’, ‘resources’, ‘good’, ‘information’.

In 2017 Active Learning methods were implemented in the courses, including Light Weight Teams where students work in small groups and have little direct impact on each other’s final grades, with significant long-term socialization, and Flipped Classroom where students watch lecture videos and work on related in-class activities. Therefore, we claim that the implementation of Light Weight Teams and Flipped Classroom Active Learning methods increase positive emotions among students and improve their learning experience.

Figure 3. Temporal Distribution of Word Cloud - Most frequent words appear with largest font. Positive words in green and Negative words in red.

5. CONCLUSIONS

We perform emotion detection on the qualitative feedback provided by students in end of course evaluations for computer science course. The student comments are labeled with eight human emotions: ‘anger’, ‘fear’, ‘joy’, ‘surprise’, ‘anticipation’, ‘disgust’, ‘sadness’, and ‘trust’ along with the two sentiment polarities ‘positive’ and ‘negative’. The emotions are analyzed to identify the impact of active learning methods incorporated in the courses during the years 2016, 2017, and 2018. Results show that words associated with positive emotions have increased in the recent years. Therefore, we claim that the implementation of Light Weight Teams and Flipped Classroom Active Learning methods increase positive emotions among students and improve their learning experience. In the future we plan to extend this work to focus on women and minorities in computing discipline.

6. REFERENCES

[1] Latulipe, C., Long, N. B., & Seminario, C. E. (2015, February). Structuring flipped classes with lightweight teams and gamification. In Proceedings of the 46th ACM Technical Symposium on Computer Science Education (pp. 392-397). ACM.

[2] Dorodchi, M., & Dehbozorgi, N. (2017, June). Addressing the Paradox of Fun and Rigor in Learning Programming. In Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education (pp. 370-370). ACM.

[3] Gehringer, E. F., & Peddycord III, B. W. (2013, March). The inverted-lecture model: a case study in computer architecture. In Proceeding of the 44th ACM technical symposium on Computer science education (pp. 489-494). ACM.

[4] MacNeil, S., Latulipe, C., Long, B., & Yadav, A. (2016, February). Exploring lightweight teams in a distributed learning environment. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education (pp. 193-198). ACM.

[5] Silpasuwanchai, C., Ma, X., Shigemasu, H., & Ren, X. (2016, June). Developing a comprehensive engagement framework of gamification for reflective learning. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems (pp. 459-472). ACM.

[6] Tucker, C., Pursel, B. K., & Divinsky, A. (2014). Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes. The ASEE Computers in Education (CoED) Journal, 5(4), 84.

[7] Ortigosa, A., Mart ń, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in human behavior, 31, 527-541.

[8] Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. Journal of interactive Media in education, 2012(3).

[9] Rivard, R. (2013). Measuring the MOOC dropout rate. Inside Higher Ed, 8, 2013.

[10] Wen, M., Yang, D., & Rose, C. (2014, July). Sentiment Analysis in MOOC Discussion Forums: What does it tell us?. In Educational data mining 2014.

[11] Nichols, M. (2003). A theory for eLearning. Educational technology & society, 6(2), 1-10.

[12] Cross, J. (2004). An informal history of eLearning. On the Horizon, 12(3), 103-110.

[13] Kechaou, Z., Ammar, M. B., & Alimi, A. M. (2011, April). Improving e-learning with sentiment analysis of users' opinions. In 2011 IEEE Global Engineering Education Conference (EDUCON) (pp. 1032-1038). IEEE.

[14] Mart ń, J. M., Ortigosa, A., & Carro, R. M. (2012, October). SentBuk: Sentiment analysis for e-learning environments. In 2012 International Symposium on Computers in Education (SIIE) (pp. 1-6). IEEE.

[15] Song, D., Lin, H., & Yang, Z. (2007, September). Opinion mining in e-learning system. In 2007 IFIP International Conference on Network and Parallel Computing Workshops (NPC 2007) (pp. 788-792). IEEE.

[16] Colace, F., De Santo, M., & Greco, L. (2014). SAFE: a sentiment analysis framework for e-learning. International Journal of Emerging Technologies in Learning (iJET), 9(6), 37-41.

[17] Binali, H. H., Wu, C., & Potdar, V. (2009, June). A new significant area: Emotion detection in e-learning using opinion mining techniques. In 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies (pp. 259-264). IEEE.

[18] Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Learning sentiment from students’ feedback for real-time interventions in classrooms. In Adaptive and Intelligent Systems (pp. 40-49). Springer, Cham.

[19] Leong, C. K., Lee, Y. H., & Mak, W. K. (2012). Mining sentiments in SMS texts for teaching evaluation. Expert Systems with Applications, 39(3), 2584-2589.

[20] Altrabsheh, N., Gaber, M., & Cocea, M. (2013, June). SA-E: sentiment analysis for education. In International Conference on Intelligent Decision Technologies (Vol. 255, pp. 353-362).

[21] Dhanalakshmi, V., Bino, D., & Saravanan, A. M. (2016, March). Opinion mining from student feedback data using supervised learning algorithms. In 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC) (pp. 1-5). IEEE.

[22] Jagtap, B., & Dhotre, V. (2014). SVM and HMM based hybrid approach of sentiment analysis for teacher feedback

assessment. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 3(3), 229-232.

[23] Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-based sentiment analysis of teachers' evaluation. Applied Computational Intelligence and Soft Computing, 2016, 1.

[24] Stoyanov, V., Cardie, C., & Wiebe, J. (2005, October). Multi-perspective question answering using the OpQA corpus. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 923-930). Association for Computational Linguistics.

[25] Mac Kim, S., & Calvo, R. A. (2010, June). Sentiment Analysis in Student Experiences of Learning. In EDM (pp. 111-120).

[27] Angelina, A.T., & Jaishree, R. (2018 September). Emotion Mining from Student Comments a Lexicon Based Approach for Pedagogical Innovation Assessment. In European Journal of Education and Applied Psychology (issue. 3, pp. 3-13). DOI: https://doi.org/10.29013/EJEAP-18-3-3-13

[28] Hedley, J. (2009). jsoup: Java html parser. 2009-11-29)[2015-06-12] http://jsoup. org.

[29] Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

[30] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. Computational Intelligence, 29(3), 436-465.

[31] Mohammad, S. M., & Turney, P. D. (2010, June). 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 (pp. 26-34). Association for Computational Linguistics.

[32] Cleveland, W. S. (1993). Visualizing data. Hobart Press.