

Text Mining Online Discussions in an Introductory Physics Course

Patrick Kelley,¹ Rebecca Lindell,² and A. Gavrin¹

¹*Department of Physics, Indiana University - Purdue University Indianapolis, 402 N. Blackford St., Indianapolis, IN, 46202*

²*Tiliadal STEM Education Solutions, 5 N 10th Suite A, Lafayette, IN, 47901*

We implemented a social networking platform called Course Networking (CN) in IUPUI's introductory calculus based mechanics course and recorded three semesters of online discussions. We used the Syuzhet package in R to evaluate sentiment in the recorded discussions, and to quantify the incidence of eight basic emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. We applied this text mining method to over nine thousand posts and replies to identify and analyze student sentiment during three semesters. We also investigated the variation of these emotions throughout the semester, the role played by the most vocal students as compared to the least frequent posters, and gender differences. With an abundance of students' online discussions, text mining offers an expedient and automated means of analysis, providing a new window into students thinking and emotional state during semester-long physics courses.

I Introduction

Social media has become widespread and is particularly common among college-aged students. Facebook alone has over 1.65 billion users worldwide [1]. Further, there is growing evidence that social media may be used effectively in an educational setting [2]. One strength of social media is in the potential for students to form communities. Students can offer help to their peers in learning the course content, as well as provide one another with emotional support, encouragement, and reassurance, leading ultimately to a sense of belonging in the class. These benefits can occur naturally in small classes, and on residential campuses; however, they can be elusive in large classes, especially on commuter campuses.

We have established a social-media platform in our courses, and found that it provides a window into students' thinking and behavior with respect to the course. Faculty reviewing the transcripts of students' interactions can gain insight into what problems students are having, and how they are seeking solutions. However, the quantity of information can be unwieldy, particularly with a large class. One possible solution is to use text mining techniques.

Text mining is a method used to extract useful information from written words. Text mining applies pattern recognition and statistical analysis to natural language, which is often not structured. Common methods of data mining are network analysis [4, 5], topic modeling [6–9], and cluster analysis [10]. Many of these methods may be useful, even when data is not numerical.

The power of using these tools is the ability to handle quantities of textual data too large to be manually read and analyzed. Thus, text mining offers a solution that goes beyond numerical data mining. Of course, there are concomitant difficulties. Because text mining averages over large quantities of text, some nuances are lost. Whether this trade-off is acceptable depends on the research context.

In this paper, we use text mining, particularly sentiment analysis [11, 12, 19], a tool that allows us to characterize the emotions expressed in our corpus, the body of text comprised of student submissions to our social network. The characterization is given in terms of sentiment, which may be positive or negative, and eight emotional categories. Other studies have used sentiment analysis to evaluate a massive open online course

(MOOC) [12]. However, the literature on social media analysis in physics education is limited.

A compelling reason to perform sentiment analysis is the role that emotion has on learning. Pekrun, Goetz, Wolfram, and Perry [3] did a comprehensive study of the impact of emotion on education, finding strong interrelationship between emotion and students' learning strategies, cognitive resources, self-regulation, motivation, and academic achievement. With text mining, the process of extracting sentiment and emotion can now be automated and quantitatively assessed.

II Research Questions

The investigation by means of text mining the recorded online discussions attempts to address three central questions:

Research Question 1 (RQ1): Can we use text mining methods to gain insight into class sentiment?

Research Question 2 (RQ2): Are there gender differences in the observed sentiment?

Research Question 3 (RQ3): Are there differences in sentiment between students who participate heavily and those who do not?

III Educational Setting

Our data is derived from three semesters of an introductory physics course offered at Indiana University - Purdue University Indianapolis (IUPUI). PHYS 152 is a calculus-based mechanics course with a typical enrollment of 150-200 students. The course is comprised of two lectures, two recitations, and one lab each week. We analyze the fall 2014, 2015, 2016 data. The same instructor (AG) taught the course during each of these semesters, and implemented Course Networking (CN), a free cloud-based social media platform designed for academic use [13]. Students were encouraged to communicate with one another using CN, to share both academic help and emotional support. All semesters' courses also used Turning Technologies clickers for implementation of Peer Instruction [14]. There were also several significant differences among the semesters. The fall 2014 semester utilized the Smart Physics system, now called FlipItPhysics [15], for both text and homework. In the

fall of 2015 and 2016, we switched to Webassign [16] and its incorporation of the e-book version of *Physics for Scientists and Engineers* [17]. In addition, the exam schedule was changed. Fall 2014 had 3 exams and a final whereas fall 2015 and 2016 had 5 exams and a final.

The social networking platform we used (CN) is intended as a transition from conventional course or learning management systems (LMSs), to a model focused on student-to-student communication. Traditional LMSs are faculty-centered. In contrast, CN is student-centered. Its interface is similar to a social networking platform such as Facebook, in that its primary function is to facilitate communications among participants. Most of the screen space is occupied by a running list of students' "posts" and "reflections". Posts can include embedded links, videos, and downloadable documents. Participants can also "like" others' posts. In addition, students obtain award points called "Anar Seeds" (pomegranate seeds) for posting, commenting, and liking other class members' posts. In all semesters studied, students earning 350 anar seeds by the end of the course received 5% "extra credit".

The power of CN is its ability to offer students a means to form connections. IUPUI is largely a commuter school; less than 10% of students live on campus. As a result students often find it difficult to connect with one another, form study groups, etc. Using CN, students can easily collaborate and communicate with the entire class. This allows students to schedule meetings, express concerns about the class, and exchange reassurance with classmates.

By the end of each semester, a large number of students' comments and replies are recorded. This "corpus" of posts and replies represents a detailed instantiation of our students' impressions, opinions, and connections in the course. This corpus thus represents a new window into student interactions and learning.

IV Methodology

Out of the multiple free text mining tools is R [18], an open source statistical and graphical programming language. We used "Syuzhet", an R package that extracts sentiment from text using three sentiment dictionaries [19].

We begin by importing the posts and replies. To gain a chronological account of the course, the posts and subsequent replies were sorted by time stamp. After sorting, the data was imported into three separate row tables, with sizes of 3311, 3165, 3241 for fall 2014, 2015, and 2016, respectively. We then converted the text corpus into an array of words and used our primary package, Syuzhet, to do the sentiment analysis. The algorithm runs over the vector and returns a summary of sentiments present. We used the "nrc" lexicon [22], which is named after the National Research Council Canada (NRC).

The algorithm compares words in the corpus to the NRC Sentiment and Emotion Lexicon, associating words with sentiment (positive or negative) and with eight emotions. Eight standard emotions (anticipation, trust, joy, surprise, fear, sadness, anger, and disgust) have been used as a set of prototypical emotions for many years [20]. We investigated other lexicons, and found no significant differences in our results. Figure 1 gives an illustration of this process.

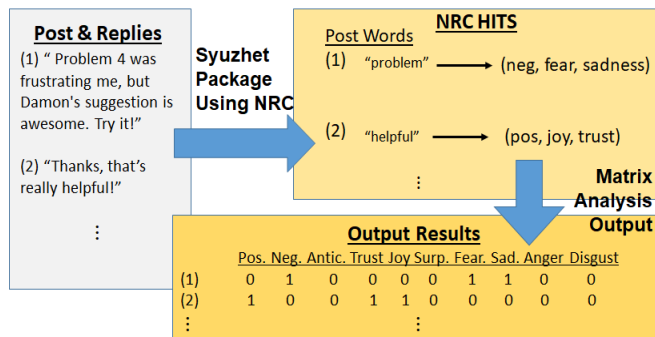


FIG. 1: Schematic sentiment analysis procedure

As can be seen in Fig. 1, the word "problem" elicits a negative sentiment although the usage in this example isn't meant as such. Such "mis-assignments" are an example of lost nuance, but make little difference after averaging over the full corpus. We discuss this further in section VI.

V Results and Analysis

RQ1: We applied the Syuzhet sentiment analysis to the corpus of the three semesters. Overall, we found that the class sentiment and emotion were generally positive, and quite similar over these three semesters, as illustrated in Figure 2a. As a result, we will focus the remainder of our analysis and discussion on the fall 2015 semester.

The large number of posts and replies allows us to carry out this analysis for subsets of the data. In particular, we explore breaking down the data as a time series at the one week level, by gender and by subsets of the students based on numbers of posts.

Figures 2b, 2c, and 2d show an analysis of the Fall 2015 semester as a time series. The number of posts in each week is shown in Fig. 2b, the sentiment is shown in Fig. 2c, and Fig. 2d shows the breakdown by emotion. We have listed the emotions in order of decreasing negative connotation: disgust, anger, fear, sadness, surprise, anticipation, trust and joy (our own ranking). The weeks were defined as Mondays through Sundays beginning a week before classes begins. The observed period then lasts 17 weeks including the final exam week.

We were surprised to see how little structure is apparent in the time series. The sentiment begins highly positive and decays slowly over the first seven weeks of the semester. After some fluctuation, sentiment recovers slowly as the semester nears its end. There are no clear features associated with exams. Further work will show if this result depends on the instructor, course structure, or other variables. It is possible to automate this analysis, giving the instructor weekly updates on class sentiment during the semester.

RQ2: We also wished to determine if there were gender related differences in sentiment. We acknowledge that the binary view of gender is at best a first-order model, simplifying a wide range of sociocultural factors and nuanced gender identities into two categories. Figure 3 shows the positive and negative sentiments expressed during fall 2015 by gender. We observe no significant gender-based differences between the num-

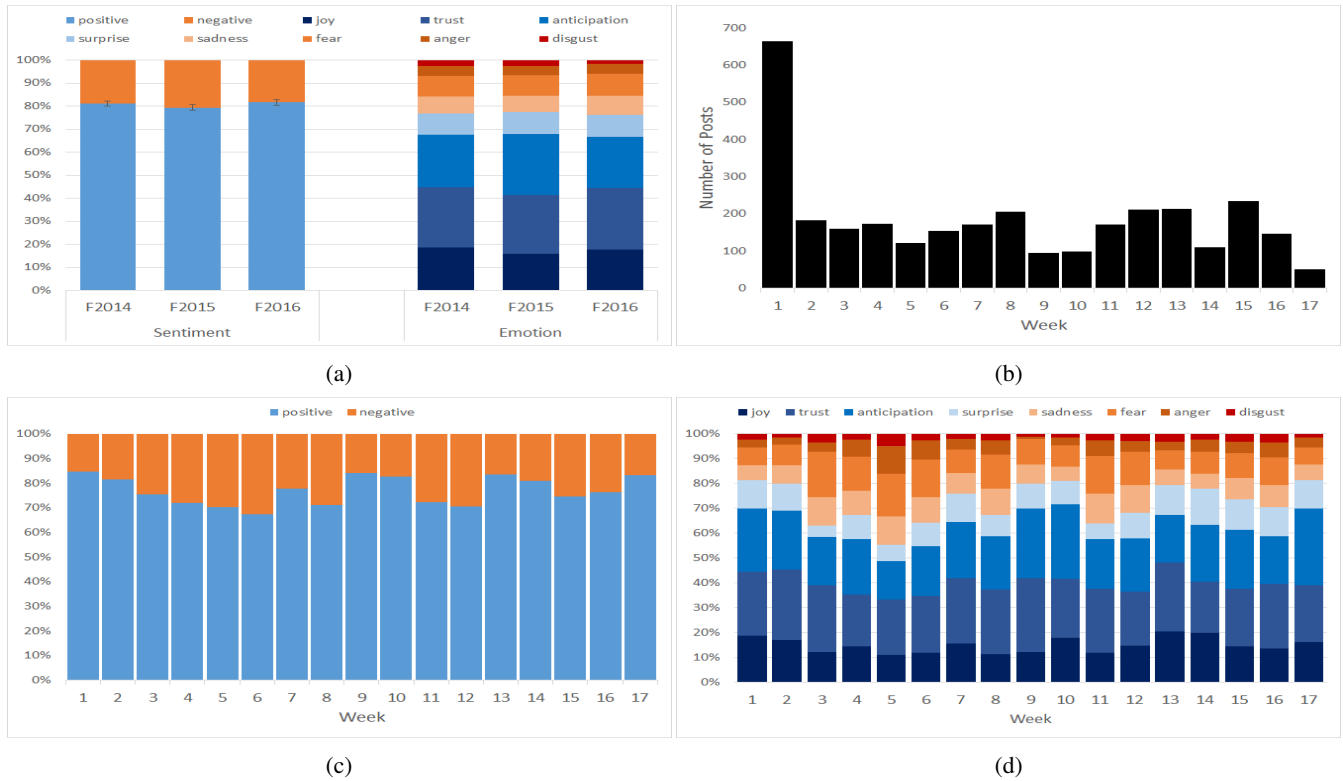


FIG. 2: (2a) Cumulative sentiment & emotion by semester and Fall 2015 breakdown by weekly inspection: (2b), (2c), and (2d) displays the number of posts, the sentiment distribution, and the distribution of eight emotions in the semester, respectively

bers of contributions or sentiments in this data (p-values: 0.96 and 0.56, respectively) or in the two semesters not shown.

RQ3: We were also interested in differences in sentiment between frequent and infrequent users. To determine if this is the case, we broke the class down into quartiles by frequency. Quartiles designated Q1, Q2, Q3, and Q4, had 32, 32, 31, and 32 students and a range of 1-4, 4-14, 16-33, and 34-122 posts respectively. Figure 4 displays the mean sentiment in each of these quartiles. Running a Tukey Significant Difference Test [21] comparing each pair of quartiles, we found a significant difference only between Q1 and Q2. This may be due to the fact that self-introductions were encouraged during the first week

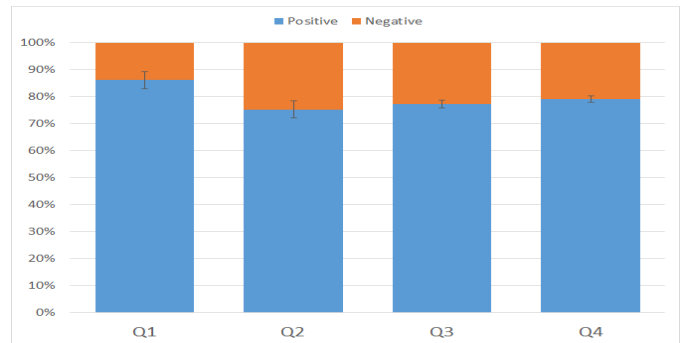


FIG. 4: The sentiment for each quartile of students

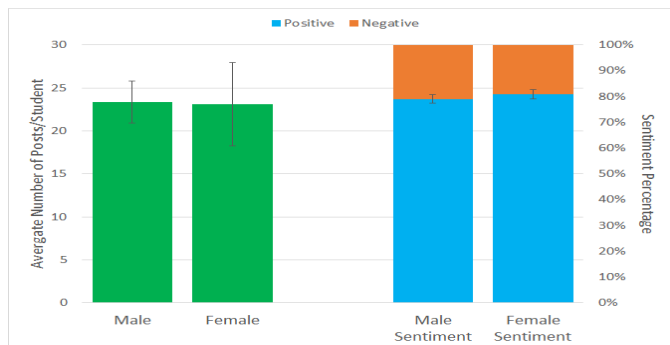


FIG. 3: Gender analysis of the average posts/person (left) and average sentiment (right)

of class, so many students who only posted once or twice may have only posted about themselves. We may also conclude that frequent posters do not dominate the class sentiment. As a side note, dropping the top three posters (with 90, 102, 122 posts made) as outliers made no difference in this conclusion.

VI Discussion

We were pleased to see the overall positive sentiment, and the generally positive emotions expressed. In all semesters, some structures are apparent. For instance, all semesters show a lull in activity associated with the Thanksgiving holiday in week 14. It is also apparent that changes in sentiment are re-

flected in the mix of emotions expressed in Figure 2c and 2d because of the correlation between sentiment and emotion which is how we interpreted ranking the emotions by sentimental connotation as listed in (our) order in the Results and Analysis section. But it is important to know that there is a distinction between sentiment and emotion. The word "rocket" has no sentiment but is associated with the emotion anger.

Sentiment analysis tools are relatively new, and have some shortcomings. The NRC Word-Emotion Association Lexicon lists over 14,000 words, but it is by no means complete. For instance, "fault" is included in the lexicon, but "faults" is not. We attempted to remedy this by removing such endings as "s", "ed", and "ing" from verbs and nouns-"stemming" the text corpus. Unfortunately, the stemming process also alters some associations. For instance, the word "acquiring" is associated with anticipation, but "acquire" is not. We made a comparison, and found that stemming the corpus made no discernible difference in our results. Only the results from the stemmed corpus are reported here.

We wish to note that some words commonly used in physics classes have sentiment or emotional associations connected with their use in common language. As examples other than the word "problem" in Fig. 1, the word "force" is associated with negative sentiment, and with the emotions fear and anger. We also found that masking a glossary of text book physics words produced negligible effects, with "force" as the only masking word for the fall 2015 semester appearing only 72 times out of the roughly 5000 sentimental word associations. These results reflect the strength of mining a large corpus. Although some nuances are lost, the overall results reflect averages over a large data set and are thus quite robust.

VII Conclusions & future work

We find that we can use text mining to gain insight into the sentiment of our classes (positive vs negative) as well the emo-

tions our students expressed. Analyzing our data as a function of time allowed us to gain a rough depiction of the variable nature of students' sentiment and emotion throughout the semester. Additionally, we found no significant gender differences in sentiment on frequency of contributions, a reassuring sign. Finally, we find that course sentiment is not related to frequency of use, suggesting that the most frequent posters do not dominate the measured sentiment. Any sufficiently large course could use this text mining tool in conjunction with CN or a similar social media platform. This could potentially allow instructors to monitor course sentiment in real time.

This analysis far exceeds what we could learn from the brief snapshot typically provided by the end-of-semester student evaluations. In the future, we expect the combination of social-media and text mining to provide a rich new window on the academic and emotional development of students in our classes. We have several planned research projects using this combination of social-media and text mining software.

We were surprised to see how little structure was evident in our time series data, and wish to see if this relative stability depends on the instructor, course structure, or other variables. We will also work to quantify the benefits of using a social media tool in our courses. Since we have found no initial gender-related differences, it may be possible to explore the use of social media to promote identity development in female students and to explore these questions with respect to underrepresented racial and ethnic groups. A deeper analysis of these sets of data as a specific function of time, as in weeks before exams or when a specific topic is taught may indicate frustration or confidences with various topics. Because CN is relatively new, further development of CN itself is possible. We will work with the development team to develop analytic tools that will allow faculty to gauge their students' mood, degree of engagement, and learning in real time.

-
- [1] Constine, J., <https://techcrunch.com/2016/04/27/facebook-q1-2016-earnings/> retrieved 6/28/16.
- [2] Gregory, Peter, Gregory, K. and Eddy, E., J. Computers in Math. And Sci. Teaching, **33** (1), 5 (2014).
- [3] Pekrun, R., et al. "Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research." Educational psychologist 37.2 (2002): 91-105.
- [4] Traxler, A., Proc. of the Physics Educ. Research Conf. 2015, 331 (2015).
- [5] Brewe, E. Proc. of the Physics Educ. Research Conf. 2009, **1179**, 105 (2009).
- [6] Hong, L., and Davison, B. D., Proc. of the 1st Workshop on Social Media Analytics (SOMA '10), 80 (2010).
- [7] Alghamdi, R. and Alfalqi, K., Intern. Journ. of Adv. Comp. Sci. and Appl. (IJACSA), **6**(1), 147 (2015).
- [8] Ramage, D., Rosen, E., Chuang, J., Manning, C. D., and McFarland, D. A., Workshop on Applications for Topic Models, NIPS (2009).
- [9] Hu, X., and Liu, H. *Mining Text Data*, 385-414, Springer US (2012).
- [10] JafariAsbagh, M., Ferrara, E., Varol, O., Menczer, F., and Flammini, A. Social Network Analysis and Mining, **4**(1), 1 (2014).
- [11] Altrabsheh, N., Gaber, M., and Cocca, M. *Proc. of the 5th KES Int. Conf. on Intelligent Decision Technologies*, 26 (2013).
- [12] Wen, M., Yang, D., and Rosé, C. P., *Proc. of the 7th Int. Conf. on Edu. Data Mining*, 130 (2014).
- [13] <http://www.thecn.com>
- [14] Mazur, E., *Peer Instruction* Pearson, Upper Saddle River, NJ (1997).
- [15] <https://www.flipitphysics.com>
- [16] <https://www.webassign.net>
- [17] Tipler, Paul A., and Mosca, G., *Physics for Scientists and Engineers, 6th edition* (W. H. Freeman, New York, 2008).
- [18] <http://www.R-project.org/>
- [19] Mohammad, S. M., and Turney, P. D., *Computational Intelligence*, **29** (3), 436 (2013).
- [20] Plutchik, R., Emotion: Theory, research, and experience, **1**(3), 3 (1980).
- [21] Tukey, John W. "Comparing individual means in the analysis of variance." *Biometrics* (1949): 99-114.
- [22] The NRC Word-Emotion Association Lexicon was "crowd-sourced", with over 2000 people participating in the process. The volunteers established the associations of sentiment (negative vs. positive) and emotions with words. As an example of the associations, the word "abandon" has negative sentiment and two emotional associations: fear and sadness.