Sentiment Analysis in MOOC Discussion Forums: What does it tell us?

Miaomiao Wen
School of Computer Science
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA
mwen@cs.cmu.edu

Diyi Yang
School of Computer Science
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA
diyiy@cs.cmu.edu

Carolyn Penstein Rosé
School of Computer Science
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA
cprose@cs.cmu.edu

ABSTRACT
Sentiment analysis is one of the great accomplishments of the last decade in the field of Language Technologies. In this paper, we explore mining collective sentiment from forum posts in a Massive Open Online Course (MOOC) in order to monitor students' trending opinions towards the course and major course tools, such as lecture and peer-assessment. We observe a correlation between sentiment ratio measured based on daily forum posts and number of students who drop out each day. On a user-level, we evaluate the impact of sentiment on attrition over time. A qualitative analysis clarifies the subtle differences in how these language behaviors are used in practice across three MOOCs. Implications for research and practice are discussed.

Keywords
Sentiment analysis, Opinion mining, Massive Open Online Course, MOOC, Forum posts

1. INTRODUCTION
Working towards improving MOOCs, it is important to know students' opinions about the course and also the major course tools. Based on opinions extracted from students' reviews, previous work illustrates that the most important factor to students is who is teaching the course [1]. However, for a given MOOC that will be offered again by the same instructor team, it is more critical to know what can be improved in the course. Recent research on social media use has demonstrated that sentiment analysis can reveal a variety of behavioral and affective trends. For example, collective sentiment analysis has been adopted to find the relationship between Twitter mood and consumer confidence, political opinion [20], and stock market fluctuations [5]. Course forums provide students with the chance to engage in social learning in MOOCs [6]. Analyzing the data from this part of the course, we can infer important information about attitudes prior to and even in the absence of post-course surveys [31]. The contribution of this paper is an investigation into what sentiment analysis can tell us about the students' opinions towards the course. We also analyze the impact of sentiment on attrition over time in MOOCs using a survival modeling technique.

Despite the great potential, the current generation of MOOCs has so far failed to produce evidence that the potential is being realized. Of particular concern is the extremely high rate of attrition that has been reported. Much of this research focuses specifically on summative measures of attrition. They seek to identify factors that predict completion of the course, for example, by conducting correlational analysis between course completion and click stream evidence of engagement with course activities [12]. However, what we see is that attrition happens over time. While a large proportion of students who drop out either fail to engage meaningfully in the course materials at all or drop out after the first week of participation, a significant proportion of students remain in the course longer than that but then drop out along the way. This suggests that there are students who are struggling to stay involved. Supporting the participation of these struggling students may be the first low hanging fruit for increasing the success rate of these courses. Before we can do so, we need to understand better their experience of participation along the way as they struggle and then ultimately drop out. Thus, in this paper we employ a survival modeling technique to study various factors' impact on attrition over course weeks.

As a reflection of student experience communicated through their posts, we investigate sentiment expressed in course forum posts. While the association between sentiment with summative course completion has been evaluated in prior work [24], and while the impact of other linguistic measures and social factors on attrition over time has been published as well[31, 26], this is the first work we know of that has brought this lens to explore what sentiment can tell us about drop out along the way in this type of environment. In particular, we explore this connection across three MOOCs in order to obtain a nuanced view into the ways in which sentiment is functioning similarly and differently or signaling similar and different things across these three courses. Our goal is for this analysis to reflect some of the flexibility in how these linguistic constructs are used in practice in order to inform application of such techniques in future analysis in this community.

In the remainder of the paper, we begin by describing our dataset and discussing related work. Next, we explain how a collective sentiment analysis can reflect students' attitudes towards the course and course tools. In light of the collective sentiment analysis, we continue with a survival analysis that shows what sentiment can tell us about drop out along the way in MOOC environments. Finally, we conclude with a summary and possible future work.

2. COURSERA DATASET

The data used for the analysis presented here was extracted from three courses by permission from Coursera.org using a screen scraping protocol. The three courses cover a wide range of topics. Our dataset consists of three courses: one social science course, *Accountable Talk: Conversation that works*[1], offered in October 2013. We refer to this course as the *Teaching* course; one literature course, *Fantasy and Science Fiction: the human mind, our modern world*[2], offered in June 2013. We refer to this course as the *Fantasy* course; one programming course, *Learn to Program: The Fundamentals*[3], offered in August 2013. We refer to this course as the *Python* course. Statistics about the three courses are listed in Table 1.

MOOC	Active Users	Total Days	Total Posts	Avg. Posts Per Day
Teaching	1,146	53	5,107	96
Fantasy	771	43	6520	152
Python	3,590	49	24963	510

Table 1: Statistics of the three Coursera MOOCs. Active users refer to those who post at least one post in a course forum.

3. RELATED WORK

3.1 Sentiment Analysis for Social Media

Affect mined from Facebook and Twitter posts is known to be reflective of public behavior and opinion trends [20, 5]. The results generated via the analysis of collective mood aggregators are compelling and indicate that accurate public mood indicators can be extracted from online materials. Sentiment analysis has been used as an invaluable tool for identification of markers of affective responses to crisis [10], as well as depression [9], anxiety, and other psychological disorders [8] from social media sites. Using publicly available online data to perform sentiment analyses requires far less cost in terms of effort and time than would be needed to administer large-scale public surveys and questionnaires. Most MOOCs offer course forums as a communication and learning tool. While only a small percentage of students actively participate in the threaded discussions, if course instructors can use automated analysis of those posts as a probe that indicates whether things are going well in the course, and the analysis reveals something about what the issues are, they will be better prepared to intervene as necessary.

3.2 Sentiment Analysis for Educational Data Mining

Mackness et al. [15] posed the question of how to design a MOOC that can provide participants with positive experiences. Most of the prior work that addresses this question involved conducting surveys and interviews [25, 2]. In contrast, in some prior E-learning research, automatic text analysis, content analysis and text mining techniques have been used to mine opinions from user-generated content, such as reviews, forums or blogs [27, 4, 11]. Attitude is important to monitor since learners with a positive attitude have been demonstrated to be more motivated in E-learning settings [18]. Correspondingly, previous work reveals that boredom was associated with poorer learning and problematic behavior. In contrast, frustration was less associated with poorer learning [3]. Based on user-generated online textual reviews submitted after taking the courses, Adamopoulos [1] has applied sentiment analysis to collect students' opinions towards MOOC features such as the course characteristics and university characteristics. In that work, the goal was to determine which factors affect course completion, it is also important to address the related but different question of what can be improved when the course is offered again.

Given the recent work on MOOC user dropout analysis, very little has attempted finer-grained content analysis of the course discussion forums. Brinton et al. [6] identified high decline rate and high-volume, noisy discussions as the two most salient features of MOOC forum activities. Ramesh et al. [24] use sentiment and subjectivity of user posts to predict engagement/disengagement. However, neither sentiment nor subjectivity was strongly predictive of engagement in that work. One explanation is that engaged learners also post content with negative sentiment on the course, such as complaints about peer-grading. Thus, the problem is more complex than the operationalization used in that work. Taking the analysis a step further to explore such nuances is the goal of this paper.

4. METHOD

This work reflects on how sentiment analysis can be useful in a MOOC context. On the course-level, we use collective sentiment analysis, which has been successfully applied in many social media investigations, to explore the relation between opinions expressed by students and the students' dropout rate. To help MOOC instructors collect students' opinions towards various course tool designs, we extract the positive and negative sentiment words that are associated most with the course tool topic keywords. In order to understand the impact of sentiment on the user-level, we adopt survival analysis to to examine how sentiment that members have expressed and are exposed to in a particular week predicted their continued participation in the forum discussion.

4.1 Course-level Sentiment Analysis: Collective Sentiment Analysis

In this section we first describe how we use collective sentiment analysis to study students' attitudes towards the course and course tools based on forum posts. To improve MOOC design, it is important to obtain feedback from students. Most MOOCs conduct post-course surveys where students' opinions towards the course are elicited. However, only a very limited portion of students who registered for the course will actually fill out the survey. A discussion forum is

[1]https://www.coursera.org/course/accountabletalk
[2]https://www.coursera.org/course/fantasysf
[3]https://www.coursera.org/course/programming1

MOOC	Course	Lecture	Assig- nment	Peer- assessment
Teaching	820	725	904	97
Fantasy	731	327	2515	375
Python	1430	2492	3700	-

Table 2: Number of course tool-related posts in the three courses' forums. The Python course did not implement peer-assessment.

a natural place where students convey their satisfaction or dissatisfaction with the course. Can we analyze forum posts to infer students' opinions towards the course in the same manner that post-course surveys elicit such feedback from the students? If so, then tracking students' opinion based on daily forum post content could be a far more timely alternative to post-course surveys, and may provide a less biased view of the course because it would have the opportunity to capture the attitude of students who drop out before the post-course survey is administered.

Sentiment polarity analysis techniques applied to individual messages may make many errors, partly due to the extent to which the text is taken out of context. However, with a large number of such measurements aggregated together, the errors might cancel, and the resulting composite indicator may be a more faithful indicator of public opinion. In previous work on text-based social media sites, summary statistics derived from similarly simple sentiment analysis are demonstrated to correlate with many objective measures of population level behaviors and opinions [20, 21]. In this section, we use sentiment analysis to understand students' opinion towards the course and course tools.

4.1.1 Setting

To extract students' aggregate opinions on a topic, we need to first identify posts relating to a topic (*post retrieval step*). Then we need to estimate the opinion of these posts (*opinion estimation step*). Following the same methodology used in previous work [20], in the *post retrieval step*, we only use messages containing a topic keyword. To decide which topic keywords to use for each course tool of interest, we run a distributional similarity technique called Brown clustering [7, 13] on all three courses' posts in order to identify clusters of words that occur in similar contexts. It is conceptually similar to Latent Semantic Analysis, but is capable of identifying finer grained clusters of words. From the results, we construct keyword lists for topics by starting with hand-selected keywords, and then finding the clusters that contain those words and manually choosing the words that are human-identified as being related to the same topic. The numbers of posts retrieved for each topic are shown in Table 2.

- For *Course* topic, we use "the course", "this course", "our course" and the name of each MOOC.
- For *Lecture* topic, we use "lecture" and "video".
- For *Assignment* topic, we use "assignment", "essay", "reading" and "task".
- For *Peer-assessment* topic, we use "peer assessment", "peer grading", "assess your peer", "peer score", "peer feedback", "peer review" and "peer evaluation".

In the *opinion estimation step*, for each set of posts that are related to a topic, we define the **topic sentiment ratio** x_t on day t as the ratio of positive versus negative words used in that day's post set. Positive and negative terms used in this paper are defined by the sentiment lexicon from [14], a word list containing about 2,000 and 4,800 words marked as positive and negative, respectively.

$$x_t = \frac{Total\, poistive\, terms}{Total\, negative\, terms}$$

Day-to-day, the topic sentiment ratio rapidly rises and falls each day. In order to derive a more consistent signal, and following the same methodology used in previous work [20], we smooth the sentiment ratio with one of the simplest possible temporal smoothing techniques, a moving average over a window of the past k days:

$$MA_t = \frac{1}{k}(x_{t-k+1} + x_{t-k+2} + ... + x_t)$$

The moving average of sentiment ratio MA_t is our estimation of collective opinion expressed by the students in the course forum during day t.

4.1.2 Results

Part 1. Opinion towards the course

In this section, we explore the correlation between collective opinions mined from the forum posts and objective measures related to students' actual opinions.

To objectively measure students' opinions towards the course, we count how many students drop out of the course each day based on the students' login information. Here we consider that a student drops the course on day t if the student's last login date of the course is on day t. There are many students who just register for a course to see what it is about, without serious intention of actually taking the course. The number of students who drop out in the first week is much larger than the other weeks. We calculate the correlation between the number of users who drop the course and the Course sentiment ratio (MA_t) starting from course week 2 until the last course day in order to avoid the analysis being muddied by the very different factors that affect dropout in the first week.

In the Teaching course we can operationalize drop out in two different ways because we have both forum data and login data. In the other two courses, we have only forum data. Thus, we are able to measure the correlation between sentiment and dropout two different ways in the Teaching course, which enables us to understand how these two operationalizations may reveal different patterns, and then we can use that understanding to interpret what we find in the other two courses.

First we explore how sentiment shifts over the seven weeks of each course. In Figure 1, we show how the number of students who drop out and the Course topic sentiment ratio vary from day-to-day. In all three courses, Course sentiment ratio is much higher during the last course week.

		Course	Lecture	Assignment	Peer-assessment
Teaching	**Pos**	incredibly,benefits enjoyment,richer,greatest	incredibly,benefits,richer guarantee,gaining	substantive,benefits rich,soft,gaining	smart,kudos,praise prominent,mastery
	Neg	missed,negative,low-rated taxing,superficial	breaking,worry,missed challenging,thoughtless	struggles,taxing,poor struggled,unacceptable	riled,worry,missed challenging,conflict
Fantasy	**Pos**	avidly,incredibly,substantive benefits,guarantee	incredibly,kudos,smart beauteous,consistent	masterfully,avidly,substantive guarantee,admiration	consistent,benefits,richer competitive,richer,balanced
	Neg	damnation,lie,missed belated,negative	lie,breaking,wrong anxious,worry	shortcomings,confuse creeping,menace,flaky	negative,frustrating,wrong invasive,hate
Python	**Pos**	self-satisfaction,impresses providence,kudos,smart	remedy,convenient,merit gaining,smartest	incredibly,guarantee,richer benefits,proud	- -
	Neg	forbidden,unforeseen,worry breaking,challenging	embarrassing,worry, challenging,missed	unforeseen,confuse,bug swamped,shock	- -

Table 3: Sentiment words associated most with each course tool.

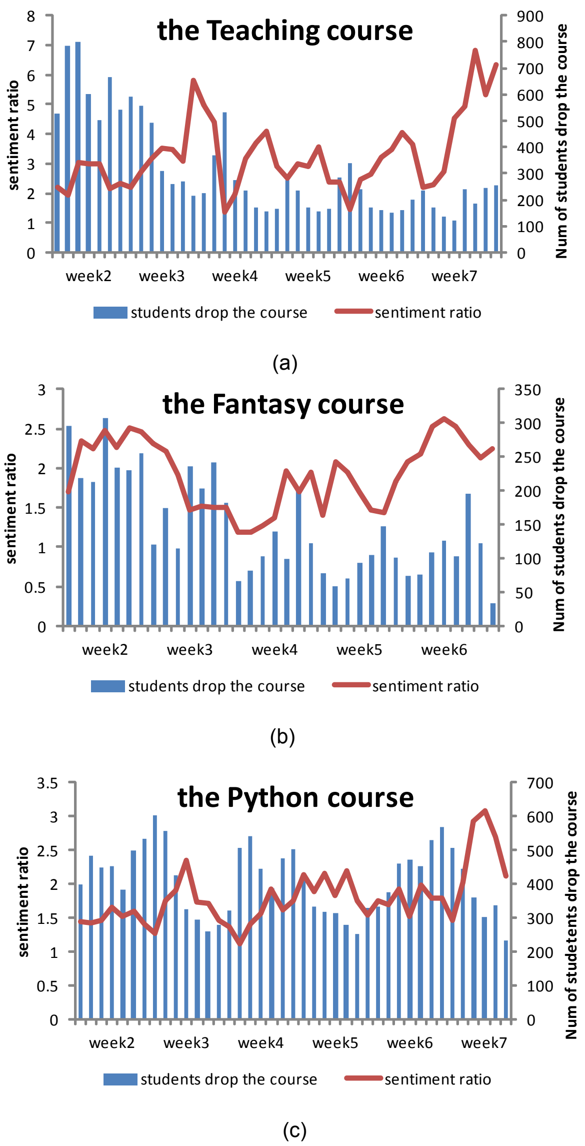

Figure 1: Moving average MA_t of Course topic sentiment ratio and number of students who drop out over course weeks. Window size k equals 3.

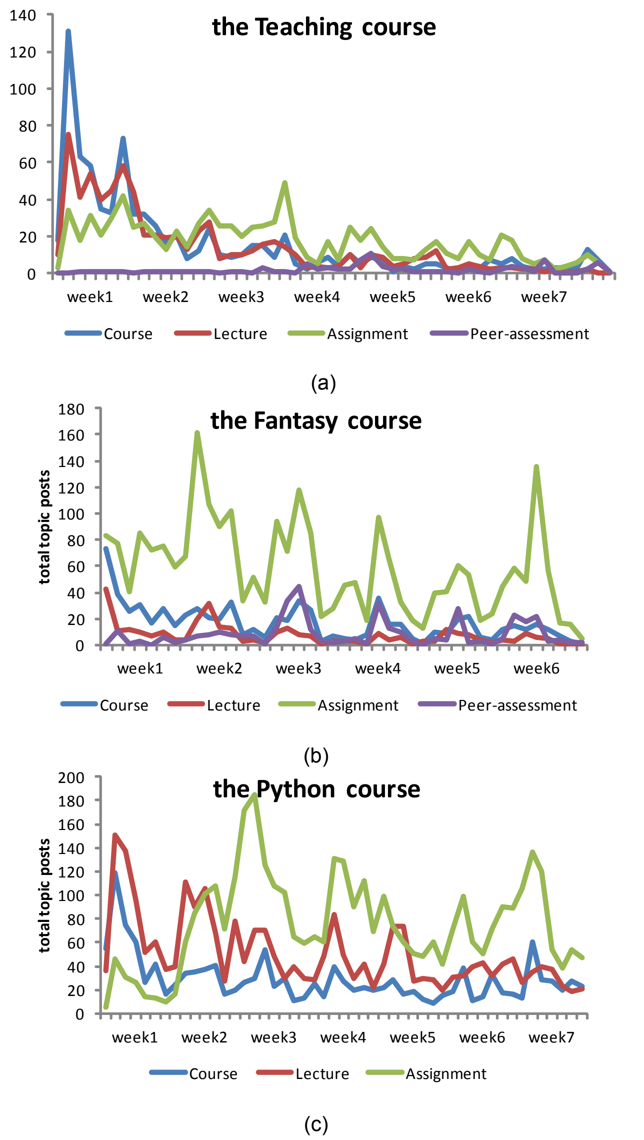

Figure 2: Trends of total number of course tool related posts.

Because many students post "thank you" and positive feedback towards the course during the last course week. First we consider a student to drop out from the course forum if the student posts his/her last post on day t. Across all three courses, we observe a trend in the expected direction, namely that higher sentiment ratios are associated with fewer dropouts, which is significant in two out of the three courses(r = -0.25, $p < 0.05$ for the Teaching course; r = -0.12 for the Fantasy course(Figure 1(b)); r = -0.43, $p < 0.01$ for the Python course(Figure 1(c))). In the Teaching course, where we can determine dropout more precisely from login information, we see a stronger correlation(r = -0.39, $p < 0.01$). This analysis serves as a validation that the collective sentiment extracted from these posts can partly reflect the opinion of the entire student population towards the course.

Part 2. Opinion towards the course tools

As important components of the course, the course tools have a big impact on a student's experience in a MOOC. For example, peer-assessment serves as a critical tool for scaling the grading of complex, open-ended assignments to MOOCs with thousands of students. But it does not always deliver accurate results compared to human experts [22]. In the Fantasy course, we see heated discussions about peer-assessment during course week 3 and week 4 when peer-assessment was conducted. One discussion thread with the title "Why I'm dropping out (the fundamental issue is the peer grading)" got more than 50 comments and many students expressed their opinions after receiving their peer grades in that thread.

Though one could be generally happy about the course, he/she might not be satisfied with a certain course tool. In many MOOCs' post-course surveys, students are required to separately rate the course tools such as lecture, assignment and peer-assessment. Then the instructors are able to obtain summative feedbacks on various course components. In the course discussion forums, students naturally express their opinions towards these course tools. We show the total number of related posts for each course in Table 2. It is impossible for course instructors to read hundreds or even thousands of potentially related-posts. In this session, we try to extract the most prominent opinions associated with each course tool from these posts.

In Figure 2, we show the number of topic-related posts on each course day. Across the three courses, all topics have a weekly cyclical structure, occurring more frequently on weekdays, especially in the middle of the week, compared to weekends. Talk about assignments is the most frequent topic since TAs post weekly discussion threads for students to discuss assignments.

For each course tool, we extract the positive and negative sentiment words that associate most frequently with the course tool topic keywords. We rank the sentiment words by the Pointwise Mutual Information (PMI) [16] between the word and the topic keyword:

$$PMI_{w,TopicKeyword} = \frac{P(w, TopicKeyword)}{P(w)P(TopicKeyword)}$$

Where $P(w, keyword)$ is the probability of the sentiment word w and a topic keyword appears in the same post; $P(w)$ is the probability of a sentiment word w appears in a post; $P(TopicKeyword)$ is the probability that at least one topic keyword appears in a post. We show the top five sentiment words that are most frequently associated with the course tool topic keywords in Table 3. We can see that some of the words are wrongly identified as sentiment words, such as "lie", "missed" and "soft". From the table we can identify some of the merits and problems of a course tool. These representative sentiment words can complement the rating obtained from the post-course survey.

4.2 User-level Sentiment Analysis: Survival Analysis

In the previous section, we measured sentiment and dropout on a course-level. Our goal in this section is to understand how expression of sentiment relates to attrition over time in the MOOCs on a user-level. We apply survival analysis to test if students' attitudes as expressed in their posts or the ones they are exposed to correlate with dropout from the forum discussion. Recent work has questioned the impact of course forum flaming on students in MOOC courses [29]. We explore sentiment as it relates to an individual's own posts during a week as well as to the other posts that appear on the same thread as that individual's posts during the same period of time. While we cannot be sure that the student read all and only the other posts appearing on the same threads where that student posted during a week, this provides a reasonable proxy for what conversational behavior a student was exposed to within a period of time in the absence of data about what students have viewed. A similar approach was used in a previous analysis of social support in an online medical support community [30].

4.2.1 Survival Analysis Setting

In our survival model, the dependent measure is **Dropout**, which is 1 on a student's last week of active participation unless it is the last course week (i.e. the seventh course week), and 0 on other weeks. In our sentiment analysis, we separate the measure of positivity and negativity rather than operationalizing them together as a single scale. For each week, we measure a student's expressed sentiment to see if the sentiment a student expressed in his/her posts is correlated with drop out. To study if a student would be influenced by the sentiment expressed in their peers' posts, we measure the amount of sentiment a student is exposed to during that week.

In our data, we find across the three courses correlations with R value less than .13 between a measure of positivity and of negativity. Thus, we separate these measures and evaluate them separately in our survival models.

Individual Positivity (Indiv. Positivity): average positivity in the user's posts that week

$$Indiv.\ Positivity = \frac{Total\ positive\ terms}{Total\ number\ of\ words}$$

Individual Negativity (Indiv. Negativity): average negativity in the user's posts that week

$$Indiv.\ Negativity = \frac{Total\ negative\ terms}{Total\ number\ of\ words}$$

Thread Positivity: this variable measures the average positivity a user was exposed to in a week. It was calculated by dividing the total number of positive words in the threads in a week where the user had posted by the total number of words in those threads.
Thread Negativity: this variable measures the average negativity a user was exposed to in a week. It was calculated by dividing the total number of negative words in the threads in a week where the user had posted by the total number of words in those threads.

4.2.2 Modeling
Survival analysis is a statistical modeling technique used to model the effect of one or more indicator variables at a time point on the probability of an event occurring on the next time point. In our case, we are modeling the effect of certain language behaviors (i.e., expression or exposure to expression of sentiment) on probability that a student drops out of the forum participation on the next time point. Survival models are a form of proportional odds logistic regression, and they are known to provide less biased estimates than simpler techniques (e.g., standard least squares linear regression) that do not take into account the potentially truncated nature of time-to-event data (e.g., users who had not yet ceased their participation at the time of the analysis but might at some point subsequently). In a survival model, a prediction about the probability of an event occurring is made at each time point based on the presence of some set of predictors. The estimated weights on the predictors are referred to as hazard ratios. The hazard ratio of a predictor indicates how the relative likelihood of the failure (in our case, student dropout) occurring increases or decreases with an increase or decrease in the associated predictor. A hazard ratio of 1 means the factor has no effect. If the hazard ratio is a fraction, then the factor decreases the probability of the event. For example, if the hazard ratio was a number n of value .4, it would mean that for every standard deviation greater than average the predictor variable is, the event is 60% less likely to occur (i.e., 1 - n). If the hazard ratio is instead greater than 1, that would mean that the factor has a positive effect on the probability of the event. In particular, if the hazard ratio is 1.25, then for every standard deviation greater than average the predictor variable is, the event is 25% more likely to occur (i.e., n - 1).

4.2.3 Quantitative Analysis
Intuitively, we might expect that positive sentiment indicates that students are enjoying or benefitting from a course whereas negative sentiment might indicate that a student is frustrated with a course. The results from our quantitative analysis are not consistent with our intuition. A qualitative analysis of how these features play out across the three courses, which is provided in Section 5, will offer a more nuanced view.

A summary of the results of the survival analysis are presented in Table 4. Typically, we observe lexical accommodation in discussions, including threaded discussion forums [19]. Consistent with this, we find low but significant correlations between individual level sentiment scores and thread level sentiment scores. The correlations are low enough that they are not problematic with respect to including these variables together within the survival models. Including them together allows us to compare the effect of a student's behavior with the effect of exposure to other students' behavior. As we see in Table 4, not only do we see differential effects across courses, we also see differential effects between behavior and exposure.

Indep. Variable	Teaching	Fantasy	Python
Indiv. Positivity	1.03	0.97	1.04*
Indiv. Negativity	0.99	0.84**	1.05**
Thread Positivity	0.95	0.99	1.02
Thread Negativity	1.06*	0.82**	0.98

Table 4: Hazard ratios of sentiment variables in the survival analysis(*: p<0.05, **: p<0.01).

Specifically, in the Python course, we see a significant association between both positive and negative expression and student dropout. In particular, students who express a standard deviation more positive emotion than average are 4% more likely to drop out of the course by the next time point than students who express an average level of positive emotion. Similarly, students who express a standard deviation more negativity than average are 5% more likely to drop out by the next time point than students who express an average amount of negative emotion. Exposure to emotion makes no significant prediction about dropout in this course.

In the Fantasy course, the pattern is different. Negative emotion, whether expressed by an individual or present on the threads that student participated in, is associated with less attrition. In particular, students who either express a standard deviation more negativity or are exposed to a standard deviation more negativity on a time point are nearly 20% less likely to drop out on the next time point than students who express or are exposed to an average amount of negativity. However, positivity has no significant effect.

In the Teaching course, again the pattern is different. There is no effect of expressing negativity or positivity. But students who are exposed to a standard deviation more negativity are 6% more likely to drop out on the next time point than students who are exposed to an average amount of positivity.

We might expect that differences in effect might be related to differences in norms of behavior between courses. For example, positivity and negativity might have more of an effect where they are unusual. However, while one important difference across courses is the average level of positivity and negativity that is present in the discussions, this pattern is not consistent with what we would expect if differences in behavioral norms was the explanation for the differences in effect. The qualitative analysis in the discussion section will again elucidate some potential explanations for differences in behavioral norms.

We also tried to measure individual and thread sentiment based on topic post set retrieved in Section 4.1.1. However, as users typically have too few posts that contain a topic keyword in each course week, the positivity/negativity scores are not available for most of the users. So the survival analysis results on each topic post set might not be meaningful.

5. DISCUSSION: QUALITATIVE ANALYSIS

The results of the survival analysis were not completely consistent either across courses or with an initial naive expectation. In this section, we elucidate those quantitative results with qualitative analysis.

In the Fantasy course, negative comments were ones where people are describing some characters in the fiction. They use some strong negative words which should be very rare in usual conversation, such as "destroy, "devil", "evil", "wicked", "death", "zombie", "horror", etc. One example post is shown below, the negative words are underlined. These messages got high negativity scores because of words taken out of context, which seemed to happen more for negative words than positive ones. The negative word use in this course is actually a sign of engagement because messages with negative words are more likely to be describing science fantasy related literature or even posting their own essay for suggestions.

- Indiv. Negativity = 0.23, the Fantasy course
  "The Death Gate Cycle was such a haunting story!"

In the Python course, forum posts are mostly problem-solving. Both very positive and very negative messages are predictive of more dropout. The most positive messages were thanking for a response. E.g. "Cool!" or "Thank you!". Messages rated as very negative were mainly reporting problems in order to get help or commiserate on an issue already posted. E.g. "It's my error." or "Same problem here." Users who post messages like this are more in the role of forum content consumer. They may only be browsing the forum to look for answers for their particular problems without the intention of contributing content, such as solving the other's problems.

The Teaching course was a social science course about good communication skills. In that course, most forum posts are discussing course-related concepts and techniques. Nevertheless, negativity was low on average, perhaps because the community had a higher politeness norm. A lot of messages contain negative words because of discussion about problem solving. One example post is shown below. It is important to note that in this course, discussion of problems takes on a different significance than in the Python course because changing your interpersonal practices takes time. Whereas in Python you can get your question answered and move on, when it comes to behavior change, discussion of such personal questions signals more intimacy.

- Indiv. Negativity = 0.22, the Teaching course
  A lot of people got crushed by their overloaded work pressure, so why bother yourself talking so complex, complicated, irrelevant and non-rewarding topics while you can spare yourself in those funny little talks and relax a little.

The important take home message here is that the explanation for the pattern goes beyond simple ideas about sentiment and what it represents. We see that expressions of sentiment are being used in different kinds of contexts to serve different functions, and thus this operationalization of attitude is not picking up on the same things across the three courses. With respect to sentiment, we cannot afford to make intuitive assumptions about what it means when variables related to sentiment achieve high predictive value in models that predict choices that people make.

6. LIMITATIONS

We use a relatively simple sentiment detector to explore the uses of sentiment analysis in MOOC context. The sentiment lexicon we utilize is designed for predicting sentiment polarity of product reviews. Creating a more comprehensive lexicon specifically for a MOOC context could improve the system [23]. We associate the opinion to a topic term co-existing in the same context. If we have enough posts with annotated sentiment and topic, many machine learning approaches could capture the mixture of document topics and sentiments simultaneously and substantially improve the accuracy of opinion tracking [17, 28].

7. CONCLUSIONS AND IMPLICATIONS FOR PRACTICE

In this paper, we utilize sentiment analysis to study drop out behavior in three MOOCs. Using a simple collective sentiment analysis, we observe a significant correlation between sentiment expressed in the course forum posts and the number of students who drop the course. Through a more detailed survival analysis, we did not observe consistent influence of expressed sentiment or sentiment a student is exposed to on user dropout. This analysis suggests that sentiment analysis should be used with caution in practice, especially when the texts are very noisy and limited in quantity. However, we see that within a specific course, the relationship between sentiment and dropout makes sense once one examines practices for expressing sentiment within that specific course context. Thus, reports of sentiment could be valuable if they also provide users with examples of how the sentiment words are typically used in that course.

8. ACKNOWLEDGMENTS

We want to thank Ryan Carlson, David Adamson and Tanmay Sinha, who helped provide the data for this project. The research reported here was supported by National Science Foundation grant 0835426.

9. REFERENCES

[1] P. Adamopoulos. What makes a great mooc? an interdisciplinary analysis of student retention in online courses. In *Proceedings of the 34th International Conference on Information Systems, ICIS*, volume 2013, 2013.

[2] C. e. a. Alario-Hoyos. Analysing the impact of built-in and external social tools in a mooc on educational technologies. In *Scaling up learning for sustained impact*, pages 5–18. Springer, 2013.

[3] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser. Better to be frustrated than bored: The incidence, persistence, and impact of learnersŠ cognitive–affective states during interactions with three different computer-based learning environments.

International Journal of Human-Computer Studies, 68(4):223–241, 2010.

[4] H. H. Binali, C. Wu, and V. Potdar. A new significant area: Emotion detection in e-learning using opinion mining techniques. In *Digital Ecosystems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on*, pages 259–264. IEEE, 2009.

[5] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[6] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong. Learning about social learning in moocs: From statistical analysis to generative model. *arXiv preprint arXiv:1312.2159*, 2013.

[7] P. F. Brown, P. V. Desouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479, 1992.

[8] M. De Choudhury, S. Counts, and E. Horvitz. Major life changes and behavioral markers in social media: case of childbirth. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1431–1442. ACM, 2013.

[9] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. In *AAAI Conference on Weblogs and Social Media*, 2013.

[10] M. De Choudhury, A. Monroy-Hernandez, and G. Mark. "narco" emotions: Affect and desensitization in social media during the mexican drug war. In *CHI*. ACM, 2014.

[11] A. El-Halees. Mining opinions in user-generated contents to improve course evaluation. In *Software Engineering and Computer Systems*, pages 107–115. Springer, 2011.

[12] R. F. Kizilcec, C. Piech, and E. Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pages 170–179. ACM, 2013.

[13] P. Liang. *Semi-supervised learning for natural language*. PhD thesis, Massachusetts Institute of Technology, 2005.

[14] B. Liu. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666, 2010.

[15] J. Mackness, S. Mak, and R. Williams. The ideals and reality of participating in a mooc. In *Networked Learing Conference*, pages 266–275. University of Lancaster, 2010.

[16] C. D. Manning and H. Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.

[17] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web*, pages 171–180. ACM, 2007.

[18] J. Moshinskie. How to keep e-learners from e-scaping. *Performance Improvement*, 40(6):30–37, 2001.

[19] A. Nenkova, A. Gravano, and J. Hirschberg. High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 169–172. Association for Computational Linguistics, 2008.

[20] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11:122–129, 2010.

[21] M. J. Paul and M. Dredze. You are what you tweet: Analyzing twitter for public health. In *ICWSM*, 2011.

[22] C. Piech, J. Huang, Z. Chen, C. Do, A. Ng, and D. Koller. Tuned models of peer assessment in moocs. *EDM*, 2013.

[23] G. Qiu, B. Liu, J. Bu, and C. Chen. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27, 2011.

[24] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé III, and L. Getoor. Modeling learner engagement in moocs using probabilistic soft logic. In *Workshop on Data Driven Education, Advances in Neural Information Processing Systems 2013*, 2013.

[25] C. O. Rodriguez. Moocs and the ai-stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-Learning*, 2012.

[26] C. Rosé, R. Carlson, D. Yang, M. Wen, L. Resnick, P. Goldman, and J. Sheerer. Social factors that contribute to attrition in moocs. In *ACM Learning at Scale*, 2014.

[27] D. Song, H. Lin, and Z. Yang. Opinion mining in e-learning system. In *Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP International Conference on*, pages 788–792. IEEE, 2007.

[28] I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on World Wide Web*, pages 111–120. ACM, 2008.

[29] Y. Wang and R. Baker. Mooc learner motivation and course completion rate. *MOOC Research Initiative Conference*, 2013.

[30] Y.-C. Wang, R. Kraut, and J. M. Levine. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 833–842. ACM, 2012.

[31] M. Wen, D. Yang, and C. Rosé. Linguistic reflections of student engagement in massive open online courses. *In International AAAI Conference on Weblogs and Social Media*, 2014.