Exploring the Effect of Confusion in Discussion Forums of Massive Open Online Courses

Diyi Yang, Miaomiao Wen, Iris Howley, Robert Kraut, Carolyn Rosé
Carnegie Mellon University, Pittsburgh, PA 15213
{diyiy, mwen, ihowley}@cs.cmu.edu, robert.kraut@cmu.edu, cprose@cs.cmu.edu

ABSTRACT
Thousands of students enroll in Massive Open Online Courses (MOOCs) to seek opportunities for learning and self-improvement. However, the learning process often involves struggles with confusion, which may have an adverse effect on the course participation experience, leading to dropout along the way. In this paper, we quantify that effect. We describe a classification model using discussion forum behavior and clickstream data to automatically identify posts that express confusion. We then apply survival analysis to quantify the impact of confusion on student dropout. The results demonstrate that the more confusion students express or are exposed to, the lower the probability of their retention. Receiving support and resolution of confusion helps mitigate this effect. We explore the differential effects of confusion expressed in different contexts and related to different aspects of courses. We conclude with implications for design of interventions towards improving the retention of students in MOOCs.

Author Keywords
Massive Open Online Courses (MOOC); Confusion; Survival Analysis.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
With the recent boom in development of educational resources in industry and academia, Massive Open Online Courses (MOOCs) have rapidly moved into a place of prominence in the media. MOOCs enable thousands upon thousands of students to register for courses to learn at their convenience and with no cost. Despite all the potential, MOOCs have so far suffered from extremely high rates of attrition. The reasons behind the dropout are various, but evidence suggests that a potential culprit is learners' confusion and frustration evidenced in their learning process either explicitly or implicitly [6]. Previous explorations of student affect in educational contexts have demonstrated a strong connection between affect, engagement, and learning. However, little systematic research has been conducted to specify how confusion affects student commitment to their continued participation in MOOCs.

Confusion has a complex influence on learning and engagement. Confusion experienced during learning processes is not always associated with negative outcomes. In certain circumstances, a prompt response to confusion (e.g., support from an instructor or help from other sources), or simply the experience of overcoming confusion through a student's own efforts, may lead to a beneficial effect [8]. On the other hand, students who experience confusion may struggle to stay involved in a course and ultimately drop out, especially if students are already vulnerable to dropout for other reasons. Thus, the focus of our work is to better understand the student participation experience as they struggle, express confusion, and ultimately stop participating in a MOOC course.

Behavior traces during periods of confusion may provide clues that offer the opportunity to observe the process of student struggles with confusion [11, 2]. For example, students may become confused while watching lecture videos or working on assignments. They might express their confusion via the discussion forum with detailed descriptions of the problem (e.g. "I'm stuck on this question.." and "I'm very lost on this concept.."), or through traceable interactions with course affordances, for example, re-watching a video or slowing the video speed. Without receiving help from other students or teaching staff, students might remain confused, and persistent confusion might promote negative attitudes toward the course and ultimately dropout. On the other hand, monitoring the behavior trace for such evidence offers the opportunity to offer just-in-time support, and supporting the participation of these struggling students might increase the overall course success.

Our work focuses on exploring different levels of confusion and their corresponding impact on student retention, which affects the health of the MOOC learning community. Based on two MOOC datasets, we first describe a classifier that predicts the degree of confusion associated with student posts in the MOOC discussion forum. The model takes into account student video watching patterns and the linguistic choices displayed in their posts. We then apply this model to measure the quantity of confusion over all posts. With reasonable accuracy we are able to distinguish different types of confusion,

L@S 2015, March 14–18, 2015, Vancouver, BC, Canada.

http://dx.doi.org/10.1145/2724660.2724677

such as confusion that appears in different contexts or confusion in response to different aspects of courses. This allows us to examine the different effects of this array of confusion experience types on students' continued participation.

The contribution of our work is two-fold. Theoretically, by exploring the impact of confusion on student survival, we have identified several important connections between student confusion and their dropout of MOOC courses. Practically, our findings provide guidance for development of potential interventions that might promote retention by providing just-in-time help and support for confused students.

RELATED WORK

Analysis of Attrition in MOOCs

The recent boom of MOOC platforms such as Coursera and edX raise hopes related to the potential of distance learning and lifelong learning to reach the masses, while at the same time sparking a debate about the extremely high attrition. The level of concern related to MOOC dropout rates is reflected in the attention the research community has dedicated to it [29]. For example, Rose et al. [17] investigate the influence of social positioning factors on student commitment to MOOC courses, and Yang et al. [30] demonstrate that students who share similar behavior patterns influence other students' commitment to the courses through their interaction with them. These analyses focus on social connection and social positioning, but do not deeply explore student affect. Work from Ramesh [16] incorporates content analysis to investigate the correlation between sentiment and subjectivity of user posts and student engagement/disengagement. In finer-grained forum content analysis work [25], the authors explore sentiment mining from forum posts to monitor students' trending opinions towards the course, and find a correlation between measured sentiment and student attrition. Additionally, Wen et al. [24] have used computational linguistic models to measure learner motivation and cognitive engagement from forum posts and their association with student survival rates. Here we see an expected pattern of motivation and cognitive engagement consistently associated with commitment, but weak or inconsistent effects associated with simple measures of sentiment. Prior work provides a multi-faceted understanding of student engagement through examination of discussion forum posts. However, little effort appears to have dealt specifically with understanding students' experience of confusion while participating in MOOC courses.

Confusion and Learning

Student affect, of which confusion is one example, has been investigated in many scenarios related to student learning [2, 27]. It has been pointed out that accurate understanding of student affect has the potential not only to contribute towards improved affect detection, but also to contribute insights towards design of interventions that might increase engagement. Students experience confusion when they are confronted with an anomaly, contradiction, or an impasse; and are uncertain about how to proceed. On the positive side, confusion, as one of the most frequently studied forms of student affect [13], has been found to correlate with learning [12], particularly with learning at deeper levels of comprehension. Confusion causes students to stop, reflect, and begin active problem solving to resolve their confusion. Struggling with confusion as a cognitive activity may enable learners to acquire a deeper understanding of complex topics [12, 8].

The positive effect of confusion occurs when students can effectively regulate their confusion or if the learning environment provides sufficient scaffolding to help them do so [8, 12]. On the other hand, prolonged confusion leads to poorer student achievement [11]. [8] also emphasizes the importance of rapidly resolving confusion when it arises. That is, when a learner becomes confused, a quick instructional strategy is to provide explanations and other scaffolds to alleviate the confusion. Furthermore, confusion associated with failure to resolve an impasse [7] might transition into frustration, boredom, and become a crucial point at which the learner disengages from the learning process [10].

When it comes to the MOOC context, the distant nature and the size of MOOCs introduce limitations on opportunities for students to interact with others as effectively as in traditional classroom learning or intelligent tutor situations [21]. Lacking immediate feedback, interactive communication, or timely support increases the likelihood of members leaving such communities [22, 23], especially when members get confused in the learning process. As we look towards the more successful MOOCs of the future, we require new interventions to provide confused students with feedback, interaction and even instructor interventions [16]. Such interventions will depend on an ability to monitor behavior traces to identify the experience of confusion as well as an understanding of how confusion plays out in the context.

Modeling Confusion and its Relationship to Dropout

As a step towards being able to regulate student confusion and ultimately to increase their engagement, first we must be able to identify confusion, as we track and monitor learner experiences [12]. Building on that, we must model the connection between the experience of confusion and dropout.

Considerable prior research has been directed towards automatic identification of confusion in the context of educational software such as intelligent tutoring systems [2]. Researchers have been investigating student affect changes, including confusion, over extended periods of time by developing models using approaches such as classification or knowledge engineering [14]. Some of the most successful affect detectors are built through physiological sensors, identifying vocal patterns [4] or through log files[2], and conversation cues [9]. For instance, Lee et al. [11] explored the relationship between novice programmer confusion and achievement based on students' compilation logs from their programming courses, and concluded that prolonged confusion is associated with poorer student achievement. Similarly, Conati [5], developed a detector based on a combination of questionnaire and log data to predict self-reported student affect. Their model was better at identifying focused and curious students, but less successful at identifying students who were confused. However, approaches relying upon sensors or self-reported affect are limited in application to datasets for which such things are present, and also have limitations in terms of large-scale deployment. We address these limitations in our work in the

MOOCs by relying only on information readily available in that context.

Our analysis is conducted in the following two steps. First, we train and validate a confusion detection classifier to measure the degree of student confusion revealed in their forum post and click behavior. Then we leverage survival analysis to explore how student confusion predicts continued participation by itself and in connection with receiving support or a resolution to confusion. We also use a similar modeling approach to explore the differential effect of confusion experienced in connection with different course aspects.

DATA PREPARATION

In our investigation, we partnered with faculty at a well-known West Coast state university, which provided the data from two Coursera MOOCs. Our dataset for this paper consists of two Coursera courses: one mathematics course, "Algebra" and one economics course "Microeconomics". Algebra had 2,126 active users (active users refers to those who post at least once in a course forum) and 7,994 forum posts; Microeconomics had 2,155 active users and 4,440 forum posts. A flag that indicates whether the issue was resolved or not was also provided for each thread. The duration of each of the two courses was 12 weeks. Besides forum records, each student's interaction (student clicks) with the course materials is also recorded in a clickstream. Algebra has 8,686,230 student clicks, and Microeconomics has 2,709,053 clicks. This clickstream data provides us with the opportunity to investigate the relationship between click patterns and student confusion. Around 3.4% students in Algebra course and 8.3% students in the Microeconomics have posted at least once in the forum.

CONFUSION PREDICTION

As we have argued, confusion is an important factor in the experiences of students in MOOCs. In order to offer the opportunity to identify students who are experiencing confusion, we built machine learning models to automatically identify the level of confusion expressed in students' posts in the course discussion forums. Such models use statistical procedures to map a set of input features to a set of output categories. In our work, we extract the input features both from students' click behaviors and their posting behaviors, including clicking patterns in courses, presence of domain-specific content words, and high-level linguistic features. The output is a numerical value indicating the degree of confusion expressed in each post. We began our work by creating a human annotated corpus for developing our automated measure of student confusion. Second, we represented each post as a set of features as input for our machine learning models. Finally, we constructed a statistical model based on the hand-coded data and evaluated its performance.

Creating the Human-Coded Dataset: MTurk

We used Amazon's Mechanical Turk (MTurk) to construct a reliable, hand-coded dataset to automatically measure student confusion. Amazon Mechanical Turk[1] is a crowdsourcing Internet marketplace, which allows for requesters to post tasks known as HITs (Human Intelligence Tasks) and workers (Turkers) to browse among existing tasks and complete them for a monetary payment set by the requester. Buhrmester et al. [3] demonstrated that MTurk data quality matched or surpassed psychometric standards of traditional studies, for simple tasks. In addition, Snow et al. [20] showed that the combined ratings of five to seven Turkers yielded judgments of textual content comparable to judgments made by experts, such as expressions of emotion, event timing, similarity of words, disambiguation of word meaning, and language-based entailment/implication.

To increase the annotation quality, we required Turkers to have a United States location and a 98% approval rate for their previous work on MTurk. We randomly sampled 522 posts from the Algebra course forum, and 584 posts form the Microeconomics course forum. Non-English posts were filtered out, and personal identifiers were verified to have been replaced by anonymous unique identifiers. Personal details of student experiences were obscured in order to preserve the privacy of the students in the course. For each post, Turkers judged the level of confusion contained in the message on a 1-4 Likert scale ranging from "No Confusion", "Slightly Confused", "Moderately Confused" to "Seriously Confused". We provided them with explicit definitions and examples to use in helping their judgments. Each post was labeled by five different workers. To encourage workers to take the rating task seriously, we asked Turkers to highlight the portion of the post's text that supported their judgments. Turkers received $0.06 for rating each post.

We aggregated the 5 workers' responses for each post by averaging their ratings. Thus, each post has a 1-4 scale numerical value indicating the level of confusion contained in the message. Following are two examples from our final hand-coded datasets, with one example illustrating serious confusion and another indicating slight confusion.

- Student Confusion = 4.0 (Seriously Confused)

  I am completely lost on quiz 3. (I've never been awesome at math of any kind) Quiz 1 &2 were easy breezy, but 3 is throwing me. I feel like the video barely touched on what this is asking for. Or maybe I am just *that* bad at this...? Anyone want to help me understand better?

- Student Confusion = 2.0 (Slightly Confused)

  Hi, I am having problems running coursera on Chrome from my laptop. I do not have this problem using Explorer or Chrome from another PC. Have anyone else has the same issues?

To evaluate the reliability of the annotations we calculated the intra-class correlation coefficient for the confusion annotation. Intra-class correlation is appropriate to assess the consistency of quantitative measurements when all objects are not rated by the same judges. The intra-class correlation coefficient is 0.745 for Algebra, and 0.801 for Microeconomics, which indicates good agreement. To assess the validity of their ratings, we also had the workers code 50 posts, which had been previously coded by three experts. The correlation between MTurkers' average ratings and the experts' average

[1] `https://www.mturk.com/mturk/welcome`

ratings were 0.86 and 0.80 on the two courses, which is indicates moderate agreement. We divided the set of annotated data into two balanced groups based on the confusion scores: "confused" posts and "unconfused" posts, acting as positive and negative instances respectively.

Classifier Construction: Feature Space Design

Students expressed their confusion explicitly or implicitly via different approaches. For example, post writers could use different language strategies when expressing different levels of confusion in their messages. Students could also demonstrate their learning difficulties via their video watching behavior implicitly, outside of discussion forums. In the following, we describe how we capture these different cues to build a classifier, which can identify the degree of confusion that students express in their messages. This collection of features (e.g. linguistic choices, usage of question marks, behavior as reflected in logfile data, etc.) are given as input into a classifier algorithm which determines how to weight the features to predict confusion. The process requires a test set of human-coded data to determine the accuracy of the feature weighting. Once accuracy is achieved, the algorithm can be applied to un-coded data, allowing us to automatically predict the confusion levels in a nearly infinite number of students based solely on the feature variables and their coefficients. Taking advantage of previous literature, we designed our three types of features as described below.

Click Patterns

Clickstream data in MOOCs describes students' complete interaction with each course component, and can be regarded as a shallow but ever present indication of student experience. This trace log data reveals students' learning patterns, however, researchers are just beginning to investigated how click patterns in MOOCs correlate with student engagement in a meaningful way. [19] focuses on investigating students' information processing behaviors while interacting with video lectures. Effective analysis of such click patterns helps to better identify students' confusion when combined with their language choices in discussion forum posts, and also may eventually provide the means for us to analyze confusion in students enrolled in MOOCs even without their participating in discussion boards. For example, imagine a student becomes confused about a concept when watching class videos. Even though he/she did not make an explicit post about the issue, his/her click behaviors that may have involved re-watching and referring to course lectures may suggest potential confusion.

Based on clickstream data, we distinguish instances where students are taking quizzes (quiz), watching lectures (lecture), participating in forums (forum), and viewing other course materials (course). The complete record of click behaviors within a 3 hour window before the student has made a post are collected to analyze patterns that might be associated with confusion. N-grams of behaviors, with a maximum length of 4, are then extracted from these collections of behaviors. A ceiling of 4 was selected in order to avoid sparsity. The top twenty click patterns are then extracted as predictors of confusion. For example, we observed a student having a specific click pattern such as "quiz-quiz-forum" before making a post "*I try to submit the quiz w1: [...] but it says wrong. Could someone tell me what is wrong here?? Thank you*". In this case, accessing the forum after taking the quiz reflects that the student encountered problems during the quiz and turned to the forum to seek help from peers. Similarly, another pattern "quiz-lecture-quiz-lecture" might indicate that this student referred to the lectures for clarification and help after working on a quiz.

Linguistic Features

The Linguistic Inquiry and Word Count (LIWC), designed by Pennebaker et al. [15] calculates the degree to which people use different categories of words across a wide array of texts, including emails, speeches, poems, or transcribed daily speech. LIWC dictionaries were selected based on their relevance to confusion affect. For example, "I, my" reflects that the student is describing something related to herself, whereas negative terms such as "depress, struggle" express a negative affect related to learning. Negation words or phrases might act as a proxy for potential confusion, such as not, shouldn't, or did not. Similarly, insight words such as consider or think reflect students' active orientation towards learning. To summarize, the LIWC features we incorporated into our machine learning models may be categorized into the following: (1) Pronouns : "I, we, you, she/he", impersonal pronouns; (2) Sentiment : affective processes (e.g. "happy, cried, abandon"), positive emotion, negative emotion ("anxiety, anger, sadness"); (3) Spoken Language: negation (e.g., no, not, never), disfluencies (e.g., rr, hm, umm), insight words (e.g., "think, know, consider"), assent ("agree, ok, yes"), adverbs (e.g., "very, much"), certainty (e.g., "always, never"), discrepancy (e.g., "should, would, could").

Question Features

Confusion is often conveyed by asking questions [26]. Thus, we counted the number of question marks in the sentences as a question marker feature. In addition, since not all questions are asked directly and end with a question mark, we took advantage of a heuristic rule to detect question sentences (i.e., whether the sentences start with a modal verb or a question word [23]). For example, sentences beginning with "is, was, had, does, can, may, which, why", etc., might provide cues for potential confusion. Additionally, we calculated whether sentences begin with a confusion related fixed expression such as " I am confused", "I was stuck", "I am struggling with" etc.

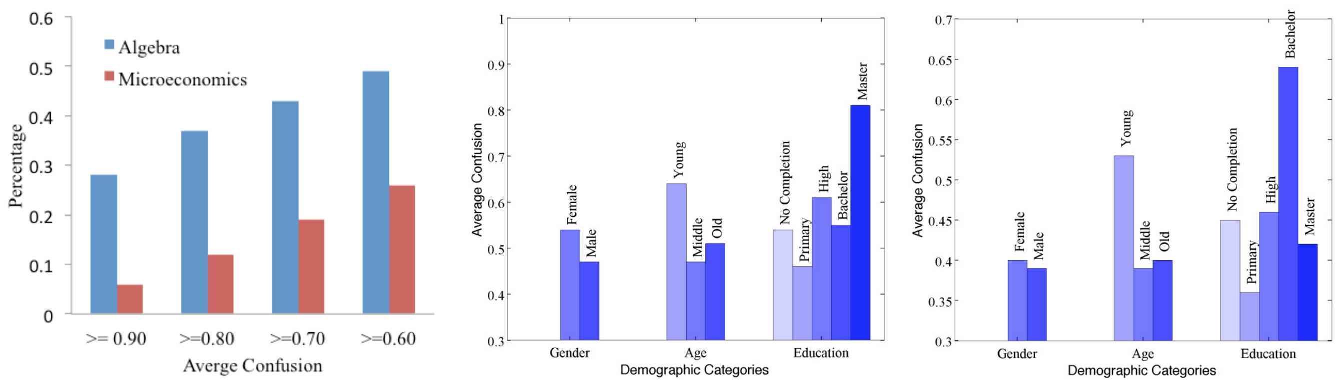
Classification Performance Discussion

Quantitative Discussion

We began by examining the classifiers' performance on identifying the degree of confusion expressed in each discussion forum post. We trained a Logistic Regression model for these text classification experiments and evaluated performance with a 10-fold cross validation. Accuracy and Cohen's Kappa are used to evaluate model performance. The labeled data for both Algebra and Microeconomics courses were divided into two balanced sets. Thus the baseline performance of random guessing is around 50%. We compared several different models: a unigram feature representation as a baseline feature set (**Unig**), classifiers using click patterns **Click**,

Courses	Algebra Course			Microeconomics Course		
Classifier	Accuracy	Kappa	#Features	Accuracy	Kappa	#Features
Unig	0.791	0.581	2967	0.698	0.375	4921
Click	0.587	0.176	20	0.567	0.049	20
Ling	0.647	0.293	15	0.594	0.056	15
Question	0.683	0.370	4	0.652	0.333	4
CLQ	0.723	0.448	39	0.678	0.309	39
CLQ + Unig	0.796	0.582	3006	0.706	0.388	4960
Reduced(CLQ + Unig)	0.803	0.606	632	0.712	0.403	650

Table 1. Performance of Confusion Classifiers on Two Courses



(a) Course Difference (b) Algebra Course (c) Microeconomics Course

Figure 1. Qualitative Confusion Comparison over Courses and Demographics

Courses	Algebra	Microeconomics
Most Important Features (Feature Weight)	question marker count(1.16) 1st pers singular (1.31) question word count(0.52) click pattern (0.38) impersonal pronouns (-0.15) certainty (-0.17) negation (-0.19) adverbs (-0.20)	question marker count(1.30) start with modal words (1.09) 1st pers singular(0.73) question word count(0.17) adverbs (-0.17) affect (-0.18) click pattern(-0.19) negation (-0.20) insight (-0.28)

Table 2. Top Ranked Features for Confusion Detection

linguistic features **Ling**, and question features **Question** individually, a classifier **CLQ** combining Click, Ling, and Question feature sets, and a classifier **CLQ+Unig** taking all the combination of previous feature sets. We also conducted feature selection to reduce the high dimensionality. The reduced model is denoted as **Reduced(CLQ+Unig)**. We present the results in Table 1. **CLQ+Unig** achieves better performance compared to models with individual types of features. In contrast, the performance of the reduced CLQ+Unig model is superior to that of the full CLQ+Unig feature set. Given the Reduced(CLQ+Uing) model achieves adequate levels of accuracy of 80.3% on the Algebra data set and 71.2% on the Microeconomics data, we then applied the model to the task of predicting the confusion degree of all remaining posts in the two courses. Table 2 shows some important features identified by the **CLQ+Unig** model on the two different datasets.

This is consistent with our initial analysis that confusion is closely related to viewing quizzes and exams.

Findings and Observations

Course Differences: Considering that Algebra and Microeconomics can be considered technical and less technical courses respectively, we compared the average confusion students expressed in the two forums in Figure 1(a). We found that students expressed more confusion in Algebra and less confusion in Microeconomics. For example, 28% posts in the Algebra course have an average confusion score larger than 0.90, while that score is 6% in Microeconomics. Besides, 54.6% posts in the Alegebra course have expressed confusion (having a confusion score larger than 0.5) and that of Microeconomics is 35.1. A possible explanation might be that technical courses focus more on problem solving, where confusion is natural to discuss as students work through the problems. Nontechnical courses require less problem solving, and thus confusion might be more indirectly expressed via discussion rather than through direct questions.

Demographic Differences: We also examined how confusion varies across different demographic groups, in terms of *Gender*, *Age* and *Highest Education Level*. The average confusion degree comparisons over different demographics are presented in Figure 1(b) and Figure 1(c). We found that females tend to express more confusion than males; young people have the highest expressed confusion; and middle-aged people are the least likely to express confusion. In addition, students who achieved higher education degrees shared a relatively low level of confusion compared to students whose highest education were primary school or no completion.

Courses	Algebra				Microeconomics			
	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max
TotalPost	8.41	17.18	1	232	6.01	9.56	1	172
Starter	0.26	0.57	0	5	0.16	0.46	0	5
ExprConfusion	0.29	0.34	0	1	0.23	0.28	0	1
UserExpoConfusion	0.10	0.20	0	0.69	0.05	0.14	0	0.69
OtherExpoConfusion	0.21	0.23	0	0.69	0.15	0.17	0	0.69
Resolved	0.36	0.46	0	1	0.25	0.42	0	1
Reply	0.40	0.47	0	1	0.26	0.42	0	1

Table 3. Descriptive Statistics for the Variables in the Survival Analysis

ASSESSING THE INFLUENCE OF CONFUSION

Our hypothesis in this work is that in the MOOC context, different kinds of student confusion may lead to different effects on students' survival in courses. For this purpose, we first investigate how expressing confusion and being exposed to others' confusion influences students' continued participation and how such effects might at least partly be mitigated. Then we explore what students are confused about and examine the influences of confusion expressed within different course contexts. We use survival analysis to explore these quantitatively, controlling for other forum behaviors. For example, the number of posts students contributes within a time period is a strong indicator of a priori level of commitment and thus makes an effective control variable. The constructed models are used to quantify the impact of confusion.

Survival Analysis Design

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. In our case, we investigate how measures of student confusion at a time point is associated with propensity of the student to drop out of the course after that time point. More specifically, we want to understand whether our automatic measure of confusion predicts the length of student participation in the course. Survival analysis is known to provide less biased estimates than simpler standard regression techniques that do not take into account the potentially truncated nature of time-to-event data. The Hazard Ratio (HZ) measures the effect of an explanatory variable on the probability of participant drop out. We used the statistical software package Stata [1] to conduct our experiments. Specifically, we used parametric regression survival analysis and assumed a Weibull distribution of survival times. We included all the active students who contributed at least one post to the course forums in the two courses. The time interval is defined as student participation weeks. We considered the timestamp of the first post by each student as the starting point for that student's active participation in the course discussion forums, and the date of the last post as the end of participation unless it is the last course week.

Dependent Variables

- **Dropout**: we consider a student dropping out from participation in the course community if that student has no activities in the course forum (even if they may engage in other forms of engagement with the materials). This dependent variable is a binary indicator, with 1 on a student's last week of active participation unless it is the last course week, and 0 otherwise.

Control Variables

- **TotalPost**: This is the number of posts a student contributes to the forums in one week. It could be regarded as a basic effort measurement of student engagement [24].
- **Starter**: Students who start a discussion thread tend to have a higher probability of being confused compared to those who participate in discussions initiated by others. This variable is defined as the number of threads a student has initiated in a week.

Independent Variables

- **Expressed Confusion** (ExprConfusion):
  This measures the average confusion per post a student has expressed in a week. It was calculated by averaging confusion scores of an individual's posts in that week.
- **User Exposed Confusion** (UserExpoConfusion):
  This measures the average confusion per post a student was exposed to by averaging confusion scores of posts in the threads that student initiated during the time period.
- **Others Exposed Confusion** (OthersExpoConfusion):
  This measures the average confusion a student was exposed to by averaging the measured confusion of posts in all the threads he/she participated in those he/she initiated.
- **Confusion Resolved** (Resolved):
  This variable indicates how many threads are initiated by a student and are later resolved. Students sometimes express confusion through initiating threads with questions. Others providing satisfactory help to such threads might relieve the confusion of those students. Whether a thread was resolved or not is provided in the datasets.
- **Reply** (Reply):
  This variable indicates how many threads a student initiated that have received a response from others. Student communication in the discussion forums is a vital component in MOOCs where personalized interaction is limited.

Table 3 reports the descriptive statistics for the variables entered into the survival regression models.

Commitment Analysis

In addition to validating how students' expressed confusion and confusion exposure affect their participation in the courses, we also examine how their continued participation is affected by the interaction of being confused, experiencing confusion resolution, and receiving instructor support.

Confusion on Commitment

Results of the two survival models from the two courses are shown in Table 4. We investigated the influences of student

	Algebra Model 1		Algebra Model 2		Microeco Model 1		Microeco Model 2	
Variables	HZ	(Std. Err)S.E	HZ	S.E	HZ	S.E	HZ	S.E
TotalPost	0.60^{***}	0.02	0.47^{***}	0.02	0.71^{***}	0.03	0.60^{***}	0.03
Starter	1.29^{***}	0.02	1.38^{***}	0.05	1.19^{***}	0.02	1.45^{***}	0.11
ExprConfusion			1.17^{***}	0.03			1.18^{***}	0.05
UserExpoConfusion			1.06	0.04			1.05	0.04
OthersExpoConfusion			1.85^{***}	0.06			1.51^{***}	0.06

Table 4. Results of Confusion on Students' Survival (*: $p < 0.05$, **: $p < 0.01$, *** $p < 0.001$)

	Algebra						Microeco					
	Model 0		Model 1		Model 2		Model 0		Model 1		Model 2	
Variables	HZ	S.E	HZ	S.E	HZ	S.E	HZ	S.E	HZ	S.E	HZ	S.E
TotalPost	0.58^{***}	0.02	0.59^{***}	0.02	0.60^{***}	0.02	0.66^{***}	0.03	0.67^{***}	0.03	0.68^{***}	0.03
Starter	1.11^{***}	0.02	1.22^{***}	0.03	1.30^{***}	0.04	1.08^{***}	0.03	1.18^{**}	0.04	1.21^{***}	0.04
ExprConfusion	1.64^{***}	0.04	1.69^{***}	0.05	1.59^{***}	0.04	1.54^{***}	0.04	1.54^{***}	0.04	1.54^{***}	0.04
Resolved			0.83^{***}	0.03					0.87^{**}	0.04		
ExprConfusion $\times$Resolved			0.87^{*}	0.04					0.93^{**}	0.03		
Reply					0.79^{***}	0.03					0.83^{***}	0.04
ExprConfusion $\times$Reply					0.91^{***}	0.02					0.93^{**}	0.02

Table 5. Results of Survival Analysis for Interaction Effects(*: $p < 0.05$, **: $p < 0.01$, *** $p < 0.001$)

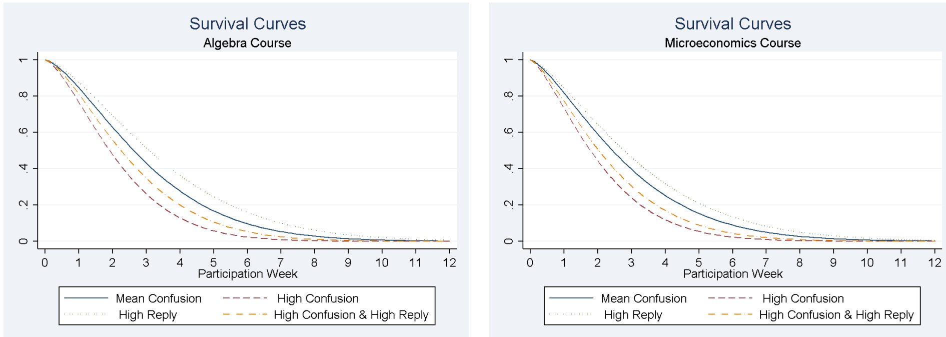
expressed confusion and confusion exposure when controlling for other important distinguishing characteristics of students such as the total number of posts they contributed at the time point. Model 1 reports the effects of the control variables of TotalPost and Starter (i.e., number of threads started by the student per week) on student dropout. The hazard ratio (HZ) for TotalPost is 0.60, indicating that students who contribute a standard deviation more posts than average are 40% more likely to survive compared to students who have lower post counts. HZ 1.29 for Starter indicates that survival rates are $29\%((100\% \times 1.29) - 100\%)$ more likely to dropout for those who have started one standard deviation more thread starters than average. This finding is consistent in Microeconomics. Model 2 shows when controlling for control variables, student expressed confusion and confusion exposure both influence the survival rate negatively. Confusion exposure on threads started by other students has the strongest influence (HZ=1.85) on dropout compared to students expressed confusion (HZ=1.17). The reason behind it might be, expressing confusion reflects students' difficulty in the study process but also indicates their desire to learn and overcome an impasse. Exposure to others' confusion might just communicate to the student that the course materials are low quality, especially when seeing that the lecture/quiz is confusing to other students. For example, a student in the Algebra course is confused about a quiz and the following discussion respond to it by complaining about the same problem: "A: *I dont understand how to write square root for my answer. i saw it put it's been rejected.* B: *I don't understand how to write it either.... Help..* C: *-16-7sqrt(14). Is this correct? Because it's not working.* D: *I am still not getting it to go. Through hand i know that I'm correct*". In the Microeconomics course, cases such as poor presentation in course materials are common, for instance, "A: *Hello! Just trying to watch the lecture now but it always stops in the middle for some reason. Does anyone have the same problem as mine?* B:*my biggest issue is also that the video freezes when I am trying to replay a section to take notes* C: *I think the quality of videos is very bad. This may lead to lost the interest in the course*".

Resolving Threads and Receiving Replies

The research literature shows that confusion might negatively impact learning, and resolvable confusion could enrich the learning experience and might encourage deeper engagement [21]. This motivates us to explore the influence of the interaction between a variable indicating being confused and a variable indicating problems getting resolved. Similarly we explore the interaction between a variable indicating being confused and a variable indicating getting others' replies. We explore both of these interaction effects on student commitment to courses, which results in two variables: ExprConfusion $\times$ Resolved and ExprConfusion $\times$ Reply. The first variable measures the interaction effect of expressed confusion and confusion resolution and the latter measures the interaction effect of expressed confusion and receiving replies.

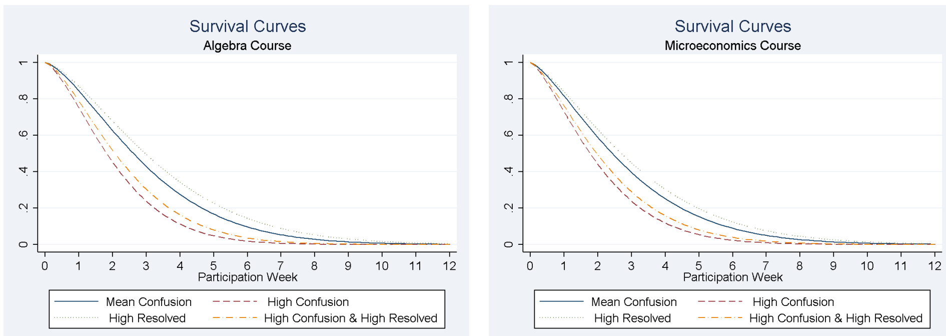
The results are presented in Table 5. Taking the Algebra course as an example, the interaction between the problem being resolved and being confused has a Hazard Ratio (HZ) of 1.22, one third of that of simply being confused(1.64) [2]. It indicates that students whose ExprConfusion X Resolved are one standard deviation higher than average are 22% more likely to dropout. Students whose ExprConfusion X Reply are one standard deviation higher have a Hazard Ratio of 1.14, which means that they are one quarter as likely to drop out as the full set of students who express confusion at a level of 1 standard deviation higher than average (1.64). Specifically, students who are measured at 1 standard deviation higher expressed confusion are 14% more likely to dropout. Thus we conclude that, even though being confused

[2] $1.22 = 1.69 * 0.83 * 0.87$



(a) Algebra Course (b) Microeconomics Course

Figure 2. Survival Curves for Students Exposed to Different Levels of Confusion Being Replied



(a) Algebra Course (b) Microeconomics Course

Figure 3. Survival Curves for Students Exposed to Different Levels of Confusion Being Resolved

in courses has a negative influence on one's commitment, providing replies to such confusion and resolving the underlying issues at least partly mitigates that negative effect. Especially in a MOOC context, where instructors provide very limited guidance (Only 13% threads have instructor intervention in the Algebra course and 18% in Microeconomics), how to encourage students to offer help to others and even help unresolved threads get answered is a direction for potential positive impact. Figure 3 and Figure 2 illustrate these results graphically. *High* means that a variable is 1 standard deviation higher than average, and *Low* describes the case of 1 standard deviation lower than average.

Confusion Content Identification

The above correlational confusion analysis demonstrates the statistical association between confusion and student dropout. To take the next step and automatically recommend help, learning partners, and even experts in that specific field to confused students [28], we not only need to know who is confused, but also what students are confused about. For this purpose, first we designed a simple and effective method for identifying what students are talking about by utilizing their recent click behavior. That is, the content of current post is determined by the type of behavior indicated by the previous click before making that post. The intuition is that a student is more likely to express confusion about recently browsed course material or a recent course experience. The targeted types of contents in the posts and clicks can be divided into three categories: (1) *Course*: content about course tools, course settings, course resources, etc; (2) *Lecture*: content about lectures/videos or discussion on problems/content in lectures; (3) *Quiz*: content about quiz questions, such as quiz answers, submission, and deadlines. Post content that does not belong to any of the above types is categorized into *Other*. To validate the effectiveness of recent click type in determining post content, we randomly sampled 300 posts from each course, asked experts to judge what students are talking about, and checked the correlation between human labeling and our method. Results demonstrated that the automated method that refers to recent click behavior can correctly identify 60% and 64% on Algebra and Microeconomics courses respectively. Phi-Coefficient between human labeling and automatically identification is 0.405 with Kappa 0.300 on Algebra; on Microeconomics course, Phi-Coefficient is 0.632, with Kappa 0.308. We conclude that utilizing the previous click behavior before a post was made captures what students are talking about well enough to offer some visibility into the differential effects of confusion related to different course aspect, so we proceed with the analysis.

Once the context of each post has been identified, we then measure the confusion towards Course, Quiz, Lecture and Other by computing the average confusion scores in a user's Course/Quiz/Lecture/Other posts; denoted as *Course Confusion*, *Quiz Confusion*, *Lecture Confusion*, *Other Confusion* separately. We show the average confusion related to the four categories in Figure 4. We conducted the survival analysis again to check how confusion within different course contexts influences dropout, as results shown in Table 6. It can be concluded that, even though confusion distributions are similar across the two courses (quiz is associated with highest confu-

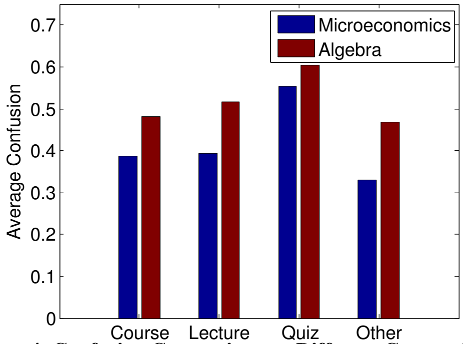


Figure 4. Confusion Comparison on Different Content Types

	Algebra		Microeco	
Variables	HZ	S.E	HZ	S.E
TotalPost	0.51***	0.02	0.63***	0.03
Starter	1.06*	0.02	1.08**	0.03
Course Confusion	1.39***	0.03	1.40***	0.04
Quiz Confusion	1.51***	0.03	1.00	0.03
Lecture Confusion	1.15***	0.02	1.26***	0.03
Other Confusion	1.01	0.03	1.03	0.03

Table 6. Confusion towards Different Course Aspects on Survival

sion), confusion towards different content types do not have similar influence on dropout. Instead, quiz confusion has the biggest impact on dropout in the Algebra course while confusion towards the course more generally leads to the highest dropout in the Microeconomics course. One explanation is that this pattern is a side effect of the fact that mathematical courses emphasize problem solving whereas less technical courses require much more discussion and opinion exchange. Thus we conclude the influence of different types of confusion on dropout is determined by the specifics of course settings, and thus we cannot make strong claims generally about specific types of confusion being more or less damaging than others across courses.

DISCUSSION AND CONCLUSION

Increasing student engagement and facilitating learning is of great importance to future MOOC deployment. In this work, we investigated the effect of student confusion on survival in MOOC courses. We first built a confusion detection classifier to measure the confusion contained in students' forum posts in the course's discussion forums and as indicated by click behavior. Then we examined the effect of different kinds of confusion on students' continued participation in the MOOC courses. Our results demonstrate that: (1) the more students express their confusion and are exposed to confusion in the MOOC forums, the less likely students are to remain active in the learning community; (2) helping resolving or providing responses to student confusion reduces their dropout in the courses; (3) the extent to which different types of confusion affect dropout is determined by specific courses. In our data, quiz confusion leads to the highest dropout in a technical course while confusion towards the course in general contributes more to dropout in a less technical course.

More specifically, compared to traditional classroom learning and many intelligent tutoring systems, students in MOOC courses are less likely to receive immediate feedback when confused. MOOCs have a large number of students in each course and do not often enable face-to-face interaction with instructors or other well-performing students. Such interaction could play an important role in maintaining student engagement because of its directive, facilitative and motivational functions [18]. However, once students get confused, they must choose to either remain confused or attempt to continue course learning, or make their confusion explicit and wait for resolution. Remaining in a confused state without support from instructors and others might easily transition into dropout, while expressing confusion in the forum may not receive a timely response, which might eventually cause disengagement. The effect of being confused and then being given support (e.g. resolving their problems or providing students interaction) has a demonstrated effect to partly mitigate the effect of confusion on dropout. Therefore, in the context of MOOCs, how to encourage students offer help to confused students and even help confusion get resolved, is beneficial for students, instructors and creators of MOOC courses.

Implications and Limitations

Our findings provide guidance for future MOOC deployment. This study sheds light on how to identify student confusion during the learning process. Tracking and monitoring student confusion helps instructors give appropriate feedback to students. Our classifier is able to give a reasonably accurate estimation of students' confusion expressed in the discussion forums. The LIWC features, question features, and click pattern features may be able to be applied to confusion detection in other contexts since they are not tailored to any specific domain. Secondly, our results show that getting students' confusion resolved and providing help reduces their dropout. More interventions that incorporate resolving problems and providing support for students could be developed and deployed, especially tailored to the distant nature and the sheer size of MOOCs.

This study is also subject to the limitation that, all measures of student expression, including expressed confusion, exposed confusion, as well as interaction effect of confusion with reply or resolution, are all judged by Turkers who are not participants and possibly have no experience in MOOC course forums. A natural follow-up is to conduct surveys to collect self-reported confusion when students make posts to see whether such data offer a confirming view.

ACKNOWLEDGEMENT

This research was funded in part from a grant from Google.

REFERENCES

1. 2011, S. C. Stata statistical software release. *7.0: Programming. Stata Corporation* (2011), 418–443.
2. Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., Aleven, V., Kusbit, G. W., Ocumpaugh, J., and Rossi, L. Towards sensor-free affect detection in cognitive tutor algebra. *International Educational Data Mining Society* (2012).
3. Buhrmester, M., Kwang, T., and Gosling, S. D. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science 6*, 1 (2011), 3–5.

4. Calvo, R. A., and D'Mello, S. Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on 1*, 1 (2010), 18–37.
5. Conati, C., and Maclaren, H. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction 19*, 3 (2009), 267–303.
6. Conole, G. Moocs as disruptive technologies: strategies for enhancing the learner experience and quality of moocs. *Revista de Educación a Distancia 39* (2013), 1–17.
7. DMello, S., and Graesser, A. Dynamics of affective states during complex learning. *Learning and Instruction 22*, 2 (2012), 145–157.
8. DMello, S., Lehman, B., Pekrun, R., and Graesser, A. Confusion can be beneficial for learning. *Learning and Instruction 29* (2014), 153–170.
9. DMello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., and Graesser, A. Automatic detection of learners affect from conversational cues. *User modeling and user-adapted interaction 18*, 1-2 (2008), 45–80.
10. Larson, R. W., and Richards, M. H. Boredom in the middle school years: Blaming schools versus blaming students. *American Journal of Education* (1991), 418–443.
11. Lee, D. M. C., Rodrigo, M. M. T., d Baker, R. S., Sugay, J. O., and Coronel, A. Exploring the relationship between novice programmer confusion and achievement. In *Affective Computing and Intelligent Interaction*. Springer, 2011, 175–184.
12. Lehman, B., DMello, S., and Graesser, A. Interventions to regulate confusion during learning. In *Intelligent Tutoring Systems*, Springer (2012), 576–578.
13. Lehman, B., Matthews, M., DMello, S., and Person, N. What are you feeling? investigating student affective states during expert human tutoring sessions. In *Intelligent Tutoring Systems*, Springer (2008), 50–59.
14. Pekrun, R., Goetz, T., Titz, W., and Perry, R. P. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist 37*, 2 (2002), 91–105.
15. Pennebaker, J. W., Francis, M. E., and Booth, R. J. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates 71* (2001), 2001.
16. Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., and Getoor, L. Modeling learner engagement in moocs using probabilistic soft logic. In *NIPS Workshop on Data Driven Education* (2013).
17. Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., and Sherer, J. Social factors that contribute to attrition in moocs. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM (2014), 197–198.
18. Shute, V. J. Focus on formative feedback. *Review of educational research 78*, 1 (2008), 153–189.
19. Sinha, T., Jermann, P., Li, N., and Dillenbourg, P. Your click decides your fate: Leveraging clickstream patterns in MOOC videos to infer students' information processing and attrition behavior. *CoRR abs/1407.7131* (2014).
20. Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *EMNLP*, Association for Computational Linguistics (2008), 254–263.
21. VanLehn, K., Siler, S., Murray, C., Yamauchi, T., and Baggett, W. B. Why do only some events cause learning during human tutoring? *Cognition and Instruction 21*, 3 (2003), 209–249.
22. Vlahovic, T. A., Wang, Y.-C., Kraut, R. E., and Levine, J. M. Support matching and satisfaction in an online breast cancer support community. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 1625–1634.
23. Wang, Y.-C., Kraut, R., and Levine, J. M. 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*, CSCW '12, ACM (New York, NY, USA, 2012), 833–842.
24. Wen, M., Yang, D., and Rosé, C. P. Linguistic reflections of student engagement in massive open online courses. In *Proceedings of the International Conference on Weblogs and Social Media* (2014).
25. Wen, M., Yang, D., and Rosé, C. P. Sentiment analysis in mooc discussion forums: What does it tell us? *Proceedings of Educational Data Mining* (2014).
26. Wilson, N. Learning from confusion: Questions and change in reading logs. *English Journal* (1989), 62–69.
27. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., and Picard, R. Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology 4*, 3 (2009), 129–164.
28. Yang, D., Adamson, D., and Rose, C. Question recommendation with constraints for massive open online courses. In *Recsys*, ACM (2014).
29. Yang, D., Sinha, T., Adamson, D., and Rosé, C. P. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-Driven Education Workshop* (2013).
30. Yang, D., Wen, M., and Rose, C. Peer influence on attrition in massive open online courses. *Proceedings of Educational Data Mining* (2014).