Diverse Learners, Diverse Motivations: Exploring the Sentiment of Learning Objectives

Nigel Bosch
University of Illinois at Urbana-Champaign
pnb@illinois.edu

R. Wes Crues
University of Illinois at Urbana-Champaign
crues2@illinois.edu

Najmuddin Shaik
University of Illinois at Urbana-Champaign
shaik@illinois.edu

ABSTRACT
Massive Open Online Courses (MOOCs) have become increasingly popular in recent years, enabling millions of students worldwide to pursue their educational objectives in new ways. However, little is known about the nature of the reasons why students enroll in courses, and how those reasons differ across demographic groups. In this paper we explore the connection between student engagement and the sentiment of their self-reported reasons for enrolling in MOOCs. We found that there were significant differences in sentiment between demographic groups, and that sentiment of enrollment reasons had small—but consistent—power to predict future course engagement level (Spearman's $\rho = .102$). Finally, we discuss the implications of these findings for future student modeling research in MOOC contexts, particularly for students with different backgrounds.

Categories and Subject Descriptors
H.4 [**Computers and Education**]: Computer Uses in Education

Keywords
Sentiment, MOOC enrollment, demographic differences

1. INTRODUCTION
Recent years have seen an explosion in the availability of massive open online courses (MOOCs), and millions of new students have enrolled in them [8]. MOOCs offer worldwide access to high-quality courses offered by prestigious institutions, and as such they attract a diverse group of learners from around the globe. These students are often even more diverse than typical university student bodies, in part because of the unique affordances MOOCs provide: low cost and flexible schedules.

However, reasons students enroll are complex and especially difficult to define in MOOC-style learning contexts. Furthermore, enrollment reasons may systematically differ between different courses and student demographics [5]. In this poster, **we focus specifically on the sentiment aspect of students enrollment reasons**. Particularly, we explore whether sentiment of reasons for enrolling differs across course and demographic dimensions, and question how sentiment of enrollment reasons relates to students' levels of engagement in a course. Sentiment in this context refers to aspects such as positivity, fear, certainty, and others that can be inferred from text. One might also expect, for example, that the amount of positivity inherent in students' stated reasons for enrolling relates to how long they persevere in a course.

Previous work has also found that demographics relate to MOOC outcomes (e.g., gender relates to persistence in MOOCs [4]). Given that there were 78 million students in almost 10,000 MOOCs [8] in the year 2017, it is tremendously important to understand all potential indicators of student engagement and student needs. This includes the emotional cues contained in the goals they may express when enrolling in a course.

Wladis et al. analyzed approximately 27,800 students who enrolled in online STEM (science, technology, engineering, and math) classes [11]. They found key differences in enrollment rates, concluding that non-traditional students were significantly more likely to enroll in online courses than their peers. Furthermore, they found that female students performed less well in online environments than in face-to-face learning, but older students enjoyed greater success online [12].

Robinson et al. [7] asked students enrolling in a MOOC to "provide one or two specific examples of how you think what you will learn in this class will apply to your life." They extracted frequent word unigrams and bigrams, and trained a logistic regression classifier to predict whether students would drop out of the MOOC or not. Their model was statistically better than chance as measured by area under the receiver operating characteristic curve (AUC)—specifically, AUC = .564 (versus .500 chance level). Additionally, they found that including student demographics as predictors improved model accuracy to AUC = .598. This study demonstrated that linguistic aspects of the reasons students state for enrolling in a class can modestly predict course outcomes, but that student demographics are also worth considering.

In this study we apply non-parametric statistics and machine learning methods to explore the relationships between the sentiment of students' reasons for enrolling in MOOCs, student demographics, and course engagement. We cross-validate analysis across courses and demographic variables to answer three research questions: 1) How does sentiment of enrollment reasons differ across student demographics? 2) Does sentiment of enrollment reasons predict the level of course engagement? and 3) Is sentiment of enrollment reasons equally predictive of engagement across different courses and demographics?

2. METHOD

We analyzed data from five different MOOCs offered on the Coursera platform[1]. These included *Creative, Serious, and Playful Science of Android Apps*, *Introductory Organic Chemistry*, *Subsistence Marketplaces*, *Introduction to Sustainability*, and *E-Learning Ecologies*. We queried students for demographic information, including age range and gender[2], and asked them to provide their reasons for enrolling in the course by writing an answer to the open-ended prompt "Why are you taking this course? What do you hope to get out of it?" Of 37,178 students who enrolled and responded to at least one question, 9,327 responded in English to all questions.

2.1 Sentiment of Enrollment Reasons

We extracted sentiment from students' written reasons for enrolling in MOOCs with the SEANCE (SEntiment ANalysis and Cognition Engine) tool [3]. SEANCE provides indices of sentiment derived from a collection of eight different databases of words, where each word is associated with a sentiment. SEANCE also provides 20 *component scores*, which are derived from principal components analysis and have interpretable labels based on the indices the components are derived from. These component scores compose the sentiment-based feature space in which we represent students' reasons for enrolling. Given the large ratio of students to features (approximately 450:1), we did not perform feature selection.

Sentiment components provided by SEANCE were not normally distributed. Thus, for comparisons involving sentiment we calculated non-parametric statistics. To compare sentiment across genders, we coded gender as a number and computed Spearman's rho (ρ) correlations between gender and sentiment components. This analysis permits testing for significant differences between genders as well as providing an estimate of the effect size (ρ ranging from -1 to 1). Age groups are categorical, but strictly ordered, so ρ is an appropriate measure for the relationships between age groups and sentiment components as well. Geographical areas are not strictly orderable in a meaningful way, so we could not measure ρ across all geographical areas together.

2.2 Prediction of Engagement from Sentiment

In this study we adopt a multi-level engagement definition to distinguish students who are only active during a few weeks of the course (≤ 2 weeks), versus those who engage with the course for some time but not the entire set of content ($3-5$ weeks), and those who complete essentially all of the course ($6-8$ weeks).

We predicted engagement from sentiment components by training a random forest [2] machine learning model using *scikit-learn* [6]. Random forests work by training a large number of small tree models (i.e., a forest) on random subsamples of data. Random forest models make no assumptions about the distribution of the data, as a Gaussian model does, for instance. This is a key consideration given the non-normal distributions of sentiment components. Our definition of engagement is also multi-level, and is thus a multiclass problem for which random forests are suited.

Predictive student models are frequently evaluated with accuracy metrics suited for binary classification problems (e.g., Cohen's κ, F_1). However, in this study the prediction target (engagement) has three strictly-ordered levels. Therefore, we evaluate model accuracy with Spearman's ρ.

We utilized different cross-validation approaches to answer the research questions in this paper. In each approach, we split data into training and testing data, trained a random forest model (optimized on training data only), and evaluated the model by its ability to predict the unseen the testing data. We repeated this process iteratively until every student (data point) had been in the testing data exactly once.

[1]https://www.coursera.org

[2]Gender responses included female, male, and other, but after filtering the dataset (as described in Section 2) the only responses were female and male.

3. RESULTS

In this section we present results for our three research questions, with explanatory methods for the first research question and predictive models for the second and third questions.

RQ1: How does sentiment of enrollment reasons differ across student demographics? The number of students analyzed was large (9,327). Thus, many correlations between sentiment components and demographic variables were highly statistically significant (25 of 60 correlations with $p < .001$) even after Benjamini-Hochberg corrections for multiple tests [1]. Therefore, we report only the largest five correlations for the sake of conciseness (Table 1).

In general, females expressed more sentiment in their stated reasons for enrolling. In fact, mean $\rho = .047$ across all 20 sentiment components. The largest difference between genders was in the SEANCE economy component, which consists of words from manually-curated lists of nouns and adjectives related to economical concerns [9]. Females were coded as 1, so the positive correlation ($\rho = .106$) indicates that females expressed more economy-focused words than males.

Both female students and older students expressed more fear and disgust in their reasons for enrolling ($\rho = .104$ and .101 respectively). Older students also appeared to express more sentiment than younger students, based on the largest five correlations in Table 1. However, mean correlation across all 20 sentiment components for age groups was just $\rho = .007$, indicating that the larger sentiment differences in Table 1 were offset by many smaller negative correlations (12 of 20 correlations were negative).

RQ2: Does sentiment of enrollment reasons predict the level of course engagement? We trained predictive models with four-fold cross-validation to answer this research question. Predictions were significantly better than chance ($\rho = .102$, $p < .001$), confirming the hypothesis of the research question. Additionally, accuracy was consistent across folds, ranging from $\rho = .093$ to $\rho = .116$. This serves as a baseline for research question 3, which explores prediction variance across demographics and courses to quantify generalization.

Table 1: Differences between enrollment reason sentiment components for students with different demographics.

Sentiment component	Spearman's ρ
Gender (female = 1)	
Economy	.106
Fear and disgust	.104
Joy	.085
Politeness	.082
Virtue adverbs	.081
Age group	
Fear and disgust	.101
Respect	.066
Certainty	-.057
Politeness	.056
Objects	.052

Overall accuracy was modest. It is, however, notable that the prediction was better than chance, given the difficulty of the problem—predicting student engagement before the course even begins. In comparison, Robinson et al. [7] trained models to predict course dropout from extensive text features. They achieved a similar degree of accuracy (AUC = .564 versus .500 chance level), despite using features capturing all types of words and word pairs—not just sentiment words.

RQ3: Is sentiment of enrollment reasons equally predictive of engagement across different courses and demographics? We re-trained the classification model in research question 2 to measure generalization by cross-validating across courses and demographics instead of four-fold cross-validation. Table 2 details the results.

Course-level cross-validation resulted in notably lower accuracy than four-fold cross-validation (overall ρ = .066 versus .102), indicating that sentiment of students' enrollment reasons was less predictive across courses. Accuracy, when testing on the *Android Apps* course, was particularly notable, in that it was not significantly above chance despite having 3,050 students. Conversely, engagement prediction did generalize well from other courses to the *Subsistence Marketplaces* course (ρ = .119).

Models did not generalize well across genders compared to the four-fold model that ignored gender (overall ρ = .073 versus .102). However, female and male results were similar (ρ = .085 and .067 respectively).

Conversely, predictive models generalized well across age groups (overall ρ = .103). Accuracy was consistent as well, ranging from ρ = .078 to .133. Because there was little fluctuation in ρ across age groups, it follows that age group and sentiment were unrelated, at least with respect to engagement (though there were differences in sentiment overall; see Table 1).

There was a large degree of variation in prediction accuracy across different geographical regions, ranging from ρ = -.019 to .368. However, several of these regions were represented by only a few students (as low as 12), so results should be approached with an appropriate degree of caution. Overall accuracy was notably lower than the four-fold model (ρ = .070 versus .102), indicating that sentiment of enrollment reasons was unequally predictive across regions.

Table 2: Classification accuracy (Spearman's ρ) when predicting course engagement generalizing across courses and demographics.

Cross-validation approach	ρ	p-value	N
Leave one course out			
Android Apps	-.009	.637	3,050
E-Learning Ecologies	.091	.009	830
Organic Chemistry	.021	.565	782
Subsistence Marketplaces	.119	.001	728
Sustainability	.052	.001	3,937
Overall result	*.066*	*.000*	*9,327*
Leave one gender out			
Female	.085	.000	4,061
Male	.067	.000	5,266
Overall result	*.073*	*.000*	*9,327*
Leave one age group out			
< 18	.133	.175	105
18-24	.105	.000	1,484
25-29	.078	.001	1,931
30-39	.082	.000	2,471
40-49	.112	.000	1,457
50-59	.108	.000	1,096
> 59	.119	.001	783
Overall result	*.103*	*.000*	*9,327*
Leave one region out			
Africa	.171	.177	64
Asia	.081	.089	444
Australia	-.019	.863	81
Central and South America	.124	.110	167
Europe	.101	.003	865
North America	.074	.000	7,694
Other	.368	.239	12
Overall result	*.070*	*.000*	*9,327*

4. GENERAL DISCUSSION

We expected gender, age, and geographical variation among students would relate to the sentiment they express in their reasons for enrolling. For example, there were clear differences in rates of enrollment for females and males depending on course topic, especially for the *Android Apps* course (much higher male enrollment). Such enrollment differences could be driven, in part, by sentiment at the time of enrollment. In fact, we found differences in the sentiment of students' reasons for enrolling among students of differing genders and ages, though less difference across geographical regions. Both female and older students shared an increased expression of fear and disgust compared to their male and younger student peers, respectively (Table 1).

We also expected sentiment of enrollment reasons to be predictive of course engagement, though not to a large degree since there are other possible factors at play (e.g., individual differences, unexpected life events, quality of instruction). Indeed, we found a predictive random forest classification model based on sentiment was significantly better than chance when predicting three levels of engagement. However, prediction accuracy was greatly impacted by geographical region (Table 2). It is possible that the results are indicative of regional differences between students. For example, cultural expectations could impact expression of sentiment, as could use of English as a second language—as is likely the case for many students outside North America.

Our findings suggest design choices for MOOCs with data-driven interventions to improve retention (e.g., [10]). Students' enrollment sentiment could be analyzed to predict engagement or enrollment with the goal of driving interventions. Given modest accuracy, such interventions should be "fail-soft", but would also be combined with existing models in an ensemble to target interventions more accurately.

4.1 Limitations and Future Work

In this study we explored the sentiment of students' reasons for enrolling in MOOCs. However, some students might stay in MOOCs for different reasons than why they enrolled. In other words, they might discover unexpected value in a MOOC that extends or replaces the original reasons they had for enrolling. Future work should extend this research to consider how students' reasons for remaining in a MOOC evolve over time, and in particular how sentiment of their reasons changes in response to successes and failures they experience.

Some of our results were also limited by sample size despite the large number of students considered. Certain comparisons between demographic groups and predictive model generalization across demographics would have benefited from more data. For instance, there were only 64 students from Africa (Table 2) in our data, and even though results suggest the engagement prediction model generalized well to these students, it is unclear without additional data. Future work should focus on groups underrepresented in MOOCs so that they are not "left behind" by models and analyses tuned for traditional majority students.

5. CONCLUSION

In this paper we examined the sentiment of students' self-reported reasons for enrolling in MOOCs, and found that there were demographic differences. Although those differences were small, they consistently predicted some of the variation in course achievement across five different MOOCs. Our findings will lead to future work understanding students' learning objectives, especially with respect to better understanding how learners from different backgrounds approach courses differently. It is our objective that this will eventually lead to MOOCs that are designed to support the needs of all students.

References

[1] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[2] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct. 2001. DOI: 10.1023/A:1010933404324.

[3] S. A. Crossley, K. Kyle, and D. S. McNamara. Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, 49(3):803–821, June 2017. DOI: 10.3758/s13428-016-0743-z.

[4] R. W. Crues, G. M. Henricks, M. Perry, S. Bhat, C. J. Anderson, N. Shaik, and L. Angrave. How do gender, learning goals, and forum participation predict persistence in a computer science MOOC? *ACM Transactions on Computing Education.*

[5] R. F. Kizilcec and E. Schneider. Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(2):6:1–6:24, Mar. 2015. DOI: 10.1145/2699735.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, Nov. 2011.

[7] C. Robinson, M. Yeomans, J. Reich, C. Hulleman, and H. Gehlbach. Forecasting student achievement in MOOCs with natural language processing. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 383–387, New York, NY, USA. ACM, 2016. DOI: 10.1145/2883851.2883932.

[8] D. Shah. By The Numbers: MOOCS in 2017, Jan. 2018. URL: https://www.class-central.com/report/mooc-stats-2017/.

[9] P. J. Stone, D. C. Dunphy, and M. S. Smith. *The general inquirer: A computer approach to content analysis.* MIT Press, Oxford, England, 1966.

[10] J. Whitehill, J. J. Williams, G. Lopez, C. A. Coleman, and J. Reich. Beyond prediction: First steps toward automatic intervention in MOOC student stopout. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*, pages 171–178. International Educational Data Mining Society, 2015.

[11] C. Wladis, K. M. Conway, and A. C. Hachey. The online STEM classroom—Who succeeds? An exploration of the impact of ethnicity, gender, and non-traditional student characteristics in the community college context. *Community College Review*, 43(2):142–164, Apr. 2015. DOI: 10.1177/0091552115571729.

[12] C. Wladis, A. C. Hachey, and K. Conway. Which STEM majors enroll in online courses, and why should we care? The impact of ethnicity, gender, and non-traditional student characteristics. *Computers & Education*, 87:285–308, Sept. 2015. DOI: 10.1016/j.compedu.2015.06.010.