M-CAFE: Managing MOOC Student Feedback with Collaborative Filtering

Mo Zhou
IEOR Department
UC Berkeley
mzhou@berkeley.edu

Alison Cliff
IEOR Department
UC Berkeley
alisoncliff@berkeley.edu

Allen Huang
IEOR Department
UC Berkeley
hunallen@berkeley.edu

Sanjay Krishnan
EECS Department
UC Berkeley
sanjaykrishnan@berkeley.edu

Brandie Nonnecke
CITRIS Data&Democracy Initiative
UC Berkeley
nonnecke@berkeley.edu

Kanji Uchino
Fujitsu Laboratories of America
Kanji@us.fujitsu.com

Sam Joseph
Hawaii Pacific University
sjpseph@hpu.edu

Armando Fox
EECS Department
UC Berkeley
fox@cs.berkeley.edu

Ken Goldberg
IEOR & EECS Department
UC Berkeley
goldberg@berkeley.edu

L@S 2015, Mar 14-18, 2015, Vancouver, BC, Canada
ACM 978-1-4503-3411-2/15/03.
http://dx.doi.org/10.1145/2724660.2728681

Abstract
Ongoing student feedback on course content and assignments can be valuable for MOOC instructors in the absence of face-to-face-interaction. To collect ongoing feedback and scalably identify valuable suggestions, we built the MOOC Collaborative Assessment and Feedback Engine (M-CAFE). This mobile platform allows MOOC students to numerically assess the course, their own performance, and provide textual suggestions about how the course could be improved on a weekly basis. M-CAFE allows students to visualize how they compare with their peers and read and evaluate what others have suggested, providing peer-to-peer collaborative filtering. We evaluate M-CAFE based on data from two EdX MOOCs.

Author Keywords
Course assessment; Collaborative filtering; Instructor support; MOOCs

ACM Classification Keywords
K.3.1 Computers Users in Education: Distance learning. H.5.3 Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces

Introduction
Student engagement with course material is important for effective teaching [1]. However, assessing engagement is particularly challenging in the context of MOOCs because of the lack of face-to-face interaction.

The online structure of MOOCs encourages development of novel course evaluation methods and timely feedback systems.

Apart from quantitative data from course evaluations, instructors often find qualitative (textual) data to be a valuable supplement [4]. Stephens-Martinez et al. suggest that qualitative data such as discussion forums, class surveys, and discussions with students are considered the most useful [2]. However, the vast quantity of qualitative data produced by these sources impedes professors from obtaining valuable information efficiently. New tools must be developed that allow MOOC instructors to efficiently collect and analyze data.

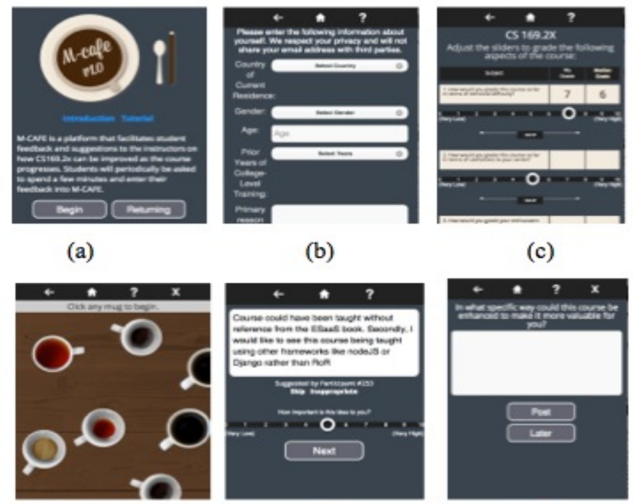
In this work-in-progress report, we present a new platform, the MOOC Collaborative Assessment and Feedback Engine (http://opinion.berkeley.edu/m-cafe-2), which encourages students to check in weekly to numerically assess the course, provide qualitative (textual) suggestions on course improvements, and rate each other's suggestions.

As a pilot study, we evaluate M-CAFE with 2 EdX MOOCs by testing the following hypotheses:

1. The ranking of textual input using peer-to-peer collaborative filtering will be significantly closer to expert human performance than NLP methods.

2. The temporal trends of quantitative input will be significantly correlated with events and changes in the course (i.e. material difficulty, homework assignments, quizzes, etc.).

M-CAFE User Interface



(a) (b) (c)

(d) (e) (f)

Figure1: Screenshots of M-CAFE

Upon entering M-CAFE (Figure 1 (a)), new students are required to register by email and are given the option to provide their age, gender, home country, years of college-level education, and the primary reason why they are taking the course (Figure 1 (b)). Returning students can login with their email address. Students are then directed to a page (Figure 1 (c)) where they rate the following 5 quantitative assessment topics (QAT) about the course on a scale of 1-10 (1 represents "Very Low" and 10 represents "Very High"):

(1) How would you grade this course so far in terms of technical difficulty? (Course Difficulty)
(2) How would you grade this course so far in terms of usefulness to your career? (Course Usefulness)
(3) How would you grade your enthusiasm so far for this course? (Self-Enthusiasm)

(4) How would you grade your performance so far in this course? (Self-Performance)
(5) How would you grade the effectiveness of course assignments to help you develop your skills? (HW Effectiveness)

Next, students are directed to a virtual “cafe” interface (figure 1 (d)) where mugs representing other students are arranged on a coffee table [18]. In the cafe, qualitative (textual) suggestions from students are solicited around a Discussion Question: “In what specific ways could this course be enhanced to make it more valuable for you?”

Students click on mugs to view the textual suggestions of their peers and then evaluate how valuable the suggestions are on a scale of 1 -10 (Figure 1 (e)). After rating two suggestions, a new mug appears in the middle of the table, prompting the student to enter his or her own suggestion (Figure 1 (f)).

Case Study

We utilized M-CAFE during the June-July 2014, 6-week EdX MOOC course CS 169.2x and the October-December 2014, 8-week EdX MOOC course CS 169.1x: Engineering Software as a Service. Each week students were invited via email to visit or revisit M-CAFE to enter feedback about the course.

Qualitative Textual Feedback

Data

Students provided a total of 83 suggestions for CS 169.1x and 132 suggestions for CS 169.2x to answer the question “In what specific way could this course be enhanced to make it more valuable for you?” We also observed a total of 1,691 and 3,564 peer-to-peer ratings for the two courses respectively.

We ranked the suggestions using the lower bound of the binomial proportion confidence interval (also called the Wilson Score). This score incorporates both the mean and the variance of the ratings received. We took the mean grade **g** and then calculated the 95% confidence interval of **g** using standard error: **g +/- 1.96*SE(g)**. We ranked the comments by the lower bound **g - 1.96*SE(g)** since it is more robust when suggestions are contentious and receive a different number of ratings.

Collaborative Filtering

Various online platforms with crowd-sourcing ability have adopted collaborative filtering as an alternative approach to Natural Language Processing (NLP) [3]. While NLP of qualitative data hints at the important words or phrases, it fails to identify the underlying meaning of the sentences. Given the complicated nature of natural language, collaborative filtering has gained popularity for content-based items, such as customer recommendations and solution rankings on online QA platforms. Unlike most cases where the items are rated dichotomously, in M-CAFE, we adopted a 10-point scale to better capture rating differences between items in a setting with hundreds of participants.

We hypothesize that peer-to-peer collaborative filtering is capable of identifying valuable suggestions in M-CAFE. Our work will assess the performance of collaborative filtering based on three questions:

1. Are novel comments receiving higher ratings?

2. Is the top-rated set of comments representative of the entire dataset in terms of topics covered?
3. Are comments with higher quality (e.g. conciseness, clarity and completeness) receiving higher ratings?

If the answer is yes to all of the above questions, looking at the top-rated comments would provide us most of the information contained in the entire set, thus addressing the scale issue of qualitative data by reducing the dataset to a representative subset.

Demographic Influence

We will also evaluate the demographic influence on the students' behavior in M-CAFE. How does gender, age, country, or years of training affect the number of peer-to-peer ratings provided? Which types of students are more likely to provide insightful suggestions and what aspects of the course are students from similar backgrounds mostly concerned with?

Quantitative Feedback

There are numerous studies on the danger of purely quantitative course evaluations [3]. Stark and Freishtat [4] argue that quantitative rating data from course evaluations are subject to significant self-selection bias. Though our data nonetheless suffer from self-selection bias, the temporal nature of M-CAFE allows for relative analysis of quantitative data (i.e. increases and decreases) that may be less susceptible to biases.

We will explore the relative relationship of the QATs between weeks throughout each course and attempt to link the quantitative observations to the qualitative feedback in the corresponding weeks to ensure the validity of quantitative trends. In addition, we will study the inter-relationship between the QATs and the correspondence of the QATs to the schedule of the course, aiming to identify other aspects of the course that could explain and validate the trends presented by the quantitative data.

Acknowledgment

This work is supported in part by the Blum Center for Developing Economies and the Development Impact Lab (USAID Cooperative Agreement AID-OAA-A-12-00011) as part of the USAID Higher Education Solutions Network, UC Berkeley's Algorithms, Machines, and People (AMP) Lab, a grant from Fujitsu, and the UC CITRIS Data and Democracy Initiative. Thanks to all students who participated and colleagues who provided feedback on earlier drafts.

References

[1] Angelo, T.A. Cross, P.K. Classroom assessment techniques: A handbook for faculty. National Center for Research to Improve Teaching and Learning 1993. Ann Arbor, MI, 1993.

[2] Kristin, S.M., Hearst, M.A., and Fox, A. Monitoring MOOCs: which information sources do instructors value? *Proceedings of the First ACM conference on Learning @ Scale Conference.* (2014), 79-88.

[3] Marsh, H.W., and Roche, L.A. Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52(11).* (1997), 1187-1197.

[4] Stark, P.B. and Richard F. An Evaluation of Course Evaluations. *UC Berkeley Technical Report; Center for Teaching and Learning.* (2014)