

EMNLP 2014

**The 2014 Conference on Empirical Methods  
In Natural Language Processing  
Workshop on Modeling Large Scale Social Interaction  
In Massively Open Online Courses**

**Proceedings of the Workshop**

October 25, 2014  
Doha, Qatar

Production and Manufacturing by  
*Taberg Media Group AB*  
*Box 94, 562 02 Taberg*  
*Sweden*

©2014 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-937284-96-1

## Introduction

Welcome to the EMNLP 2014 Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses. An emerging area for real world impact of technology for analysis of social media at a large scale is online education in Massively Open Online Courses (MOOCs). The goal of this workshop is to explore what the language technologies community has to offer this endeavor. At this one day workshop organized around a shared task related to analysis of large scale social interaction in MOOCs, we will grapple with the competing images of the inner workings of massive learning communities provided by alternative computational approaches.

With the recent press given to online education and increasing enrollment in online courses, the need for scaling up quality educational experiences online has never been so urgent. Current offerings provide excellent materials including video lectures, exercises, and some forms of discussion opportunities. One important hurdle that prevents MOOCs from reaching their transformative potential is that they fail to provide the kind of social environment that is conducive to sustained engagement and learning, especially as students arrive in waves to these online learning communities. While limited, current affordances for social interaction in MOOCs have already shown some value for providing students with connection to others that provides some needed motivational benefits.

Computational modeling of massive scale social interaction has the potential to yield new knowledge about the inner-workings of interaction in such environments so that support for healthy community formation can be designed and built. However, the state-of-the-art in graphical models applied to large scale social data provides representations of the data that are challenging to interpret in light of specific questions that may be asked from a learning sciences or social psychological perspective. What is needed are new methodologies for development and interpretation of models that bridge expertise from machine learning and language technologies on one side and learning sciences, sociolinguistics, and social psychology on the other side. The field of language technologies has the human capital to take leadership in making these breakthroughs. Other specific opportunities for the field associated with that enterprise are problems in assessment of student work (e.g., automated essay scoring), generation of in process feedback for students learning online independently or in groups (e.g. tutorial dialogue agents), support for large scale threaded discussions (e.g., dialogue agent based facilitation), and summarization of participation data for facilitators and course developers who revise and maintain course materials (e.g., conversation summarization).

MOOCs are especially interesting as a source of large scale social data. The unique developmental history of MOOCs creates challenges that require insight into the inner-workings of massive scale social interaction in order to meet. In particular, rather than evolving gradually as better understood forms of online communities, MOOCs spring up overnight and then expand in waves as new cohorts of students arrive from week to week to begin the course. As massive communities of strangers that lack shared practices that would enable them to form supportive bonds of interaction, these communities grow in an unruly manner. While some students may successfully find birds of a feather with whom to bond and find support, when others come they may find an overwhelming amount of communication having already been posted that they feel lost in. Others may find themselves somewhere in between these two extremes. They may begin to form weak bonds with some other students when they join, however, massive attrition may create challenges as members who have begun to form bonds with fellow students soon find their virtual cohort dwindling. Early attempts to organize the community into smaller study groups may be thwarted by such periodic growth spurts paired with attrition, as groups that initially had an appropriate critical mass soon fall below that level and then are unable to support the needs of remaining students. Can our models serve as useful lenses to offer insights into these social processes? Come to the workshop and join in the discussion!!

**Organizers:**

Carolyn Rosé, Carnegie Mellon University

George Siemens, University of Texas at Arlington

**Program Committee**

Hua Ai, Georgia Institute of Technology

Ryan Baker, Teacher's College, Columbia University

Kristy Boyer, North Carolina State University

Emma Brunskill, Carnegie Mellon University

Brian Butler, University of Maryland

Hal Daumé III, University of Maryland

Barbara Di Eugenio, University of Illinois at Chicago

Jana Diesner, University of Illinois at Urbana-Champaign

Jacon Eisenstein, Georgia Institute of Technology

Dragan Gasevic, Athabasca University

Neil Heffernan, Worcester Polytechnic

Eduard Hovy, Carnegie Mellon University

Lillian Lee, Cornell University

Alice Oh, Korea Advanced Institute of Science and Technology

Mari Ostendorf, University of Washington

Keith Sawyer, University of North Carolina

Hinrich Schuetze, University of Munich

Simon Buckingham Shum, The Open University

Yla Tausczik, Carnegie Mellon University

Stephanie Teasley, University of Michigan

Joel Tetreault, Nuance

Chong Wang, Carnegie Mellon University

Jason Williams, Microsoft Research

Alyssa Wise, Simon Fraser University

Eric Xing, Carnegie Mellon University

## Table of Contents

<i>Keynote: Data Archeology: A theory informed approach to analyzing data traces of social interaction in large scale learning environments</i>	
Alyssa Wise .....	1
<i>Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions</i>	
Tanmay Sinha, Patrick Jermann, Nan Li and Pierre Dillenbourg .....	3
<i>Identifying Student Leaders from MOOC Discussion Forums through Language Influence</i>	
Seungwhan Moon, Saloni Potdar and Lara Martin .....	15
<i>Towards Identifying the Resolvability of Threads in MOOCs</i>	
Diyi Yang, Miaomiao Wen and Carolyn Rose .....	21
<i>Point-of-View Mining and Cognitive Presence in MOOCs: A (Computational) Linguistics Perspective</i>	
Noureddine Elouazizi .....	32
<i>Keynote Talk: Analytics: climbing up the ladder of behavior control</i>	
Patrick Jermann .....	38
<i>Shared Task on Prediction of Dropout Over Time in Massively Open Online Courses</i>	
Carolyn Rose and George Siemens .....	39
<i>Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners</i>	
Tanmay Sinha, Nan Li, Patrick Jermann and Pierre Dillenbourg .....	42
<i>A Process for Predicting MOOC Attrition</i>	
Mike Sharkey and Robert Sanders .....	50
<i>Predicting Attrition Along the Way: The UIUC Model</i>	
Bussaba Amnueypornsakul, Suma Bhat and Phakpoom Chinprutthiwong .....	55
<i>Predicting MOOC Dropout over Weeks Using Machine Learning Methods</i>	
Marius Kloft, Felix Stiehler, Zhilin Zheng and Niels Pinkwart .....	60

# Conference Program

**Saturday, October 25, 2014**

## **Session 1**

09:00–09:20 *Opening Remarks*  
The organizers

09:20–10:30 *Keynote: Data Archeology: A theory informed approach to analyzing data traces of social interaction in large scale learning environments*  
Alyssa Wise

**10:30–10:50** *Coffee Break*

## **Session 2**

10:50–11:15 *Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions*  
Tanmay Sinha, Patrick Jermann, Nan Li and Pierre Dillenbourg

11:15–11:40 *Identifying Student Leaders from MOOC Discussion Forums through Language Influence*  
Seungwhan Moon, Saloni Potdar and Lara Martin

11:40–12:05 *Towards Identifying the Resolvability of Threads in MOOCs*  
Diyi Yang, Miaomiao Wen and Carolyn Rose

12:05–12:30 *Point-of-View Mining and Cognitive Presence in MOOCs: A (Computational) Linguistics Perspective*  
Noureddine Elouazizi

**12:30–14:00** *Lunch*

**Saturday, October 25, 2014 (continued)**

**Session 3**

- 14:00–15:10 *Keynote Talk: Analytics: climbing up the ladder of behavior control*  
Patrick Jermann
- 15:10–15:30 *Shared Task on Prediction of Dropout Over Time in Massively Open Online Courses*  
Carolyn Rose and George Siemens
- 15:30–16:00** *Coffe Break*

**Session 4**

- 16:00–16:20 *Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners*  
Tanmay Sinha, Nan Li, Patrick Jermann and Pierre Dillenbourg
- 16:20–16:40 *A Process for Predicting MOOC Attrition*  
Mike Sharkey and Robert Sanders
- 16:40–17:00 *Predicting Attrition Along the Way: The UIUC Model*  
Bussaba Amnueypornsakul, Suma Bhat and Phakpoom Chinprutthiwong
- 17:00–17:20 *Predicting MOOC Dropout over Weeks Using Machine Learning Methods*  
Marius Kloft, Felix Stiehler, Zhilin Zheng and Niels Pinkwart