MLA'14 – Third Multimodal Learning Analytics Workshop and Grand Challenges

Xavier Ochoa
Escuela Superior Politécnica del Litoral
Km. 30.5 Vía Perimetral
Guayaquil, Ecuador
xavier@cti.espol.edu.ec

Marcelo Worsley
Stanford University
520 Galvez Mall
Stanford, USA
mworsley@stanford.edu

Katherine Chiluiza
Escuela Superior Politécnica del Litoral
Km. 30.5 Vía Perimetral
Guayaquil, Ecuador
kchilui@espol.edu.ec

Saturnino Luz
Trinity College Dublin
College Green
Dublin, Ireland
Saturnino.Luz@cs.tcd.ie

ABSTRACT

This paper summarizes the third Multimodal Learning Analytics Workshop and Grand Challenges (MLA'14). This subfield of Learning Analytics focuses on the interpretation of the multimodal interactions that occurs in learning environments, both digital and physical. This is a hybrid event that includes presentations about methods and techniques to analyze and merge the different signals captured from these environments (workshop session) and more concrete results from the application of Multimodal Learning Analytics techniques to predict the performance of students while solving math problems or presenting in the classroom (challenges sessions). A total of eight articles will be presented in this event. The main conclusion from this event is that Multimodal Learning Analytics is a desirable research endeavour that could produce results that can be currently applied to improve the learning process.

Categories and Subject Descriptors

K.3.1 [**Computers in Education**]: Miscellaneous

Keywords

Learning analytics, educational data mining, multimodal interaction

1. INTRODUCTION

Learning Analytics is the "middle-space" where Educational Sciences, Computer Science, Learning Technologies and Data Science converge. The main goal of this new field of knowledge is to contribute to new empirical findings, theories, methods, and metrics for understanding how students learn and to use that knowledge to improve those students' learning.

One promising research line within Learning Analytics is Multimodal Learning Analytics, which emphasizes the analysis of natural rich modalities of communication during situated learning activities. This includes students' speech, writing, and nonverbal interaction (e.g., gestures, facial expressions, gaze, etc.). A primary objective of Multimodal Learning Analytics is to analyze coherent signal, activity, and lexical patterns to understand the learning process and to provide relevant feedback to participants and practitioners. The Third International Workshop on Multimodal Learning Analytics brings together researchers in multimodal interaction and systems, cognitive and learning sciences, educational technologies, and related areas to discuss the recent developments and future opportunities in this sub-field.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
ICMI'14, November 12–16, 2014, Istanbul, Turkey.
ACM 978-1-4503-2885-2/14/11.
http://dx.doi.org/10.1145/2663204.2668318.

2. PROGRAM

Following the First International Workshop on Multimodal Learning Analytics in Santa Monica in 2012 and the ICMI Grand Challenge on Multimodal Learning Analytics in Sydney in 2013, this third workshop comprises a mixture of a workshop session and two data-driven grand challenge sessions.

2.1 Workshop Session

The workshop session focuses on the presentation of multimodal signal analysis techniques that could be applied in Multimodal Learning Analytics. Instead of requiring research results, that usually are presented at the Learning Analytics and Knowledge (LAK) or Multimodal Interaction (ICMI) conferences, this workshop challenges presenters to concentrate on benefits and shortcomings of different research and technical methods used for multimodal analysis of learning signals. This session includes four articles from diverse topics: theoretical and conceptual considerations for different forms of multimodal data fusion; voice analysis to determine the level of rapport in learning exercises; video analysis of live classrooms; and the role of multimodal analysis in the service of studying complex learning environments.

2.2 Challenge Sessions

Following the successful experience of the previous Multimodal Learning Analytics Grand Challenge in ICMI 2013, this year this event will provide two data sets with a wealth of research questions to be tackled by interested participants:

2.2.1 Math Data Challenge

The Math Data Corpus from the 2013 ICMI Multimodal Learning Analytics Grand Challenge was again made available for analysis. This corpus involves 12 sessions, with groups of three students collaborating while solving mathematics problems (i.e. geometry, algebra). Data were collected of participants' natural multimodal communication and activity patterns during these problem-solving sessions. The modalities collected include: speech, digital pen input, and video. In total, approximately 16 hours of multimodal data is available. Eighteen high-school students participated in this study. Each group of three students met for two sessions. The student groups varied in performance characteristics, with some low-to-moderate performers and others high-performing students. During the sessions, students were engaged in authentic problem solving and peer tutoring as they worked on 16 mathematics problems. The sixteen problems were uniformly distributed in terms of difficulty, with four of each type of question: easy, moderate, hard, and very hard difficulty levels. The data have been segmented by start and end time of each problem, scored for solution correctness, and also scored for which student solved the problem correctly.

This year, the dataset has been expanded with full manual and automatic transcripts of the students' speech. It also contains more than 10,000 annotations of the students' notes and diagrams. The main research questions considered include: 1) automatic prediction of which math problems will be solved correctly or incorrectly; 2) which student in a group is the dominant domain expert; and 3) identification of significant precursors of performance and learning. Predictors could be based on information from unimodal or multimodal signals, lexical/representational content, individual or group dynamics, or combined information sources. One article presented in this challenge provided a detailed exploration of how to use the digital pen information to predict the expertise in the group. This work reaches high levels of accuracy (83%) when i dentifying the expert student among the 3 participants.

2.2.2 Presentation Quality Challenge

This challenge includes a data corpus that involves 40 oral presentations of Spanish-speaking students in groups of 4 to 5 members. These participants are presenting projects on entrepreneurship ideas, literature reviews, research designs, software design, etc. Data were collected on presenters' natural multimodal communication in regular classroom settings. The following data is available: speech, facial expressions and physical movements in video, skeletal data gathered from Xbox Kinect for each individual, and slide presentation files. In total, approximately 10 hours of multimodal data is available for analysis of these presentations. In addition, individual and group grades (based on the quality of the slides) grading for individuals when doing their presentations is included as well as a group-grade related to the quality of the slides used when doing each presentation. This challenge seeks to solve the following questions: a) How multimodal techniques can help in evaluating presentation quality and presenter behaviors; and b) How good is a group presentation based on the individual presentations and the quality of the slides used in a presentation.

This session is composed of three articles. The first one explores the slide presentation files and the audio features to predict the grade obtained by each student. This work found that the models created from slide features were accurate up to 65%. The most relevant features for the slide-base models were: number of words, images, and tables, and the maximum font size. The audio-based models reach 69% accuracy, with pitch and filled pauses related features being the most significant. The second work makes use of all the provided modalities (audio, video, Kinect data and slide files) and suggests that multimodal cues can predict human scores on presentation tasks. Furthermore, this second paper shows that a scoring model comprising both verbal and visual features can outperform a unimodal model. The final article uses the video and Kinect information to predict human grading. The Machine-learning evaluations resulted in models that predicted human grading of eye contact and postures, and body language, with accuracies 61% and 68%, respectively. Furthermore, the results suggest that a set of body language features, such as arm movement and smoothness, are of high significance for predicting the level of development for presentation skills.

3. ORGANIZATION

Because this community is still in its infancy, the workshop and grand challenges chairs were selected from previous organizers and participants:

- Xavier Ochoa, ESPOL, Ecuador
- Marcelo Worsley, Stanford, USA
- Katherine Chiluiza, ESPOL, Ecuador
- Saturnino Luz, Trinity College Dublin, Ireland

4. CONCLUSIONS

The third Multimodal Learning Analytics Workshop and Grand Challenges (MLA'14) was organized to bring together experts in computer science, learning sciences, learning technologies and data science. It was envisioned that this expert meeting would be the space to initiate research in this nascent subfield of Learning Analytics. New challenges and insights will arise from the convergence of practitioners, academics and researchers, which in turn will create opportunities to collaborate and to create applications and tools to assist students, teachers and the community.

5. ACKNOWLEDGEMENTS

The organizers want to express their gratitude to Incaa Designs and Escuela Superior Politécnica del Litoral (ESPOL) for the access provided to their datasets. Without their help and work, this event would not be possible.