ICMI 2013 Grand Challenge Workshop on Multimodal Learning Analytics

Louis-Philippe Morency
University of Southern California
Institute for Creative Technologies
morency@ict.usc.edu

Sharon Oviatt
Incaa Designs
oviatt@incaadesigns.org

Stefan Scherer
University of Southern California
Institute for Creative Technologies
scherer@ict.usc.edu

Nadir Weibel
University of California San Diego
Department of Computer Science and Engineering
weibel@ucsd.edu

Marcelo Worsley
Stanford University
Graduate School of Education and Department of Computer Science
mworsley@stanford.edu

ABSTRACT
Advances in learning analytics are contributing new empirical findings, theories, methods, and metrics for understanding how students learn. It also contributes to improving pedagogical support for students' learning through assessment of new digital tools, teaching strategies, and curricula. M*ultimodal learning analytics (MMLA)*[1] is an extension of learning analytics and emphasizes the analysis of natural rich modalities of communication across a variety of learning contexts. This MMLA Grand Challenge combines expertise from the learning sciences and machine learning in order to highlight the rich opportunities that exist at the intersection of these disciplines. As part of the Grand Challenge, researchers were asked to predict: (1) which student in a group was the dominant domain expert, and (2) which problems that the group worked on would be solved correctly or not. Analyses were based on a combination of speech, digital pen and video data. This paper describes the motivation for the grand challenge, the publicly available data resources and results reported by the challenge participants. The results demonstrate that multimodal prediction of the challenge goals: (1) is surprisingly reliable using rich multimodal data sources, (2) can be accomplished using any of the three modalities explored, and (3) need not be based on content analysis.

Categories and Subject Descriptors
K.3.1 [Computing Milieux]: Computers and Education-Computer Uses in Education.

General Terms
Evaluation, Human Factors, Performance, Design, Experimentation, Algorithms

Keywords
Multimodal learning analytics, Predictive data and models, Domain expertise, Empirical and machine learning techniques, Collaboration

1. INTRODUCTION
Multimodal learning analytics [1] represents an integration of multimodal learning, empirical multimodal analysis methods, and engineering techniques. Multimodal learning has long been advocated because of the diverse learning opportunities that it creates for students, and the expansive opportunities that students have for interacting with and expressing knowledge in different ways (e.g. [2], [3]). However, with the diversity of learning opportunities and the equally diverse ways that learning is evidenced, comes an even greater number of factors to consider when trying to understand what and how students are learning in that environment. Traditional approaches in the analysis of multimodal learning data consist of extensive human labeling and annotation. Furthermore, human processing of multimodal data can often be coarse, with the person coding unable to truly capture the nuances of what is taking place because of limits on human processing of micro-level events. In response to these, multimodal learning analytics proposes the use of multimodal analysis techniques in order to 1) streamline and systematize the data analysis process and 2) achieve a level of analysis that would be seemingly impossible without the tools of computation. Additionally, while prior research in learning analytics and educational data mining has demonstrated that learning is expressed across various modalities, most of this work has examined these modalities in isolation. This overlooks the interactions that exist between the different modalities, and limits our understanding of cognition and learning.

This workshop aims to bring together researchers from a broad set of expertise in order to realize the development and dissemination of novel techniques for analyzing multimodal learning data, in a way that is both technically rigorous and pedagogically-informed. It also aims to develop new learning analytics techniques that are more appropriate for the multimodal interfaces that are on modern computing devices such as smart phones and tablets, which have become the dominant educational platform worldwide.

This paper summarizes the corpus that participants used, and then presents a synopsis of the various techniques and results reported by workshop participants. Finally, it concludes with remarks on future directions of this research domain and on-going challenges that we hope the larger community will begin to tackle.

2. MATH DATA CORPUS
The Math Data Corpus used for the MMLA grand challenge consists of high-fidelity time-synchronized multimodal data recordings on collaborating groups of students as they work together to solve mathematics problems varying in difficulty. Data were collected on students' natural multimodal communication and activity patterns, including their speech, digital pen input, and

ICMI'13, December 9–13, 2013, Sydney, Australia.

http://dx.doi.org/10.1145/2522848.2534669

video. The dataset includes 12 sessions, with six three-student groups who each met twice. In total, approximately 29 student-hours of recorded multimodal data are available during the collaborative problem solving sessions. This data resource includes initial coding of problem segmentation, problem-solving correctness, and representational content on students' writing [4].

2.1 Student Participants

Participants in this study included 18 high school students, 9 female and 9 male, who ranged in age from 15 to 17 years old. All had recently completed Introductory Geometry at a local high school and represented a range of geometry skills from average to high performers. During the data collection, small groups of three students who were gender matched jointly solved problems and mentored one another.

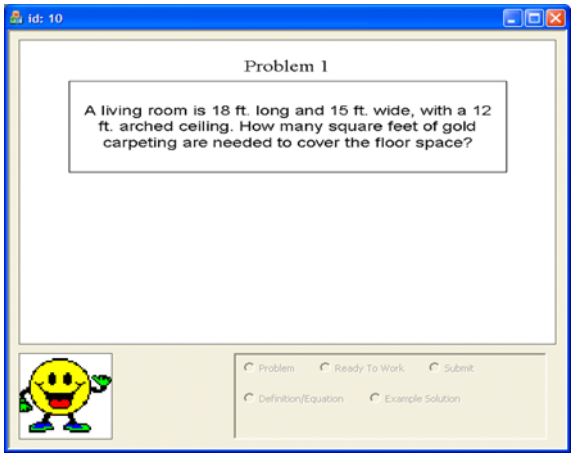

Figure 1. Interface displaying easy problem.

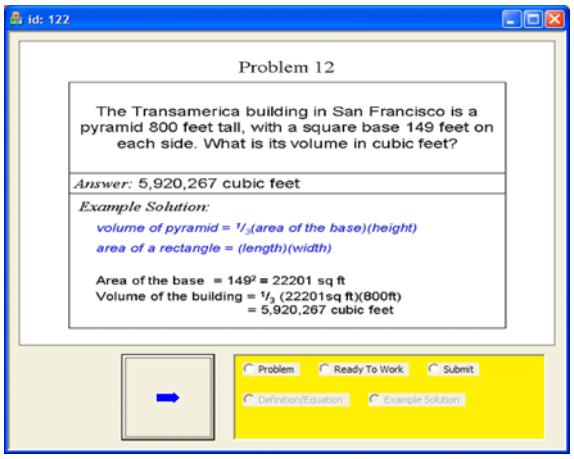

Figure 2. Interface displaying very hard problem

2.2 Math Tasks

During each session, students engaged in authentic problem solving and peer tutoring as they worked on 16 geometry and algebra problems, four apiece representing easy, moderate, hard, and very hard difficulty levels. These math problems were presented as word problems. Figure 1 shows an example of a low-difficulty problem, whereas Fig. 2 shows an example of a very high-difficulty problem. The difficulty level of the problems was validated using teacher records, pre-experimental piloting, and then confirmed with students' percentage of correct solutions in the study.

2.3 Data Collection Procedure

Each of the six student groups met for two sessions, during which students could view the math problems displayed one at a time on a tabletop computer screen (for details see [4]). One student in the group was designated as the leader for a given session. The designated leader switched on the group's second session to a different student.

Each group was instructed to exchange information as they worked on solving the problems, so everyone understood the solution and could explain it if asked by the computer system when they finished.

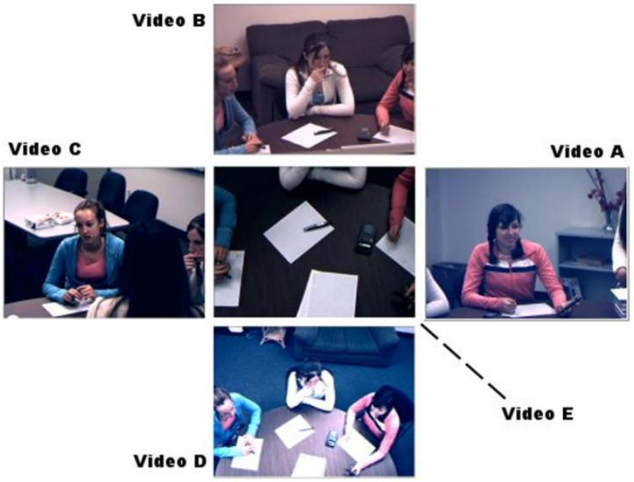

Figure 3. **Synchronized views from all five video-cameras at the same moment in time during data collection. Videos A, B and C show close-up views of the three individual students, video D a wide-angle view of all students, and video E a top-down view of student**

Each math problem solving and mentoring session lasted about an hour. While working on each of the 16 problems during a session problem, the leader would begin by asking to see the next problem. Students then typically discussed it with one another, and the leader could ask for related mathematical terms and equations on behalf of the group. All students could use pen and paper and a calculator as tools to draw diagrams, make calculations, etc. while working on each problem. One of the students usually proposed a solution to the others, which then was discussed among them. Once it was agreed upon as correct and all students understood how it had been solved, the leader then submitted the group's answer to the computer. Afterwards, the system displayed the correct answer, so students could verify their solution. If correct, one of the students was randomly called upon by the computer to explain how they had arrived at the solution. If not correct, the leader could ask to see a worked example of how the solution had been calculated, which students then discussed before the leader would ask to see the next problem.

2.4 Multimodal Data Collection

For each session, a high-resolution digital video close-up of each participant from the waist up was captured, along with a wide-

angle room view and a view of the tabletop with paper, pens, and other artifacts. Figure 3 shows video frames taken at the same moment in time during group problem solving. Figure 4 illustrates multimodal data capture, and transmission of recordings to a screen in a separate room where the display of the math problem content was controlled.

Figure 3. Integration of multimodal data capture using cameras, microphones and digital pen input, with speech and pen data collected wirelessly

Digital audio recordings were collected of each participant's speech using close-talking, high-fidelity, and unobtrusive microphones, with transmission facilitated by wireless transmitter/receivers. A fourth digital audio recording of the room was collected using an omni-directional microphone hung above the table. Finally, each participant's writing was collected using digital pens and paper based on Anoto technology[1]. For details about the data collection infrastructure, how the twelve media streams were synchronized, and the degree of fidelity in media synchronization, see [4].

2.5 Ground-Truth Data Coding

The data were segmented from each session into time phases representing the start and end of each problem-solving episode, the time when one student first proposed the problem's solution, the time of actual answer submission, and the total time to solution. Problems were also coded by difficulty level, whether they were solved correctly or not, and were further annotated indicating which student proposed the problem solution.

To assess domain expertise associated with individual students, each student's cumulative problem-solving performance was calculated across their group's two sessions. The operational definition adopted for determining a student's level of domain expertise was based on how many problems they solved correctly versus incorrectly at different difficulty levels. [5] shows that (1) students classified as "domain experts" were a reliably distinct and non-overlapping group from those identified as non-experts, and (2) significant learning occurred during the longitudinal sequence (i.e., from session 1 to session 2).

In addition to the above ground-truth coding, all digital pen input was coded for the number, type, and semantic content of all written representations (i.e., diagrams, words, numbers, symbols, marks). Written disfluencies, task irrelevant content, and degree of diagram complexity also were coded for over 10,000 written representations in the dataset.

[1] http://www.anoto.com

3. GRAND CHALLENGE PAPERS

In this section we highlight the papers presented during the Grand Challenge Workshop. For each paper, we present a brief summary of their work. We then conclude with a synthesis of the collective ideas, as well as thoughts about future research in multimodal learning analytics.

3.1 Expertise Estimation based on Simple Multimodal Features, Xavier Ochoa [6]

This work processes video, audio and digital pen information included in the Math Data Corpus to address the challenge's primary goals. It identifies individual factors that are capable of successfully discriminating between experts and non-experts in in this corpus as they solve math problems. The main finding is that several of these individual factors, such as the percentage of time spent using the calculator, speed of writing or drawing, and the percentage of time numbers or mathematical terms were mentioned, are good discriminators between expert and non-expert students. Precision levels were reported of 63% for individual problems, and up to 80% when 12 problems were aggregated to make this distinction. The authors report that the methodology used to uncover individual predictive factors could potentially be very useful to create discrimination models for other contexts.

3.2 Automatic Identification of Experts and Performance Prediction in the Multimodal Math Data Corpus through Analysis of Speech Interaction, Saturnino Luz [7]

An analysis of multiparty interaction in the problem solving sessions of the Multimodal Math Data Corpus is presented. The analysis focuses on non-verbal cues extracted from the audio recordings. Algorithms for expert identification and performance prediction (correctness of solution) are implemented based on patterns of speech activity among session participants. Both of these categorization algorithms employ an underlying graph-based representation of dialogues for individual problem solving activities. The proposed Bayesian approach to expert prediction proved quite effective, reaching accuracy levels of over 92% with as few as 6 dialogues of training data. Performance prediction was not quite as effective. Although the simple graph-matching strategy employed for predicting incorrect solutions improved considerably over a Monte Carlo simulated baseline (F1 score increased by a factor of 2.3), there is still much room for improvement in this task.

3.3 Written and Multimodal Representations as Predictors of Expertise and Problem-solving Success in Mathematics, Sharon Oviatt [8]

In this research, writing and multimodal speech and writing are analyzed from the Math Data Corpus, both in terms of activity patterns (no content analysis) and the semantic content of representations. Findings reveal that in 96-97% of cases the correctness or incorrectness of a group's solution was predictable in advance based on students' written work content. In addition, a linear regression revealed that 65% of the variance in individual students' domain expertise rankings (i.e., based on their spoken contributions during group discussion) could be accounted for based on their preceding written work content.

With respect to the second challenge task, the dominant domain expert in a group was correctly predicted 100% of the time based on multimodal content analysis of individual student's written and

spoken input, which exceeded unimodal prediction rates. However, a simple multimodal activity analysis with no content analysis whatsoever also successfully identified the domain expert in a group 100% of the time. This activity analysis was based simply on the number of times a given student contributed a problem solution, irrespective of whether it was correct. This latter multimodal predictor would be easier to automate in the short term.

Further analysis revealed a reversal between experts and non-experts in the percentage of time that a match versus mismatch was present between their oral and written answer contributions, with non-experts demonstrating higher mismatches. Implications are discussed for developing reliable multimodal learning analytics systems that incorporate digital pen input to automatically identify consolidation of domain expertise.

3.4 Using Micro-patterns of Speech to Predict the Correctness of Answers to Mathematics Problems: an Exercise in Multimodal Learning Analytics, Kate Thompson [9]

In this paper, a rich description of the processes of learning at the system level (with regards to social interaction, generation of knowledge, and discourse related to action) was generated for a subset of sessions in the Math Data Corpus. Learning analytics techniques are traditionally used on the 'big data' collected at the course or university level. The application of such techniques to data generated in complex learning environments can provide insights into the relationships between the design of learning environments, the processes of learning, and learning outcomes.

In this paper, two of the codes described as part of the Collaborative Process Analysis Coding Scheme (CPACS) were extracted from the Math Data Corpus. The codes selected were tense and pronouns in spoken language (i.e., based on lexical transcription of speech), which have been found to indicate phases of group work and the action associated with collaboration. Rather than examine these measures of social interaction in isolation, a framework for the analysis of complex learning environments was applied. This facilitated an analysis of the relationships between the social interactions, the task design and learning outcomes, as well as tool use. The generation of a successful problem solution of one expert and one non-expert group was accurately predicted in one preliminary estimate (75%-94%).

The examination of interactions between the social, epistemic and tool elements of the learning environment for one group showed that successful role differentiation and participation were related to successful problem solutions in the first meeting. In the second meeting, these were less important. The relationship between discourse properties and the correctness of problems solutions was found to be less reliant. Further analysis is needed of additional data to pursue ideas reported in this paper.

DISCUSSION

This collection of very interesting findings confirms that extending traditional notions of educational assessment to embrace new developments in multimodal analysis potentially can improve our understanding of learning, even within interactive and collaborative problem-solving sessions that appear very complex and challenging. Among the papers selected for presentation is a consistent indication that learning is evidenced across all modalities (writing, speech, physical movement patterns). It also is possible to detect learning-oriented behaviors and expertise at different levels, from signal to representational.

Importantly, joint multimodal analysis of student behavior can support more reliable prediction of domain expertise and problem-solving correctness than unimodal sources alone. While most of the papers in this workshop focused on speech-based features, the integration of speech and writing channels, for example to predict correct answer contributions, represents a concrete case in which tracking students' multimodal activity patterns potentially can lead adequately high reliabilities (96-97%) to implement for practical purposes, even in a conservative field like education. This provides an early example that multimodal analysis is a promising direction that can improve our understanding of human cognition and learning.

With respect to future work, significant opportunities remain to integrate an even greater number of modalities into our future predictive analyses. Doing this will require the technical expertise of the multimodal community, as well as the theoretical insights of education researchers. It is our desire that this workshop will foster more multidisciplinary collaboration of this type, while also motivating institutions to initiate programs that permit students to develop dual specializations in computation and education.

Another key area for ongoing research is the development of automatic real-time systems that can both detect and enhance student learning. This must include presenting teachers and practitioners with the on-demand information that they need to promote student learning. The Intelligent Tutoring and Educational Robots communities have done significant work in this area, and they are prime candidates for leveraging multimodal learning analytic techniques in the near future.

4. CONCLUDING REMARKS

The Multimodal Learning Analytics Grand Challenge was organized to bring together expertise in the computationally-oriented multimodal community, with that in the learning sciences. We anticipate that embarking on multimodal learning analytics research will create new technical challenges for machine learning and multimodal researchers, and also provide advances for an important application area that is in need of analytic techniques. A strongly forged collaboration across these disciplines can drive profound changes in education and education research by: 1) uncovering entirely new insights about human cognition and learning; 2) creating novel and powerful computational tools that can assist teachers, parents and students; and 3) providing tools for designing more natural and effective learning environments, especially to accommodate the accelerating adoption of worldwide mobile devices.

5. REFERENCES

[1] Scherer, S., Worsley, M., and Morency, L.P. 2012. 1st international workshop on multimodal learning analytics: extended abstract. In Proceedings of the 14th ACM international conference on Multimodal interaction (ICMI '12). ACM, New York, NY, USA, 609-610. DOI=10.1145/2388676.2388803 http://doi.acm.org/10.1145/2388676.2388803

[2] Kress, G., Charalampos, T., Jewitt, C., & Ogborn, J. 2006. Multimodal teaching and learning: The rhetorics of the science classroom. Continuum International Publishing Group.

[3] Jewitt, C. 2012. Multimodal teaching and learning. The Encyclopedia of Applied Linguistics.

[4] Oviatt, S., Cohen, A. & Weibel, N. 2013. Multimodal learning analytics: Description of math data corpus for ICMI grand challenge workshop with full appendices, *Second International Workshop on Multimodal Learning Analytics*, Sydney Australia: http://mla.ucsd.edu/data/MMLA_Math_Data_Corpus.pdf

[5] Oviatt, S.L. 2013. Problem solving, domain expertise and learning: Ground-truth performance results for math data corpus, *Second International Workshop on Multimodal Learning Analytics*, Sydney Australia, December 2013.

[6] Ochoa, X. 2013. Expertise Estimation based on Simple Multimodal Features. *Second International Workshop on Multimodal Learning Analytics*, Sydney Australia, December 2013.

[7] Luz, S. 2013. Automatic Identification of Experts and Performance Prediction in the Multimodal Math Data Corpus through Analysis of Speech Interaction. *Second International Workshop on Multimodal Learning Analytics*, Sydney Australia, December 2013.

[8] Oviatt, S.L. 2013. Written and Multimodal Representations as Predictors of Expertise and Problem-solving Success in Mathematics. *Second International Workshop on Multimodal Learning Analytics*, Sydney Australia, December 2013.

[9] Thompson, Kate. 2013. Using Micro-patterns of Speech to Predict the Correctness of Answers to Mathematics Problems: an Exercise in Multimodal Learning Analytics, Kate Thompson, *Second International Workshop on Multimodal Learning Analytics*, Sydney Australia, December 2013.