A Methodological Framework for the Exploratory Analysis of Multimodal Features in Learning Activities

Alejandro Andrade[1] and Marcelo Worsley[2]

[1] Indiana University, Blommington, IN 47405, USA
[2] Northwestern University, Evanston, IL 60208, USA
laandrad@indiana.edu

Abstract. We make a call for the formalization of a methodological framework that allows researchers to perform reliable and valid multimodal data analysis. We believe a first step is to guide the collection and interpretation of multimodal data. What we call the MMLA Exploratory Framework suggests that data collection needs to take place at least at two time points to account for within-subject variation and learning gains. We briefly describe the data set we are currently using to illustrate the application of this exploratory framework.

Keywords: Cognitive Disequilibrium; Complex Systems Concepts Learning; Methodology; Multimodal Learning Analytics.

1 Introduction

Collecting multimodal data from student learning activities is becoming ubiquitous in learning sciences research. Multimodal features—such as body position, facial expressions, and paralinguistic elements of speech—can be extracted from sensors and audiovisual recordings. Some of the objectives for capturing these multimodal features are (a) to better understand the learning process, (b) to develop fine-grained metrics and assessment of student learning, and (c) to improve learning experiences by providing feedback and support for pedagogical decision-making. In order to use multimodal features for these purposes, however, researchers have to carefully find and validate how such features interact with student learning. Thus, we identify two general challenges for the collection of reliable and valid multimodal learning features: (a) the collection (which includes gathering, integration, analysis, and visualization) of multimodal data, and (b) the validation of multimodal features during learning activities

Our goal in this paper is to make a call for the formalization of a methodological framework that allows researchers to identify reliable and valid multimodal data sets intended for any of the learning analytics goals listed above. As is typical in nascent research, we need to start from an exploratory analysis. Thus, we believe our first step is to develop what we call an MMLA Exploratory framework. The MMLA exploratory framework places multimodal features and learning indicators along two dimensions (see Figure 1). Exploratory MMLA examines the changes in natural groupings within students' multimodal features at various stages of the learning activity (dimension 1), and then correlates these changes with shifts in student understanding levels (dimension 2). We propose that each participants' multimodal features and understanding levels have to be measured at least at two time points to account for the within-subject variation and learning gains (measured as change in student understanding). In order to see whether such variation is meaningful in the learning context, a correlation between variation in multimodal features and variation in student learning gains has to be stablished. Our working hypothesis is that such correlation represents the interplay between multimodal features and learning. In finding evidence that such an association exists, one can then devise ways in which to use MMLA to support learning.

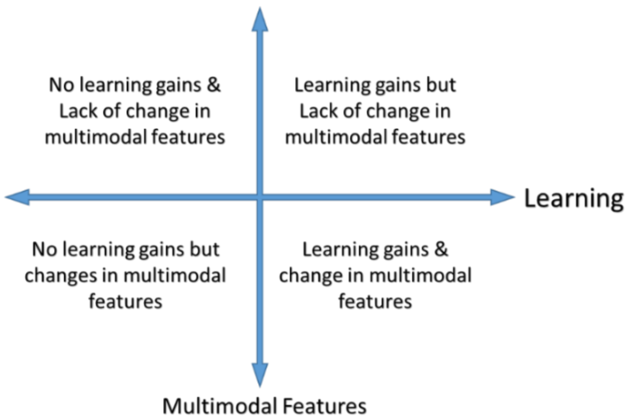

Fig. 1. MMLA Exploratory Framework examines the correlation between learning gains and changes in multimodal features.

Our current work includes illustrating the MMLA Exploratory Framework with affect data from cognitive interviews with elementary students. Specifically, we suggest an exploratory analysis of affective states, through facial expressions and speech prosodics, to understand how affect interacts with learning gains. We believe that multimodal features provide a vantage point to uncover students' affective states, experienced during a learning activity in interaction with a tool or with another person [4]. While using these learning tools, students experience some affective responses, and it has been suggested that these affective responses are correlated with student learning [2]. We collect affective indicators from two types of indicators: (a) we capture students' facial action units that reveal affective states from facial expressions, and (b) extract MFCC features that represent students' speech prosodic elements (e.g., hesitation, confidence). We correlate these affective measures with student learning gains.

2 Description of the Data Set

2.1 Learning Content

The learning content is complex systems concepts, dynamic equilibrium in interacting feedback loops in particular. feedback loops are a key concept to reason about interactions among organisms in an ecosystem [3]. Hokayem, Ma and Jin [3] describe a learning progression for feedback loops at the elementary school level (see Figure 2). In this progression, students move from an incipient understanding of one-way simple causality (see Figure 2.a) to a two-way simple causality (see Figure 2.b) to a two-way cyclical relationship that demonstrates dynamic change in both populations (see Figure 2.c).

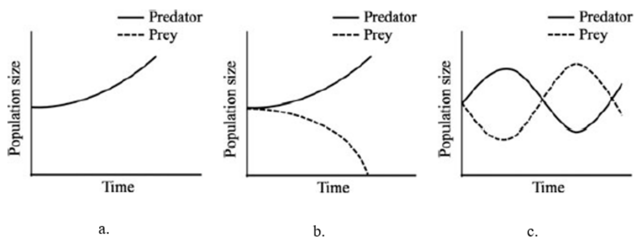

Fig. 2. Hokayem et al. [2] learning progression of feedback loop reasoning.

2.2 Participants

Participants were fifteen 3rd and 4th elementary students from a Midwest school in the United States.

2.3 Learning Activity

The activity is a cognitive interview, where a set of two questions where designed to elicit a cognitive disequilibrium. Cognitive disequilibrium is assumed to be observable in a sudden change of a student's affective state [2].

2.4 Sources of Data

Audiovisual recordings of the set of questions collected, which represent about 1 min each of the total 20 interview (see Excerpt 1 for an example). Transcripts of the conversations were created, where a student is responsible of approx. 40% of the uttered words in average (of about 300 total words per interview section). We believe the repetition of words given the similar structure of the questions and answers can make these data suitable for a cleaner analysis of the prosodic element in the students' speech. In observing the video data, changes in students' facial expressions are often observable right after the student hears the second question.

Excerpt 1

[00:03:10] **Interviewer:** So, if the number of wolves goes up, what would happen to the number of sheep?
[00:03:21] **Interviewee:** (raises red ball) the wolves would go up and (lowers yellow ball) the sheep would go down.
[00:03:23] **Interviewer:** Why?
[00:03:24] **Interviewee:** Because there would be lots of sheep for them to eat.
[00:03:46] **Interviewer:** Great. And if the number of sheep goes up?
[00:03:51] **Interviewee:** If the number of sheep goes up? Then the number of wolves would also go up (raises red ball in line with yellow) because there would be more sheep for them to eat.
[00:04:10] **Interviewer:** That was great, let's move on to the next scenario.

2.5 Measures and Multimodal Features

Student understanding was measured using a coding scheme for the analysis of students' verbal explanations of feedback loops [3]. Facial expressions were extracted as facial action units from the video. Prosodic elements, were extracted as MFCC features from the audio file.

3 Challenges thus Far

A first challenge has been the extraction of Facial Action Units. We are using the OpenFace software [1] to extract facial action units (see Figure 3). The software has some limitations such as faces have to be minimum 100 pixels long and mostly frontal. There are also some obstructions to the face when the students move their hands to answer the questions.

A second challenge is the synchronization of transcript and audio files. We are using the P2FA software, which is a Python extension that makes use of the HTK speech recognition tool kit, for aligning the words in the transcript and the audio wave file. This operation requires a clean transcription without annotations or commentaries, only the uttered words.

A third challenge is the abstract representation of student behavior from multimodal features. After the features are extracted, a statistical approach is required to model the students' behaviors. These statistical models have to account for temporal dependencies in the data. When using latent models or cluster analysis, as is common during exploratory analyses, one needs to interpret the meaning of these natural groupings post facto.

AU #	FACS Name
5	Upper Lid Raiser
15	Lip Corner Depressor
17	Chin Raiser
23	Lip Tightener
25	Lips Part

Fig. 3. OpenFace facial action units annotated video output.

4 Conclusions

MMLA analysis faces several challenges along the way. Two major challenges have been identified: (a) Data pre-processing such as synchronization and identification of features to be extracted; and (b) How to statistically represent student behavior from fine-grained multimodal features and how to account for temporally dependent data. We argue that identifying and explaining the usefulness of multimodal features requires a validation approach. To help this validation process, we suggest an MMLA Exploratory Framework to guide the collection and interpretation of data. The exploratory framework suggests data collection at least at two time points to account for within-subject variation and learning gains.

5 Acknowledgments

This work has been partially funded by an NSF Data Consortium Fellowship.

References

1. Baltrušaitis, T., Mahmoud, M., and Robinson, P., 2015. Cross-dataset learning and person-specific normalisation for automatic Action Unit detection. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on IEEE, 1-6.
2. D'mello, S. and Graesser, A., 2012. Dynamics of affective states during complex learning. Learning and Instruction 22, 2, 145-157. DOI= http://dx.doi.org/https://doi.org/10.1016/j.learninstruc.2011.10.001.
3. Hokayem, H., Ma, J., and Jin, H., 2015. A learning progression for feedback loop reasoning at lower elementary level. Journal of Biological Education 49, 3, 246-260.
4. Worsley, M. and Blikstein, P., 2015. Using learning analytics to study cognitive disequilibrium in a complex learning environment. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK'15 ACM, New York, NY, USA, 426-427. DOI= http://dx.doi.org/10.1145/2723576.2723659.Author, F.: Article title. Journal 2(5), 99–110 (2016).