1. Multimodal Analysis

By Marcelo Worsley

Method Description, Motivation, and Conjectures

Cyberlearning researchers are using innovative methods to investigate how people interact with learning designs and to measure their impact. Learning scientists rely on multimodal analysis — analysis of visual, audio, gestural, movement, and other data sources — to draw inferences about what and how students are learning. For example, using videos, researchers have looked at facial expressions, gestures, tone of voice, and discourse to understand how people learn (e.g., Barron, Pea, & Engle, 2013; Koschmann, Stahl, & Zemel, 2007). Historically, video analysis has provided a means to get a rich picture into complex human-to-human and human-to-technology interactions in learning environments. However, video analysis has often been painstakingly slow, and video does not capture everything that happens in a learning environment. New data streams (such as clickstream log data from playing a game, movement data from sensors, audio data from microphones, and visual data from cameras) and new computational tools for analysis are now transforming how researchers measure and evaluate the impact of cyberlearning projects on learners.

- **Computer science innovation:** Combining multiple forms of data and applying machine learning algorithms and probabilistic models to make sense of how people interact with technology
- **Learning science innovation:** Using the streams of data from different devices to find new patterns in how people learn in complex environments

Using these new data streams is important because the data can enable researchers to ask and answer new questions about learning. Sensors can measure stress or arousal, facial expression, eye gaze, heart rate, and many other things beyond what can otherwise be observed on a video. One example of such a sensor is electrodermal activation, commonly referred to as skin conductance. This sensor detects the increased perspiration that the body exhibits when a person is stressed, surprised, or under significant cognitive load.

This increase in perspiration is frequently hard to perceive for observers except in instances of extreme nervousness or arousal, and it is virtually impossible to quantify through human observation. Measuring this affective response along with indicators of cognitive activity can help researchers to understand how affect and cognition are related while students learning. Important new questions that can be asked and answered include: When a student learns about physics or another topic in an augmented reality experience, is some amount of tension (or stress) good? How much tension is too much? When students are stressed, how does what they look at in a learning environment change?

Examples of Projects Using Multimodal Analysis

In one of many cyberlearning projects that analyze multimodal data, James Lester is studying the relationship between student emotions (affect) and student learning. The project, **Adapting to Affect in Multimodal Dialogue-Rich Interaction with Middle School Students** (NSF #1409639) gathers data on facial expression, posture, gaze, speech, heart rate, skin conductance, and actions. Research involves developing a model of changing affect as students participate in the **Crystal Island** virtual environment, where students play the role of a medical field detective investigating a mysterious infectious disease outbreak. The multimodal information will help researchers capture each learner's experience and also understand elements of the environment

Crystal Island, a game-based learning environment for middle grade science and literacy, raises challenging issues of how emotional responses — such as feelings of calm or tension — are involved in learning.

Used with permission of James Lester.

that may elicit unexpected or surprising responses from learners. Furthermore, the analysis is helping the researchers improve the environment to better integration emotions and learning.

Other researchers are using multimodal analysis to study collaborative learning experiences that do not involve computers. Schneider and colleagues (2015, 2016) used mobile eye trackers to study students as they collaborated with 2-D and 3-D objects. The mobile eye trackers and video recordings allowed them to examine when learners were paying attention to each other and to the objects. They found this approach could accurately predict student performance and learning gains. Worsley, Scherer, Morency, and Blikstein (2015) used multimodal data to study collaborative learning in an engineering design context. Students worked with everyday materials to complete a design challenge. The researchers had hand/wrist movement, electrodermal activation and speech, head pose, and facial expression data. To analyze this data, they developed an automated process to break the stream of data into meaningful segments and then analyze learning with the segments. Ultimately, they found that the automated method outperformed traditional methods from education research and from computer science. The automated approach resulted in better models of the quality of students' engineering designs and how much they learned compared with more traditional approaches that relied on people to make observations and perform analyses.

Finally, a number of researchers are examining opportunities to use multimodal analysis to investigate embodied learning. In the **ELASTIC3S** (NSF #1441563) and **Developing Crosscutting Concepts in STEM with**

ELASTIC3S explores ways that body movement can be used to enhance learning of big ideas in science.

Used with permission of Robb Lindgren.

Simulation and Embodied Learning (NSF #1441563) projects, Robb Lindgren is using multimodal sensors to study how embodied learning can support students as they learn STEM concepts: How do students use gestures in learning science? Dor Abrahamson in **Collaborative Research: Gesture Enhancement of Virtual Agent Mathematics Tutors** (NSF #1321042) is developing a gesture-based virtual agent to support students as they learn about fractions. The virtual agent (see also the Virtual Peers and Coaches section) is being developed by analyzing teachers' and learners' gestures. The gestures being investigated go beyond simple hand movements by incorporating facial expression and posture. Going back to early scholars like George Herbert Mead (who died in 1931), social scientists have known that gesture is important to the process of learning. New multimodal analysis methods are now enabling scientists to analyze more data, more systematically to understand exactly how gesture plays into learning.

Contributions, Opportunities, and Challenges

Multimodal analysis is likely to advance both the learning sciences and computational sciences. In the learning sciences, reliable low-cost sensors enable deeper investigation of how people learn in complex environments. When coupled with machine learning, multimodal analysis can give researchers novel insights into the efficacy of a given intervention or the emergence of different patterns in their data. In computer science, researchers are developing algorithms, models, and visualization techniques that push the boundaries on how analysts engage with data. There is important work to do to achieve an analysis workflow that is sound, efficient, and integrates human insight with computational power.

Key challenges to conducting this kind of research (Blikstein & Worsley, 2016; Worsley, 2012) span every step of the workflow from data collection (collecting reliable synchronized data from several data streams) to data analysis (determining the appropriate tools and analytic techniques for processing and integrating the streams). Although the vision of using big data can make it seem easy to gain insights rapidly, the reality is that working with these new data sources and analysis techniques remains very challenging. To address these challenges, the **Catalyzing Research in Multimodal Learning Analytics** project (NSF #1548254) has been organizing workshops that introduce participants to some of the capabilities of multimodal analysis and provide them with preliminary resources for using the techniques. Additionally, the principal investigator of this project, Marcelo Worsley, is developing the **Multimodal Data Capture and Analysis Tool,** which will significantly streamline the process for collecting high-quality data and employing that data as evidence for learning.

The workshops have identified three key directions for research:

1. Making it easier or faster to incorporate new or additional data streams into existing research approaches.
2. Identifying new research questions that can be asked and answered only with new data streams.
3. Using multimodal tools to develop new learning experiences and new ways of engaging with learning data.

Many of the new research questions relate to learners' socio-emotional experiences — and the new streams are important because they can more directly collect information about emotional responses and the changes in gaze, gesture, and posture that show how people are relating socially. Society has already started to see the learning possibilities that emerge when a speech-activated device is added to a home or school. Multimodal tools can lead to learning experiences that adapt more fully to various learners' abilities, emotions, needs, and preferences.

Resources

Crystal Island: http://projects.intellimedia.ncsu.edu/crystalisland/

ELASTIC3S: Embodied Learning Augmented through Simulation Theaters for Interacting with Cross-Cutting Concepts in Science:
http://stemforall2017.videohall.com/presentations/973