**INTRODUCTION**

Collaborative learning methods have been implemented broadly by organizations at all stages, as research recommends that active human involvement in cohesive and micro group communications is critical for effective learning . In current research, an important line of inquiry focuses on finding accurate evidence and valid assessment of these micro-level interactions which supports collaborative learning. Even though there is along practice of using mathematical models for modeling human behavior, Cipresso (2015) introduced a computational psychometrics-based method for modeling characteristics of real behavior. Cipresso’s article provides us with a way to extract dynamic interaction features from multimodal data for modeling and analyzing actual situations. In this paper, we propose a three-stage method to explore and study collaborative group behaviors. The first stage integrates and processes multimodal data obtained in a collaborative learning environment (CLE) that includes sensor input, audio,video, eye tracking, facial expressions, movement, posture, gestures, and behavioral interaction log data. The second stage performs feature extraction and cloud computation using computational psychometrics (CP) and convolutional neural network (CNN)-based deep learning for skill, pattern, and trend identification. Finally, the third stage uses the parameters measured in the previous two stages to understand and model group interactions, competencies, and collaborative behavior at a micro-level. The third stage uses machine learning for effective assessment and visualization of group dynamics such as Correctly assessing the increase in the groups’ level of shared understanding of different perspectives, and ability to clarify misconceptions.

In the past few years, the Artificial Intelligence (AI) and Machine Learning (ML) communities have been putting their efforts into presenting and publishing advanced methods for processing and analyzing human behavior related multi-modal data. Due to page limitations, it is not possible to cover and cite all of these works, but we will provide brief highlights regarding our own work.

In our most recent work, Chopade et al, presented a framework which incorporates computational psychometrics (CP), Artificial Intelligence (AI), and a Machine Learning (ML)- based system architecture, methodology, and related algorithms to find patterns of interactions, learning, team relationships, and effective teamwork assessment of collaborative problem solving (CPS) and a collaborative learning environment (CLE). Khan presented an approach which uses multimodal telemetry data for two pilot studies from the domains of collaborative learning and illustrated a framework to analyze participant behavior patterns through temporal dynamics. Polyak et al.presented the application of CP for the measurement of CPS skills. They performed machine learning analysis on actual behavioral and post-game studies data. Their CPS game tasks were designed and developed based on the psychometric principles of Evidence-Centered Design (ECD) which are associated with ACT’s Holistic Framework (HF). In their experiment, they performed a cluster analysis on participants’ sub-skill performance scores and configurations of particular dialog responses obtained from participants’ gameplay data.

In this paper, we extend the foundation laid out in our recent work to implement CP, AI, and ML for effective assessment of teamwork skills. The next sections discuss in detail, the three stage architecture for data-intensive computing and efficient assessment. Massive data-intensive, high-performance, scalable computing is transforming our capabilities to gather and analyze data in different forms. This may lead to new inventions and discoveries in education, science, and technology. This may also impact learning and assessment (LAS) platforms. Data-intensive computing changes our thinking about education, science, and technology, by accelerating an ability to perform advanced data collection and computing. Data-intensive scalable computing has a high potential for unique applications. This will be more challenging when we need to scale up the platform to handle large-scale datasets (Terabyte, Petabyte, Zettabyte scale). Recent improvements in computing have led to substantial progress towards the visualization capabilities of such data. Data analytics and visualization will serve as a vital tool for the validation of expected results by accurately identifying patterns and relationships in data. Visualization may play an essential role in understanding the big picture in group interactions within the CLE and may assist in detecting hidden factors. Convolutional Neural Network (CNN) and one of its approaches – deep learning (DL) may require the use of a highly efficient Graphical Processing Unit (GPU) implementation or for training on multiple GPUs or for applications of this architecture to substantial learning populations.

1) Data Integration and Processing

Establishing identities from vast volumes of CLE interaction multimodal data obtained from different sources is an essential task of data analytics and computation architectures. Large amounts of CLE multimodal interaction data that provide the identities of humans, machines, sensors, etc. collected from different sources will be processed through a set of solutions built upon the Hadoop data analytics platform. This arrangement considers the individual’s identity, such as username, email, real name, gender, eye color, fingerprints and user’s input data such as eye tracking and models of behavior. A data cluster would enable massive data integration, processing, and performing of such tasks as data collection, information extraction, and storing large-size distributed datasets for long-term access. Data collection is an essential phase of acquiring data from multiple sources, categorizing it, and passing it to the next stage in the process Data on humans, machines, and other entities can be categorized as structured or unstructured and incorporated into a distributed Hadoop infrastructure.

2) Massive data intensive CNN (deep learning) based cloud

Computing and Computational Psychometrics (CP) Once data has been categorized (as described in part ‘A’), we can use a computation cluster to analyze the data on a cloud platform to understand individual and group abilities. We plan to use Python/R to run deep learning in the cloud using ACT’s enterprise learning analytics platform (LEAP). We plan to deploy a feature extraction algorithm including CNN based deep learning for skill, pattern, trend identification, and for achieving state-of-the-art accuracy in feature classification. We plan to update this network structure over time to make this a dynamic system. Cloud computing moves computation closer to the data. The main advantage to this process is that this approach is scalable to hundreds of computing nodes, each providing at least a modest performance. Data-intensive cloud computing platform consists of 3 layers, i.e., map/reduce on top of Hadoop, HPC (high-performance computing) infrastructure for massive data processing and CNN deep learning-based cloud computing. For HPC, we use a method for dynamic partitioning of processes.

Network updater adds new network specific data entries under the situation of any real-time events such as changes in the team and their activities. Computational Psychometrics (CP): Collaborative problem solving (CPS) is identified as cross-cutting capabilities which is part of ACT’s Holistic Framework- a comprehensive description of the knowledge and 21st-century skills individuals need to know and be able to succeed at school and work. Advanced development in computational techniques and analytical tools has produced new pathways in CPS research. Simultaneously, psychometrics researchers started developing assessments using advanced computational techniques and analytical tools which have emerged as a novel interdisciplinary field of prominent research called, “Computational Psychometrics (CP)”. CP is a new area of learning and assessment (LAS) research, which consist of data-driven machine learning and information querying computer science methods, theory-driven psychometrics, and stochastic theory – all used in order to measure learner’s latent abilities in real times.