MUTLA: A Large-Scale Dataset for Multimodal Teaching and Learning Analytics

Fangli Xu[1*], **Lingfei Wu**[2], **KP Thai**[1], **Wei Wang**[1], **Richard Tong**[1]
[1]Squirrel AI Learning by Yixue Education Inc., New Jersey
[2]IBM T.J. Waston Research Center, Yorktown Heights, New York
{lili, kp.thai, wei, richard}@yixue.us, lwu@email.wm.edu

Abstract

Automatic analysis of teacher and student interactions is crucial for improving quality of teaching and the engagement of student learning. Despite some recent progress of utilizing multimodal data for teaching and learning analytics, there is little to no effort to collect and gather a rich set of multimodal data in the complex real learning environment. To bridge this gap, in this paper we present MUTLA, a large-scale dataset for MUltimodal Teaching and Learning Analytics. This dataset includes time-synchronized multimodal data records(learning logs, videos, EEG brainwaves)of students as they work on Squirrel AI Learning System(SAIL) to solve problems that vary in subjects and difficulty levels. The dataset resources include user records from learner records store of SAIL, brainwaves collected by EEG headset devices, and video data captured by web camera while student working on SAIL. The primary aim is to analyze student learning activity, facial expression, body movement, and brainwave pattern that can be used as a real-time proxy of engagement, which will allow researcher to either predict student learning outcomes or learner behavior, or make system adaptivity and so on. An additional goal is to provide a dataset gathered from the real-world educational activities versus those from the controlled lab environment to benefit educational learning community.

1 Introduction

Recent huge advancements of Artificial Intelligence with Deep Learning [LeCun *et al.*, 2015] have rendered fast growing interests in applying advanced machine learning/deep learning techniques in education domain. In particular, these challenges and opportunities are particularly abundant in multimodal teaching and learning analytics (MUTLA), a research methodology that aims to bring together Educational Data Mining and Learning Analytics to multimodal learning environments by directly working on data from multimodalities [Worsley and Martinez-Maldonado, 2018; Worsley *et al.*, 2016]. Over last ten years, researchers have exploited to apply machine learning/ data mining techniques on multimodal data for various tasks, including communicative interaction [MacWhinney, 2004], online education [Thomas, 2018], student's uncertainty modeling [Jraidi and Frasson, 2013], and emotional responses recognition in Children [Nojavanasghari *et al.*, 2016].

Despite some recent progress of collecting multimodal data and utilizing them in learning science [MacWhinney, 2004; Antoniadou, 2017; Oviatt *et al.*, 2013; Nojavanasghari *et al.*, 2016], the ability of performing teaching and learning analytics is largely limited by the quality and quantity of multimodal data that is publicly accessible. In this paper, we bridge this gap between enthusiastic AI researchers and challenging teaching and learning analtyics problems, by presenting MUTLA, a large-scale multimodal dataset for MUltimodal Teaching and Learning Analytics.

Our dataset includes time-synchronized multimodal data recordings(learning logs, videos, EEG brainwaves) on collaborating groups of students as they work on Squirrel AI learning system to solve problems that vary in subjects and difficulty levels. The dataset resources include user records from learner records store of Squirrel AI learning system, brainwave data collected by a EEG headset device, and video data captured by web camera while student working on the learning system. The primary aim is to analyze student learning activity, facial expression, body movement, and brainwave pattern that can be used as a real-time proxy of engagement, which will allow researcher to either predict student learning outcomes or learner behavior, or make system adaptivity and so on. An additional goal is to provide a dataset gathered from the real-world educational activities versus those from the controlled lab environment to benefit educational learning community. Our dataset can be publicly accessed via the following link. [1]

*Corresponding Author

[1]https://yixueus-my.sharepoint.com/:f:/g/personal/kp_thai_yixueus_onmicrosoft_com/Eg94llMFSPJBivTjaGr7GIUBU0tgiycJ6RalWXfIZ9MNZw?e=n0xNng

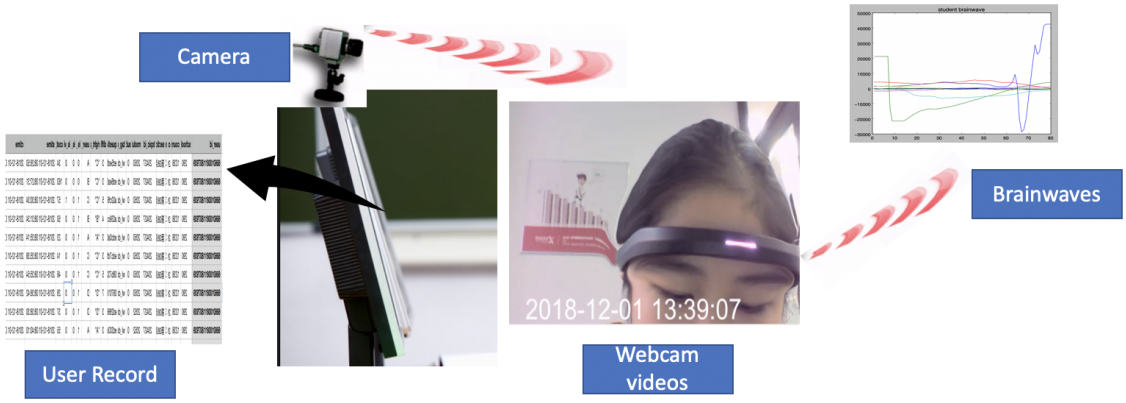

Figure 1: Integration of multimodal data capture using camera and EEG headband. During the problem solving process, we collect student's learning record, EEG brainwave signals, and a set of video data.

2 Method

2.1 Participants

For this dataset, participants include students in 2 offline schools of Yixue Education in China, 34 person time from school G and 233 person time from school N involved. Students worked on 6 Subjects in total. For Chinese and English, students come from primary school and middle school. For Math, Physics Chemistry and English reading students only come from middle school.

2.2 Tasks and data collection procedure

Students come to Yixue offline schools to do after school session. In school G after school tutoring session is approximately 1.5 hours long and in school N is 2 hours long. In each session, students focus on a particular set of knowledge points (tag_code) for a specific course/subject. They typically start with a pretest or review, then view instructional videos and work on learning and practice questions, until the knowledge points have been "mastered". Pretest questions do not have immediate corrective feedback, but students can view the answer and the explanations of the answer after they completed the pretest. The learning and practice questions contain hints and corrective feedback. Students can activate different levels of hints and view the answer explanation for each question, before or after submitting the response.

After each learning and practice question, SAIL computes and updates each student's proficiency level on each knowledge point to determine the next question to present. When the proficiency estimates reach a certain threshold, that knowledge point is considered "learned" and will be queued for review at a later time.

During each after school learning session, while students were using the system to study they were asked to wear a headband(headbands are from BrainCo company) and to have the webcam camera turned on. Thus, there are three main data sources collected for each students. The user records are generated by the SAIL system and can be fetched from SAIL learner record store. Students brainwave data while working on the after school session are collected by the headband then using FocusEDU platform from Brainco to store. The video data captured by the webcam installed on each computer stored by the Debut software. The whole procedure and data source flows are shown in figure 1.

User record

The user records are collected from Squirrel AI learning's learner record store. They contain all the question level logs of student responses while student working on the after school tutoring sessions for particular subject. Each item is a question. Students answering questions through out the whole session except watching videos for particular knowledge component.

Brainwaves

For students, while they are working on each learning session, the headset they wearing will generate three data files: attention, EEG and events.

The attention data contains a Unix timestamp followed by its attention value. The timestamp needs to be converted to Beijing time to synchronized to other data source.

The EEG file contains the raw EEG data. Each row has the timestamp, sequence Num, battery level, logging lable and EEG array. Each point represents the difference in potential between the EEG reference point and the acquisition point. There are 160 such points in a minute for uA. The vector in square brackets [ ] are the electrical signal output values of the sensors. These electrical signals can be transformed into frequency domain signals or waveforms (alpha, beta, gamma, etc.) by Fourier transform, and the average energy of each wave band can be calculated.

The event file contains the raw events data. Each data point has a Unix timestamp followed by the device stage. It indicates whether the device is connected or not in the corresponding time.

Webcam videos

While students working on the after school session on SAIL system. The teacher inside the classroom will set up the computer and turn on the camera and the timestamp on the computer. In each video, students upper body include face was

Algorithm 1 Syncing Brainwave with User Record

Input: Table of all User record , Table of brainwave data
Parameter:user_id, subject, end_time_with_date, ctime, class_start_time, class_end_time, brainwave_file_path
Output: brainwave triplet file paths

```
while user_id = 52027 and subject = "E" in table Brainwave do
    find the end_time_with_date and class_start_time
    while user_id = 52027 in English User record do
        find ctime
        if ctime is between class_start_time and end_time_with_date then
            assign the brainwave_file_path to the item which user_id = 52027 in User record
        end if
    end while
end while
return User record with brainwave file path appended
```

recorded by the camera. After the end of each session, the camera will turned off.

2.3 Multimodal data synchronization and processing

User records

Since we record students facial and body movement by the camera and the brainwave captured by the headset while they are using the SAIL system working on the after school session, we sync these three data sources together by time.

Brainwave data and synchronization

There are two steps in brainwave data synchronization. The first step is to find the right students who own the brainwave. The syncing schema is using student user id, session time, headset number to match the brainwave data to each student. The second step is to accomplish synchronization between brainwave and user record. The algorithm for syncing these two data source is shown in Algorithm 1. In this algorithm, ctime represents the answer submission time of corresponding question. The main task is to find the right student brainwave triplets while the student working on the question.

Video processing and synchronization and segmentation

Webcam video data has the same fields as the brainwave data. We used the same algorithm to sync webcam video stream to user record. After the synchronization, each question item in User record has the original video file path.

After we sync the video with user record, the data were segmented from each session into time phases representing the start and end of each problem. Each problem in User record has a stime which means the time to get the question and a ctime which equals to the submitted time. Based on answer submission time, the total time required by the student to solve the question was summarized(ctime-stime). We extract the video pieces according to the time window of the question to be fetched to the time of the answer was submitted.

The extracted question segments have two files associated with each segment: a json file with metadata of the question and an npy (python Numpy) file containing the tracking metadata which should be very easy to load into Python.

1. Question segments filenames name as "school name_video ids_segment numbers.

2. Json file with metadata of the question can be retrieved from user records.

3. File named "npy_key.md" describing the meanings of the various data point included in the numpy file.

After the webcam videos lined up with user record, the total video length of different Subjects are shown in table 1.

Subject	total time(hours)
Chemistry	10
Chinese	23.18
English	189.11
Math	248.92
Physics	155

Table 1: Total time for different subjects

Percentage of valid segments	Percentage(2170 segments)
At least 50%	61.54%
At least 70%	49.88%
100%	18.30%

Table 2: valid tracking about segments

After filtering, it contains 2170 segments. Table 2 showed the percentage of valid tracking of the segments for different percentile. We define that fully visible student face as valid tracking. At least 50% means at least half of the frames have valid tracking information.

3 Conclusion and Future Work

In this paper, we present MUTLA, a large-scale multimodal dataset for enabling researchers to study teaching and learning analtyics to improve the effectiveness of teaching and the level of engagements of student learning. Our dataset is collected in the real complex learning environment on Squirrel AI learning system. The dataset includes time-synchronized multimodal data recordings(learning logs, videos, EEG brainwaves) on various groups of students when they work on Squirrel AI learning system to solve problems that vary in subjects and difficulty levels. Our main aim is to promote the research of topics using MUTLA in educational learning science community and beyond.

Some future works are summarized as follows: i) use advance machine learning and deep learning to perform a first study on this multimodal data; ii) define and solve more interesting problems using currrent multimodal dataset.

4 Acknowledgement

The data described were collected as part of the MIBA system of Yixue Education. Employees from Squirrel AI learning by Yixue Education are the major help of this dataset. Students and Teachers in two offline school campus of the Company set up the equipment. Funding by Yixue Education and supported this research. Preparation of this dataset for the workshop were also supported by SRI international.

References

[Antoniadou, 2017] Victoria Antoniadou. Collecting, organizing and analyzing multimodal data sets: The contributions of caqdas. *Research-publishing. net*, 2017.

[Jraidi and Frasson, 2013] Imène Jraidi and Claude Frasson. Student's uncertainty modeling through a multimodal sensor-based approach. *Journal of Educational Technology & Society*, 16(1):219–230, 2013.

[LeCun *et al.*, 2015] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[MacWhinney, 2004] Brian MacWhinney. Talkbank: Building an open unified multimodal database of communicative interaction. 2004.

[Nojavanasghari *et al.*, 2016] Behnaz Nojavanasghari, Tadas Baltrušaitis, Charles E Hughes, and Louis-Philippe Morency. Emoreact: a multimodal approach and dataset for recognizing emotional responses in children. In *Proceedings of the 18th acm international conference on multimodal interaction*, pages 137–144. ACM, 2016.

[Oviatt *et al.*, 2013] Sharon Oviatt, Adrienne Cohen, and Nadir Weibel. Multimodal learning analytics: description of math data corpus for icmi grand challenge workshop. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 563–568. ACM, 2013.

[Thomas, 2018] Chinchu Thomas. Multimodal teaching and learning analytics for classroom and online educational settings. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pages 542–545. ACM, 2018.

[Worsley and Martinez-Maldonado, 2018] Marcelo Worsley and R Martinez-Maldonado. Multimodal learning analytics' past, present, and potential futures. In *CrossMMLA@ LAK*, 2018.

[Worsley *et al.*, 2016] Marcelo Worsley, Dor Abrahamson, Paulo Blikstein, Shuchi Grover, Bertrand Schneider, and Mike Tissenbaum. Situating multimodal learning analytics. In *12th International Conference of the Learning Sciences: Transforming Learning, Empowering Learners, ICLS 2016*, pages 1346–1349. International Society of the Learning Sciences (ISLS), 2016.