The Big Five: Addressing Recurrent Multimodal Learning Data Challenges

Daniele Di Mitri
Open University of The Netherlands
daniele.dimitri@ou.nl

Jan Schneider
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
schneider.jan@dipf.de

Marcus Specht
Open University of The Netherlands
marcus.specht@ou.nl

Hendrik Drachsler
Open University of The Netherlands – DIPF
hendrik.drachsler@ou.nl

ABSTRACT: The analysis of multimodal data in learning is a growing field of research, which has led to the development of different analytics solutions. However, there is no standardised approach to handle multimodal data. In this paper, we describe and outline a solution for five recurrent challenges in the analysis of multimodal data: the data collection, storing, annotation, processing and exploitation. For each of these challenges, we envision possible solutions. The prototypes for some of the proposed solutions will be discussed during the Multimodal Challenge of the fourth Learning Analytics & Knowledge Hackathon, a two-day hands-on workshop in which the authors will open up the prototypes for trials, validation and feedback.

Keywords: multimodal learning analytics, wearables, CrossMMLA, sensor-based learning

1 BACKGROUND

The Learning Analytics & Knowledge (LAK) community has acknowledged the necessity of taking into account physical and co-located learning activities as much as practice-based skills training; it is undeniable that in the classroom and at the workplace these "offline moments" still represent the bulkiest set of learning activities. Bringing these moments into account requires extending the data collection to additional data sources which go beyond the conventional ones, such as online learning systems, Massive Online Open Courses (MOOCs) platforms or student information systems. With the term *multimodal data*, we refer to the learning data sources collected "beyond user-computer interaction", i.e. those data sources collected during learning moments alternative to the classic desktop-based learning scenario. Although user-computer interaction data could still hold some relevant information, they can be complemented by additional multimodal data; these data can be classified into 1) data describing the learner's behaviour: including motoric and physiological data; 2) data regarding the learning situation, including social context, learning environment and learning activity. Most of these aspects can be monitored through wearable sensors, cameras or Internet of Things (IoT) devices. These tools can capture only what is "visible" by a generic sensor, meaning they generally do not have the ability to reason on the meaning behind the collected data. The observability line – i.e. what is visible by sensors and what not, conceptually separates multimodal data by human-driven qualitative interpretations, like expert reports or teacher assessments. The latter, that are more qualitative and human-driven, describe dimensions that the sensors cannot directly observe, such as learning outcomes, cognitive aspects or affective states.

Bridging the gap between learner's complex behavioural patterns with learning theories and other unobservable dimensions is the paramount challenge for multimodal analysis of learning (Worsley, 2014). Multimodal data can be used as historical evidence for the analysis and the description of the learning process: this field of research is called *Multimodal Learning Analytics* (Blikstein, 2013). The related literature shows the potential to apply a multimodal approach in a variety of learning settings including dialogic learning in teacher-student discourse (D'mello et al., 2015); computer-supported collaborative learning during knowledge-sharing and group discussions (Martinez-maldonado et al., 2017; Schneider & Blikstein, 2015); in practice-based and open-ended learning tasks, when understanding and executing a practical learning tasks (Ochoa et al., 2013).

The potential benefits of multimodal data are not only limited to analytics, e.g. human interpretation of dashboards or other visual metaphors. If multimodal data are reliable and correctly addressed and exploited, they can be used as the base to drive machine intelligence and achieve better personalisation and adaptation during learning. Multimodal data is expanding the horizon of the Learning Analytics community and its moving towards the intelligent tutoring and the artificial intelligence in education communities. For decades the long-term goal of these communities consisted in designing intelligent computer agents empathic to the learners which work as an *instructor in the box*, and that can implement strategies to reduce the difference between experts and student performance (Polson, Richardson, & Soloway, 1988). Multimodal data can facilitate achieving this goal, by equipping intelligent tutors with action-based recognition and reasoning, so that they can deal with open-ended learning tasks in uncontrolled environments.

2 MULTIMODAL CHALLENGES

The analysis of multimodal data in learning is a fairly new but a steadily growing field of research. As the interest tracing learning through the use of multimodal data grows, the opportunities stemming from it become more evident. As some authors have pointed out, the field of MLA faces a set of open challenges that create research gaps that need to be filled (Blikstein & Worsley, 2016). For instance, the LAK community (and its CrossMMLA interests group) still lacks a standardised approach for modelling of the evidence extracted from the learning process and producing valuable feedback with multimodal data. In contrast, multiple tailored ad-hoc solutions have been developed in related researches. A standardised approach to MMLA, in our understanding, should help researchers in setting-up their multimodal experiments by clarifying how the collection, storage, analysis and exploitation of the multimodal data takes place in a pragmatic and scalable manner that can be adopted into real-life educational settings. To contribute filling this gap, in this paper, we outline five main challenges stemming from the feedback loop empowered by multimodal data and learning analytics. For each of these challenges, we describe possible solutions or approaches. The prototyping, testing and validation of the proposed solutions, coincide with the agenda of the *multimodal challenge* proposed for the Fourth Learning Analytics Hackathon[1]. In these two-days, hands-on, pre-conference event, we will roll-out the first prototypes like the *LearningHub* or the *Visual Inspection Tool*; we will test their usability and validity and we will open them up for discussion with experts in the field.

2.1 Data collection

The first step of the journey is the data collection, that being the creation of datasets through multiple sensors and external data sources. The sensors employed are most likely to be produced by different vendors, hence to have different specifications and support. The approach used for data collection must be flexible and extensible to different sensors, it should allow the collection of data at different frequencies and formats. Strongly connected to the collection is the data synchronisation.

Proposed solution: to address this challenge, we introduce the *LearningHub*, a software prototype whose purpose is to synchronise and fuse different streams of multimodal data generated by the multiple sensor-applications. The *LearningHub*'s main role is to deal with the low-level specifications of every sensor offering a customisable interface to start and stop the capturing of a meaningful part of a learning task, i.e. moments clearly definable by atomic actions; we call this an *Action Recording*. The *LearningHub* is responsible to collect the updates for every sensor, organising and synchronising them chronologically.

2.2 Data storing

The second step is the data storing that encompasses the serialisation, storing and logic for retrieval of the action recordings. This step is crucial to organise the complexity of multimodal data which has multiple formats and big sizes.

Proposed solution: The *LearningHub* channels the data from multiple sensors and provides as output multiple JSON files, which serialise and synchronise the sensor values for each sensor application. The JSON files allow for sensors having multiple attributes with different time frequencies and formats; they work as exchange format documents and provides also the logic to facilitate the action recording for storing and later retrieval.

2.3 Data annotation

The *data annotation* challenge consists in finding a seamless and unobtrusive approach for labelling the learning process, i.e. triangulating the multimodal action recordings with the evidence (e.g. video clips) of the learning activities. The annotation step is rather crucial, as most of the time the meaning of a recording is not trivial to derive just by looking at the sensor values. The format chosen for assigning the semantics to the action recordings is also a relevant issue.

---

[1] LAK Hackathon 2018, Sydney, Australia, March 5-6, 2018, https://lakhackathon.wordpress.com/

Proposed solution: to address this challenge, we propose the *Visual Inspection Tool (VIT)*. The VIT is a web-application prototype for the retrospectively analysis and annotation of multimodal action recordings. The VIT allows to load multimodal datasets, plot them on a common time scale and triangulate them with video recordings of the learning activity. It allows to select a particular timeframe and annotate the multimodal data slice with an Experience API (xAPI) triplet, assigning an actor, a verb and an object. The VIT offers a human-computer interface which helps to deal with the complexity of multimodal data.

2.4 Data processing

The data processing steps consist in extracting and aligning the relevant attributes from the "raw" multimodal data and transforming them into a new data representation suitable for exploitation. The data processing steps depend tightly on the data exploitation which is discussed in next section. Common steps for data processing include data cleaning (e.g. handling missing values, resampling and realigning the time series), feature extraction, dimensionality reduction and normalisation. The challenging side of the data processing for multimodal data is given by the size of the multimodal datasets, the need to process them periodically and the need to process as close to real-time as possible, a relevant aspect especially in the case of immersive feedback generation.

Proposed solution: the idea is to have a Pipeline for multimodal data for learning, a cloud-based application which allows to plan and execute data processing routines (e.g. Spark jobs). These routines should query the Learning Record Store and fetch the all recent/relevant xAPI statements and load into memory all the action recordings connected to each xAPI statement. The raw action recordings will be transformed according to the set of operations specified which will output a transformed action recording which is saved and ready to be fed into a data mining algorithm.

2.5 Data exploitation

Through an analysis of the related experiments in the literature using multimodal data in learning settings, we concluded that there are different *use cases* generally used for enhancing and facilitating the learning process with multimodal data.

Proposed solution: we classify the different use cases into five *exploitation strategies:*

1. *light-weight feedback*: hardcoded rules and feedback based on heuristics of the form "if sensor value is x then y";
2. *replica*: replays of the action recordings, e.g. ghost-tracks of motoric sensors data;
3. *historical reports*: aggregated visualisations in forms of analytics dashboard, a group of data visualisations that show the historical progress of the sensor recordings in condensed form;
4. *frequent patterns*: mining of recurrent sensor values occurrences within one or multiple sensor recordings;
5. *predictions*: estimation of the human annotated labels during similar action recordings.

The strategies can be used for different purposes and applications. They differ in the level of data processing used and consequently by the methods used for data analysis; these include descriptive statistics, supervised or unsupervised machine learning. For example, *light-weight feedback* requires simple hardcoded rules; *historical reports* require visualisations that can be grouped into analytics dashboard; *frequent patterns* or *predictions* require training either machine learning models, store them into memory, and use them to estimate the value or the class of a particular target attribute. Historical reports also differ by the effort required by human experts, for example in collecting the labels or for interpreting the visualisations; in a similar way, the strategies differ by the level of machine reasoning, e.g. between those using machine learning and those which use heuristics.

3 CONCLUSIONS

In this paper, we have introduced five main challenges connected to the use of multimodal data in learning. These challenges deal with the data collection, storing, annotation, processing and exploitation and constitute important research questions for all the CrossMMLA community. Along with these challenges, we briefly explained some practical solutions. Being these ideas preliminary, we use them as agenda points and research questions to the Multimodal Challenge of the LAK Hackathon, a hands-on workshop which will take place during the eight Learning Analytics & Knowledge Conference in Sydney. We hope that pointing out these challenges can raise interest and awareness in the current research endeavours in the area of multimodal learning analytics.

REFERENCES

Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, 102. http://doi.org/10.1145/2460296.2460316

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining : using computational technologies to measure complex learning tasks, *7750*(x), 220–238. http://doi.org/http://dx.doi.org/10.18608/jla.2016.32.11

D'mello, S., Olney, A., Blanchard, N., Sun, X., Ward, B., Samei, B., & Kelly, S. (2015). Multimodal Capture of Teacher-Student Interactions for Automated Dialogic Analysis in Live Classrooms. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 557–566. http://doi.org/10.1145/2818346.2830602

Martinez-maldonado, R., Power, T., Hayes, C., Abdiprano, A., Vo, T., & Shum, S. B. (2017). Analytics Meet Patient Manikins : Challenges in an Authentic Small-Group Healthcare Simulation Classroom, 1–5. http://doi.org/10.1145/3027385.3027401

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. *Proceedings of the 15th ACM International Conference on Multimodal Interaction (ICMI '13)*, 583–590. http://doi.org/10.1145/2522848.2533789

Polson, M. C., Richardson, J. J., & Soloway, E. (1988). *Foundations of intelligent tutoring systems. Foundations of intelligent tutoring systems.* Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1988-97172-000&lang=de&site=ehost-live

Schneider, B., & Blikstein, P. (2015). Unraveling Students' Interaction Around a Tangible Interface using Multimodal Learning Analytics. *JEDM - Journal of Educational Data Mining*, *7*(3), 89–116.

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. *3rd Multimodal Learning Analytics Workshop and Grand Challenges, MLA 2014*, 1–4. http://doi.org/10.1145/2666633.2666634