Towards Novel Researcher Tooling Based on Multimodal Analytics

Luis P. Prieto[1] and María Jesús Rodríguez-Triana[1],[2]

[1] Tallinn University, 10120 Tallinn (Estonia)
[2] École Polytechnique Fédérale de Lausanne (Switzerland)
lprisan@tlu.ee, mjrt@tlu.ee

Abstract. Much of educational research today employs a complex mixture of qualitative and quantitative analyses, both during the exploratory and confirmatory phases. However, researchers are still stuck with tools that were developed mainly for single-perspective research. The problem is even more acute in emergent areas with complex, high-frequency data, such as multimodal learning analytics (MMLA). In this position paper we posit that a new generation of researcher tools are needed to account for this new complexity of research processes. We also anticipate that such new wave of tools should leverage recent advances in computing technology, while keeping humans in the loop. The paper presents a proof-of-concept ongoing study focusing on one of the main "points of friction" in social sciences research: the manual coding of audio/video. The results of this study, to be presented in the workshop, will illustrate some of the advantages and unsolved challenges in the development of computationally-enhanced researcher tools that can lead to MMLA solutions usable in the real world.

Keywords: research tools, multimodal learning analytics, video coding, content analysis, automation

1 Researcher Tooling in Modern Educational Research

Most of the research in the learning sciences and other adjacent fields (e.g., learning analytics) employs a complex mixture of qualitative and quantitative analyses, both during initial explorations of the data, and during confirmatory or inferential phases of the research [16]. For instance, in a recent study, Dornfeld and collegues used discourse analysis, Markov and topic modeling, along with nonparametric statistical tests, to study both the outcomes and the process of a computer-supported collaborative learning activity [6]. This mixing of analyses is often compounded by the fact that iterative research methodologies (such as design-based research) are becoming increasingly commonplace in educational technology research [10]. The complexity of data pre-processing and analysis is even more acute in recently-emerged areas with multi-source, high-frequency data, such as multimodal learning analytics (MMLA). This has lead many learning analytics researchers to characterize their work and datasets as "messy" [4].

Despite this increasing complexity, as researchers we seem to be stuck with tools quite similar to those we were using 10 years ago, designed mainly for single-perspective, single-data-source research. This includes computer-assisted qualitative data analysis (CAQDAS) tools like NVivo[3] and ATLAS.ti[4] for qualitative analyses, or statistical packages (e.g., SPSS[5], R[6]) for quantitative analyses. Most of the recent developments in these tools seem to be focused in basic usability enhancements, with only timid attempts at automating the most time-consuming tasks, e.g., auto-coding based on pieces of text from transcribed dialogue (which has already provoked some concerns and discussion in the qualitative research community [5, 15]). Independently, MMLA researchers themselves have started experimenting with the automation of certain steps in their analyses, using the latest advances in computing technology (e.g., the use of cloud-based automated speech-to-text recognition to detect questions in [2]).

In this position paper, we contend that a new generation of researcher tools is needed, if we are to make sense of increasingly complex and heterogeneous data in our iterative research processes. We also posit that, while such new tools should take advantage of the latest advances in computing technology, they should also keep the researcher "in the loop" (in contrast with a fully automated approach). Below, we look at one of the most common and time-consuming processing tasks we often encounter in MMLA: the coding of an audiovisual recording. The following sections outline a proof-of-concept approach to support researchers in such video coding, and an ongoing exploratory case study set in an authentic, current research effort. We end the paper with open questions and a future outlook in this line of work, which can be discussed during the workshop.

2 The Case of Video Coding

The assignment of codes to data (i.e., markers identifying a piece of evidence as representative of a certain idea) is a very common step in any qualitative analysis of data from interviews, observations, the video recording of a lesson, etc. Very often, this coding is done on the basis of a text transcription of the evidence (e.g., a recorded interview) – which also has advantages in terms of data storage and privacy of the informants. However, the availability of almost unlimited storage, the importance of nonverbal information, and the fact that many research processes are iterative and may require coming back to the raw sources to re-interpret or re-code according to new dimensions, is making the direct coding from audio/video a quite common endeavor. Tools like ELAN[7] are often used in our community for this purpose. In any case, the coding of audiovisual sources is recognized as one of the most time-consuming processes in qualitative data analysis [5, 1] and in any learning sciences research effort.

[3] http://www.qsrinternational.com/nvivo-product

[4] http://atlasti.com/

[5] https://www.ibm.com/us-en/marketplace/spss-statistics

[6] https://www.r-project.org/

[7] http://tla.mpi.nl/tools/tla-tools/elan/

Depending on the kinds of codes being extracted from the video, and whether a textual transcription is needed, between three and ten times the length of the media are often cited as necessary to process such data (not counting later higher-level hierarchical coding or revision/iteration of the coding). As mentioned above, researchers within the MMLA community have started to experiment with new ways of bringing this processing time down. For instance, attempting to code whether a certain moment in the (audio) recording of a lesson is a question or not [3], or at what social plane a teacher is interacting with students (combining audiovisual and wearable accelerometer data) [12].

However, it should be noted that these initial attempts at automated coding have been moderately successful for codes with *low semantic value*. Video coding about more abstract or complex notions (e.g., the level of participation in collaborative learning, as shown in [8]) still have to be performed manually by human researchers. Hence, multiple open questions still remain: Are these initial automation efforts reliable enough to be used in a researcher's everyday practice? What kinds of coding tasks (or constructs) are best suited for them? How much researcher time is really saved by applying these video coding automations? The next section describes our initial exploration of these questions.

3 Next-Generation Researcher Tools: A Proof-of-Concept For Video Coding

3.1 A Human-in-the-loop Approach

The concerns voiced by experts in qualitative data analysis about the dangers of automating coding and how that automation shapes data analysis [1, 15], prompt us to adopt a "human in the loop" approach to the automation of video coding (in contrast with full automation). In such an approach, a human would provide initial examples of coding a subset of the data, which would then be fed into machine learning algorithms as training data; on the basis of this initial training, the algorithm would suggest automatically a coding for a subsequent subset of the data; the human would verify this first automated coding, making the necessary modifications to the suggested coding; this new verified coding would also be fed into the algorithm as additional training data; thus, the rest of the dataset would be coded in iterative human-machine cycles, in which the human labor would (hopefully) represent a diminishing proportion of the effort (see Fig. 1). Indeed, initial tool prototypes following this kind of approach are starting to appear, e.g., to code objects being looked at in mobile eye-tracking studies [7].

This manner of operation not only stems from methodological concerns, but also from our own experience in learning sciences and MMLA research. Given that the features that can normally be extracted automatically from audio, video or sensor data are rather low-level and relatively semantic-less, and considering the wildly varying nature of potential coding tasks, *initial human input* is crucial, to map these low-level signals with the current context of the recording, as the

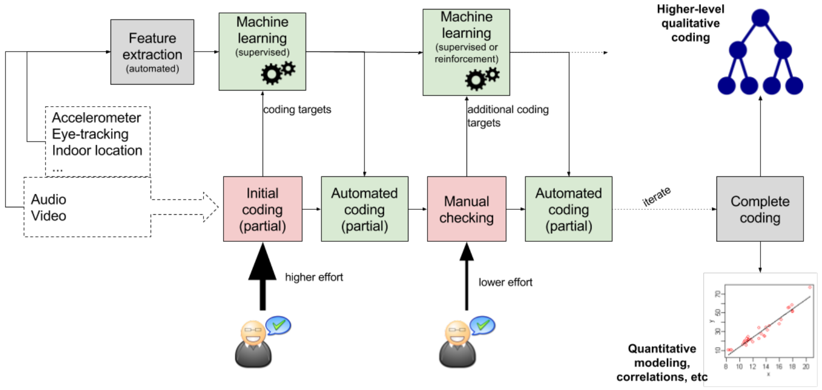
Feature extraction (automated)
Machine learning (supervised)
Machine learning (supervised or reinforcement)
Higher-level qualitative coding
coding targets
additional coding targets
Accelerometer
Eye-tracking
Indoor location
...
Audio
Video
Initial coding (partial)
Automated coding (partial)
Manual checking
Automated coding (partial)
iterate
Complete coding
higher effort
lower effort
Quantitative modeling, correlations, etc
10
30
50
70
y
8
10
12
14
16
18
20
x

Fig. 1. Proposed approach for computationally-enhanced multimodal coding researcher tools.

researcher understands it. Similarly, later *iterative human checkpoints* are needed to ensure that the algorithms are improving their performance (as the human understands it), not over-fitting to some meaningless features of the data.

From a MMLA perspective, it is also important to note that having *multiple sources of data* available in the dataset (including other sensors aside from the audiovisual feed) is very important, too. Albeit current studies have shown reasonable performance in using just audiovisual [12], or even aural-only data [2] for particular coding tasks, supporting a wide variety of coding tasks may require many different input channels (e.g., nervousness during periods of silence can be undetectable to audio or even video analysis, but can be easily picked up by an accelerometer worn by a student).

3.2 Exploratory Study

With this approach in mind, we are conducting an exploratory study within an ongoing mixed-methods research process, to explore the following research questions:

1. What features and machine learning models perform best in this approach, taking into account the limited amount of labeled data available for training?
2. How accurate is this automated coding? How does performance vary depending on the kind of coding task? How does performance evolve over multiple human-machine iterations?
3. What is the actual gain in terms of human effort, versus manual coding?

The context of the study is our investigation about how social media can be used effectively in the classroom. We are studying the usage of a social media tool

(SpeakUp[8]) in an authentic university setting. The study, whose initial results were presented in [13], focuses on the tool's impact on classroom engagement, attention and social interaction. To explore these aspects, we use a mixed method approach combining quantitative and qualitative data coming from teachers, students, an observer, and the SpeakUp tool itself. That is, we combine data from questionnaires, observations, video recordings, content and activity tracking (via SpeakUp logs). So far, we gathered data during six 90-minute face-to-face sessions, handled by three teachers, in which 145 students participated.

In the initial stage of our study [13], a classroom video of the first session was manually coded, along the following coding dimensions: which *actor* was speaking (e.g., each of the three teachers present, or one of the students); what *action* was being performed at that moment (e.g., presentation/lecturing, asking questions, providing answers, noting technical or other kinds of problems); who was the *target* of the interaction, if any (e.g., a teacher, students, or all the class); and finally, what supporting *resources* were being used, if any (e.g., slides, videos, SpeakUp). This systematic annotation allowed us to synchronize the observations of the face-to-face interaction with the evidence gathered from the computer-mediated activities in SpeakUp.

In our continuation of this study, we would like to diminish the human effort involved in the video coding of the other five sessions. As of this writing, we are extracting automatically low-level audio and video features (similarly as it was done in our previous automated coding efforts [12]) from the available classroom videos. After training different algorithms on the first, already-labeled session, we will automatically code a second session, while the researcher codes it manually as well. Then, after evaluating the model's initial performance, subsequent iterations will take place to code the rest of the sessions, measuring the successive performances as well as the time needed to "correct" the automatically-generated coding. The results of these measurements can be discussed face-to-face during the workshop.

4 Outlook

We started the paper contending that we are on the brink of a revolution in researcher tools that can help researchers in making sense of increasingly complex and messy data. In fact, the MMLA community has already started to push in this direction, with proposals such as the Structured TRansactional Event Analysis of Multimodal Streams (STREAMS) tool for the alignment of data sources [9]. We ourselves are starting to tackle another dreaded task: that of manual coding of audiovisual data. We hope to present in the workshop the results and lessons learned from our ongoing study using the presented human-in-the-loop approach.

However, developing better researcher tools for MMLA is still quite a novel direction in our field, with many open questions that we hope to explore during the workshop:

[8] SpeakUp: https://web.speakup.info

- The fact that we lack clear descriptions and understanding of *how MMLA research is actually done* (at the ethnographic level of who does what, when and using which tools, and in contrast to the "clean" post-hoc descriptions provided in publications). This understanding, which could be the subject of a workshop in itself, is essential to the design of better MMLA researcher tools.
- The technical issues of what kind of *machine learning* should be used in developing such researcher tools: whether "black box" algorithms are acceptable or not (taking into account that we keep humans in the loop); what is the right way to partition and train our automated coding models (including the use of one-shot learning and other techniques that are still rarely used in MMLA); etc.
- In a similar vein, the brief account of our approach above has been expressed in terms of supervised machine learning, which is quite widespread in learning analytics. However, by its very iterative nature, our approach lends itself to representation in terms of online or *reinforcement learning* [14] techniques. While much less widespread in our field, we should not ignore the advances done in this area of machine learning research.
- Our text above also glossed over the issue of the *level of abstraction* of the codes we want to automatically obtain. In some cases, these codes can be very far removed from the features available to the machine learning models. Although certain progress can probably be achieved by working our way up via intermediate-level features and codes, we should also keep an eye on another emergent machine learning sub-field, that of transfer learning [11], which can provide techniques that help weed out the generally-applicable features from the ones that work only in a particular context.
- How can the new researcher tools be developed from a *software engineering* perspective, including toolkits to base them upon, strategies to enable integration between the different ongoing projects and with the variety of data gathering setups being used in the community... all while keeping the tools usable and effective (i.e., the human-computer interaction perspective, tying back with the ethnographic description of our research processes).

In line with the theme of the workshop emphasizing "real world" learning support, we should note that, even if we have taken the researcher perspective throughout this paper, the approach and tools we envision are a crucial step towards the development of MMLA solutions that can be applied in the real world. It is unrealistic to expect solutions that are usable by teachers and students, when expert researchers consider themselves swamped in "messy data" and struggle with the definition of useful features and metrics, and their transfer to new data gathering contexts. Human-in-the-loop automations like the ones we presented here, once refined, could be applied to student support mechanisms, or to aid live observations by teachers in the context of professional development. Once we solve the problems of how to merge and improve the automated coding models trained by different researchers in different contexts, taking advantage of the thousands of hours that researchers spend coding worldwide, we will have more reliable mid-level metrics on which our real-world MMLA solutions can be based, in a way that actually works across the wide variety of contexts out there.

Acknowledgements

This research has been partially funded by the European Union in the context of CEITER and the Next-Lab (Horizon 2020 Research and Innovation Programme, grant agreements no. 669074 and 731685).

References

1. Basit, T.: Manual or electronic? the role of coding in qualitative data analysis. Educational research 45(2), 143–154 (2003)
2. Blanchard, N., Brady, M., Olney, A.M., Glaus, M., Sun, X., Nystrand, M., Samei, B., Kelly, S., DMello, S.: A study of automatic speech recognition in noisy classroom environments for automated dialog analysis. In: International Conference on Artificial Intelligence in Education. pp. 23–33. Springer (2015)
3. Blanchard, N., D'Mello, S., Nystrand, M., Olney, A.M.: Automatic classification of question & answer discourse segments from teacher's speech in classrooms. In: Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015), International Educational Data Mining Society (2015)
4. Dawson, S., Gašević, D., Siemens, G., Joksimovic, S.: Current state and future trends: A citation network analysis of the learning analytics field. In: Proceedings of the fourth international conference on learning analytics and knowledge. pp. 231–240. ACM (2014)
5. Dey, I.: Qualitative data analysis: A user friendly guide for social scientists. Routledge (2003)
6. Dornfeld, C., Zhao, N., Puntambekar, S.: A mixed-methods approach for studying collaborative learning processes at individual and group levels. Philadelphia, PA: International Society of the Learning Sciences. (2017)
7. Fong, A., Hoffman, D., Ratwani, R.M.: Making sense of mobile eye-tracking data in the real-world: A human-in-the-loop analysis approach. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. vol. 60, pp. 1569–1573. SAGE Publications Sage CA: Los Angeles, CA (2016)
8. Isohätälä, J., Järvenoja, H., Järvelä, S.: Socially shared regulation of learning and participation in social interaction in collaborative learning. Int J Educ Res 81, 11–24 (2017), http://linkinghub.elsevier.com/retrieve/pii/S0883035516305298
9. Liu, R., Stamper, J.: Multimodal data collection and analysis of collaborative learning through an intelligent tutoring system. In: Prieto, L.P., Martinez-Maldonado, R., Spikol, D., Hernandez-Leo, D., Rodríguez-Triana, M.J., Ochoa, X. (eds.) Joint Proceedings of the Sixth Multimodal Learning Analytics (MMLA) Workshop and the Second Cross-LAK Workshop (MMLA-CrossLAK). pp. 47–52. No. 1828 in CEUR Workshop Proceedings, Aachen (2017), http://ceur-ws.org/Vol-1828/#paper-07
10. Meyer, P., Kelle, S., Ullmann, T.D., Scott, P., Wild, F.: Interdisciplinary cohesion of tel–an account of multiple perspectives. In: European Conference on Technology Enhanced Learning. pp. 219–232. Springer (2013)

11. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on knowledge and data engineering 22(10), 1345–1359 (2010)
12. Prieto, L.P., Sharma, K., Dillenbourg, P., Rodríguez-Triana, M.J.: Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. pp. 148–157. ACM (2016)
13. Rodríguez-Triana, M.J., Holzer, A., Prieto, L.P., Gillet, D.: Examining the effects of social media in co-located classrooms: A case study based on speakup. In: European Conference on Technology Enhanced Learning. pp. 247–262. Springer (2016)
14. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction, vol. 1. MIT press Cambridge (1998)
15. Tummons, J.: Using software for qualitative data analysis: Research outside paradigmatic boundaries. In: Big data? Qualitative approaches to digital research, pp. 155–177. Emerald Group Publishing Limited (2014)
16. Yoon, S.A., Hmelo-Silver, C.E.: What do learning scientists do? a survey of the isls membership. Journal of the Learning Sciences 26(2), 167–183 (2017)