Editorial Acknowledgments and Introduction to the Special Issue for the EDM Journal Track

Olga C. Santos (aDeNu Research Group - UNED, Spain) and Amal Zouaq (LAMA-WeST Lab, Ecole Polytechnique de Montréal, Canada)

INTRODUCTION

The 14th International Conference on Educational Data Mining (EDM 2021) took place in a hybrid format, both online and in-person, to facilitate participation and networking for all. It was held from June 29th to July 2nd, 2021, in Paris (France). Educational Data Mining is a leading international forum for high-quality research that mines datasets to answer educational research questions, including exploring how people learn and how they teach. The overarching goal of the Educational Data Mining research community is to support learners and teachers more effectively, by developing data-driven understanding of the learning and teaching processes in a wide variety of contexts and for diverse learners.

For the seventh time, EDM 2021 held a Journal track allowing papers submitted to JEDM to be presented at the conference. A summary was included in the EDM 2021 proceedings, and the full texts are included in this Special Issue of the Journal of Educational Data Mining (JEDM).

The JEDM Track at EDM 2021 received 7 submissions, among which 3 made it to the final stage in time for publication. We are pleased to publish the following works in this issue.

In the first paper, **Paaßen et al.** propose a solution to allow the application of data mining techniques to student activity in the context of computer programming where computer programs have the form of syntax trees instead of using the vector-shaped input expected in many data mining techniques. Their approach is called ast2vec and consists in a neural network autoencoder for trees that maps Python syntax trees to vectors in Euclidean space and back, thereby facilitating data mining on computer programs as well as the interpretation of data mining results. Ast2vec has been trained on almost half a million programs of novice programmers and is designed to be applied across learning tasks without re-training, meaning that users can apply it without any need for (additional) deep learning.

The second paper focuses on affective computing. **Chen et al.** introduce a methodological framework to support causal inference and discovery of causal relationships among groups of variables that describe parents' coaching decisions, children's cognitive-affective experiences, and children's individual stable factors. In this way, they investigate the existing complex multimodal interactions in human one-on-one coaching going beyond verbal interactions and incorporate the rich aspects of multimodal cognitive-affective experiences. Their solution combines statistical approaches in Sparse Multiple Canonical Correlation, causal discovery and inference methods for observations. The proposed analytical framework is explainable and amenable to incorporating domain knowledge and it is demonstrated using a multimodal one-on-one math problem-solving coaching dataset collected at naturalist home environments involving parents and young children.

Finally, **Choffin et al.** present a data mining solution to help learners organize their study time and improve their memory retention applying a spaced repetition learning strategy that consists in temporally distributing learning episodes instead of learning in a single massed study session. As discussed in the paper, there exists some adaptive spacing schedulers that support pure learning and memorization of simple pieces of knowledge, such as factual knowledge (e.g., vocabulary words) by personalizing the temporal distribution of retrieval practice, but they do not help students schedule their reviewing sessions for learning a set of skills. To address this open issue, these authors propose and compare the strengths and weaknesses -in terms of performance, robustness and complexity- of five heuristics that extend three existing adaptive spacing algorithms. Three of the proposed extensions select the best skill to review at any timestamp given a student's past study history and the other two use a multi-skills version in which they select the most promising subset of skills to review with a greedy procedure.