Industry Panel: The Future of Practical Applications of EDM at Scale

Ryan Baker
Teachers College
Columbia University
baker2@exchange.tc.columbia.edu

John Carney
Carney Labs LLC
john.carney@carneylabs.com

Piotr Mitros
edX
pmitros@edx.org

Bror Saxberg
(moderator)
Kaplan Inc.
bror.saxberg@kaplan.com

John Stamper
Carnegie Mellon University
and PSLC DataShop
john@stamper.org

ABSTRACT

This mixed panel of different professionals working in EDM will be a conversation about increasing the connection between research and real-world applications. What's going on now to scale techniques for use "out there" in the field? What should researchers, funders, regulators, publishers, trainers, schools/universities and others be doing to get ready for practical work? What's in the way that we can usefully start work to address? We'll ask the audience to engage in this conversation as well - what's in your way to moving work from research environments to practically help learners at scale - and to generate more useable data at scale?

Ethics and Privacy in EDM

Dragan Gasevic
University of Edinburgh
dragan.gasevic@ed.ac.uk

Taylor Martin (moderator)
National Science Foundation
htmartin@nsf.gov

Zach Pardos
UC Berkeley
pardos@berkeley.edu

Mykola Pechenizkiy
TU Eindhoven
m.pechenizkiy@tue.nl

John Stamper
CMU and PSLC DataShop
john@stamper.org

Osmar Zaiane
University of Alberta
zaiane@ualberta.ca

ABSTRACT

Educational data mining is inherently falls into the category of the so-called secondary data analysis. It is common that data that have been collected for administrative or some other purposes at some point is considered as valuable for other (research) purpose. Collection of the student generated, student behavior and student performance related data on a massive scale in MOOCs, ITSs, LMS and other learning platforms raises various ethical and privacy concerns among researches, policy makers and the general public. This panel is aimed to discuss major challenges in ethics and privacy in EDM and how they are addressed now or should be addressed in the future to prevent any possible harm to the learners. Several experts are invited to discuss the potential and challenges of privacy-preserving EDM, ethics-aware predictive learning analytics, and availability of public benchmark datasets for EDM among others.

Grand Challenges for EDM and Related Research Areas

Ryan Baker (moderator)
Teachers College
Columbia University
baker2@exchange.tc.columbia.edu

Peter Brusilovsky (UM Inc)
School of Information Sciences
Pittsburgh University
peterb@pitt.edu

Dragan Gasevic (SoLAR)
Schools of Education and Informatics
University of Edinburgh
dragan.gasevic@ed.ac.uk

Neil T. Heffernan (AIED)
Department of Computer Science
Worcester Polytechnic Institute
nth@wpi.edu

Mykola Pechenizkiy (IEDMS)
Department of Computer Science
TU Eindhoven
m.pechenizkiy@tue.nl

Alyssa Wise (ISLS)
Faculty of Education
Simon Fraser University
alyssa_wise@sfu.ca

ABSTRACT

Educational data mining (EDM) and Learning analytics are still rather young research areas. The goal of this panel is to share different views on what major challenges researches need to address in EDM, learning analytics and related research areas including but not limited to User modeling, AI in Education, and Learning Sciences. The representatives of the corresponding communities are invited to discuss what grand challenges we should aim to address for the next five years, and which of these challenges are old and which are new, which of them peculiar to one distinct research area and which of them are shared across two or more research areas.

JEDM TRACK PAPERS

(abstracts)

Metrics for Evaluation of Student Models

Radek Pelánek
Masaryk University Brno
pelanek@fi.muni.cz

ABSTRACT

Researchers use many different metrics for evaluation of performance of student models. The aim of this paper is to provide an overview of commonly used metrics, to discuss properties, advantages, and disadvantages of different metrics, to summarize current practice in educational data mining, and to provide guidance for evaluation of student models. In the discussion we mention the relation of metrics to parameter fitting, the impact of student models on student practice (over-practice, under-practice), and point out connections to related work on evaluation of probability forecasters in other domains. We also provide an empirical comparison of metrics. One of the conclusion of the paper is that some commonly used metrics should not be used (MAE) or should be used more critically (AUC).

Multi-Armed Bandits for Intelligent Tutoring Systems

Benjamin Clement
Inria, France
benjamin.clement@inria.fr

Didier Roy
Inria, France
didier.roy@inria.fr

Pierre-Yves Oudeyer
Inria, France
pierre-yves.oudeyer@inria.fr

Manuel Lopes
Inria, France
manuel.lopes@inria.fr

ABSTRACT

We present an approach to Intelligent Tutoring Systems which adaptively personalizes sequences of learning activities to maximize skills acquired by students, taking into account the limited time and motivational resources. At a given point in time, the system proposes to the students the activity which makes them progress faster. We introduce two algorithms that rely on the empirical estimation of the learning progress, RiARiT that uses information about the difficulty of each exercise and ZPDES that uses much less knowledge about the problem.

The system is based on the combination of three approaches. First, it leverages recent models of intrinsically motivated learning by transposing them to active teaching, relying on empirical estimation of learning progress provided by specific activities to particular students. Second, it uses state-of-the-art Multi-Arm Bandit (MAB) techniques to efficiently manage the exploration/exploitation challenge of this optimization process. Third, it leverages expert knowledge to constrain and bootstrap initial exploration of the MAB, while requiring only coarse guidance information of the expert and allowing the system to deal with didactic gaps in its knowledge. The system is evaluated in a scenario where 7–8 year old schoolchildren learn how to decompose numbers while manipulating money. Systematic experiments are presented with simulated students, followed by results of a user study across a population of 400 school children.

Developing Computational Methods to Measure and Track Learners' Spatial Reasoning in an Open-Ended Simulation

Aditi Mallavarapu
University of Illinois at Chicago
amalla5@uic.edu

Leilah Lyons
University of Illinois at Chicago
llyons@uic.edu

Tia Shelley
University of Illinois at Chicago
tshell2@uic.edu

Brian Slattery
University of Illinois at Chicago
bslatt2@uic.edu

Moira Zellner
University of Illinois at Chicago
mzellner@uic.edu

Emily Minor
University of Illinois at Chicago
eminor@uic.edu

ABSTRACT

Interactive learning environments can provide learners with opportunities to explore rich, real-world problem spaces, but the nature of these problem spaces can make assessing learner progress difficult. Such assessment can be useful for providing formative and summative feedback to the learners, to educators, and to the designers of the environments. This work adds to a growing body of research that is applying EDM techniques to more open-ended problem spaces.

The open-ended problem space under study here is an environmental science simulation. Learners were confronted with the real-world challenge of effectively placing green infrastructure in an urban neighborhood to reduce surface flooding. Learners could try out different spatial arrangements of green infrastructure and use the simulation to test each solution's impact on flooding. The learners' solutions and the solutions' performances were logged during a controlled experiment with different user interface designs for the simulation. As with many open-problem spaces, analyzing this data was difficult due to the large state space, many good solutions, and many alternate paths to those good solutions.

This work proposes a procedure for reducing the state space of solutions defined by spatial patterns while maintaining their critical spatial properties. Spatial reasoning problems are a problem class not yet examined by EDM, so this work sets the stage for further research in this area. This work also details a procedure for discovering effective spatial strategies and solution paths, and demonstrates how this information can be used to give formative feedback to the designers of the interactive learning environment.

Move your lamp post: Recent data reflects learner knowledge better than older data

April Galyardt
University of Georgia
galyardt@uga.edu

Ilya Goldin
Pearson
ilya.goldin@pearson.com

ABSTRACT

In educational technology and learning sciences, there are multiple uses for a predictive model of whether a student will perform a task correctly or not. For example, an intelligent tutoring system may use such a model to estimate whether or not a student has mastered a skill. We analyze the significance of data recency in making such predictions, i.e., asking whether relatively more recent observations of a student's performance matter more than relatively older observations. We investigate several representations of recency, such as the count of prior practice in the AFM model, and the proportion of recent successes with exponential and box kernels. We find that an exponential decay of a proportion of successes provides the summary of recent practice with the highest predictive accuracy over alternative models. As a secondary contribution, we develop a new logistic regression model, Recent-Performance Factors Analysis, that leverages this representation of recent performance, and has higher predictive accuracy than existing logistic regression models.