Introduction to the Issue on Signal Processing and Machine Learning

MACHINE learning and data science have already been used in education for many years to enable the assessment of students at scale, for example, by largely automating some grading processes and by adaptive testing. But these are early days. As the technology matures, we will see machine learning used to construct systems that connect students with the materials they need and to other students with similar interests, shared difficulties, or complementary skills. These systems will require an understanding of student engagement and will need to learn what actions to take in order to keep students motivated. Eventually algorithms will even be able to automatically provide rich feedback to students on complex and open ended assignments, enabling the delivery of a high quality experience to anyone with an internet connection.

Human learning can also benefit from machine learning. Teachers and students, from K-12 to higher education, from in-class to online, generate a large amount of data as they learn. Conversely, they also respond to signals of all kinds. The possibility of capturing, analyzing, visualizing, and leveraging all of this data and using it for prediction, recommendation, and individualization, presents tremendous opportunities to make education more effective, more scalable, and less costly.

The surge in popularity of Massive Open Online Courses (MOOCs) and other online and blended learning environments has demonstrated the potential of the Internet for scaling education. While advances in technology have enabled content delivery to massive numbers of students, these platforms remain limited in their ability to provide an effective learning experience for individual students. Recent advances in machine learning and signal processing provide promising new avenues to move beyond this "one size fits all" educational approach. The key is that today's learning technology platforms can capture personal learning data about students as they proceed through courses: performance on homework and exams, click actions made while watching lecture videos or interacting with simulations, the social learning networks formed among the students, the content posted on discussion forums, and so on. This data enables new opportunities to study the process of student learning and to design systems that improve learning at scale by closing the learning feedback loop.

This special issue of showcases research from the signal processing community that is advancing effective learning at scale. The paper "Behavior-Based Grade Prediction for MOOCs via Time Series Neural Networks" develops new methods to predict learning outcomes in massive open online courses (MOOCs). The system is based on a time series neural network that predicts a MOOC student's overall course grade as they progress through the course material, taking as input their prior performance on assessments and their video-watching behavior.

The paper "Context-Aware Recommendation-Based Learning Analytics Using Tensor and Coupled Matrix Factorization" develops a collaborative filtering system for learning analytics that predicts student grades in college courses. The system takes into account not just the courses each student has taken, but also their timing, to improve the prediction.

The paper "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs" considers the important problem of ensuring that students not only succeed in individual courses but also persist through a sequence of courses and complete an entire degree program. This problem is complicated by the high variance in the backgrounds of students entering a program, the fact that different courses are not equally informative for prediction purposes, and by the need to track students' progress over multiple years. The authors develop two new tools for this purpose: a new predictor based on ensemble learning and a data-driven clustering method based on probabilistic matrix factorization to cluster courses.

The paper "BLAh: Boolean Logic Analysis for Graded Student Response Data" extends the machine learning models that form the core of personalized learning systems. Student response models have been used to analyze tests for decades now and allow us to decompose student understanding across the multiple concepts being assessed. Traditionally these models have been linear and thus have faced fundamental limitations — for example they cannot naturally handle questions that require a student to master multiple concepts in order to answer a test item correctly. The BLAh model goes beyond these traditional assumptions, modeling concept understanding via more general Boolean logic. Importantly, model expressivity is gained without sacrificing model interpretability; and this makes the BLAh approach a rarity today as black box approaches such as neural networks have become pervasive in the machine learning literature.

Over the coming years, as we increasingly leverage the massive scale data collected from learners, a challenge we will face is how to minimize the biases of our automated systems to ensure equal opportunity for all learners. Even disembodied machine learning algorithms are not completely objective. Indeed, it is far too easy for biases to creep in – even unintentionally – during data collection or algorithm design. A system that grades or teaches based on stereotypes or biased assumptions could have long lasting negative consequences. Recognizing and minimize

Digital Object Identifier 10.1109/JSTSP.2017.2723218

these biases is critical; this is a problem that our community must address starting now and not in the distant future.

Throughout this endeavor, technology and pedagogy should advance hand in hand. We, as a research community, need to develop access to as much data as possible while preserving the privacy and integrity of the data. We also need feedback loops that: turn data analysis into actions, and evaluate competing methodologies based on the effect of such actions. Understanding the different ways in which different human beings learn is a challenging task, but a crucial one.

MIHAELA VAN DER SCHAAR, *Editor*
University of Oxford
Oxford OX2 6ED, U.K.
mihaela.vanderschaar@eng.ox.ac.uk

RICHARD G. BARANIUK, *Editor*
Rice University
Houston, TX 77005 USA
richb@rice.edu

MUNG CHIANG, *Editor*
Purdue University
West Lafayette, IN 47907 USA
chiangm@princeton.edu

JONATHAN HUANG, *Editor*
Google, Inc.
Seattle, WA 98103 USA
jonathanhuang@google.com

SHENGDONG ZHAO, *Editor*
National University of Singapore
Singapore 117418
Huawei Technologies, Co. Ltd.,
Shenzhen 518129, China

Mihaela van der Schaar (F'10) is Man Professor at the University of Oxford and the Chancellor's Professor of electrical and computer engineering at the University of California Los Angeles. She has received multiple best paper awards, including the Darlington Best Paper award. She holds 33 U.S. patents. Her recent interests are on machine learning for medicine and for education.

Richard G. Baraniuk (F'02) is the Cameron Professor of electrical and computer engineering at Rice University and the Founder and Director of OpenStax. He is a member of the American Academy of Arts and Sciences and a Fellow of the National Academy of Inventors and American Association for the Advancement of Science. He received the DOD Vannevar Bush Faculty Fellow Award, the IEEE James H. Mulligan, Jr., Education Medal, and the IEEE Signal Processing Society Best Paper, Best Column, Education, and Technical Achievement Awards.

Mung Chiang is the John A. Edwardson Dean of the College of Engineering and the Roscoe H. George Professor of electrical and computer engineering at Purdue University. He was previously the Arthur LeGrand Doty Professor at Princeton University. His research on networking received the 2013 NSF Alan T. Waterman Award and his MOOC "Networked Life" and textbooks reached over a quarter million students. He cofounded several startup companies in mobile, IoT and AI, and the global nonprofit OpenFog Consortium.

Jonathan Huang received the Ph.D. degree in 2011 from the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. He is a Senior Research Scientist in Google Research and Machine Intelligence. Prior to Google, he was an NSF Computing Innovation Postdoctoral Fellow at the Geometric Computing Group at Stanford University. His research interests are diverse, ranging from algorithmic and statistical problems that arise when reasoning with combinatorially structured data to large-scale analysis of educational data and most recently to deep learning with applications to computer vision and code analysis (particularly in educational settings). His research has resulted in a number of publications in premier machine learning and vision conferences and journals and most recently led a team of researchers at Google to win the 2016 COCO detection challenge.

Shengdong Zhao is an Associate Professor in the Department of Computer Science, National University of Singapore, and a senior consultant with Huawei Consumer BG. He established the NUS-HCI Research Lab. He is the recipient of the ICACHI Outstanding Young Leaders in HCI, Best iPad App of Year, and the NUS Young Investigator Awards.