Guest editorial

Enhancement of teaching and learning: applications in learning analytics and educational data mining

Learning analytics (LA) and educational data mining (EDM) are highly related subjects that overlap in definition and scope. Although both communities of researchers within LA and EDM have similarities where learning science and analytic techniques intersect, there are some significant differences between them in terms of origins, techniques, fields of emphasis and types of discovery (Siemens and Baker, 2012; Chatti *et al.*, 2012; Romero and Ventura, 2013). EDM refers to computerized methods and tools for automatically detecting and extracting meaningful patterns and information from large collections of data from educational settings (Kumar and Sharma, 2017). LA is focused on understanding and optimizing learning and learning environments by measuring, gathering, analyzing and reporting of data about learners and learning contexts (Siemens and Baker, 2012).

The aim for this special issue on Applications in LA and EDM covers all aspects of data analytics in supporting teaching, learning and administration for researchers in P-16 education, and the development of technology-enriched formats of instructional delivery, such as various categories of blended and online learning. Traditionally, the main data sources of LA and EDM research reply on the database in the learning management system (LMS). The developments of Internet of Things (IoT) or sensors, at some levels, make up the gap of activity tracking outside the LMS. The special issue endeavors to publish research and practice which explores the applications of LA and EDM by including data sources outside the LMS, such as open data, in classroom devices, IoT, mobile devices, academic data warehouse and other devices which can track, diagnose and store learning activities.

This special issue is also interested in innovative approaches of feature extraction, pattern identification/recognition, data anonymization, modeling and intervention to support innovative applications of machine learning and deep learning in education.

All presentations in this special edition were referred through a double-blinded procedure before being accepted for publication. Each manuscript submitted was reviewed by two to three invited reviewers. The review criteria were:

- importance of the subject;
- originality of the approach;
- soundness of scholarship displayed;
- level of interest and pertinence for readers;
- depth and strength of argument; and
- clarity of expression.

The first article, Youngjin Lee's "Estimating student ability and problem difficulty using item response theory (IRT) and TrueSkill", examined item response theory (IRT) and TrueSkill applied to simulated and real problem-solving data to estimate the ability of students solving homework problems in the massive open online course (MOOC). Based on the estimated ability, data mining models predicting whether students can correctly solve homework and quiz problems in the MOOC were developed. This study found that the correlation between students' ability estimated from IRT and TrueSkill was strong. In addition, IRT- and TrueSkill-based data mining models showed a comparable predictive power when the data included a large number of students. While traditional IRT research has been focused on the assessment design and development, this study explores the application of IRT in measuring student's capability by taking advantage of MOOC.

The second article, Dimić Gabrijela, Dejan Rančić, Nemanja Maček, Petar Spalević and Vida Drąsutė's "Improving the Prediction Accuracy in Blended Learning Environment using Synthetic Minority Oversampling Technique", studied the prediction accuracy of student's activity patterns in the blended learning environment. Classification models were used in the comparison for different cardinality subsets. The results showed the opposite with the fact that reducing the number of features leads to prediction accuracy increasing. The authors argued that improving the prediction accuracy in the described learning environment is based on applying synthetic minority oversampling technique what had affected on results on correlation-based feature selection method. While highly imbalanced data is a challenge for predictive modeling, this study proposed a method to enhance prediction accuracy via synthetic minority oversampling.

The third study, Brian Wright's "Analysis of supportive campus environments and first-generations-student learning outcomes", intends to analyze the relationship between supportive campus measures and student learning outcomes for first-generation students and non-first-generation students to determine if variances are present via EDM. Data were gathered through cluster sampling of pre-existing datasets on undergraduate seniors generated by issuances of the National Survey on Student Engagement. It found that clear pattern differences are present between first-generation and non-first-generation students in terms of supportive campus environment factors contributing to learning outcomes. While server log is the major data source in LA and EDM (Ihantola *et al.*, 2015; Schwendimann *et al.*, 2017), this study analyzed a longitudinal investigation and identify useful insights for higher education administrators.

The fourth article, Riccardo Pecori, Vincenzo Suraci and Pietro Ducange's "Efficient computation of key performance indicators in a distance learning university", proposed a framework to compute efficiently key performance indicators, summarizing the trends of students' academic careers, by using EDM. The parallel computation of the indicators through Map and Reduce nodes is carefully described, together with the workflow of data, from the educational sources to a NoSQL database and to the LA engine. The framework tested in an Italian distance learning institution. It concluded the framework was able to significantly reduce the amount of time needed to

The current issue and full text archive of this journal is available on Emerald Insight at: **www.emeraldinsight.com/2398-6247.htm**

Information Discovery and Delivery
47/2 (2019) 65–66
© Emerald Publishing Limited [ISSN 2398-6247]
[DOI 10.1108/IDD-05-2019-074]

compute Key Performance Indicators. While LA and EDM aim to support data-driven decision- making, this study revealed the efforts of a large higher education institution in real-time data processing and reporting.

The fifth study, Dennis Foung and Julia Chen's "Discovering disciplinary differences: blending data sources to explore the student online behaviors in a university English course", explored disciplinary differences in completing blended learning tasks in an academic literacy course and the feasibility of using a blended LA approach to explore disciplinary differences. This study blended data from the LMS and timetabling arrangements. Results suggested that the online behaviors of design students and accounting students are different in terms of starting day and completion rate of online activities. While online course design aims to accommodate learning needs from learners with assorted backgrounds, this study provides valuable insights by analyzing data collected at a higher education institution.

Based on the editing process and the selected five articles, we observed the data-driven decision-making is crucial for various learning environments. Possible research aspects can be obtained from data collection, processing, analysis and application. However, most studies in data-driven decision-making have been focused on the learner's aspects. As teaching and learning are two inseparable and interacting parts, improving learning performance could be achieved from these two aspects. We recommend future research can focus more on the assessment of teaching process to provide feedback to teachers.

As guest editors, we would like to express our sincere gratitude to the review panelists. Without their professional reviews and participation, this special issue would not converge invaluable collective and collaborative knowledge.

Xu Du, Jui-Long Hung and Chih-Hsiung Tu

References

Chatti, M.A., Dyckhoff, A.L., Schroeder, U. and Thüs, H. (2012), "A reference model for learning analytics", *International Journal of Technology Enhanced Learning*, Vol. 4 Nos 5/6, pp. 318-331.

Ihantola, P., Vihavainen, A., Ahadi, A., Butler, M., Börstler, J., Edwards, S.H., Isohanni, E. and Korhonen, A. (2015), "Educational data mining and learning analytics in programming: literature review and case studies", *ITiCES WGR'16, Vilnius, ACM*, pp. 41-63.

Kumar, R. and Sharma, A. (2017), "Data mining in education: a review", *International Journal of Mechanical Engineering and Information Technology*, Vol. 5 No. 1, pp. 1843-1845.

Romero, C. and Ventura, S. (2013), "Data mining in education", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 3 No. 1, pp. 12-27.

Schwendimann, B.A., Rodrigueztriana, M.J., Vozniuk, A., Prieto, L.P., Boroujeni, M.S., Holzer, A. and Gillet, D. (2017), "Perceiving learning at a glance: a systematic literature review of learning dashboard research", *IEEE Transactions on Learning Technologies*, Vol. 10 No. 1, pp. 30-41.

Siemens, G. and Baker, R.S.J.D. (2012), "Learning analytics and educational data mining: towards communication and collaboration", *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC*, pp. 252-254.

About the editors

Xu Du received his PhD in Communication Engineering from Huazhong University of Science and Technology, China, in 2002. He is currently a Professor with National Engineering Research Center for E-Learning, Central China Normal University, China. His research interests include smart environment and mobile learning, resource distribution, scheduling and recommendation, learning analytics and educational data mining. Xu Du is the corresponding editor and can be contacted at: duxu@mail.ccnu.edu.cn

Jui-Long Hung received his EdD in Instructional and Educational Technology with a minor in Information System at Texas Tech University in 2007. His research focuses on educational data and text mining and learning analytics. He is a Professor in the Department of Educational Technology at Boise State University and a Researcher of the following research centers: Center for Advanced Analytics and Business Intelligence at Texas Tech University, USA, Data Mining Lab at University of Central Florida, USA, and National Engineering Laboratory for Educational Big Data at Central China Normal University, China.

Chih-Hsiung Tu PhD, is a Professor at Northern Arizona University, Flagstaff, AZ, USA and an Educational/Instructional Technology Consultant with experience in online learning, open network learning, technology training in teacher education, online learning community, personal learning environment, social network analysis (SNA), data-driven instruction and digital lifelong learning. His research interests are distance education, socio-cognitive learning, socio-cultural learning, online learning community, social media, personal learning environments and network learning environments.