# Guest Editorial of the FGCS Special Issue on Advances in Intelligent Systems for Online Education

Geoffray Bonnin[a], Danilo Dessì[b,c], Gianni Fenu[d], Martin Hlosta[e], Mirko Marras[f,*], Harald Sack[b,c]

[a] *Université de Lorraine, Nancy, France*
[b] *FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Karlsruhe, Germany*
[c] *Karlsruhe Institute of Technology, Institute AIFB, Karlsruhe, Germany*
[d] *University of Cagliari, Cagliari, Italy*
[e] *The Open University, Milton Keynes, United Kingdom*
[f] *EPFL, Lausanne, Switzerland*

## 1. Introduction

The education sector is increasingly relying on online learning. Educational and training institutions are being motivated to endorse online learning strategies thanks to its technical, economic, and operational feasibility. This has become even more important after the lockdown caused by the breakout of COVID-19, when even universities revised their educational strategy, moving to online teaching. Learners and teachers are benefiting from the flexibility, accessibility, and costs of learning and teaching online. Nonetheless, moving education online is bringing unprecedented challenges. For instance, learners may feel isolation online, massive content alternatives often overload learners and teachers who look for educational resources, and institutions are being challenged to ensure academic integrity in online exams. The increasing amount of learning-related data and high performance computing are enabling intelligent systems that can successfully support stakeholders. Bringing this intelligence to online education leads to a wide range of advantages, e.g., avoiding manual error-prone tasks or providing learners with personalized guidance [1].

As artificial intelligence research and development is getting more mature, and the corresponding outputs are being deployed at scale in real-world contexts, the crucial role of using automated systems, as an additional support for stakeholders during decision making processes, becomes more evident nowadays. Current research has greatly expanded our understanding on such artificial intelligence, but there are less investigations on how it applies to online education. Data, methods, tools, and applications in this area are still limited, though they promise to proliferate in the next years. Further, more research and questions remain to be answered to bridge technological, social, pedagogical, and ethical aspects within intelligent systems.

This special issue brings together high-quality original research results reporting advances in the state of the art of online education systems empowered by artificial intelligence, covering different levels of the experimental pipeline, including but not limiting to data collection, computational models, and applicative systems. The rest of this article is structured as follows: Section 2 briefly presents the articles included in this special issue, and Section 3 provides concluding remarks.

## 2. Content of This Special Issue

The submissions received in the context of this special issue are peer-reviewed on the basis of relevance for the special issue, novelty, originality, significance, technical quality and correctness, quality and clarity of presentation, quality of references and reproducibility. In the end, this special issue contains 12 articles that advance the state-of-the-art, considering different tasks, ranging from community analysis, student risk and success prediction to course recommendation, cognitive state analysis and classification, and teaching content analysis and classification. These articles target a wide range of educational scenarios, including massive open online courses (MOOCs) and higher university. The richness of the research in this special issue shows how learning analytics and educational technologies research have many different perspectives that join in one goal: understanding and improving student learning.

**Community Analysis and Detection**. The authors in this special issue address the analysis and detection of communities in educational settings from several perspectives.

The article "*Maximal Cliques Based Method For Detecting And Evaluating Learning Communities in Social Networks*", by Adraoui et al. [2], examines the efficacy of using a Maximal Cliques based method for detecting learning communities, including a dynamic evaluation (social interactions of learners) and a static evaluation (e.g., using demographic information). The idea behind the authors' approach involves the evaluation of learners by communities, instead of using their individual production. The detection of communities proved to be relevant in order to support all the stakeholders (e.g., teachers and administrators) in understanding learners' needs and to finally improve the educational process. The reliability and efficiency of the approach are shown using three real-world scenarios.

Ruipérez-Valiente et al., in their article "*Data-driven Detection and Characterization of Communities of Accounts Collab-*

*Corresponding author and managing guest editor
*Email address:* mirko.marras@acm.org (Mirko Marras)

*orating in MOOCs*" [3], propose a clustering-based approach for the detection and characterization of different collaboration types in MOOCs based on student interactions with the learning platform, without having prior knowledge about the existence of such collaborations. The authors also investigate the behavioral characteristics of the detected communities of accounts (e.g., fruitful, free-riding, and illicit collaboration). Experiments on two MOOCs show that the proposed approach is a viable way for automatically identifying community of accounts and that, while collaborations in MOOCs are generally positive for the learning process, not all the identified students' collaborations can be considered as good or beneficial (e.g., collaborating in dishonest ways to facilitate course completion).

In their article "*Learning Behaviours Data in Programming Education: Community Analysis and Outcome Prediction with Cleaned Data*", Tan Mai et al. [4] focus on identifying communities of learners based on their behaviour in the event logs, instead of analysing the behaviour of each individual student. A further analysis contrast the behaviour of the communities and the final results of students. The experiments are conducted on two university programming courses with around 400 students show that similar behaviour groups have similar performance measure by the final grade. While high performing communities are characterised by more practical related activities, low performing ones mostly passively engage with the course lesson materials.

**Student Success and Dropout Prediction**. This special issue sheds light on advances in the formalization, design, and development of models for student success and dropout prediction.

In their article "*Student Success Prediction using Student Exam Behaviour*", Kuzilek et al. [5] study the impact of using temporal information from exam-taking behaviour on student's success prediction. They propose a new method for encoding students' exam states and use these encoded states as input of different typical predictive models. The experiments show that their proposed approach not only leads to a significant improvement in terms of performance prediction, but also allows to identify critical exam-taking patterns.

Prenkaj et al., in their article "*Hidden Space Deep Sequential Risk Prediction on Student Trajectories*" [6], present an end-to-end deep-learning approach based on raw time series and embedded information (e.g., navigational, forum-based, video-based and homework-based e-tivities) to address the student dropout prediction task. The proposed neural architecture exploits autoencoders to mitigate feature sparsity and stacked embedded gated recurrent units to extract latent information and temporal dependencies even in presence of long temporal gaps. Experiments show that the proposed approach has superior performances in capturing long-term behavioral dependencies and outperforms a range of state-of-the-art shallow and deep learning baselines on two standard benchmarking datasets of online courses (XuetangX and KDDCup15) and a novel released dataset on online university degrees (Unitelma).

The article "*Deep Cognitive Diagnosis Model for Predicting Students' Performance*", by Gao et al. [7], proposes a new generative model of the relations between problems and skills, with the aim of improving students' performance prediction and educational recommendations. The authors represent skills and problems as vectors whose dimensions correspond to difficulty and keywords, and rely on students' mastery of skills and skill interaction to model the problems' proficiency. They then compare their proposed model with two classical cognitive diagnosis models, two latent factor models and two deep learning based models on two datasets, showing that the proposed model has a better predictive performance and interpretability.

**Course Recommendation**. The design and evaluation of intelligent systems able to recommend courses of interest to students are the focus of two articles in this special issue.

The article "*ACMF: An Attention Collaborative Extended Matrix Factorization Based Model for MOOC Course Service via a Heterogeneous View*", by Sheng et al. [8], focuses on the problem of personalized course recommendation in a MOOC environment. The authors propose a new model, based on recent methodologies from the domains of graph learning and recommender systems, that captures structures and semantics. The experiments compare the proposed model against a number of node-embedding baselines and recommendation methods on two datasets. The results show a good robustness against data sparsity and imbalance.

The article "*Enabling Cross-continent Provider Fairness in Educational Recommender Systems*" [9] by Gomez et al., deals with the concept of provider fairness for demographic groups of teachers that share the same continent of provenience. The authors consider the notion of fairness based on equity, which compare the share of recommendations of a group with its representation in the data. They assess the presence of disparities on data coming from a MOOC platform by applying state-of-the-art collaborative filtering approaches (ranging from point-wise to pair-wise as well as from memory-based to model-based algorithms). The authors find out that there are disparities in the visibility and exposure at expenses of smaller demographic groups. These disparities are mitigated with a novel multi-class approach that regulates the visibility and exposure given to each group, without affecting recommendation effectiveness for the learners, and thus allowing cross-continent provider fairness.

**Cognitive State Analysis and Classification**. Two articles of this special issue focus on supporting a better understanding of students' cognitive states from diverse viewpoints.

Abate et al., in their article "*Attention Monitoring for Synchronous Distance Learning*" [10], introduce a system for providing indicators on the didactic efficacy of the lecture to the teacher, avoiding the need of activating learner cameras while using a video conference system for synchronous distance lectures. A software module runs in background and locally on the learners' computers for tracing their blinks, gaze and expressions (no sensitive information is shared through the network), while another software module automatically analyzes and aggregates information in a user interface provided to the teacher, including a heat-map that highlights the parts of the slide the learners are focusing on the most and the distribution of the classified expressions. Experiments with volunteers, ranging from university students to employees engaged in training activities, result in positive feedback in terms of gaze tracking and evaluation questionnaires, from both learner and teacher sides.

The article "*Classifying Students based on Cognitive State in Flipped Learning Pedagogy*", by Shaw and Patra [11], investigates students' learning while preparing for a lesson in a flipped classroom scenario by analysing their cognitive state based on their brain signals. Focusing on the EEG signal, the authors designed a method, based on a siamese neural network, that can classify students into three categories according to their attention. Experiments show that this approach outperforms other state-of-the-art classification methods. Once these sensors are more spread , this approach can help teachers identify students who are not paying enough attention and not learning despite putting effort into studying.

**Teaching Content Analysis and Classification**. Finally, the papers of this special issue investigate the automated analysis and classification of teaching material and strategies.

In their article "*Is it a Good Move? Mining Effective Tutoring Strategies from Human-Human Tutorial Dialogues*", Lin et al. [12] investigate automated approaches for detecting both effective and ineffective tutoring strategies, by mining a large-scale dialogue corpus of online human-human tutoring. First, the authors adopt a widely-used educational dialogue act scheme to describe the action behind the utterance (e.g., asking/answering a question, providing hints). Then, they apply a sequence analysis on the inferred actions to identify prominent patterns closely related to students' problem-solving performance. Lastly, they use these labelled tutorial actions as input to a well-established machine learning approach to predict students' problem-solving strategies. Experiments on a large dataset of tutoring sessions contribute to a better understanding of tutors' as well as students' behavior in human-human online tutoring and show that the proposed approach can be practically used to locate (un)successful dialogues in an automated way.

Gasparetti, in his article "*Discovering Prerequisite Relations from Educational Documents through Word Embeddings*" [13], presents a novel method to discover prerequisites in documents, taking advantage of latent representations for their identification in a binary classification setting. The proposed methodology employs various machine learning algorithms and is tested on four different datasets which vary in terms of covered subjects and amount of instances. The results indicate that, although the underlying architecture is simple, the proposed approach achieves an accuracy close to complex deep learning architectures based on recurrent neural networks and transformers.

## 3. Conclusions

Intelligent systems empowered by data mining and machine learning and tailored to online education are fostering increasingly intense research to ensure that these systems can support stakeholders in an effective, efficient, adaptive and timely way. The articles included in this special issue cover a range of interesting topics and highlight important next steps in this active and rapidly evolving field of research and development. We hope that the readers will find this selection of articles informative and helpful in keeping themselves up-to-date on the current challenges and directions, and that the content of these articles can contribute to timely drive their future research progresses.

## Acknowledgments

We thank all the authors for considering this special issue as an outlet to publish their research results in the area of intelligent systems for online education. We express our deepest gratitude to all the reviewers who devoted much of their precious time providing useful and thoughtful comments on the articles submitted to this special issue. We warmly thank the Editor-in-Chief, the Special Content Editor, and the journal office team of FGCS, for their kind support, advice, and encouragements throughout the management of this special issue.

## References

[1] L. Chen, P. Chen, Z. Lin, Artificial intelligence in education: A review, IEEE Access 8 (2020) 75264–75278. doi:10.1109/ACCESS.2020.2988510.
URL https://doi.org/10.1109/ACCESS.2020.2988510

[2] M. Adraoui, A. Retbi, M. K. Idrissi, S. Bennani, Maximal cliques based method for detecting and evaluating learning communities in social networks, Future Generation Computer Systems 126 (2022) 1–14. doi:https://doi.org/10.1016/j.future.2021.07.034.

[3] J. A. R. Valiente, D. Jaramillo-Morillo, S. Joksimovic, V. Kovanovic, P. J. Muñoz-Merino, D. Gasevic, Data-driven detection and characterization of communities of accounts collaborating in moocs, Future Gener. Comput. Syst. 125 (2021) 590–603. doi:10.1016/j.future.2021.07.003.

[4] T. T. Mai, M. Bezbradica, M. Crane, Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data, Future Generation Computer Systemsdoi:https://doi.org/10.1016/j.future.2021.08.026.

[5] J. Kuzilek, Z. Zdráhal, V. Fuglik, Student success prediction using student exam behaviour, Future Gener. Comput. Syst. 125 (2021) 661–671. doi:10.1016/j.future.2021.07.009.

[6] B. Prenkaj, D. Distante, S. Faralli, P. Velardi, Hidden space deep sequential risk prediction on student trajectories, Future Gener. Comput. Syst. 125 (2021) 532–543. doi:10.1016/j.future.2021.07.002.

[7] L. Gao, Z. Zhao, C. Li, J. Zhao, Q. Zeng, Deep cognitive diagnosis model for predicting students' performance, Future Generation Computer Systems 126 (2022) 252–262. doi:https://doi.org/10.1016/j.future.2021.08.019.

[8] D. Sheng, J. Yuan, Q. Xie, L. Li, Acmf: An attention collaborative extended matrix factorization based model for mooc course service via a heterogeneous view, Future Generation Computer Systems 126 (2022) 211–224. doi:https://doi.org/10.1016/j.future.2021.08.001.

[9] E. Gómez, C. S. Zhang, L. Boratto, M. Salamó, G. Ramos, Enabling cross-continent provider fairness in educational recommender systems, Future Generation Computer Systemsdoi:https://doi.org/10.1016/j.future.2021.08.025.

[10] A. F. Abate, L. Cascone, M. Nappi, F. Narducci, I. Passero, Attention monitoring for synchronous distance learning, Future Gener. Comput. Syst. 125 (2021) 774–784. doi:10.1016/j.future.2021.07.026.

[11] R. Shaw, B. K. Patra, Classifying students based on cognitive state in flipped learning pedagogy, Future Generation Computer Systems 126 (2022) 305–317. doi:https://doi.org/10.1016/j.future.2021.08.018.

[12] J. Lin, S. S. Singh, L. Sha, W. Tan, D. Lang, D. Gasevic, G. Chen, Is it a good move? mining effective tutoring strategies from human-human tutorial dialogues, Future Gener. Comput. Syst.

[13] F. Gasparetti, Discovering prerequisite relations from educational documents through word embeddings, Future Generation Computer Systemsdoi:https://doi.org/10.1016/j.future.2021.08.021.