

## Preface to the special issue on data mining for personalised educational systems

Cristóbal Romero · Sebastián Ventura

Received: 5 January 2011 / Accepted: 5 January 2011  
© Springer Science+Business Media B.V. 2011

1 Educational data mining (EDM) is an emerging interdisciplinary research area which  
2 deals with the development of methods for the exploration of data which have origi-  
3 nated in an educational context (Baker and Yacef 2009). EDM uses statistical, machine  
4 learning and data mining (DM) approaches to analyse educational data in order to  
5 study educational issues. On the one hand, DM has many applications, and many tasks  
6 in educational environments have been resolved through DM (Romero and Ventura  
7 2010), such as the analysis and visualisation of data, providing feedback in order  
8 to support instructors, providing recommendations for students, student modelling,  
9 detecting undesirable student behaviours, predicting student performance, grouping  
10 students, social networks analysis, developing concept maps, constructing course-  
11 ware, planning and scheduling, etc. In particular, this special issue of User Modelling  
12 and User-Adapted Interaction explores recent developments and applications of data  
13 mining techniques in various aspects of user modelling and the personalisation, rec-  
14 ommendation and adaptation of educational systems. On the other hand, there are  
15 different types of educational systems, such as traditional classrooms, test/quiz sys-  
16 tems, e-learning systems, learning management systems (LMS), adaptive educational  
17 hypermedia systems (AEHS), intelligent tutoring systems (ITS), etc. (Romero and  
18 Ventura 2007). In particular, AEHS and ITS are adaptive and intelligent educational  
19 systems which attempt to be more adaptable by building a model of the goals, prefer-  
20 ences and knowledge of each individual student and using this model throughout the  
21 interaction with the student in order to adapt to their individual needs (Brusilovsky and  
22 Peylo 2003). These specific types of educational systems aim to fulfill the ultimate goal

---

C. Romero (✉) · S. Ventura  
Department of Computer Sciences and Numerical Analysis, University of Córdoba, Córdoba, Spain  
e-mail: cromero@uco.es

S. Ventura  
e-mail: sventura@uco.es

of any user-adaptive system, which is to provide users with what they need without being explicitly asked to do so. Therefore, these systems must be able to infer what a user requires based on previous or current interactions with that user, and possibly other users (Mobasher and Tuzhilin 2009). In order to resolve this issue, data mining can be used to develop user/student models as well as automatic educational systems for personalisation and recommendation (Romero and Ventura 2010). For example, student modelling can be used to develop cognitive models of human users/students, including a model of their skills and declarative knowledge. Automatic personalisation can help to adapt teaching methods to fit the needs of each particular student, and educational recommendation systems can make suggestions to the students with regard to the next chapter or subject they should visit, the next task or problem to be addressed, the what learning material to read, etc. Different data mining techniques and methods have been used for all of these tasks, such as clustering, classification, prediction, association, etc. In fact, this special issue presents six papers which represent a number of emerging trends in data mining for personalised educational systems.

The first paper by Su, Tseng, Lin and Chen propose a personalised learning content adaption mechanism which applies data mining techniques (including clustering and decision trees). Their objective is to efficiently manage a large number of historical learner's requests, and to intelligently deliver suitable personalized learning content from a learning object repository. They consider not only the mobile device's hardware capabilities and the network conditions of a mobile learning environment, but also the diverse needs of the learners.

The paper by Köck and Paramythis presents an approach based on the modelling of learners' problem solving activity sequences in ITS, and on the use of models in targeted, and ultimately automated, multilevel clustering. This results in the discovery of new, semantically meaningful information about the learners. They show how monitoring and interpreting sequential learner activities has the potential to improve adaptivity and personalisation within educational environments.

Muldner, Burleson, Van de Sande and VanLehn describe a high school student's gaming detector in an ITS with the purpose of automatically labelling data, and apply several data mining techniques, including machine learning of Bayesian network parameters. They try to reduce gaming by understanding when and why it occurs, what causes gaming and what impact it has on students, and by analysing the predictors of gaming. Their primary research question is whether the student or problem features is a better predictor of gaming.

The paper by Chi, VanLehn, Litman and Jordan presents a detailed methodology for using reinforcement learning in order to optimise the use of effective and adaptive pedagogical strategies based on limited interactions with human students. These pedagogical strategies help to decide which action to take next when there are multiple options available. They empirically validate the utility of the approach in the context of highly interactive e-learning environments, ITS and natural language tutoring systems.

Anaya and Boticario propose an approach for modelling students' collaboration in terms of general features which provides students with timely relevant information and an evaluation of their collaborative act. This approach analyses the students' interactions in order to evaluate their collaboration using two DM methods named clustering and collaboration metrics with decision tree algorithms. They try to help

69 clarify which user modelling issues should be considered in computer-supported col-  
70 laborative learning (CSCL).

71 Finally, Vialardi, Chue, Peche, Alvarado, Vinatea, Estrella and Ortigosa describe  
72 a recommender system designed to support the enrollment process using previously  
73 acquired knowledge about the students. These researchers propose a methodology  
74 based on data preparation and data mining using classification, in order to suggest  
75 elements that can help for better decision-making. Their objective is to guide students  
76 during the regular enrollment process by offering an additional criterion to help them  
77 to make their decisions.

78 We hope that the reader will find the approaches and research ideas described in  
79 these papers interesting, informative, and that they will contribute to further developed  
80 knowledge in this research area about personalized educational systems.

## 81 References

- 82 Baker, R., Yacef, K.: The state of educational data mining in 2009: a review and future visions. *J. Educ.*  
83 *Data Min.* **1**(1), 3–17 (2009)
- 84 Brusilovsky, P., Peylo, C.: Adaptive and intelligent web-based educational systems. *Int. J. Artif. Intell.*  
85 *Educ.* **13**, 156–169 (2003)
- 86 Mobasher, B., Tuzhilin, A.: Preface to the special issue on data mining for personalization. *User Model.*  
87 *User-Adapt. Interact.* **19**, 1–3 (2009)
- 88 Romero, C., Ventura, S.: Educational data mining: a survey from 1995 to 2005. *Expert Syst. Appl.*  
89 **1**(33), 135–146 (2007)
- 90 Romero, C., Ventura, S.: Educational data mining: a review of the state-of-the-art. *IEEE Trans. Syst. Man*  
91 *Cybern. C* **40**(6), 601–618 (2010)

## 92 Author Biographies

93

94 **Cristóbal Romero** is currently an Associate Professor in the Department of Computer Science and Numerical  
95 Analysis at the University of Córdoba. He received his B.Sc. and Ph.D. degrees in computer science  
96 from the University of Granada, Spain, in 1996 and 2003 respectively. He has published more than 40  
97 international papers, 15 of which have been published in international journals. He is the co-editor of two  
98 books specifically regarding EDM. His current research interest is focussed on the application of data min-  
99 ing in e-learning systems. Dr. Romero is a member of the IEEE Computer Society, the International EDM  
100 Working Group and the steering committee of several conferences about education, personalisation and  
101 data mining.

102 **Sebastián Ventura** is currently an Associate Professor in the Department of Computer Science and Numerical  
103 Analysis at the University of Córdoba, where he is the head of the Knowledge Discovery and Intelligent  
104 Systems Research Laboratory. He received his B.Sc. and Ph.D. degrees in sciences from the University of  
105 Córdoba, Spain, in 1989 and 1996 respectively. He has published more than 80 international publications,  
106 30 of which have been published in international journals. He has also been engaged in eleven research  
107 projects (being the coordinator of two of them) supported by the Spanish and Andalusian governments  
108 and the European Union. His current main research interests are in the fields of soft-computing, machine  
109 learning, data mining and their applications. Dr. Ventura is a senior member of the IEEE Computer, the  
110 IEEE Computational Intelligence and the IEEE Systems, Man and Cybernetics Societies as well as the  
111 Association of Computing Machinery (ACM).