International Conference on Educational Data Mining (EDM) 2011
Proceedings of the 4th International Conference on Educational Data Mining
Mykola Pechenizkiy, Toon Calders, Cristina Conati, Sebastian Ventura, Cristobal Romero and John Stamper (eds.)
Eindhoven, 6-8 July, 2011

A catalogue record is available from the Eindhoven University of Technology Library


Cover: Flying pins with the background the Chamber of Commerce located at the TU/e campus.
Printing and binding: TU/e printservice
Preface

The 4th International Conference on Educational Data Mining (EDM 2011) brings together researchers from computer science, education, psychology, psychometrics, and statistics to analyze large datasets to answer educational research questions. The conference, held in Eindhoven, The Netherlands, July 6-9, 2011, follows the three previous editions (Pittsburgh 2010, Cordoba 2009 and Montreal 2008), and a series of workshops within the AAAI, AIED, EC-TEL, ICALT, ITS, and UM conferences.

The increase of e-learning resources such as interactive learning environments, learning management systems, intelligent tutoring systems, and hypermedia systems, as well as the establishment of state databases of student test scores, has created large repositories of data that can be explored to understand how students learn. The EDM conference focuses on data mining techniques for using these data to address important educational questions. The broad collection of research disciplines ensures cross fertilization of ideas, with the central questions of educational research serving as a unifying focus.

This year’s conference includes short papers as a new submission category targeting original and unpublished research with merit in terms of originality and importance rather than maturity and technical validation. In the paper track, we received 60 long and 20 short papers, each of which was reviewed by three experts in the field, resulting in 20 long and 17 short papers accepted for presentation at the conference (some of the long paper submissions have been accepted as short paper). We also received 22 posters, targeting work in progress and last minute results with high potential to foster new developments and interesting discussions during the conference’s poster presentation sessions. These sessions included the presentation of 30 posters, 14 from the original pool of poster submissions and the reminder from the pool of paper submissions.

All accepted submissions appear in these proceedings. The conference also includes invited talks by Barry Smyth (University College, Dublin, Ireland), John Stamper (Carnegie Mellon University, USA) and Erik-Jan van der Linden (MagnaView B.V., the Netherlands), with abstract in these proceedings.

We would like to thank Eindhoven University of Technology for the sponsorship and hosting of EDM’2011. We would like to thank the Netherlands Organization for Scientific Research (NWO), Belgium-Netherlands Association for Artificial Intelligence (BNVKI) and the Dutch Research School for Information and Knowledge Systems (SIKS), University of Cordoba and PSLC DataShop.

We also want to acknowledge the amazing work of the program committee members and additional reviewers, who with their enthusiastic contributions gave us invaluable support in putting this conference together.

Our special thanks to the local organizing team and additional thanks to Evgeny Knutov and Jorn Bakker for their technical support on putting these proceedings together.

Last but not least we would like to thank Arnon Hershkovitz who has served as the Web Chair of EDM series from its first edition.

June 2011

Cristina Conati and Sebastian Ventura – Program Chairs
Mykola Pechenizkiy and Toon Calders – Conference Chairs
Cristobal Romero and John Stamper – Posters Chairs
## Organization

### CONFERENCE CHAIRS
- Mykola Pechenizkiy  Eindhoven University of Technology, The Netherlands
- Toon Calders  Eindhoven University of Technology, The Netherlands

### PROGRAM CHAIRS
- Cristina Conati  University of British Columbia, Canada
- Sebastian Ventura  University of Cordoba, Spain

### POSTERS CHAIRS
- Cristobal Romero  University of Cordoba, Spain
- John Stamper  Carnegie Mellon University, USA

### WEB CHAIR
- Arnon Hershkovitz  Tel Aviv University, Israel

### LOCAL ORGANIZING TEAM
- Jorn Bakker
- Paul De Bra
- Lam Hoang
- Evgeny Knutov
- Riet van Buul
- Ekaterina Vasilyeva
- Yongming Luo
- George Fletcher
- Ine van der Ligt

### STEERING COMMITTEE
- Esma Aïmeur  University of Montreal, Canada
- Ryan Baker  Worcester Polytechnic Institute, USA
- Tiffany Barnes  University of North Carolina at Charlotte, USA
- Joseph E. Beck  Worcester Polytechnic Institute, USA
- Michel C. Desmarais  Ecole Polytechnique de Montreal, Canada
- Neil Heffernan  Worcester Polytechnic Institute, USA
- Cristobal Romero  Cordoba University, Spain
- Kalina Yacef  University of Sydney, Australia

### PROGRAM COMMITTEE
- Esma Aimeur  University of Montreal, Canada
- Elizabeth Ayers  Carnegie Mellon University, USA
- Ryan S.J.d. Baker  Worcester Polytechnic Institute, USA
- Tiffany Barnes  University of North Carolina at Charlotte, USA
- Joseph Beck  Worcester Polytechnic Institute, USA
- Bettina Berendt  Katholieke Universiteit Leuven, Belgium
- Gautam Biswas  Vanderbilt University, USA
- Jesús G. Boticario  U.N.E.D., Spain
- Min Chi  University Of Pittsburgh, USA
- Cristophe Choquet  Université du Maine, France
- Richard Cox  University of Sussex, UK
- Michel Desmarais  Ecole Polytechnique de Montreal, Canada
- Sidney D'Mello  University of Memphis, USA
- Mingyu Feng  SRI International, USA
- Davide Fosatti  Carnegie Mellon University, Qatar
- Eva Gibaja  Universidad de Córdoba, USA
- Daniela Godoy  Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina
Neil Heffernan  Worcester Polytechnic Institute, USA
Arnon Hershkovitz  Worcester Polytechnic Institute, Israel
Roland Hubscher  Bentley University, USA
Sebastian Iksal  Université du Maine, France
Juday Kay  University of Sydney, Australia
Jihie Kim  University of Southern California, USA
Mirjam Köck  Johannes Kepler University, Austria
Kenneth Koedinger  Carnegie Mellon University, USA
Vanda Luengo  Université Joseph Fourier Grenoble, France
Tara Madhyastha  University of Washington, USA
Brent Martin  Canterbury University, New Zealand
Noboru Matsuda  Carnegie Mellon University, USA
Manolis Mavrikis  The University of Edinburgh, UK
Riccardo Mazza  University of Lugano/University of Applied Sciences of South. Switzerland
Gordon McCalla  University of Saskatchewan, Canada
Agathe Merceron  Beuth University of Applied Sciences, Germany
Julia Mingullon Alfonso  Universitat Oberta de Catalunya, Spain
Jack Mostow  Carnegie Mellon University, USA
Rafi Nachmias  Tel Aviv University, Israel
Roger Nkambou  Université du Québec à Montréal (UQAM), Canada
Alvaro Ortigosa  Universidad Autónoma de Madrid, Spain
Alexandros Paramythits  Johannes Kepler University, Austria
Philip I. Pavlik  Carnegie Mellon University, USA
Mykola Pechenizkiy  Eindhoven University of Technology, Netherlands
Cristobal Romero  Cordoba University, Spain
Carolyn Rose  Carnegie Mellon University, USA
Erin Shaw  University of Southern Californi, USA
John Stamper  Carnegie Mellon University, USA
Jun-Ming Su  National Chiao Tung University, Taiwan
Steven Tanimoto  University of Washington, USA
Sebastian Ventura  Cordoba University, Spain
Stephehn Weibelzahl  National College of Ireland, Ireland
Kalina Yacef  University of Sydney, Australia
Michael Yudelson  University of Pittsburgh, WPI
Amelia Zafras  Universidad de Córdoba, Spain
Osmar Zaiane  University of Alberta, Canada

ADDITIONAL REVIEWERS
Yue Gong  Sujith Gowda  John Kinnebrew  Daniel Mack
Zach Pardos  Terry Peckham  Michael Sao Pedro  Soo Won Seo
Benjamin Shih  Vilaythong Southavilay  Fodé Touré  Jianfei Wu
Jaebong Yoo
Sponsors

http://www.tue.nl
http://www.unimaas.nl/bnvki
http://www.nwo.nl
http://www.siks.nl
https://pslcdatashop.web.cmu.edu/
http://www.uco.es
## Table of Contents

### Invited Talks (abstracts)

Social Information Discovery  
*Barry Smyth*  
3

On exploration and mining of data in educational practice  
*Erik-Jan van der Linden, Martijn Wijffelaars, Thomas Lammers*  
5

EDM and the 4th Paradigm of Scientific Discovery - Reflections on the 2010 KDD Cup Competition  
*John Stamper*  
7

### Full Papers

Factorization Models for Forecasting Student Performance  
*Nguyen Thai-Nghe, Tomáš Horváth and Lars Schmidt-Thieme*  
11

Analyzing Participation of Students in Online Courses Using Social Network Analysis Techniques  
*Reihaneh Rabbany Khorasgani, Mansoureh Tukaffoli and Osmar Zaïane*  
21

A Machine Learning Approach for Automatic Student Model Discovery  
*Nan Li, William Cohen, Kenneth R. Koedinger and Noboru Matsuda*  
31

Conditions for effectively deriving a Q-Matrix from data with Non-negative Matrix Factorization  
*Micel Desmarais*  
41

Student Translations of Natural Language into Logic: The Grade Grinder Translation Corpus Release 1.0  
*Dave Barker-Plummer, Richard Cox and Robert Dale*  
51

Instructional Factors Analysis: A Cognitive Model For Multiple Instructional Interventions  
*Min Chi, Kenneth Koedinger, Geoff Gordon, Pamela Jordan and Kurt Vanlehn*  
61

The Simple Location Heuristic is Better at Predicting Students Changes in Error Rate Over Time Compared to the Simple Temporal Heuristic  
*Adaeze Nwaigwe and Kenneth Koedinger*  
71

Items, skills, and transfer models: which really matters for student modeling?  
*Yue Gong and Joseph Beck*  
81

Avoiding Problem Selection Thrashing with Conjunctive Knowledge Tracing  
*Kenneth Koedinger, Philip I. Pavlik Jr., John Stamper, Tristan Nixon and Steven Ritter*  
91

Less is More: Improving the Speed and Prediction Power of Knowledge Tracing by Using Less Data  
*Bahador Nooraei B., Zachary Pardos, Neil Heffernan and Ryan Baker*  
101
Analyzing frequent sequential patterns of collaborative learning activity around an interactive tabletop

Roberto Martinez Maldonado, Kalina Yacef, Judy Kay, Ahmed Kharrufa and Ammar Al-Qaraghuli

Acquiring Item Difficulty Estimates: a Collaborative Effort of Data and Judgment

Kelly Wauters, Piet Desmet and Wim Van Den Noortgate

Spectral Clustering in Educational Data Mining

Shubhendu Trivedi, Zachary Pardos, Gábor Sárközy and Neil Heffernan

Does Time Matter? Modeling the Effect of Time with Bayesian Knowledge Tracing

Yumeng Qiu, Yingmei Qi, Hanyuan Lu, Zachary Pardos and Neil Heffernan

Learning classifiers from a relational database of tutor logs

Jack Mostow, José González-Brenes and Bao Hong Tan

A Framework for Capturing Distinguishing User Interaction Behaviors in Novel Interfaces

Samad Kardan and Cristina Conati

How to Classify Tutorial Dialogue? Comparing Feature Vectors vs. Sequences

José González-Brenes, Jack Mostow and Weisi Duan

Automatically Detecting a Students Preparation for Future Learning: Help Use is Key

Ryan S.J.D. Baker, Sujith Gowda and Albert Corbett

Ensembling Predictions of Student Post-Test Scores for an Intelligent Tutoring System

Zachary Pardos, Sujith Gowda, Ryan S.J.D. Baker and Neil Heffernan

Improving Models of Slipping, Guessing, and Moment-By-Moment Learning with Estimates of Skill Difficulty

Sujith M. Gowda, Jonathan P. Rowe, Ryan S.J.D. Baker, Min Chi and Kenneth R. Koedinger

Short Papers

A Method for Finding Prerequisites Within a Curriculum

Annalies Vuong, Tristan Nixon and Brendon Towle

Estimating Prerequisite Structure From Noisy Data

Emma Brunskill

What can closed sets of students and their marks say?

Dmitry Ignatov, Serafima Mamedova, Nikita Romashkin, and Ivan Shamshurin

How university entrants are choosing their department? Mining of university admission process with FCA taxonomies.

Nikita Romashkin, Dmitry Ignatov and Elena Kolotova
What’s an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis

Marcelo Worsley and Paulo Blikstein

Using Logistic Regression to Trace Multiple Subskills in a Dynamic Bayes Net

Yanbo Xu and Jack Mostow

Monitoring Learners Proficiency: Weight Adaptation in the Elo Rating System

Kelly Wauters, Piet Desmet and Wim Van Den Noortgate

Modeling students activity in online discussion forums: a strategy based on time series and agglomerative hierarchical clustering

Germán Cobo, David García-Solórzano, Eugénia Santamaría, José Antonio Morán, Javier Melencón and Carlos Monzo

Prediction of Perceived Disorientation in Online Learning Environment with Random Forest Regression

Gökhan Akçapinar, Erdal Cosgun and Arif Altun

Analysing Student Spatial Deployment in a Computer Laboratory

Vladimir Ivancevic, Milan Celikovic and Ivan Lukovic

Predicting School Failure Using Data Mining

Carlos Marquez-Vera, Cristobal Romero and Sebastín Ventura

A Dynamical System Model of Microgenetic Changes in Performance, Efficacy, Strategy Use and Value during Vocabulary Learning

Philip I. Pavlik Jr. and Sue-Mei Wu

Desperately Seeking Subscripts: Towards Automated Model Parameterization

Jack Mostow, Yanbo Xu and Mdahaduzzaman Munna

Automatic Generation of Proof Problems in Deductive Logic

Behrooz Mostafavi, Tiffany Barnes and Marvin Croy

Comparison of Traditional Assessment with Dynamic Testing in a Tutoring System

Mingyu Feng, Neil Heffernan, Zachary Pardos and Cristina Heffernan

Evaluating a Bayesian Student Model of Decimal Misconceptions

George Goguadze, Sergey Sosnovsky, Seiji Isotani and Bruce McLaren

Exploring user data from a game-like math tutor: a case study in causal modeling

Dovan Rai and Joseph Beck

Posters

Goal Orientation and Changes of Carelessness over Consecutive Trials in Science Inquiry

Arnon Hershkovitz, Ryan S.J.D. Baker, Janice Gobert and Michael Wixon

Towards improvements on domain-independent measurements for collaborative assessment

Antonio R. Anaya and Jesús G. Boticario
A Java desktop tool for mining Moodle data
Rafael Pedraza Perez, Cristobal Romero and Sebastián Ventura

Using data mining in a recommender system to search for learning objects in repositories
Alfredo Zapata Gonzalez, Victor Hugo Menéndez Domínguez, Manuel Prieto and Cristobal Romero

E-learning Web Miner: A data mining application to help instructors involved in virtual courses
Diego García-Saiz and Marta Zorrilla Pantaleón

Computerized Coding System for Life Narratives to Assess Students’ Personality Adaption
Qiwei He, Bernard Veldkamp and Gerben Westerhof

Partially Observable Sequential Decision Making for Problem Selection in an Intelligent Tutoring System
Emma Brunskill and Stuart Russell

Mining Teaching Behaviors from Pedagogical Surveys
Joana Barracosa and Claudia Antunes

Variable Construction and Causal Modeling of Online Education Messaging Data: Initial Results
Stephen Fancsali

The Hospital Classrooms Environments Challenge
Carina González and Pedro A. Toledo

Combining study of complex network and text mining analysis to understand growth mechanism of communities on SNS
Osamu Yamakawa, Takahiro Tagawa, Hitoshi Inoue, Koichi Yastake and Takahiro Sumiya

Logistic Regression in a Dynamic Bayes Net Models Multiple Subskills Better!
Yanbo Xu and Jack Mostow

Studying the problem-solving strategies in the early stages of learning programming
Edgar Cambranes-Martinez and Judith Good

Brick: Mining Pedagogically Interesting Sequential Patterns
Anjo Anjewierden, Hannie Gijlers, Nadira Saab and Robert De Hoog

Intelligent evaluation of social knowledge building using conceptual maps with MLN
Lorenzo Moreno, Carina González, Román Estévez and Beatrice Popescu

Identifying Influence Factors on Students Success by Subgroup Discovery
Florian Lemmerich, Marianus Ifland and Frank Puppe

Analyzing University Data for Determining Student Profiles and Predicting Performance
Dorina Kabakchieva, Kamelia Stefanova and Valentin Kisimov

The EDM Vis Tool
Matthew Johnson, Michael Eagle, Leena Joseph and Tiffany Barnes
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Towards Modeling Forgetting and Relearning in ITS: Preliminary Analysis of ARRS Data</td>
<td>351</td>
</tr>
<tr>
<td>Yutao Wang and Neil Heffernan</td>
<td></td>
</tr>
<tr>
<td>Quality Control and Data Mining Techniques Applied to Monitoring Scaled Scores</td>
<td>353</td>
</tr>
<tr>
<td>Alina Von Davier</td>
<td></td>
</tr>
<tr>
<td>eLAT: An Exploratory Learning Analytics Tool for Reflection and Iterative Improvement of Technology Enhanced Learning</td>
<td>355</td>
</tr>
<tr>
<td>Anna Lea Dyckhoff, Dennis Zielke, Mohamed Amine Chatti and Ulrik Schroeder</td>
<td></td>
</tr>
<tr>
<td>Predicting graduate-level performance from undergraduate achievements</td>
<td>357</td>
</tr>
<tr>
<td>Judith Zimmermann, Kay H. Brodersen, Jean-Philippe Pellet, Elias August and Joachim M. Buhmann</td>
<td></td>
</tr>
<tr>
<td>Mining Assessment and Teaching Evaluation Data of Regular and Advanced Stream Students</td>
<td>359</td>
</tr>
<tr>
<td>Irena Koprinska</td>
<td></td>
</tr>
<tr>
<td>Investigating Usage of Resources in LMS with Specific Association Rules</td>
<td>361</td>
</tr>
<tr>
<td>Agathe Merceron</td>
<td></td>
</tr>
<tr>
<td>Towards parameter-free data mining: Mining educational data with yacaree</td>
<td>363</td>
</tr>
<tr>
<td>Jose Balcázar, Diego García-Saiz and Marta Zorrilla</td>
<td></td>
</tr>
<tr>
<td>Factors Impacting Novice Code Comprehension in a Tutor for Introductory Computer Science</td>
<td>365</td>
</tr>
<tr>
<td>Leigh Ann Sudol DeLyser and Jonathan Steinhart</td>
<td></td>
</tr>
<tr>
<td>Investigating the Transitions between Learning and Non-learning Activities as Students Learn Online</td>
<td>367</td>
</tr>
<tr>
<td>Paul Salvador Inventado, Roberto Legaspi, Merlin Suarez and Masayuki Numao</td>
<td></td>
</tr>
<tr>
<td>Learning parameters for a knowledge diagnostic tools in orthopedic surgery</td>
<td>369</td>
</tr>
<tr>
<td>Sébastien Lallé and Vanda Luengo</td>
<td></td>
</tr>
<tr>
<td>Problem Response Theory and its Application for Tutoring</td>
<td>371</td>
</tr>
<tr>
<td>Petr Jarušek and Radek Pelánek</td>
<td></td>
</tr>
<tr>
<td>Towards Better Understanding of Transfer in Cognitive Models of Practice</td>
<td>373</td>
</tr>
<tr>
<td>Michael Yudelson, Philip I. Pavlik and Kenneth R. Koedinger</td>
<td></td>
</tr>
</tbody>
</table>