

EDM 2011

**4th International Conference on
Educational Data Mining**

PROCEEDINGS OF THE
FOURTH INTERNATIONAL CONFERENCE ON
EDUCATIONAL DATA MINING

Eindhoven, July 6-8, 2011

**Mykola Pechenizkiy, Toon Calders,
Cristina Conati, Sebastian Ventura,
Cristobal Romero and John Stamper**

Mykola Pechenizkiy, Toon Calders, Cristina Conati, Sebastian Ventura,
Cristobal Romero and John Stamper

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

ISBN: 978-90-386-2537-9

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 AAI, 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This year's conference includes short papers as a new submission category targeting original and unpublished research with merit in terms of originality and importance rather than maturity and technical validation.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 remainder 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 Aimeur 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 Paramythis	Johannes Kepler University, Austria
Philip I. Pavlik	Carnegie Mellon University, USA
Mykola Pechenzkiy	Eindhoven University of Technology, Netherlands
Cristobal Romero	Cordoba University, Spain
Carolyn Rose	Carnegie Mellon University, USA
Erin Shaw	University of Southern California, 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
Stepehn Weibelzahl	National College of Ireland, Ireland
Kalina Yacef	University of Sydney, Australia
Michael Yudelson	University of Pittsburgh, WPI
Amelia Zafra	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	3
<i>Barry Smyth</i>	
On exploration and mining of data in educational practice	5
<i>Erik-Jan van der Linden, Martijn Wijffelaars, Thomas Lammers</i>	
EDM and the 4th Paradigm of Scientific Discovery - Reflections on the 2010 KDD Cup Competition	7
<i>John Stamper</i>	

Full Papers

Factorization Models for Forecasting Student Performance	11
<i>Nguyen Thai-Nghe, Tomáš Horváth and Lars Schmidt-Thieme</i>	
Analyzing Participation of Students in Online Courses Using Social Network Analysis Techniques	21
<i>Reihaneh Rabbany Khorasgani, Mansoureh Takaffoli and Osmar Zaiane</i>	
A Machine Learning Approach for Automatic Student Model Discovery	31
<i>Nan Li, William Cohen, Kenneth R. Koedinger and Noboru Matsuda</i>	
Conditions for effectively deriving a Q-Matrix from data with Non-negative Matrix Factorization	41
<i>Michel Desmarais</i>	
Student Translations of Natural Language into Logic: The Grade Grinder Translation Corpus Release 1.0	51
<i>Dave Barker-Plummer, Richard Cox and Robert Dale</i>	
Instructional Factors Analysis: A Cognitive Model For Multiple Instructional Interventions	61
<i>Min Chi, Kenneth Koedinger, Geoff Gordon, Pamela Jordan and Kurt Vanlehn</i>	
The Simple Location Heuristic is Better at Predicting Students Changes in Error Rate Over Time Compared to the Simple Temporal Heuristic	71
<i>Adaeze Nwaigwe and Kenneth Koedinger</i>	
Items, skills, and transfer models: which really matters for student modeling?	81
<i>Yue Gong and Joseph Beck</i>	
Avoiding Problem Selection Thrashing with Conjunctive Knowledge Tracing	91
<i>Kenneth Koedinger, Philip I. Pavlik Jr., John Stamper, Tristan Nixon and Steven Ritter</i>	
Less is More: Improving the Speed and Prediction Power of Knowledge Tracing by Using Less Data	101
<i>Bahador Nooraei B., Zachary Pardos, Neil Heffernan and Ryan Baker</i>	

Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop	111
<i>Roberto Martinez Maldonado, Kalina Yacef, Judy Kay, Ahmed Kharrufa and Ammar Al-Qaraghuli</i>	
Acquiring Item Difficulty Estimates: a Collaborative Effort of Data and Judgment	121
<i>Kelly Wauters, Piet Desmet and Wim Van Den Noortgate</i>	
Spectral Clustering in Educational Data Mining	129
<i>Shubhendu Trivedi, Zachary Pardos, Gábor Sárközy and Neil Heffernan</i>	
Does Time Matter? Modeling the Effect of Time with Bayesian Knowledge Tracing	139
<i>Yumeng Qiu, Yingmei Qi, Hanyuan Lu, Zachary Pardos and Neil Heffernan</i>	
Learning classifiers from a relational database of tutor logs	149
<i>Jack Mostow, José González-Brenes and Bao Hong Tan</i>	
A Framework for Capturing Distinguishing User Interaction Behaviors in Novel Interfaces	159
<i>Samad Kardan and Cristina Conati</i>	
How to Classify Tutorial Dialogue? Comparing Feature Vectors vs. Sequences	169
<i>José González-Brenes, Jack Mostow and Weisi Duan</i>	
Automatically Detecting a Students Preparation for Future Learning: Help Use is Key	179
<i>Ryan S.J.D. Baker, Sujith Gowda and Albert Corbett</i>	
Ensembling Predictions of Student Post-Test Scores for an Intelligent Tutoring System	189
<i>Zachary Pardos, Sujith Gowda, Ryan S.J.D. Baker and Neil Heffernan</i>	
Improving Models of Slipping, Guessing, and Moment-By-Moment Learning with Estimates of Skill Difficulty	199
<i>Sujith M. Gowda, Jonathan P. Rowe, Ryan S.J.D. Baker, Min Chi and Kenneth R. Koedinger</i>	
 Short Papers	
A Method for Finding Prerequisites Within a Curriculum	211
<i>Annalies Vuong, Tristan Nixon and Brendon Towle</i>	
Estimating Prerequisite Structure From Noisy Data	217
<i>Emma Brunskill</i>	
What can closed sets of students and their marks say?	223
<i>Dmitry Ignatov, Serafima Mamedova, Nikita Romashkin, and Ivan Shamshurin</i>	
How university entrants are choosing their department? Mining of university admission process with FCA taxonomies.	229
<i>Nikita Romashkin, Dmitry Ignatov and Elena Kolotova</i>	

What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis	235
<i>Marcelo Worsley and Paulo Blikstein</i>	
Using Logistic Regression to Trace Multiple Subskills in a Dynamic Bayes Net	241
<i>Yanbo Xu and Jack Mostow</i>	
Monitoring Learners Proficiency: Weight Adaptation in the Elo Rating System	247
<i>Kelly Wauters, Piet Desmet and Wim Van Den Noortgate</i>	
Modeling students activity in online discussion forums: a strategy based on time series and agglomerative hierarchical clustering	253
<i>Germán Cobo, David García-Solórzano, Eugènia Santamaría, Jose Antonio Morán, Javier Melenchón and Carlos Monzo</i>	
Prediction of Perceived Disorientation in Online Learning Environment with Random Forest Regression	259
<i>Gökhan Akçapınar, Erdal Cosgun and Arif Altun</i>	
Analysing Student Spatial Deployment in a Computer Laboratory	265
<i>Vladimir Ivancevic, Milan Celikovic and Ivan Lukovic</i>	
Predicting School Failure Using Data Mining	271
<i>Carlos Marquez-Vera, Cristobal Romero and Sebastin Ventura</i>	
A Dynamical System Model of Microgenetic Changes in Performance, Efficacy, Strategy Use and Value during Vocabulary Learning	277
<i>Philip I. Pavlik Jr. and Sue-Mei Wu</i>	
Desperately Seeking Subscripts: Towards Automated Model Parameterization	283
<i>Jack Mostow, Yanbo Xu and Mdahaduzzaman Munna</i>	
Automatic Generation of Proof Problems in Deductive Logic	289
<i>Behrooz Mostafavi, Tiffany Barnes and Marvin Croy</i>	
Comparison of Traditional Assessment with Dynamic Testing in a Tutoring System	295
<i>Mingyu Feng, Neil Heffernan, Zachary Pardos and Cristina Heffernan</i>	
Evaluating a Bayesian Student Model of Decimal Misconceptions	301
<i>George Gogvadze, Sergey Sosnovsky, Seiji Isotani and Bruce McLaren</i>	
Exploring user data from a game-like math tutor: a case study in causal modeling	307
<i>Dovan Rai and Joseph Beck</i>	
 Posters	
Goal Orientation and Changes of Carelessness over Consecutive Trials in Science Inquiry	315
<i>Arnon HersHKovitz, Ryan S.J.D. Baker, Janice Gobert and Michael Wixon</i>	
Towards improvements on domain-independent measurements for collaborative assessment	317
<i>Antonio R. Anaya and Jesús G. Boticario</i>	

A Java desktop tool for mining Moodle data	319
<i>Rafael Pedraza Perez, Cristobal Romero and Sebastián Ventura</i>	
Using data mining in a recommender system to search for learning objects in repositories	321
<i>Alfredo Zapata Gonzalez, Victor Hugo Menéndez Domínguez, Manuel Prieto and Cristobal Romero</i>	
E-learning Web Miner: A data mining application to help instructors involved in virtual courses	323
<i>Diego García-Saiz and Marta Zorrilla Pantaleón</i>	
Computerized Coding System for Life Narratives to Assess Students?’ Personality Adaption	325
<i>Qiwei He, Bernard Veldkamp and Gerben Westerhof</i>	
Partially Observable Sequential Decision Making for Problem Selection in an Intelligent Tutoring System	327
<i>Emma Brunskill and Stuart Russell</i>	
Mining Teaching Behaviors from Pedagogical Surveys	329
<i>Joana Barracosa and Claudia Antunes</i>	
Variable Construction and Causal Modeling of Online Education Messaging Data: Initial Results	331
<i>Stephen Fancsali</i>	
The Hospital Classrooms Environments Challenge	333
<i>Carina González and Pedro A. Toledo</i>	
Combining study of complex network and text mining analysis to understand growth mechanism of communities on SNS	335
<i>Osamu Yamakawa, Takahiro Tagawa, Hitoshi Inoue, Koichi Yastake and Takahiro Sumiya</i>	
Logistic Regression in a Dynamic Bayes Net Models Multiple Subskills Better!	337
<i>Yanbo Xu and Jack Mostow</i>	
Studying the problem-solving strategies in the early stages of learning programming	339
<i>Edgar Cambranes-Martinez and Judith Good</i>	
Brick: Mining Pedagogically Interesting Sequential Patterns	341
<i>Anjo Anjewierden, Hannie Gijlers, Nadira Saab and Robert De Hoog</i>	
Intelligent evaluation of social knowledge building using conceptual maps with MLN	343
<i>Lorenzo Moreno, Carina González, Román Estévez and Beatrice Popescu</i>	
Identifying Influence Factors on Students Success by Subgroup Discovery	345
<i>Florian Lemmerich, Marianus Ifland and Frank Puppe</i>	
Analyzing University Data for Determining Student Profiles and Predicting Performance	347
<i>Dorina Kabakchieva, Kamelia Stefanova and Valentin Kisimov</i>	
The EDM Vis Tool	349
<i>Matthew Johnson, Michael Eagle, Leena Joseph and Tiffany Barnes</i>	

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