



ISBN: 978-0-615-37529-8

## Preface

The Third International Conference on Data Mining (EDM 2010) was held in Pittsburgh, PA, USA. It follows the second conference at the University of Cordoba, Spain, on July 1–3, 2009 and the first edition of the conference held in Montreal in 2008, and a series of workshops within the AAI, AIED, EC-TEL, ICALT, ITS, and UM conferences. EDM2011 will be held in Eindhoven, Netherlands.

EDM brings together researchers from computer science, education, psychology, psychometrics, and statistics to analyze large data sets to answer educational research questions. The increase in instrumented educational software and databases of student test scores, has created large repositories of data reflecting how students learn. The EDM conference focuses on computational approaches for analyzing those 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.

We received a total of 54 full papers and 20 submitted posters from 21 countries. Paper submissions were reviewed by three or four reviewers and 23 of them were accepted as full papers (43% acceptance rate). All papers will appear both on the web, at [www.educationaldatamining.org](http://www.educationaldatamining.org), as well as in the printed proceedings. The conference also included invited talks by Professor Cristina Conati, Computer Science Professor, Computer Science Department and Laboratory for Computational Intelligence at the University of British Columbia, Canada and by Professor Osmar Zaiane Ph.D., Professor, Department of Computing Science, University of Alberta, Canada.

We would like to thank Carnegie Mellon University for their hosting of EDM2010, and thank the Pittsburgh Science of Learning Center DataShop and Carnegie Learning Inc for their generous sponsorship. We would like to thank the program committee members, local committee, web chair, the reviewers and the invited speakers for their enthusiastic help in putting this conference together.

**Ryan S.J.d. Baker,**  
**Agathe Merceron,**  
**Philip I. Pavlik Jr. (Eds.)**

## Conference Organization

**Conference Chair:** [Ryan S.J.d. Baker](#), Worcester Polytechnic Institute

**Program Chairs:** [Agathe Merceron](#), Beuth University of Applied Sciences Berlin;

[Philip I. Pavlik Jr.](#), Carnegie Mellon University

**Local Organization Chair:** John Stamper, Carnegie Mellon University

**Web Chair:** [Arnon Hershkovitz](#), Tel Aviv University

## Program Committee

Esma Aimeur, University of Montreal, Canada

Ivon Arroyo, University of Massachusetts Amherst, USA

Beth Ayers, Carnegie Mellon University, USA

Ryan 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

Cristophe Choquet, Université du Maine, France

Cristina Conati, University of British Columbia, Canada

Richard Cox, University of Sussex, UK

Michel Desmarais, Ecole Polytechnique de Montreal, Canada

Aude Dufresne, University of Montreal, Canada

Mingyu Feng, SRI International, USA

Art Graesser, University of Memphis, USA

Andreas Harrer, Katholische Universität Eichstätt-Ingolstadt, Germany

Neil Heffernan, Worcester Polytechnic Institute, USA

Arnon Hershkovitz, Tel Aviv University, Israel

Cecily Heiner, University of Utah, USA

Roland Hubscher, Bentley University, USA

Sebastian Iksal, Université du Maine, France

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

Gordon McCalla, University of Saskatchewan, Canada

Bruce McLaren, Deutsches Forschungszentrum für Künstliche Intelligenz, Germany

Julia Mingullon Alfonso, Universitat Oberta de Catalunya, Spain

Tanja Mitrovic, Canterbury University, New Zealand

Jack Mostow, Carnegie Mellon University, USA

Rafi Nachmias, Tel Aviv University, Israel

Roger Nkambou, Université du Québec à Montréal (UQAM), Canada

Mykola Pechenizkiy, Eindhoven University of Technology, Netherlands  
Steve Ritter, Carnegie Learning, USA  
Cristobal Romero, Cordoba University, Spain  
Carolyn Rose, Carnegie Mellon University, USA  
Steven Tanimoto, University of Washington, USA  
Sebastian Ventura, Cordoba University, Spain  
Kalina Yacef, University of Sydney, Australia  
Osmar Zaiane, University of Alberta, Canada

## Steering Committee

Esmā 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

## External Reviewers

Hua Ai  
Acey Boyce  
Amanda Chaffin  
Min Chi  
David G. Cooper  
Michael Eagle  
Philippe Fournier-Viger  
Ilya Goldin  
Andrew Hicks  
Matthew Johnson  
Evgeny Knutov  
Behrooz Mostafavi  
Andrea Nickel  
Zachary Pardos  
Ildiko Pelczer  
Reihaneh Rabbany  
Jose-Raul Romero  
Enrique Garcia Salcines  
Oliver Scheuer  
Benjamin Shih  
Ekaterina Vasilyeva  
Indre Zliobaite

# Invited Speakers for Educational Data Mining 2010

## Data-based Student Modeling in Exploratory Learning Environments

**Cristina Conati**

[Computing Science Department](#) & [Laboratory for Computational Intelligence, University of British Columbia](#) [[homepage](#)]



**Abstract:** Exploratory Learning Environments (ELE) are designed to help users acquire knowledge by freely experiencing a target domain. In this setting, it is often hard to identify interaction behaviours that are conducive to learning, vs. behaviours that indicate student confusion, making it hard to provide adaptive support to students who do not learn well with ELEs. In this talk, I will present our work on using data-based approaches to identify and recognize relevant behavioral patterns during interaction with ELEs, with the goal of enabling ELEs to monitor how a student works with the environment and provide adaptive support when needed.

## Social Network Analysis for the Assessment of Learning

**Osmar Zaiane**

[Department of Computing Science, University of Alberta](#) [[homepage](#)]



**Abstract:** Using computer-supported collaborative learning tools, learners interact forming relationships and complex flows of information. In a forum with very few learners it is customary to quickly collect thousands of messages in few months, and these are interrelated in intricate discussion threads. Assessing the participation and interaction between learners can become a daunting task. Social network analysis is a field of study attempting to understand and measure relationships between entities in networked information. Can social network analysis techniques and data mining techniques for information networks help examine and assess online interactions? We examine some work done in this area, particularly the application of community mining, and discuss some open problems pertaining to social network analysis in the e-learning domain.

# Table of Contents

## Regular papers

Effort-based Tutoring: An Empirical Approach to Intelligent Tutoring -- Ivon Arroyo, Hasmik Mehranian and Beverly P. Woolf	1
An Analysis of the Differences in the Frequency of Students' Disengagement in Urban, Rural, and Suburban High Schools -- Ryan S.J.d. Baker and Sujith M. Gowda	11
On the Faithfulness of Simulated Student Performance Data -- Michel C. Desmarais and Ildiko Pelczer	21
Mining Bodily Patterns of Affective Experience during Learning -- Sidney D'Mello and Art Graesser	31
Can We Get Better Assessment From A Tutoring System Compared to Traditional Paper Testing? Can We Have Our Cake (Better Assessment) and Eat It too (Student Learning During the Test)? -- Mingyu Feng and Neil Heffernan	41
Using Neural Imaging and Cognitive Modeling to Infer Mental States while Using an Intelligent Tutoring System -- Jon M. Fincham, John R. Anderson, Shawn Betts and Jennifer Ferris	51
Using multiple Dirichlet distributions to improve parameter plausibility -- Yue Gong, Joseph E. Beck and Neil T. Heffernan	61
Examining Learner Control in a Structured Inquiry Cycle Using Process Mining -- Larry Howard, Julie Johnson and Carin Neitzel	71
Analysis of Productive Learning Behaviors in a Structured Inquiry Cycle Using Hidden Markov Models -- Hogyong Jeong, Gautam Biswas, Julie Johnson and Larry Howard	81
Data Mining for Generating Hints in a Python Tutor -- Anna Katrina Dominguez, Kalina Yacef and James R. Curran	91
Off Topic Conversation in Expert Tutoring: Waste of Time or Learning Opportunity -- Blair Lehman, Whitney Cade and Andrew Olney	101

Sentiment Analysis in Student Experiences of Learning -- Sunghwan Mac Kim and Rafael A. Calvo	111
Online Curriculum Planning Behavior of Teachers -- Keith E. Maull, Manuel Gerardo Saldivar and Tamara Sumner	121
A Data Model to Ease Analysis and Mining of Educational Data -- André Krüger, Agathe Merceron and Benjamin Wolf	131
Identifying Students' Inquiry Planning Using Machine Learning -- Orlando Montalvo, Ryan S.J.d. Baker, Michael A. Sao Pedro, Adam Nakama and Janice D. Gobert	141
Skill Set Profile Clustering: The Empty K-Means Algorithm with Automatic Specification of Starting Cluster Centers -- Rebecca Nugent, Nema Dean and Elizabeth Ayers	151
Navigating the parameter space of Bayesian Knowledge Tracing models: Visualizations of the convergence of the Expectation Maximization algorithm -- Zachary Pardos and Neil Heffernan	161
Mining Rare Association Rules from e-Learning Data -- Cristóbal Romero, José Raúl Romero, Jose María Luna and Sebastián Ventura	171
Using Text Replay Tagging to Produce Detectors of Systematic Experimentation Behavior Patterns -- Michael Sao Pedro, Ryan S.J.d. Baker, Orlando Montalvo, Adam Nakama and Janice D. Gobert	181
Identifying High-Level Student Behavior Using Sequence-based Motif Discovery -- David H. Shanabrook, David G. Cooper, Beverly Park Woolf and Ivon Arroyo	191
Unsupervised Discovery of Student Strategies -- Benjamin Shih, Kenneth R. Koedinger and Richard Scheines	201
Assessing Reviewer's Performance Based on Mining Problem Localization in Peer-Review Data -- Wenting Xiong, Diane Litman and Christian Schunn	211
Using Numeric Optimization To Refine Semantic User Model Integration Of Adaptive Educational Systems -- Michael Yudelson, Peter Brusilovsky, Antonija Mitrovic and Moffat Mathews	221



## Young Researcher Track papers

An Annotations Approach to Peer Tutoring -- John Champaign and Robin Cohen	231
Using Educational Data Mining Methods to Study the Impact of Virtual Classroom in E-Learning -- Mohammad Hassan Falakmasir and Jafar Habibi	241
Mining Students' Interaction Data from a System that Support Learning by Reflection -- Rajibussalim	249
Process Mining to Support Students' Collaborative Writing -- Vilaythong Southavilay, Kalina Yacef and Rafael A. Callvo	257

## Poster Abstracts

Automatic Rating of User-Generated Math Solutions -- Turadg Aleahmad, Vincent Aleven and Robert Kraut	267
Tracking Students' Inquiry Paths through Student Transition Analysis -- Matt Bachmann, Janice Gobert and Joseph Beck	269
DISCUSS: Enabling Detailed Characterization of Tutorial Interactions Through Dialogue Annotation -- Lee Becker, Wayne H. Ward and Sarel vanVuuren	271
Data Mining of both Right and Wrong Answers from a Mathematics and a Science M/C Test given Collectively to 11,228 Students from India [1] in years 4, 6 and 8 -- James Bernauer and Jay Powell	273
Mining information from tutor data to improve pedagogical content knowledge -- Suchismita Srinivas, Muntaquim Bagadia and Anupriya Gupta	275
Clustering Student Learning Activity Data -- Haiyun Bian	277
Analyzing Learning Styles using Behavioral Indicators in Web based Learning Environments -- Nabila Bousbia, Jean-Marc Labat, Amar Balla and Issam Rebai	279

Using Topic Models to Bridge Coding Schemes of Differing Granularity -- Whitney L. Cade and Andrew Olney	281
A Distillation Approach to Refining Learning Objects -- John Champaign and Robin Cohen	283
A Preliminary Investigation of Hierarchical Hidden Markov Models for Tutorial Planning -- Kristy Elizabeth Boyer, Robert Phillips, Eun Young Ha, Michael D. Wallis, Mladen A. Vouk, and James C. Lester	285
Higher Contributions Correlate with Higher Learning Gains -- Carol Forsyth, Heather Butler, Arthur C. Graesser, Diane Halpern	287
Pinpointing Learning Moments; A finer grain P(J) model -- Adam Goldstein, Ryan S.J.d. Baker and Neil T. Heffernan	289
Predicting Task Completion from Rich but Scarce Data -- José P. González-Brenes and Jack Mostow	291
Hierarchical Structures of Content Items in LMS -- Sharon Hardof-Jaffe, Arnon HersHKovitz, Ronit Azran and Rafi Nachmias	293
Is Students' Activity in LMS Persistent? -- Arnon HersHKovitz and Rafi Nachmias	295
EDM Visualization Tool: Watching Students Learn -- Matthew M. Johnson and Tiffany Barnes	297
Inferring the Differential Student Model in a Probabilistic Domain Using Abduction inference in Bayesian networks -- Nabila Khodeir, Nayer Wanas, Nevin Darwish and Nadia Hegazy	299
Using LiMS (the Learner Interaction Monitoring System) to Track Online Learner Engagement and Evaluate Course Design -- Leah P. Macfadyen and Peter Sorenson	301
Observing Online Curriculum Planning Behavior of Teachers -- Keith E. Maull, Manuel Gerardo Saldivar and Tamara Sumner	303
When Data Exploration and Data Mining meet while Analysing Usage Data of a Course -- André Krüger, Agathe Merceron and Benjamin Wolf	305

AutoJoin: Generalizing an Example into an EDM query -- Jack Mostow and Bao Hong (Lucas) Tan	307
Conceptualizing Procedural Knowledge Targeted at Students with Different Skill Levels -- Martin Možina, Matej Guid, Aleksander Sadikov, Vida Groznic, Jana Krivec, and Ivan Bratko	309
Data Reduction Methods Applied to Understanding Complex Learning Hypotheses -- Philip I. Pavlik Jr.	311
Analysis of a causal modeling approach: a case study with an educational intervention -- Dovan Rai and Joseph E. Beck	313
Peer Production of Online Learning Resources: A Social Network Analysis -- Beijie Xu and Mimi M. Recker	315
Class Association Rules Mining from Students' Test Data -- Cristóbal Romero, Sebastián Ventura, Ekaterina Vasilyeva and Mykola Pechenizkiy	317
Modeling Learning Trajectories with Epistemic Network Analysis: A Simulation-based Investigation of a Novel Analytic Method for Epistemic Games -- Andre A. Rupp, Shauna J. Sweet and Younyoung Choi	319
Multiple Test Forms Construction based on Bees Algorithm -- Pokpong Songmuang and Maomi Ueno	321
Can Order of Access to Learning Resources Predict Success? -- Hema Soundranayagam and Kalina Yacef	323
A Data Driven Approach to the Discovery of Better Cognitive Models -- Kenneth R. Koedinger and John C. Stamper	325
Using a Bayesian Knowledge Base for Hint Selection on Domain Specific Problems -- John C. Stamper, Tiffany Barnes and Marvin Croy	327
A Review of Student Churn in the Light of Theories on Business Relationships -- Jaan Ubi and Innar Liiv	329

Towards EDM Framework for Personalization of Information Services in RPM Systems -- Ekaterina Vasilyeva, Mykola Pechenizkiy, Aleksandra Tesanovic, Evgeny Knutov, Sicco Verwer and Paul De Bra	331
A Case Study: Data Mining Applied to Student Enrollment -- César Vialardi, Jorge Chue, Alfredo Barrientos, Daniel Victoria, Jhonny Estrella, Juan Pablo Peche and Álvaro Ortigosa	333
Representing Student Performance with Partial Credit -- Yutao Wang, Neil T. Heffernan and Joseph E. Beck	335
Where in the World? Demographic Patterns in Access Data -- Mimi M. Recker, Beijie Xu, Sherry Hsi, and Christine Garrard	337
Pundit: Intelligent Recommender of Courses -- Ankit Ranka, Faisal Anwar, Hui Soo Chae	339
Author Index	341