EDM 2012

5th International Conference on Educational Data Mining

PROCEEDINGS OF THE FIFTH INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING

Chania, Greece, June 19-21 2012

Kalina Yacef, Osmar Zaïane, Arnon Hershkovitz, Michael Yudelson, John Stamper
Preface

The 5th International Conference on Educational Data Mining (EDM 2012) is held in picturesque Chania on the beautiful Crete island in Greece, under the auspices of the International Educational Data Mining Society (IEDMS). The EDM 2012 conference is a leading international forum for high quality research that mines large data sets of educational data to answer educational research questions. These data sets may come from learning management systems, interactive learning environments, intelligent tutoring systems, or any system used in a learning context.

We received a total of 50 full papers, 19 short papers and 20 posters from authors of 30 different countries and these submissions covered the full range of the conference topics specified in the call for papers. The papers were distributed to the 56 international program committee members, who called upon additional 17 external reviewers. All papers were reviewed by 3 reviewers and any paper submitted by organisers was sent to anonymous reviewers unbeknownst to the involved organisers. After receiving a total of 235 reviews, 17 full papers and 15 short papers were accepted to be presented at the conference and included in these proceedings, giving an acceptance rate of 34% for full papers, and 46% overall. Furthermore, 17 posters were accepted and their summaries appear in the proceedings. We have also included the abstracts of the 3 invited talks by (i) Professor Myra Spiliopoulou, Professor of Business Information Systems, Computer Science, Otto-von-Guericke University Magdeburg, Germany, (ii) Professor Danielle S. McNamara, Learning Sciences Institute Psychology Department, Arizona State University and (iii) Dr Bob Dolan, Senior Research Scientist, Assessment and Information, Pearson.

We would like to thank our sponsors Carnegie Learning (Gold Level), Pearson (Gold Level) and LearnLab (Silver Level) for their generous support. We would also like to thank the program committee members, the additional reviewers, the local committee, the web chair and the invited speakers for their invaluable help in putting this program and conference together.

We hope that the papers contained in these proceedings will be stimulating to most of the readers, provide thought-provoking new ideas to motivate new research, and prove helpful towards improving the literacy in educational data mining.

Enjoy the conference.

Kalina Yacef
Osmar Zaïane
Arnon Hershkovitz
Michael Yudelson
John Stamper
Organization

CONFERENCE CHAIR
John Stamper Carnegie Mellon University, USA

PROGRAM CHAIRS
Kalina Yacef The University of Sydney, Australia
Osmar Zaiane University of Alberta, Canada

POSTER AND DEMO CHAIRS
Arnon Hershkovitz Worcester Polytechnic Institute, USA
Michael Yudelson Carnegie Mellon University, USA

FINANCE AND WEB CHAIR
Michael Bett Carnegie Mellon University, USA

LOCAL ORGANIZING TEAM
Kitty Panourgia Neoanalysis, Greece

STEERING COMMITTEE
Ryan Baker Worcester Polytechnic Institute, USA
Tiffany Barnes University of North Carolina at Charlotte, USA
Joseph E. Beck Worcester Polytechnic Institute, USA
Michel Desmarais Ecole Polytechnique de Montreal, Canada
Neil Heffernan Worcester Polytechnic Institute, USA
Agathe Merceron Beuth University of Applied Sciences, Germany
Mykola Pechenizkiy Eindhoven University of Technology, Netherlands
Cristobal Romero Cordoba University, Spain
Kalina Yacef The University of Sydney, Australia

PROGRAM COMMITTEE
Esma Aimeur University of Montreal, Canada
Mirjam Augstein Kock Upper Austria University of Applied Sciences, Austria
Elizabeth Ayers University of California, 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
Jesus G. Boticario U.N.E.D., Spain
Rafael A. Calvo The University of Sydney, Australia
Min Chi Stanford University, USA
Christophe Choquet University of Maine, USA
Richard Cox University of Edinburgh, Scotland
Sidney D’Mello The University of Memphis, USA
Michel Desmarais Ecole Polytechnique de Montreal, Canada
Mingyu Feng SRI International, USA
ADDITIONAL REVIEWERS
Iiya Goldin          Hamman Samuel
I-Han Hsiao         Michael Sao Pedro
Nan Li              Mayya Sharipova
Ming Liu            Carlos Soares
Collin Lynch        Mansoureh Takaffoli
Roberto Martinez Maldonado  Eric Verbeek
Terry Peckham       Qingyang Wang

IEDMS OFFICERS AND BOARD OF DIRECTORS
Ryan Baker (President)  Worcester Polytechnic Institute, USA
Michel Desmarais (Treasurer)  Ecole Polytechnique de Montreal, Canada
Tiffany Barnes         University of North Carolina at Charlotte, USA
Joseph E. Beck         Worcester Polytechnic Institute, USA
Neil Heffernan         Worcester Polytechnic Institute, USA
Agathe Merceron       Beuth University of Applied Sciences, Germany
Mykola Pechenizkiy    Eindhoven University of Technology, Netherlands
Cristobal Romero      Cordoba University, Spain
Kalina Yacef          The University of Sydney, Australia
# Table of contents

## Invited Talks (abstracts)

**Stream Mining in Education? Dealing with Evolution**  
*Professor Myra Spiliopoulou*  
3

**From Text to Feedback: Leveraging Data Mining to Build Educational Technologies**  
*Danielle S. McNamara*  
5

**Five Aspirations for Educational Data Mining**  
*Bob Dolan and John Behrens*  
7

## Full Papers

**Assisting Instructional Assessment of Undergraduate Collaborative Wiki and SVN Activities**  
*Jihie Kim, Erin Shaw, Hao Xu and Adarsh G V*  
10

**Automated Student Model Improvement**  
*Kenneth R. Koedinger, Elizabeth A. McLaughlin and John C. Stamper*  
17

**Automatic Discovery of Speech Act Categories in Educational Games**  
*Vasile Rus, Arthur Graesser, Cristian Moldovan and Nobal Niraula*  
25

**Co-Clustering by Bipartite Spectral Graph Partitioning for Out-of-Tutor Prediction**  
*Shubhendu Trivedi, Zachary Pardos, Gábor Sárközy and Neil Heffernan*  
33

**Comparison of methods to trace multiple subskills: Is LR-DBN best?**  
*Yanbo Xu and Jack Mostow*  
41

**Dynamic Cognitive Tracing: Towards Unified Discovery of Student and Cognitive Models**  
*José González-Brenes and Jack Mostow*  
49

**Identifying Learning Behaviors by Contextualizing Differential Sequence Mining with Action Features and Performance Evolution**  
*John Kinnebrew and Gautam Biswas*  
57

**Identifying Students' Characteristic Learning Behaviors in an Intelligent Tutoring System Fostering Self-Regulated Learning**  
*François Bouchet, John Kinnebrew, Gautam Biswas and Roger Azevedo*  
65
Learner Differences in Hint Processing
Ilya Goldin, Kenneth Koedinger and Vincent Aleven

Methods to find the number of latent skills
Behzad Beheshti, Michel Desmarais and Rhouma Naceur

Mining Student Behavior Patterns in Reading Comprehension Tasks
Terry Peckham and Gordon McCalla

Model-Based Collaborative Filtering Analysis of Student Response Data:
Machine-Learning Item Response Theory
Yoav Bergner, Stefan Droschler, Gerd Kortemeyer, Saif Rayyan, Daniel Seaton and David Pritchard

Predicting drop-out from social behaviour of students
Tomas Obsivac, Lubos Popelinsky, Jaroslav Bayer, Jan Geryk and Hana Bydzovska

Searching for Variables and Models to Investigate Mediators of Learning from Multiple Representations
Martina Rau and Richard Scheines

The Impact on Individualizing Student Models on Necessary Practice Opportunities
Jung In Lee and Emma Brunskill

Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra
Ryan S.J.D. Baker, Sujith Gowda, Michael Wixon, Jessica Kalka, Angela Wagner, Aatish Salvi, Vincent Aleven, Gail Kusbit, Jaclyn Ocumpaugh and Lisa Rossi

Using Edit Distance to Analyse Errors in a Natural Language to Logic Translation Corpus
Dave Barker-Plummer, Robert Dale, Richard Cox and Alex Romanczuk

Short Papers

Calculating Probabilistic Distance to Solution in a Complex Problem Solving Domain
Leigh Ann Sudol, Kelly Rivers and Thomas K. Harris

Classification via clustering for predicting final marks starting from the student participation in Forums
Manuel Ignacio Lopez, Cristobal Romero, Sebastián Ventura and J.M. Luna

Development of a Workbench to Address the Educational Data Mining Bottleneck
Ma. Mercedes Rodrigo, Ryan S. J. D. Baker, Bruce McLaren, Alejandra Jayme and Thomas Dy
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Prediction of Student Self-Regulation Strategies by Combining Multiple Models</td>
<td>156</td>
</tr>
<tr>
<td>Jennifer Sabourin, Bradford Mott and James Lester</td>
<td></td>
</tr>
<tr>
<td>Identifying Successful Learners from Interaction Behaviour</td>
<td>160</td>
</tr>
<tr>
<td>Judi Mccuaig and Julia Baldwin</td>
<td></td>
</tr>
<tr>
<td>Interaction Networks: Generating High Level Hints Based on Network Community Clusterings</td>
<td>164</td>
</tr>
<tr>
<td>Michael Eagle, Matthew Johnson and Tiffany Barnes</td>
<td></td>
</tr>
<tr>
<td>Interleaved Practice with Multiple Representations: Analyses with Knowledge Tracing Based Techniques</td>
<td>168</td>
</tr>
<tr>
<td>Martina Rau and Zachary Pardos</td>
<td></td>
</tr>
<tr>
<td>Learning Gains for Core Concepts in a Serious Game on Scientific Reasoning</td>
<td>172</td>
</tr>
<tr>
<td>Carol Forsyth, Philip Pavlik Jr, Arthur Graesser, Zhiqiang Cai, Mae-Lynn Germany, Keith Millis, Heather Butler, Diane Halpern and Robert Dolan</td>
<td></td>
</tr>
<tr>
<td>Leveraging First Response Time into the Knowledge Tracing Model</td>
<td>176</td>
</tr>
<tr>
<td>Yutao Wang and Neil Heffernan</td>
<td></td>
</tr>
<tr>
<td>Meta-learning Approach for Automatic Parameter Tuning: A case of study with educational datasets</td>
<td>180</td>
</tr>
<tr>
<td>Maria De Mar Molina, Cristobal Romero, Sebastián Ventura and J.M. Luna</td>
<td></td>
</tr>
<tr>
<td>Mining Concept Maps to Understand University Students’ Learning</td>
<td>184</td>
</tr>
<tr>
<td>Jin Soung Yoo and Moon-Heum Cho</td>
<td></td>
</tr>
<tr>
<td>Policy Building – An Extension To User Modeling</td>
<td>188</td>
</tr>
<tr>
<td>Michael Yudelson and Emma Brunskill</td>
<td></td>
</tr>
<tr>
<td>The real world significance of performance prediction</td>
<td>192</td>
</tr>
<tr>
<td>Zachary Pardos, Qing Yang Wang and Shubhendu Trivedi</td>
<td></td>
</tr>
<tr>
<td>The Rise of the Super Experiment</td>
<td>196</td>
</tr>
<tr>
<td>John Stamper, Derek Lomas, Dixie Ching, Steven Ritter, Kenneth Koedinger and Jonathan Steinhart</td>
<td></td>
</tr>
<tr>
<td>Using Student Modeling to Estimate Student Knowledge Retention</td>
<td>200</td>
</tr>
<tr>
<td>Yutao Wang and Joseph Beck</td>
<td></td>
</tr>
<tr>
<td>Posters</td>
<td></td>
</tr>
<tr>
<td>A promising classification method for predicting distance students’ performance.</td>
<td>206</td>
</tr>
<tr>
<td>Diego García-Saiz and Marta Zorrilla</td>
<td></td>
</tr>
</tbody>
</table>
Analyzing paths in a student database
Donatella Merlini, Renza Campagni and Renzo Sprugnoli

Analyzing the behavior of a teacher network in a Web 2.0 environment
Eliana Scheihiing, Carolina Aros and Daniel Guerra

Automated Detection of Mentors and Players in an Educational Game
Fazel Keshtkar, Brent Morgan and Arthur Graesser

Categorizing Students’ Response Patterns using the Concept of Fractal Dimension
Rasil Warnakulasooriya and William Galen

CurriM: Curriculum Mining
Mykola Pechenizkiy, Nikola Trcka, Paul De Bra and Pedro Toledo

Data mining techniques for design of ITS student models
Ritu Chaturvedi and Christie Ezeife

Deciding on Feedback Polarity and Timing
Stuart Johnson and Osmar Zaiane

Finding Dependent Test Items: An Information Theory Based Approach
Xiaoxun Sun

Fit-to-Model Statistics for Evaluating Quality of Bayesian Student Ability Estimation
Ling Tan

Inferring learners’ knowledge from observed actions
Anna Rafferty, Michelle Lamar and Thomas Griffiths

Learning Paths in a Non-Personalizing e-Learning Environment
Agathe Merceron, Liane Beuster, Margarita Elkina, Albrecht Forstenbacher, Leonard Kappe, Andreas Pursian, Sebastian Schwarzrock and Boris Wenzlaff

Similarity Functions for Collaborative Master Recommendations
Alexandru Surpatean, Evgueni Smirnov and Nicolai Manie

Social Networks Analysis for Quantifying Students’ Performance in Teamwork
Pedro Crespo and Claudia Antunes

Speaking (and touching) to learn: a method for mining the digital footprints of face-to-face collaboration
Roberto Martinez Maldonado, Kalina Yacef and Judy Kay

Stress Analytics in Education
Rafal Kocielnik, Mykola Pechenizkiy and Natalia Sidorova
Variable Construction and Causal Discovery for Cognitive Tutor Log Data: Initial Results
Stephen Fanesali

List of Authors