Workshop on Educational Data Mining @ ICALT07 (EDM@ICALT07)

Joseph E. Beck[1], Toon Calders[2], Mykola Pechenizkiy[2] and Silvia Rita Viola[3]
[1]Carnegie Mellon University, USA
[2]Eindhoven University of Technology, The Netherlands
[3]Universita' Politecnica delle Marche, Italy
joseph.beck@gmail.com, m.pechenizkiy@tue.nl, t.calders@tue.nl, sr.viola@ieee.org

1. Introduction to EDM@ICALT07

The Educational Data Mining Workshop[1] will be held in conjunction with the 7th IEEE International Conference on Advanced Learning Technologies (ICALT) in Niigata, Japan on July 18-20, 2007. EDM@ICALT07 continues the series of Workshops organized by the International Working Group on Educational Data Mining during 2007. For upcoming events in Educational Data Mining and for information on past workshops, please see the web site www.educationaldatamining.org.

Recently, the increase in dissemination of interactive learning environments has allowed the collection of huge amounts of data. An effective way of discovering new knowledge from large and complex data sets is data mining. The EDM workshop aimed for papers that study how to apply data mining to analyze data generated by learning systems or experiments, as well as how discovered information can be used to improve adaptation and personalization.

Interesting problems data mining can help to solve are: determining what are common learning styles or strategies, predicting the knowledge and interests of a user based on past behavior, partitioning a heterogeneous group of users into homogeneous clusters, etc.

Typically, educational data sources are quite heterogeneous (e.g., web log files, interaction logs, source code, text and dialogue data, etc.), and have a variety of different scales, grain-sizes, and spatial and temporal resolution. Though the many types of educational data often differ considerably from one another, they provide multiple types of insight on a single domain or context and, above all, share the potential to reveal unexpected and useful knowledge concerning learners and/or the process of learning - if correctly and coherently analyzed.

Applying methods to mine the complex data that we can collect on educational situations requires the development of new approaches that build upon techniques from a combination of areas, including statistics, psychometrics, machine learning, and scientific computing.

The EDM workshop at ICALT'07 is aimed at providing a focused international forum for researchers to present, discuss and explore the state of the art of mining educational data and evaluating usefulness of discovered patterns for adaptation and personalization, as well as to outline promising future research directions.

2. Main topics addressed

The topics of special interest of the workshop include:
- Methods and approached for EDM
- Characteristics of educational data and how to deal with them
- Learning browsing behavior; e.g., searching for patterns in log-data
- Data mining for predicting user (potentially changing) interests
- Mining differences in user's learning behavior (e.g. between two systems)
- Mining data from A/B tests
- Application of discovered patterns for personalization and adaptation
- Description of applications
- Case studies and experiences

3. Workshop contributions

A Outliers Analysis of Learner's data based on User Interface Behaviors by Yong Se Kim, Tae Bok Yoon, Hyun Jin Cha, Young Mo Jung, Eric Wang and Jee Hyong Lee. In this paper, the development of a robust model for identifying learners' tendencies is discussed. In the paper, the outliers in the learner's data are eliminated by a k-NN method. Elimination of outliers is important, because if the collected data reflects inconsistent learner behaviors or unpredictable

---

[1] http://www.educationaldatamining.org/ICALT2007.html

learning tendencies, then the reliability of the learner model is degraded. The elimination method is then applied to an experimental data set obtained using DOLLS-HI, a learner diagnosis system that uses housing interior learning contents to diagnose learning styles. The resulting diagnosis model shows improvement in reliability by eliminating the outliers.

A framework for using web usage mining to personalize e-learning by Hafidh Ba-Omar, Ilias Petrounias, and Fahad Anwar. In this paper, the Authors aim at exploiting the huge amount of data collected by e-learning systems. Accordingly, they propose a web usage mining framework for personalizing e-learning that pays special attention towards individual learning styles. The goal is to identify learning styles by examining learning patterns and the order in which learning resources are selected. A prototype for an adaptive web based course was developed.

User session Models for Educational Systems based on Multiple Knowledge Structures by Judit Jasso and Alfredo Milani. This work introduces a framework for a general representation of web-pages. The representation is done by means of Multiple Knowledge Structures in order to improve semantic usage modeling. Their approach generalizes existing models by adding and integrating features of hypertext structured educational systems. The purpose of the Multiple Knowledge Structure Representation is to allow multiple data mining analyses, based on symbolic objects and atomic application events in a single integrated framework. Still, it is able to take into account relevant features such as structured semantics and absolute time information.

Analyzing the data collected by Programming Tutors that Provide Post-Practice Reflection by Amruth Kumar and Peter Rutigliano. In this paper, megavariate analysis techniques are used to analyze the data collected by programming tutors in order to find the relationships, if any, between the correctness of a student's answer, and several independent variables including the number of steps in the correct answer, the time taken by the student to solve each problem, the number of problems solved by the student during the tutoring session, etc. After eliminating outliers determined through combined variance of all the independent variables, the Authors found that the correct solving of the problem is closely related to the number of attempts needed by the learner to identify the underlying concept during post-practice reflection. Therefore, the reflection questions as presented by the tutors correctly assess understanding of the concept underlying each problem.

4. Workshop Chairs and PC

- Joseph E. Beck, Carnegie Mellon University, USA
- Toon Calders, Eindhoven, University of Technology, the Netherlands
- Mykola Pechenizkiy, Eindhoven University of Technology, the Netherlands
- Silvia Rita Viola, Universita' Politecnica delle Marche, Italy

- Ivon Arroyo, University of Massachusetts Amherst, USA
- Ari Bader-Natal, Brandeis University, USA
- Ryan Baker University of Nottingham, UK
- Rahel Bekele, Addis Ababa University, Ethiopia
- Mária Bieliková, Slovak University of Technology, Slovakia
- Hao Cen, Carnegie Mellon University, USA
- Raquel M. Crespo Garcia Carlos III University of Madrid, Spain
- Christophe Choquet, Université du Maine, France
- Rebecca Crowley, University of Pittsburgh, USA
- Paul De Bra, Eindhoven University of Technology, the Netherlands
- Mingyu Feng, Worcester Polytechnic Institute, USA
- Elena Gaudioso, Universidad Nacional de Educación a Distanzia, Spain
- Sabine Graf, Vienna University of Technology, Austria
- Wilhelmiina Hämäläinen, University of Joensuu, Finland
- Judy Kay, University of Sydney, Australia
- Manolis Mavrikis, University of Edinburgh, UK
- Agathe Merceron, University of Applied Sciences Berlin, Germany
- Maria Milosavljevic, Macquarie University, Sydney, Australia
- Kaska Porayska-Pomsta, London Knowledge Lab, UK
- Genaro Rebolledo-Mendez, University of Sussex, UK
- Cristobal Romero, Universidad de Córdoba, Spain
- Amy Soller, Institute for Defense Analyses, USA
- Alexey Tsymbal, Siemens AG, Germany
- Marie-Helene Ng Cheong Vee, Birkbeck University of London, UK

Acknowledgements

IEEE COMPUTER SOCIETY