

Educational Data Mining

and the Proceedings of the
International Conference on
Educational Data Mining

Benjamin Wiles
Graduate Student
Department of Mathematics
Kansas State University

Data Mining

Methods for discovering
information from large
data sets

- ▶ Mathematical Models
- ▶ Computer Algorithms
- ▶ Statistical Methods

Domain Driven (Applied) Data Mining

Application and modification of
data mining methods to benefit
or study a specific domain

- Search Engines
- Advertising
- Security
- Medicine

Educational Data Mining

Application of data
mining methods to the
study of educational
systems

- Enrollment Services
- Effectiveness of Instruction
- Student Learning Behavior
- Intelligent Tutoring Systems

Development of EDM

- ▶ 2004
 - First publication to mention EDM appeared and likely fewer than 30 parties identified with EDM (Winters)
- ▶ 2006
 - 20+ publications published in two workshops
- ▶ 2008
 - First International Conference on EDM
 - Journal of EDM soon to be accepting submissions

Who Is Involved?

- ▶ Academics from various disciplines who take an interest in improving teaching in their field
- ▶ Statisticians
- ▶ Educational Psychologists
- ▶ Computer Scientists
- ▶ Administrators

Sources of Data

- ▶ Intelligent Tutoring Systems
 - Based in artificial intelligence
 - Designed to model the student and adapt an effective teaching strategy
 - Interdisciplinary research between cognitive psychologists, educationalists, and computer scientists

Sources of Data

- ▶ School records
 - Courses taken/Major
 - Grades
 - Standardized testing
 - Extra-curricular activities
 - Socio-economic factors
- ▶ Course Records
 - Individual scores on assignments/items

Sources of Data

- ▶ Online Homework Systems
 - Questions correct
 - Number of attempts
 - Time between transactions
 - Time before looking at assignment
 - Number of requests for help

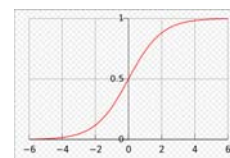
The Meeting

- ▶ The First International Conference on Educational Data Mining
 - <http://www.educationaldatamining.org/EDM2008/index.php?page=proceedings>

Common Themes

- ▶ Item Response Theory
- ▶ Bayesian Networks
- ▶ Clustering
- ▶ Analysis of “Transaction Logs”
- ▶ Assumptions on Properties of Student Learning and Behavior
- ▶ Much less focus on standard experimental design

Item Response Theory



- ▶ Modeling data from assessments
 - Use of Single Factor Item Characteristic Curve to model student ability versus probability of correctly completing an item

Bayesian Networks

- ▶ Pattern Recognition
- ▶ Parameter Estimation
- ▶ Analysis of Student Learning State
- ▶ Predicting Student Behavior

Clustering

- ▶ Grouping objects in “interesting” ways according to attribute
- ▶ Example Objects/Attributes
 - Students
 - Homework/Test Questions
- ▶ Example Attributes Values
 - Students’ Scores on Homework/Test Questions

DataShop

- ▶ <https://learnlab.web.cmu.edu/datashop/>

Selected Readings

A Response Time Model for Bottom-Out Hints as Worked Examples

Shih, B., Koedinger, K., and Scheines, R.

Computational Infrastructures for School Improvement: A Way to Move Forward

Shapiro, R.B., Petry, H., and Gomez, L.M.

Selected Readings

Machine Classification of Peer Comments in Physics
Kwangsue Cho

Mining Student Behavior Models in Learning-by-Teaching Environments
Hogyong Jeong and Gautam Biswas

Questions

- ▶ Initial reactions?
- ▶ How would you characterize the objectives and methodology in each of these papers?
- ▶ What are the implications of this research?
- ▶ What do these papers say about EDM at-large?

Questions

- ▶ Are there connections to your existing work?
- ▶ What are opportunities could applying EDM offer?
- ▶ What further information would be useful before starting EDM projects?