Educational Technologies
WS2008-09

Educational Datamining
Plan of the Course

28.10.2008 Introduction and overview
04.11.2008: Intelligent tutoring systems (1) - Cognitive Tutors
11.11.2008: Intelligent tutoring systems (2) – ActiveMath
18.11.2008: Student modelling
25.11.2008: Student modelling
02.12.2008: Pedagogical components, instructional planning
09.12.2008: Meta-cognitive support (1) – Help
16.12.2008: Error diagnosis and feedback
06.01.2009: Error diagnosis and feedback
13.01.2009: Collaborative learning technologies
20.01.2009: Multi-Media Learning principles
27.01.2009: Web-based systems
03.02.2009: Educational data mining
24.02.2009: Project presentations by students
Machine Learning Techniques

- Decision Tree learning
- Artificial Neuronal Networks
- Instance-based learning
- Bayes Network reasoning and learning
- Naive Bayes
- Genetic Algorithms
- Support Vector Machines
- (Hidden) Markov Models
- Reinforcement learning
- Explanation-based learning
- Inductive logic programming
- Boosting
Overview

▶ General
  ▶ Motivation
  ▶ Questions to be answered

▶ Some Educational Data Mining techniques
  ▶ Decision tree learning (classification)
  ▶ Learning curve analysis

▶ Applications of EDM
Motivation: Macro-level analysis

Instructional Method
$I_T$ vs. $I_C$?

Novice knowledge state

Learning Processes

Expert knowledge state

Robust Learning measures:
- Transfer
- Long-term retention
- Accelerated future learning

Pre-test

Normal Post-test

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Macro-level analysis treats knowledge states & learning processes as black boxes

Instructional Method $I_T$ vs. $I_C$

Pre-test

Normal Post-test

Robust Learning measures:
Transfer  Long-term retention  Accelerated future learning

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Unpacking Learning into States and Events

Novice knowledge state

Instructional Event

Learning Event

Assessment event

Pre-test

Instructional Event

Learning Event

Assessment event

Normal Post-test

Expert (desired) knowledge state

Instructional Event

Learning Event

Robust Learning measures
Motivation

- Making good use of the raw data collected by e-Learning and educational technology
  - Motivated by:
    - Proliferation of data from many Internet-based educational systems
    - Development on real data rather than conjecture and intuition

- Use educational data to evaluate student behavior, support teachers, improve systems

- Interactive Learning Environments
  - Large-scale and longitudinal analysis
  - How do educational technologies work
What is Educational Data Mining? Techniques

▶ **Machine Learning**
  ▶ Many techniques available -- and have been largely prepackaged, e.g.,
    ▶ Decision Trees
    ▶ Support Vector Machines
    ▶ Boosting algorithms
  ▶ Off-the shelf tools
    ▶ WEKA (A flightless bird, found in New Zealand)
    ▶ YALE (Yet Another Learning Environment)

▶ **Statistical Techniques**
  ▶ Bayesian analysis of data
  ▶ Learning curve analysis
Uses of Educational Data Mining: general

- Make use of available logs of learner actions
- Generate higher-level information

- **for system designers**
  - To improve models/parameters underlying ITS
  - To make content more effective

- **for students**: provide information or even support
- **for teachers/tutors**: provide information to analyze and react

- **for research**: provide tools and data
Uses: Examples

- Find common errors committed or help requests, so that subsequent versions of system can better address them.
- Find learning factors that are more/less relevant, so that content/curriculum can be adjusted.
- Find factors and actions that are relevant for adapting to students’ attitudes, learning styles, etc.
- Discover ways that students “game” the system and how to react to this.
- Collect/aggregate/analyze data of groups.
Some more Data Mining Questions

▶ What’s the nature of knowledge students are learning? How can we discover cognitive models of student learning?

▷ LFA

▶ What features of a tutor lead to the most learning?

▷ Learning Decomposition, extends LFA to explore different rates of learning due to different forms of instruction

▶ How to extract reliable inferences about causal mechanisms from correlations in data?

▷ Causal modeling using Tetrad
What Makes ITS Data Analyzable?

- **Multiple grain sizes**
- **Data machine-readable**
  - replacing freehand drawing with a limited palette of graphical objects and operations
  - replacing free-form responses with menu selections or NLU
  - Add annotations by humans
- **Timing**
  - how long they take to read pages, sections, or sentences
  - time spent on specific activities
  - response time to multiple choice questions to detect guessing
Action Analysis Process

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock

EdTechs WS 2008/09
Related Fields

▶ Student Modeling
  ▶ Assessing students’ knowledge (learner model), no analysis of students’ behavior

▶ Web Server Log Analysis
  ▶ Focus on user navigation, no pedagogical concern

▶ Data mining, Machine Learning
  ▶ Technology for analyzing large databases, no pedagogical concern
General: Log Analysis as Data Mining

- **Data collection**
  - Server-side data collection
  - Client-side data collection
  - Integration of additional data, such as ontological information on learning material and user registration information
Data preparation

- **Fixed transformations** for example timestamp conversions, learning material lookup, and extraction of URL parameters
- **User identification** using heuristics if data collection does not provide explicit user identification
- **Session identification** for example beginning and end of sessions, pauses, etc
- **Flexible transformations** for example cumulating information from single clicks to summarize session information
- **Data cleaning** for example removing demo sessions, detect user name changes, etc.
Log Analysis as Data Mining

- Reporting facilities
  - **Access statistics** such as hits, page impressions, peak visit times, duration of sessions, average amount of pages seen
  - **User statistics** such as first time users, returning users, number of sessions per user, average time between user sessions
  - **Session statistics** ranging from number and duration of sessions up to information of referrers, entry points and exit points
  - **More statistics** provide information on effectiveness of hints, click through rates, or failure reports among others
Log Analysis as Data Mining

Data mining

- **Association analysis** such as analyzing typical navigation paths
- **Sequence analysis** e.g. for controlling these typical navigation paths for specific users or user groups over time
- **Cluster analysis** e.g. for grouping users according to their behavior and their characteristics
- **Classification analysis** for instance in order to try to describe these clusters with classification rules such as decision trees
Log Analysis as Data Mining

- Data Collection
- Data Preparation
- Reporting Facilities
- Data Mining
- Result Deployment

- **Result deployment**
  - **Profile generation** such as profiles of specific users
  - **Teacher reports** such as generation of some high-level report
  - **Personalization** such as providing personalized links or contents to specific users or user groups
Ordering of sequences and limits

Let $x_n$ and $y_n$ denote convergent real sequences with $x = \lim_{n \to \infty} x_n$ and $y = \lim_{n \to \infty} y_n$, as well as $x_n \leq y_n \in \mathbb{R}$ for all $n$. This implies $x \leq y$. 
System Architecture

Interactive Learning Environment (ActiveMath) → User Logs → Updater → Log Database → Analyzer → Learning material (mBase)
ActiveMath Log (1/2)

<ActivemathEvent type="UserLoggedIn" ts="1115212152486" source="org.activemath.webapp.controller.Login">
  <User id="Caroline18"/>
  <Session id="BDC43AC6CEAE7C1553C6F9EC03E62157"/>
  <UserLoggedIn remoteAddr="134.96.236.41" userAgent="Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"/>
</ActivemathEvent>

<ActivemathEvent type="UserPropertyChanged" ts="1115214804505" source="org.activemath.webapp.user.User">
  <User id="Caroline18"/>
  <UserPropertyChanged property="language" oldValue="en" newValue="de"/>
</ActivemathEvent>

<ActivemathEvent type="ItemPresented" ts="1115214852848" source="org.activemath.webapp.controller.ViewBook">
  <User id="Caroline18"/>
  <Session id="BDC43AC6CEAE7C1553C6F9EC03E62157"/>
  <Item type="symbol" id="mbase://LeAM_calculus/basics/Q_plus"/>
</ItemPresented>
</ActivemathEvent>
ActiveMath Log (2/2)

<ActivemathEvent type="ExerciseStarted" ts="1115215182332" source="org.activemath.webapp.exercises.ExerciseController">
  <User id="Caroline18"/>
  <Session id="BDC43AC6CEAE7C1553C6F9EC03E62157"/>
  <Item type="exercise" id="mbase://LeAM_calculus/diffquot/exer_slope_all"/>
  <ExerciseStarted/>
</ActivemathEvent>

<ActivemathEvent type="ExerciseStep" ts="1115215224918" source="org.activemath.webapp.exercises.ExerciseController">
  <User id="Caroline18"/>
  <Session id="BDC43AC6CEAE7C1553C6F9EC03E62157"/>
  <Item type="exercise" id="mbase://LeAM_calculus/diffquot/exer_slope_all"/>
  <ExerciseStep input="1/40;"/>
</ActivemathEvent>

<ActivemathEvent type="MasteryChanged" ts="1115215225775" source=""">
  <User id="Caroline18"/>
  <Item type="definition.simple" id="mbase://LeAM_calculus/diffquot/def_average_slope"/>
  <MasteryChanged from="{KN=0.02, AP=0.02, CP=0.02}" to="{AP=0.0, KN=0.9, CP=0.0}"/>
</ActivemathEvent>
Log Database Schema

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock

EdTechs WS 2008/09
Log Database Schema

![Database Schema Image]

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Engine</th>
<th>Rows</th>
<th>Data length</th>
<th>Update time</th>
</tr>
</thead>
<tbody>
<tr>
<td>event</td>
<td>MyISAM</td>
<td>24019</td>
<td>211 kB</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventSearched</td>
<td>MyISAM</td>
<td>29</td>
<td>0.9 kB</td>
<td>2005-05-30 14:32:14</td>
</tr>
<tr>
<td>eventExercised</td>
<td>MyISAM</td>
<td>49</td>
<td>3.3 kB</td>
<td>2005-05-30 14:32:15</td>
</tr>
<tr>
<td>eventStarted</td>
<td>MyISAM</td>
<td>108</td>
<td>6.6 kB</td>
<td>2005-05-30 14:32:15</td>
</tr>
<tr>
<td>eventFocusChanged</td>
<td>MyISAM</td>
<td>175</td>
<td>1.5 kB</td>
<td>2005-05-30 14:32:15</td>
</tr>
<tr>
<td>eventHappiness</td>
<td>MyISAM</td>
<td>0</td>
<td>0 B</td>
<td>2005-05-30 14:32:15</td>
</tr>
<tr>
<td>eventRating</td>
<td>MyISAM</td>
<td>2235</td>
<td>1.4 MB</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingRemoved</td>
<td>MyISAM</td>
<td>127</td>
<td>1.1 kB</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingViewed</td>
<td>MyISAM</td>
<td>127</td>
<td>1.1 kB</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingViewed</td>
<td>MyISAM</td>
<td>925</td>
<td>35.5 kB</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingChanged</td>
<td>MyISAM</td>
<td>305</td>
<td>11.9 kB</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingDeleted</td>
<td>MyISAM</td>
<td>0</td>
<td>0 B</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingPlanned</td>
<td>MyISAM</td>
<td>0</td>
<td>0 B</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingRemained</td>
<td>MyISAM</td>
<td>0</td>
<td>0 B</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingGoal</td>
<td>MyISAM</td>
<td>64</td>
<td>6.2 kB</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>eventRatingProperty</td>
<td>MyISAM</td>
<td>26</td>
<td>1.1 kB</td>
<td>2005-05-30 14:32:14</td>
</tr>
<tr>
<td>goals</td>
<td>MyISAM</td>
<td>0</td>
<td>0 B</td>
<td>2005-05-30 14:32:20</td>
</tr>
<tr>
<td>assessment</td>
<td>MyISAM</td>
<td>2775</td>
<td>68 kB</td>
<td>2005-05-30 14:32:15</td>
</tr>
<tr>
<td>assessment</td>
<td>MyISAM</td>
<td>2775</td>
<td>68 kB</td>
<td>2005-05-30 14:32:15</td>
</tr>
</tbody>
</table>

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock

EdTechs WS 2008/09
Sample Query (1/3)

```
select
    hour (datetime) as hour, count(*) as hourcount
from eventext
group by hour
```

How many user activities at what time of the day?

- **select**: start of query
- **hour**: extracts hour information
- **as**: define new field name
- **from**: specify table from which to select the information
- **group by**: make a bin for each different value
- **count(*)**: count the number of entries in each bin
## Result Set

<table>
<thead>
<tr>
<th>hour</th>
<th>hourcount</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>59</td>
</tr>
<tr>
<td>8</td>
<td>4959</td>
</tr>
<tr>
<td>9</td>
<td>16366</td>
</tr>
<tr>
<td>10</td>
<td>10664</td>
</tr>
<tr>
<td>11</td>
<td>10826</td>
</tr>
<tr>
<td>12</td>
<td>132925</td>
</tr>
<tr>
<td>13</td>
<td>50168</td>
</tr>
<tr>
<td>14</td>
<td>4742</td>
</tr>
<tr>
<td>15</td>
<td>6093</td>
</tr>
<tr>
<td>16</td>
<td>4801</td>
</tr>
<tr>
<td>17</td>
<td>1057</td>
</tr>
<tr>
<td>18</td>
<td>281</td>
</tr>
<tr>
<td>19</td>
<td>461</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>22</td>
<td>95</td>
</tr>
</tbody>
</table>

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock

EdTechs WS 2008/09
Sample Query (2/3)

```
select page, count(*) as pagecount
from eventpagepresented
group by page
order by pagecount desc
```

What are the most popular pages?

- **order by**: list entries in ascending order
- **desc**: descending order instead of ascending order
## Result Set

<table>
<thead>
<tr>
<th>page</th>
<th>pagecount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1534</td>
</tr>
<tr>
<td>2</td>
<td>704</td>
</tr>
<tr>
<td>50</td>
<td>588</td>
</tr>
<tr>
<td>20</td>
<td>469</td>
</tr>
<tr>
<td>3</td>
<td>432</td>
</tr>
<tr>
<td>32</td>
<td>401</td>
</tr>
<tr>
<td>4</td>
<td>321</td>
</tr>
<tr>
<td>21</td>
<td>311</td>
</tr>
<tr>
<td>28</td>
<td>302</td>
</tr>
<tr>
<td>31</td>
<td>255</td>
</tr>
<tr>
<td>25</td>
<td>203</td>
</tr>
<tr>
<td>24</td>
<td>187</td>
</tr>
<tr>
<td>33</td>
<td>185</td>
</tr>
<tr>
<td>29</td>
<td>183</td>
</tr>
<tr>
<td>26</td>
<td>169</td>
</tr>
<tr>
<td>38</td>
<td>166</td>
</tr>
<tr>
<td>39</td>
<td>165</td>
</tr>
<tr>
<td>36</td>
<td>149</td>
</tr>
</tbody>
</table>
Sample Query (3/3): Duration

```sql
select e1.user,
    from_unixtime(left(e1.ts, 10)) as logintime,
    (e2.ts - e1.ts)/60000 as duration
from event e1
left join event e2 using(session)
where e1.type = 'LoggedIn'
and e2.type = 'LoggedOut'
and e2.ts - e1.ts > 0
order by duration
```

- **left join**: combine two tables by concatenating entries
- **using**: only concatenate entries where this field is identical
- **where**: define condition on selecting entries
- **and**: define further condition
- **from_unixtime(left)**: date conversion
- **-, /, =, >**: arithmetic or logical operations
### Result Set

<table>
<thead>
<tr>
<th>user</th>
<th>logintime</th>
<th>duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>bu-yannis</td>
<td>2003-03-04 12:17:30</td>
<td>27.63370</td>
</tr>
<tr>
<td>6b-Baerchen1</td>
<td>2004-11-12 12:19:21</td>
<td>27.63488</td>
</tr>
<tr>
<td>asde</td>
<td>2004-10-07 11:04:28</td>
<td>27.65200</td>
</tr>
<tr>
<td>otto</td>
<td>2004-12-08 08:33:03</td>
<td>27.65482</td>
</tr>
<tr>
<td>edgar</td>
<td>2005-01-28 12:28:15</td>
<td>27.66843</td>
</tr>
<tr>
<td>6b-pluto1</td>
<td>2004-11-26 12:23:44</td>
<td>27.69603</td>
</tr>
<tr>
<td>6d-neptun</td>
<td>2005-03-11 12:59:17</td>
<td>27.74727</td>
</tr>
<tr>
<td>6a-magic-m-m</td>
<td>2005-01-10 12:19:40</td>
<td>27.78980</td>
</tr>
<tr>
<td>6a-blume</td>
<td>2005-02-21 12:17:02</td>
<td>27.87835</td>
</tr>
<tr>
<td>6b-Drossel</td>
<td>2005-01-14 12:21:12</td>
<td>27.94512</td>
</tr>
<tr>
<td>otto</td>
<td>2004-12-07 10:33:21</td>
<td>27.95210</td>
</tr>
</tbody>
</table>
School Experiment
Visualization of Query Results
## Attributes for Further Analysis

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Generation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>Manual</td>
<td>User name (not used for the decision tree learning)</td>
</tr>
<tr>
<td>Class</td>
<td>Manual</td>
<td>Course, each comprised of about 20 students (not used for the decision tree learning)</td>
</tr>
<tr>
<td>Teacher</td>
<td>Manual</td>
<td>Each class has been split into two subgroups, with each being taught by another teacher</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(not used for the decision tree learning)</td>
</tr>
<tr>
<td>Gender</td>
<td>Manual</td>
<td>Male or female</td>
</tr>
<tr>
<td>Integration pupil</td>
<td>Manual</td>
<td>Whether the student is handicapped</td>
</tr>
<tr>
<td>Post test result</td>
<td>Manual</td>
<td>Results in the post test done in writing (binned into low, medium, and high for the decision tree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>learning)</td>
</tr>
<tr>
<td>Ex_started</td>
<td>Automatic</td>
<td>Number of exercises started</td>
</tr>
<tr>
<td>Ex_finished</td>
<td>Automatic</td>
<td>Number of exercises finished</td>
</tr>
<tr>
<td>Num_successes</td>
<td>Automatic</td>
<td>Number of successful exercises</td>
</tr>
<tr>
<td>Avg_reading</td>
<td>Automatic</td>
<td>Average number of reading actions in a session</td>
</tr>
<tr>
<td>Avg_solving</td>
<td>Automatic</td>
<td>Average number of exercise solving actions in a session</td>
</tr>
<tr>
<td>DictUsed</td>
<td>Automatic</td>
<td>Whether the student used the dictionary for searching information</td>
</tr>
<tr>
<td>WorkedOffTime</td>
<td>Automatic</td>
<td>Whether the student accessed the learning environment beyond lesson hours, e.g. from home or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>during free periods</td>
</tr>
<tr>
<td>Ex_finished_rate</td>
<td>Automatic</td>
<td>Rate of finished exercises to all started exercises</td>
</tr>
<tr>
<td>Ex_success_rate</td>
<td>Automatic</td>
<td>Rate of successful exercises to all finished exercises</td>
</tr>
</tbody>
</table>
Visualization of Dependencies
Classification Analysis

Goals

- Represent interrelations and dependencies in the data (characterize)
- For some techniques: Provide explicit and intelligible description (high-level)
- Classify new data (prediction)

Main ingredients

- Training set: Data used for machine learning
- Test set: Data used for evaluating the quality of the classifier
- (Positive/Negative) Example: Element of the training or test set characterized by the (binary) classifier

Result: Classifier = Indicator for class membership
Classification Analysis

1.) Learn classifier

One interesting (hidden) attribute ➔ Classifier ➔ Set of observed (logged) attributes

2.) Use classifier

? ➔ Classifier ➔ New set of observed (logged) attributes
Classifier: Decision Trees

<table>
<thead>
<tr>
<th>Outlook</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Windy</th>
<th>Play</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>False</td>
<td>No</td>
</tr>
<tr>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>True</td>
<td>No</td>
</tr>
<tr>
<td>Overcast</td>
<td>Hot</td>
<td>High</td>
<td>False</td>
<td>Yes</td>
</tr>
<tr>
<td>Rainy</td>
<td>Mild</td>
<td>Normal</td>
<td>False</td>
<td>Yes</td>
</tr>
</tbody>
</table>

...
Example: Decision Trees

\[ \text{temperature} = \text{hot} \land \text{windy} = \text{true} \land \text{humidity} = \text{normal} \land \text{outlook} = \text{sunny} \rightarrow \text{play} = ? \]
Constructing decision trees

- Normal procedure: top down in recursive divide-and-conquer fashion
  - First: attribute is selected for root node and branch is created for each possible attribute value
  - Then: the instances are split into subsets (one for each branch extending from the node)
  - Finally: procedure is repeated recursively for each branch, using only instances that reach the branch
- Process stops if all instances have the same class
Which attribute to select?
A criterion for attribute selection

- Which is the best attribute?
  - The one which will result in the smallest tree
  - Heuristic: choose the attribute that produces the “purest” nodes
- Popular *impurity criterion: information gain*
  - Information gain increases with the average purity of the subsets that an attribute produces
- Strategy: choose attribute that results in greatest information gain
Computing information

- Information is measured in \textit{bits}
  - Given a probability distribution, the info required to predict an event is the distribution’s \textit{entropy}
  - Entropy gives the information required in bits (this can involve fractions of bits!)
- Formula for computing the entropy:

\[
\text{entropy}(p_1, p_2, \ldots, p_n) = -p_1 \log p_1 - p_2 \log p_2 \ldots - p_n \log p_n
\]
Example: attribute “Outlook”

- “Outlook” = “Sunny”:
  \[ \text{info}([2,3]) = \text{entropy}(2/5,3/5) = -2/5 \log(2/5) - 3/5 \log(3/5) = 0.971 \text{ bits} \]

- “Outlook” = “Overcast”:
  \[ \text{info}([4,0]) = \text{entropy}(1,0) = -1 \log(1) - 0 \log(0) = 0 \text{ bits} \]

- “Outlook” = “Rainy”:
  \[ \text{info}([3,2]) = \text{entropy}(3/5,2/5) = -3/5 \log(3/5) - 2/5 \log(2/5) = 0.971 \text{ bits} \]

- Expected information for attribute:
  \[ \text{info}([3,2],[4,0],[3,2]) = (5/14) \times 0.971 + (4/14) \times 0 + (5/14) \times 0.971 = 0.693 \text{ bits} \]

Note: this is normally not defined.
Computing the information gain

- Information gain: information before splitting – information after splitting

\[
\text{gain}("Outlook") = \text{info}([9,5]) - \text{info}([2,3],[4,0],[3,2]) = 0.940 - 0.693 = 0.247 \text{ bits}
\]

- Information gain for attributes from weather data:

\[
\text{gain}("Outlook") = 0.247 \text{ bits}
\]
\[
\text{gain}("Temperature") = 0.029 \text{ bits}
\]
\[
\text{gain}("Humidity") = 0.152 \text{ bits}
\]
\[
\text{gain}("Windy") = 0.048 \text{ bits}
\]
Continuing to split

gain("Temperature") = 0.571 bits

gain("Humidity") = 0.971 bits

gain("Windy") = 0.020 bits
The final decision tree

- Note: not all leaves need to be pure; sometimes identical instances have different classes
  ⇒ Splitting stops when data can’t be split any further
## Attributes for Predicting Post-Test Results

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Generation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>Manual</td>
<td>User name (not used for the decision tree learning)</td>
</tr>
<tr>
<td>Class</td>
<td>Manual</td>
<td>Course, each comprised of about 20 students (not used for the decision tree learning)</td>
</tr>
<tr>
<td>Teacher</td>
<td>Manual</td>
<td>Each class has been split into two subgroups, with each being taught by another teacher (not used for the decision tree learning)</td>
</tr>
<tr>
<td>Gender</td>
<td>Manual</td>
<td>Male or female</td>
</tr>
<tr>
<td>Integration pupil</td>
<td>Manual</td>
<td>Whether the student is handicapped</td>
</tr>
<tr>
<td><strong>Post test result</strong></td>
<td>Manual</td>
<td>Results in the post test done in writing (binned into low, medium, and high for the decision tree learning)</td>
</tr>
<tr>
<td>Ex_started</td>
<td>Automatic</td>
<td>Number of exercises started</td>
</tr>
<tr>
<td>Ex_finished</td>
<td>Automatic</td>
<td>Number of exercises finished</td>
</tr>
<tr>
<td>Num_successes</td>
<td>Automatic</td>
<td>Number of successful exercises</td>
</tr>
<tr>
<td>Avg_reading</td>
<td>Automatic</td>
<td>Average number of reading actions in a session</td>
</tr>
<tr>
<td>Avg_solving</td>
<td>Automatic</td>
<td>Average number of exercise solving actions in a session</td>
</tr>
<tr>
<td>DictUsed</td>
<td>Automatic</td>
<td>Whether the student used the dictionary for searching information</td>
</tr>
<tr>
<td>WorkedOffTime</td>
<td>Automatic</td>
<td>Whether the student accessed the learning environment beyond lesson hours, e.g. from home or during free periods</td>
</tr>
<tr>
<td>Ex_finished_rate</td>
<td>Automatic</td>
<td>Rate of finished exercises to all started exercises</td>
</tr>
<tr>
<td>Ex_success_rate</td>
<td>Automatic</td>
<td>Rate of successful exercises to all finished exercises</td>
</tr>
</tbody>
</table>
Decision Tree for Post Test Result

\[ \text{ex\_success\_rate} = 0.76 \ & \ \text{ex\_finished\_rate} = 0.93 \ & \ \text{avg\_reading} = 56 \ \implies \ \text{post test result} = ? \]
### Confusion Matrix and Cross Validation

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>11</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Medium</td>
<td>9</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>5</td>
<td>11</td>
</tr>
</tbody>
</table>

- **Actual class**: Training set
- **Predicted class**: Test set

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock

EdTechs WS 2008/09
Learning Factors Analysis (LFA)
[Hao Cen, Kenneth Koedinger, Brian Junker]

Logistic regression, model scoring to fit statistical models to student log data

Statistics

Difficulty Factors

Combinatorial Search

A* search algorithm with “smart” operators for proposing new cognitive models based on the factors

a set of factors that make a problem-solving step more difficult for a student

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Motivation: Learning curve analysis by eye …

\[ Y = aX^b. \]
Learning Curves

Example: Steps in programming problems where the function has two parameters

[Corbett, Anderson, O’Brien, 1995]
Not a smooth learning curve -> this knowledge component model is wrong. Does not capture genuine student difficulties.
This more specific knowledge component (KC) model (2 KCs) is also wrong -- still no smooth drop in error rate.
Ah! Now we are getting a smooth learning curve. This even more specific decomposition (12 KCs) better tracks the nature of student difficulties & transfer for one problem situation to another.
Motivation: automate learning curve analysis

- Manual learning curve analysis
  - Identify blips by hand & eye
  - Manually create a new model
  - Qualitative judgment

- Need to automatically:
  - Identify blips by system
  - Propose alternative cognitive models
  - Evaluate each model quantitatively
Learning Factors Analysis, Methodology

Goal: Cognitive Model Evaluation + Improvement

Base Model – skills and additional factors

Data acquisition:
- Difficulty factors
- Operators
- Use frequency
- Success

Model search
- Logistic regression
- Model selection
Base Model: Geometry

- Cognitive model, Geometry, 15 skills:

1. Circle-area
2. Circle-circumference
3. Circle-diameter
4. Circle-radius
5. Compose-by-addition
6. Compose-by-multiplication
7. Parallelogram-area
8. Parallelogram-side
9. Pentagon-area
10. Pentagon-side
11. Trapezoid-area
12. Trapezoid-base
13. Trapezoid-height
14. Triangle-area
15. Triangle-side
Data Acquisition

- Log data
- Difficulty factors
- Operators
- Use frequency (opportunities)
- Success
### Log Data - Skills in the Base Model (of problems)

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-addition</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area</td>
<td>3</td>
</tr>
</tbody>
</table>
Difficulty Factors

Difficulty Factors - a property of the problem that causes student difficulties

e.g., factors
- Embed: alone, embed
- Backward: forward, backward
- Repeat: initial, repeat
- FigurePart: area, area-difference, area-combination, diameter, circumference, radius, side, segment, base, height

Embed factor: Whether figure is embedded in another figure or by itself (alone)

Example for skill Circle Area:
Q: Given AB = 2, find circle area in the context of the problem goal to calculate the shaded area

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
EdTechs WS 2008/09
Model Operators

- **Split Embed** ⇒ circle-area-embed, circle-area-alone
- **Add** : add skill
- **Merge Embed** ⇐ circle-area-embed, circle-area-alone
Split, example

Binary Split - splits a skill with a factor value and a skill without the factor value.

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
<th>Opportunity</th>
<th>Factor- Embed</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area</td>
<td>1</td>
<td>alone</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area</td>
<td>2</td>
<td>embed</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-addition</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area</td>
<td>3</td>
<td>alone</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student</th>
<th>Step</th>
<th>Skill</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p1s1</td>
<td>Circle-area-alone</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s1</td>
<td>Circle-area-embed</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s2</td>
<td>Rectangle-area</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p2s3</td>
<td>Compose-by-addition</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>p3s1</td>
<td>Circle-area-alone</td>
<td>2</td>
</tr>
</tbody>
</table>
Model Search: Logistic Regression model

logistic regression is a statistical model (for error rate) used for prediction of the probability of occurrence of an event by fitting data to a logistic curve

\[ f(z) = \frac{1}{1 + e^{-z}} \]

logistic function/curve

\[ p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i})}}. \]

\[ \ln \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}. \]
The Statistical Model

\[ \ln \left( \frac{p}{1-p} \right) = \sum \alpha_i X_i + \sum \beta_j Y_j + \sum \gamma_j Y_j T_j. \]

Probability of getting a step correct \((p)\) is proportional to:

- if student \(i\) performed this step = \(X_i\),
  add overall "smarts" of that student = \(\alpha_i\)
- if skill \(j\) is needed for this step = \(Y_j\),
  add easiness of that skill = \(\beta_j\)
  add product of number of opportunities to learn = \(T_j\)
  and amount gained for each opportunity = \(\gamma_j\)

Use logistic regression because response is discrete (correct or not)
Probability \((p)\) is transformed by "log odds"
"stretched out" with sigmoid curve to not bump up against 0 or 1
Model Selection: Combinatorial Search

Goal: select model from logistic regression model space

Steps:

1. Start from an initial node/model in search graph
2. Iteratively create new child nodes by operators, e.g., splitting a model using factors
3. Employ a heuristic (e.g. fit to learning curve) to rank each node
4. Expand from a new node in the heuristic order by going back to step 2
Model Selection: A* Search and Heuristics

- **best-first**, graph search algorithm for least-cost path from a given initial node to a goal node (out of one or more possible goals).

- uses a distance-plus-cost heuristic function $f(x)$ to determine order in which the search visits nodes in tree

  $$f(x) = g(x) + h(x)$$

  - path-cost function $g(x)$: cost from starting node to current node
  - admissible "heuristic estimate" of distance to goal $h(x)$

- **R-square, Log likelihood, AIC, BIC** search heuristics
  - Log likelihood measures fit between data and model
  - $\text{AIC} = -2\times\text{log-likelihood} + 2\times\#\text{parameters}$
  - $\text{BIC} = -2\times\text{log-likelihood} + \#\text{parameters} \times \log (\#\text{observations})$
Best-first Search

- an informed graph search algorithm guided by a heuristic
- Heuristics – AIC, BIC
- Start from an existing model

Original Model
AIC = 5328

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Best-first Search

- an informed graph search algorithm guided by a heuristic
- Heuristics – AIC, BIC
- Start from an existing model

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
EdTechs WS 2008/09
Best-first Search

- an informed graph search algorithm guided by a heuristic
- Heuristics – AIC, BIC
- Start from an existing model

Original Model
AIC = 5328

Split by Embed

Split by Backward

Add Formula

50+

5301

5320 5322 5313

5320 5322

5312

5320
Best-first Search

- an informed graph search algorithm guided by a heuristic
- Heuristics – AIC, BIC
- Start from an existing model

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Best-first Search

- an informed graph search algorithm guided by a heuristic
- Heuristics – AIC, BIC
- Start from an existing model

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Best-first Search

- an informed graph search algorithm guided by a heuristic
- Heuristics – AIC, BIC
- Start from an existing model

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock

EdTechs WS 2008/09
Actual / Predicted Learning Curves

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Experiments and Results:
experiment 1

Q: How can we describe learning behavior in terms of an existing cognitive model?

A: Fit logistic regression model in equation and get coefficients
Experiment 1

Results:

Higher intercept of skill -> easier skill
Higher slope of skill -> faster students learn it

<table>
<thead>
<tr>
<th>Skill</th>
<th>Intercept</th>
<th>Slope</th>
<th>Avg Opportunities</th>
<th>Initial Probability</th>
<th>Avg Probability</th>
<th>Final Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallelogram-area</td>
<td>2.14</td>
<td>-0.01</td>
<td>14.9</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>Pentagon-area</td>
<td>-2.16</td>
<td>0.45</td>
<td>4.3</td>
<td>0.2</td>
<td>0.63</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Student Intercept

<table>
<thead>
<tr>
<th>Student</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>student0</td>
<td>1.18</td>
</tr>
<tr>
<td>student1</td>
<td>0.82</td>
</tr>
<tr>
<td>student2</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Higher intercept of student -> student initially knew more

Model Statistics

<table>
<thead>
<tr>
<th>Model Statistics</th>
<th>AIC</th>
<th>BIC</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>3,950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>4,285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>0.083</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The AIC, BIC & MAD statistics provide alternative ways to evaluate models
MAD = Mean Absolute Deviation
Experiment 2

Q: How can we improve a cognitive model?
A: Run LFA on data including factors and search through model space
### Experiment 2 – Results with BIC

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Splits: 3</td>
<td>Number of Splits: 3</td>
<td>Number of Splits: 2</td>
</tr>
<tr>
<td>2. Binary split circle-radius by repeat repeat</td>
<td>2. Binary split circle-radius by repeat repeat</td>
<td>2. Binary split circle-radius by repeat repeat</td>
</tr>
<tr>
<td>Number of Skills: 18</td>
<td>Number of Skills: 18</td>
<td>Number of Skills: 17</td>
</tr>
<tr>
<td>AIC: 3,888.67</td>
<td>AIC: 3,888.67</td>
<td>AIC: 3,897.20</td>
</tr>
<tr>
<td>BIC: 4,248.86</td>
<td>BIC: 4,248.86</td>
<td>BIC: 4,251.07</td>
</tr>
<tr>
<td>MAD: 0.071</td>
<td>MAD: 0.071</td>
<td>MAD: 0.075</td>
</tr>
</tbody>
</table>

#### Splitting Compose-by-multiplication into two skills – CMarea and CMsegment, making a distinction of the geometric quantity being multiplied

Source: Erica Melis, Ken Koedinger, Martin Mühlenbrock
Experiment 3

Q: Will some skills be better merged than if they are separate skills? Can LFA recover some elements of original model if we search from a merged model, given difficulty factors?

A: Run LFA on the data of a merged model, and search through the model space
Experiment 3 – Merged Model

Merge some skills in the original model to remove some distinctions, add a difficulty factor to consider.

The merged model has 8 skills:
- Circle-area, Circle-radius => Circle
- Circle-circumference, Circle-diameter => Circle-CD
- Parallelogram-area and Parallelogram-side => Parallelogram
- Pentagon-area, Pentagon-side => Pentagon
- Trapezoid-area, Trapezoid-base, Trapezoid-height => Trapezoid
- Triangle -area, Triangle -side => Triangle
- Compose-by-addition
- Compose-by-multiplication

Add difficulty factor “direction”: forward vs. backward
Experiment 3 – Results

- Recovered three skills (Circle, Parallelogram, Triangle)
  => distinctions made in the original model are necessary
- Partially recovered two skills (Triangle, Trapezoid)
  => some original distinctions necessary, some are not
- Did not recover one skill (Circle-CD)
  => original distinction may not be necessary
- Recovered one skill (Pentagon) in a different way
  => Original distinction may not be as significant as distinction caused by another factor
Beyond Experiments 1-3

Q: Can we use LFA to improve tutor curriculum by identifying over-taught or under-taught rules?
   Thus adjust their contribution to curriculum length without compromising student performance

A: Combine results from experiments 1-3
Beyond Experiments 1-3, Results

- **Parallelogram-side is over taught.**
  - high intercept (2.06), low slope (-.01).
  - initial success probability .94, average number of practices per student is 15

- **Trapezoid-height is under taught.**
  - low intercept (-1.55), positive slope (.27).
  - final success probability is .69, far away from the level of mastery, the average number of practices per student is 4.

- **Suggestions for curriculum improvement**
  - Reducing the amount of practice for Parallelogram-side should save student time without compromising their performance.
  - More practice on Trapezoid-height is needed for students to reach mastery.
Beyond Experiments 1-3, Results

How about Compose-by-multiplication?

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>slope</th>
<th>Avg Practice Opportunities</th>
<th>Initial Probability</th>
<th>Avg Probability</th>
<th>Final Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>-.15</td>
<td>.1</td>
<td>10.2</td>
<td>.65</td>
<td>.84</td>
<td>.92</td>
</tr>
</tbody>
</table>

With final probability .92 students seem to have mastered Compose-by-multiplication.
Beyond Experiments 1-3, Results

However, after split

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>slope</th>
<th>Avg Practice Opportunities</th>
<th>Initial Probability</th>
<th>Avg Probability</th>
<th>Final Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>-.15</td>
<td>.1</td>
<td>10.2</td>
<td>.65</td>
<td>.84</td>
<td>.92</td>
</tr>
<tr>
<td>CMarea</td>
<td>-.009</td>
<td>.17</td>
<td>9</td>
<td>.64</td>
<td>.86</td>
<td>.96</td>
</tr>
<tr>
<td>CMsegment</td>
<td>-1.42</td>
<td>.48</td>
<td>1.9</td>
<td>.32</td>
<td>.54</td>
<td>.60</td>
</tr>
</tbody>
</table>

CMarea does well with final probability .96

But CMsegment has final probability only .60 and an average amount of practice less than 2

Suggestions for curriculum improvement: increase the amount of practice for CMsegment
PSLC’s DataShop

- Researchers get data access, visualizations, statistical tools

- *Learning curves* track student learning over time

- Discover what concepts & skills students need help with
Learning curves reveal over- and under-practiced knowledge components.

Rectangle-area has an initial low error rate, but is practiced often.
Other DataShop Features

- **Error Reports**
  - Identify misconceptions by looking for common student errors
  - When do students ask for hints?
  - Are there alternative correct strategies?

- **Performance Profiler**

- **Export Data**
  - Get all or part of the data in tab-delimited file
  - Use your favorite analysis tools …
Research Questions for EDM

- Assessment questions
  - Can on-line embedded assessment replace standardized tests?
  - Can assessment be accurate if students are learning during test?

- Learning theory questions
  - What are the “elements of transfer” in human learning?
  - Is learning rate driven by student variability or content variability?
  - Can conceptual change be tracked & better understood?

- Instructional questions
  - What instructional moves yield the greatest increases in learning?
  - Use learning curves rather than ANOVA to evaluate learning experiments

- Beyond cognition
  - Can student affect & motivation be detected in on-line click stream data?
  - Can student metacognitive & self-regulated learning strategies be detected in on-line click stream data?