Using the Hint Factory to Compare Model-Based Tutoring Systems

Collin F. Lynch, Thomas W. Price, Min Chi, & Tiffany Barnes

North Carolina State University, Raleigh North Carolina.

EDM2015: 6/26/2015
Can we apply data-driven methods to evaluate and augment model-based tutoring systems?
Outline

Introduction

Overview

Problem

Cold Start

Conclusions
Model-Based Tutoring Systems

- Model-Based tutoring systems based upon Expert Systems.
- They are designed by domain experts.
- Built on rich domain models.
- Paired with classical problem solvers and heuristics.
- Based upon what is ideal.
Data-Driven Methods

- Data-Driven methods draw from Machine Learning.

- They generalize from prior student solutions to identify optimal paths and common errors.

- They then guide students in the absence of domain models.

- Based not on what is *ideal* but what students *do*. 
Comparisons

- Model-based systems are *robust* for their domains.
- They are also expensive to construct and difficult to expand.
- Data-driven methods are substantially cheaper.
- But they are limited by the available data.
Question

Can we apply data-driven methods to evaluate and augment model-based tutoring systems?
Study Overview

- Collected data from two closely-related Model-Based Tutors for Probability (Andes & Pyrenees).

- Applied the Hint Factory a data-driven hint-generation system to draw comparisons between the datasets.

- Evaluated the similarity of the systems to:
  1. Highlight differences in student behavior across the systems.
  2. Assess the impact of differing design decisions.
  3. Evaluate the potential to apply hints across systems.
Andes

- Andes is an ITS for Physics and Probability originally designed at the University of Pittsburgh.

- It uses a complex multi-modal interface and provides:
  - Immediate error feedback.
  - Remediation advice.
  - Pedagogical guidance.

- Students can solve problems in any coherent order.
Andes

Problem Statement

Variable Window

Equation Window

Dialogue Window
Pyrenees

- Pyrenees is an ITS for Physics and Probability based upon Andes.

- It uses an isomorphic domain model and problem set.

- It uses a menu-driven uni-modal interface that constrains students to apply the Target Variable Strategy.

- TVS is a backward-chaining problem-solving strategy guided by some domain heuristics.

- The Andes pedagogical advice is driven by the TVS but students were not required to follow it.
Pyrenees

Problem

For an event D, we have \( p(D) = 0.05 \). Given event D, events A, B and C have \( p(\text{AnBnC}|D) = 0.80 \), \( p((\text{AnBnC})|\neg D) = 0.2 \), Find \( p(D|\text{AnBnC}) \).

Variables

\[
\begin{align*}
\text{abc} & : p(\text{AnBnC}|D) \\
\text{abc} & = 0.8; p(\text{AnBnC}|\neg D) \\
\text{abc} & = 0.05; p(D) \\
\text{abc} & = p(D|\text{AnBnC}) \\
\text{abc} & : \text{****TARGET VARIABLE****}; p(\neg D)
\end{align*}
\]

Equations

For \( nd \): Apply the complement theorem on event \( D \) and \( \neg D \).

1) \( \text{abc} \times \text{abc} = \text{abc} \times \text{abc} \times \text{abc} \times \text{abc} \times \text{abc} \times \text{abc} \)

For \( \text{abc} \): Apply the bayes' rule on \( D|\text{AnBnC} \); here, event \( D \) and \( \neg D \) are mutually exclusive and exhaustive events.
The Hint Factory

- The Hint Factory takes an MDP-based approach to automatically extract hints from prior user data.

- Prior student data is stored as an *Interaction Network*:
  - A multigraph structure.
  - Nodes represent solution states.
  - Arrows represent problem-solving steps.

- HF then applies value iteration to identify optimal solutions.
Interaction Network
Datasets

- The datasets cover 11 identical probability problems.

- Andes data was drawn from an experiment conducted at the University of Pittsburgh designed to assess the impact of Andes and Pyrenees on students’ meta-cognitive skills.
  - 66 students were included in the dataset.
  - 25 - 72 attempts per problem average 35.8.
  - 81.7% on average were successful per problem.

- Pyrenees data was drawn from a study conducted at North Carolina State University.
  - 137 students completed the study.
  - The students were not required to solve all problems.
  - 83 - 102 attempts per problem, average 90.8.
  - 83.4% on average were successful per problem.
State and Action Representations

- Problem steps were represented as interaction networks.
- Solution states were represented as unordered sets of actions.
- Incorrect actions were ignored:
  - Pyrenees forces students to correct errors immediately.
  - Andes permits errors to remain on screen.
Problem Comparison

- Our goal is to examine the impact of the TVS on student solutions.
- We examined problem-specific interaction networks.
- Conducted a case study with problem Ex242 (#10 of 11).

Events $A$, $B$ and $C$ are mutually exclusive and exhaustive events with $p(A) = 0.2$ and $p(B) = 0.3$. For an event $D$, we know $p(D|A) = 0.04$, $p(D|B) = 0.03$, and $p(C|D) = 0.3$. Determine $p(B|D)$. 
Problem: Ex242
Problem: Analysis

- Pyrenees students were divided (almost) evenly between:
  - Applications of the Conditional Probability Theorem:
    \[ P(A \cap B) = P(A|B)P(B) \]
  - Applications of Bayes’ Theorem:
    \[ p(A|B) = \left( p(B|A) \ast p(A) \right) / p(B) \]
  - The former is ideal according to the Pyrenees model and the problem was designed to teach it.
  - The latter approach is shorter and is ideal for the Hint Factory.
Problem: Analysis

- The Andes students generated a wider range of paths.
- 62 of the 126 states were unique.
- No Andes student followed the ideal CPT path.
Cross-System Hints

- We conducted a cold-start experiment to assess the general system similarity.

- For each student $i$ we calculate the average number of known states in their solution path given a prior dataset of $1, 2, \ldots, n - 1$ peers.

- We then plot the average across students and problems.

- We calculated four curves:
  - $PvP$: Pyrenees students with a Pyrenees dataset.
  - $AvA$: Andes students with an Andes dataset.
  - $AvP$: Pyrenees students with an Andes dataset.
  - $PvA$: Andes students with a Pyrenees dataset.
Cold-Start Curves

The graph shows the percent of hints available as a function of the number of previously observed students. The curves are labeled as follows:

- **PvP**: Blue line
- **AvA**: Orange line
- **PvA**: Gray line
- **AvP**: Yellow line

The x-axis represents the number of previously observed students, ranging from 1 to 21. The y-axis represents the percent of hints available, ranging from 40% to 100%.
Limitations

- This study was conducted with two closely-related systems.
- Students were drawn from two distinct studies.
- The dataset covered 11 well-circumscribed problems.
- The authors were involved in the design of Andes, Pyrenees, and the Hint Factory.
Conclusions

1. Highlight differences in student behavior across the systems.
   - Pyrenees students were generally more homogeneous.
   - The variation observed in Andes involved a substantial number of unique steps.

2. Assess the impact of differing design decisions.
   - The scaffolding provided by Pyrenees did force some, but not all, students to ideal solutions.

3. Evaluate the potential to apply hints across systems.
   - Cold-start curves showed that data-driven hints can be used to bootstrap data across systems.
   - However the curves do not reach 100%.
   - Substantial changes produce new systems.
### 3 Questions

1. **What common goals exist for graph analysis in EDM?**
   - This work highlights the use of graph analysis to evaluate design decisions.

2. **What shared resources such as tools and repositories are required to support the community?**
   - We present a general methodology that uses existing tools (HF) to evaluate existing systems.
   - The Hint Factory and Interaction Network systems are being implemented for public release.

3. **How do the structures of the graphs and the analytical methods change with the applications?**
   - The Interaction Network is a general graph structure.
   - However, the design of the state and action representation is domain specific.
¡Gracias!