Using HMMs and bagged decision trees to leverage rich features of user and skill from the 2010 KDD Cup dataset on Educational Data mining

Zachary A. Pardos
Neil T. Heffernan (advisor)

Sponsors
The Spencer Foundation

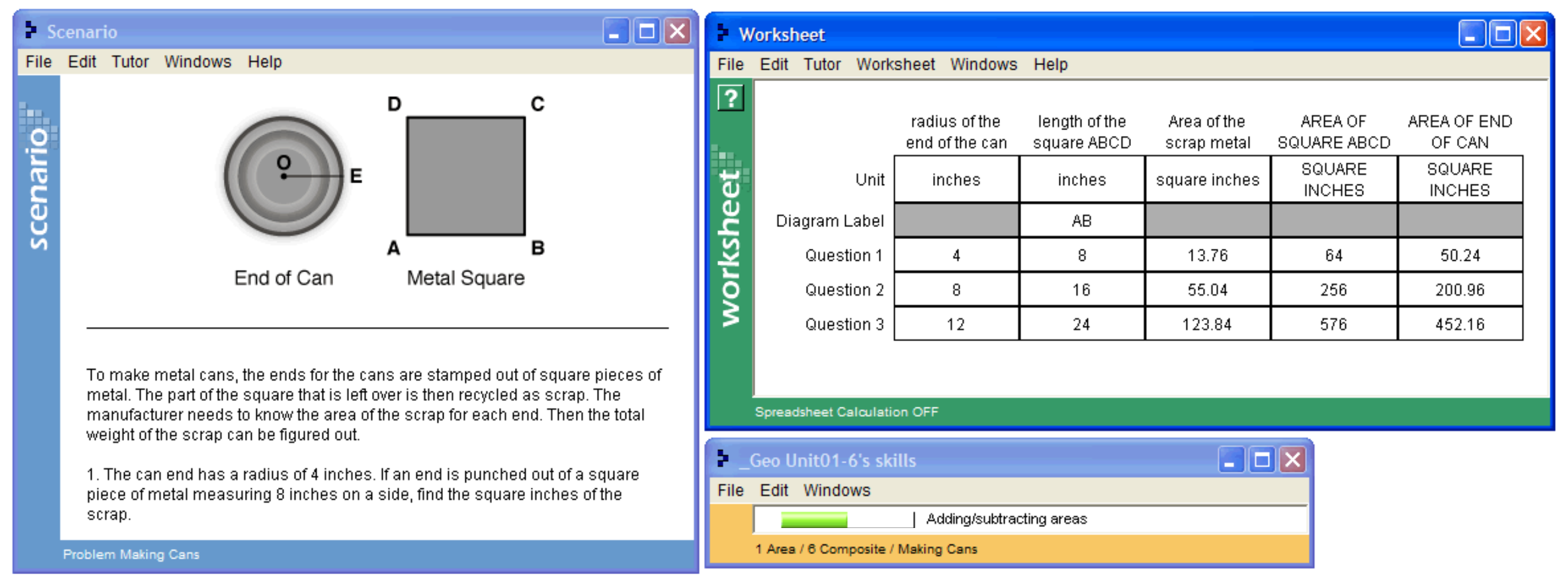
Overview

Goal: To compete in the 2010 KDD Cup challenge on educational data mining and advance the state of the art in user modeling and assessment of students in an Intelligent Tutoring System (ITS).

Background: The task of this challenge was to predict student performance on mathematical problems from logs of student interaction with an ITS. Accurate solutions can have a significant impact on education by optimizing students' time spent on task and potentially eliminating the need for standardized tests.

The Tutor and Dataset

The datasets for the competition came from student responses to the Cognitive Tutor, the largest ITS in the country, used by over 500,000 students per year.



To make metal cans, the ends for the cans are stamped out of square pieces of metal. The part of the square that is left over is then recycled as scrap. The manufacturer needs to know the area of the scrap for each end. Then the total weight of the scrap can be figured out.

1. The can end has a radius of 4 inches. If an end is punched out of a square piece of metal measuring 8 inches on a side, find the square inches of the scrap.

Problem Making Cans

	radius of the end of the can	length of the square ABCD	Area of the scrap metal	AREA OF SQUARE ABCD	AREA OF END OF CAN
Unit	inches	inches	square inches	SQUARE INCHES	SQUARE INCHES
Diagram Label		AB			
Question 1	4	8	13.76	64	50.24
Question 2	8	16	55.04	256	200.96
Question 3	12	24	123.84	576	452.16

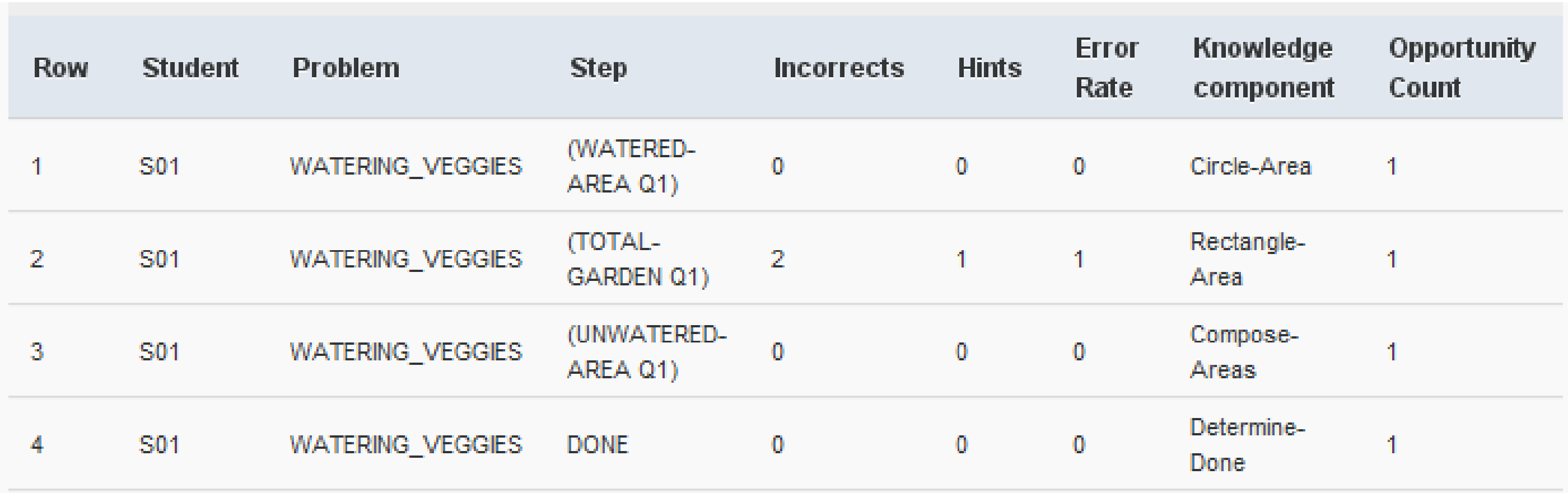
Spreadsheet Calculation OFF

Adding/subtracting areas

1 Area / 6 Composite / Making Cans

Facts about the dataset:

- Largest ever KDD Cup dataset (9 gigabytes on disk)
- Consisted of over 9 thousand students and 30 million rows of data from two algebra tutors
- Each row of the data corresponded to a student's response to the tutor and included 18 features describing attributes of the problem and of the student such as timestamp, skill name associated with the problem and number of times the student has attempted to answer problems of this skill



Row	Student	Problem	Step	Incorrects	Hints	Error Rate	Knowledge component	Opportunity Count
1	S01	WATERING_VEGGIES	(WATERED-AREA Q1)	0	0	0	Circle-Area	1
2	S01	WATERING_VEGGIES	(TOTAL-GARDEN Q1)	2	1	1	Rectangle-Area	1
3	S01	WATERING_VEGGIES	(UNWATERED-AREA Q1)	0	0	0	Compose-Areas	1
4	S01	WATERING_VEGGIES	DONE	0	0	0	Determine-Done	1

Student Modeling Approach

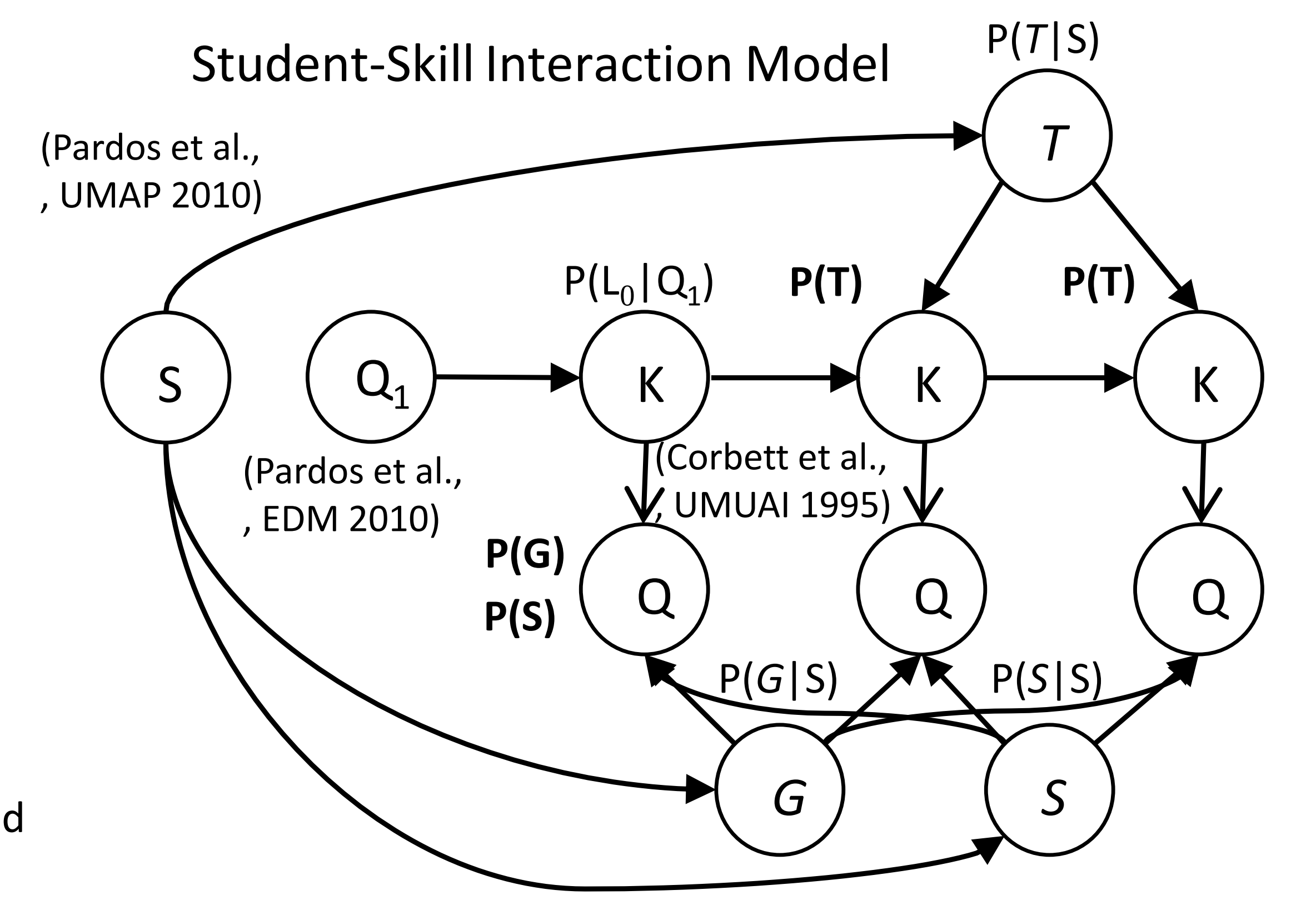
- The first approach was to create a model around the simple assumption that students' knowledge of a skill will increase with practice. An HMM was used to represent how the latent variable of knowledge impacts performance. The base model is developed from Knowledge Tracing, 1995, used in the Cognitive Tutor.

Model Parameters
$P(L_0)$ = Probability of initial knowledge
$P(L_0|Q_1)$ = Individual Cold start $P(L_0)$
$P(T)$ = Probability of learning
$P(T|S)$ = Students' Individual $P(T)$
$P(G)$ = Probability of guess
$P(G|S)$ = Students' Individual $P(G)$
$P(S)$ = Probability of slip
$P(S|S)$ Students' Individual $P(S)$

Node states
K , Q, Q_1, T, G, S = Two state (0 or 1)
Q = Two state (0 or 1)
S = Multi state (1 to N)
(Where N is the number of students in the training data)

Node representations
K = Knowledge node
Q = Question node
S = Student node
Q_1= first response node
T = Learning node
G = Guessing node
S = Slipping node

Parameters in **bold** are learned from data while the others are fixed



Student-Skill Interaction Model
(Pardos et al., , UMAP 2010)
P(T|S)
T
P(L0|Q1)
P(T)
P(T)
S
Q1
K
K
K
(Pardos et al., , EDM 2010)
(Corbett et al., UMUAI 1995)
P(G)
P(S)
Q
Q
Q
P(G|S)
P(S|S)
G
S

- The goal of the Bayesian Networks model above was to improve predictive accuracy by adapting a student's individual speed of learning to the classical model. This is the first model in the field to significantly improve predictive performance over standard knowledge tracing by using individualized parameters.

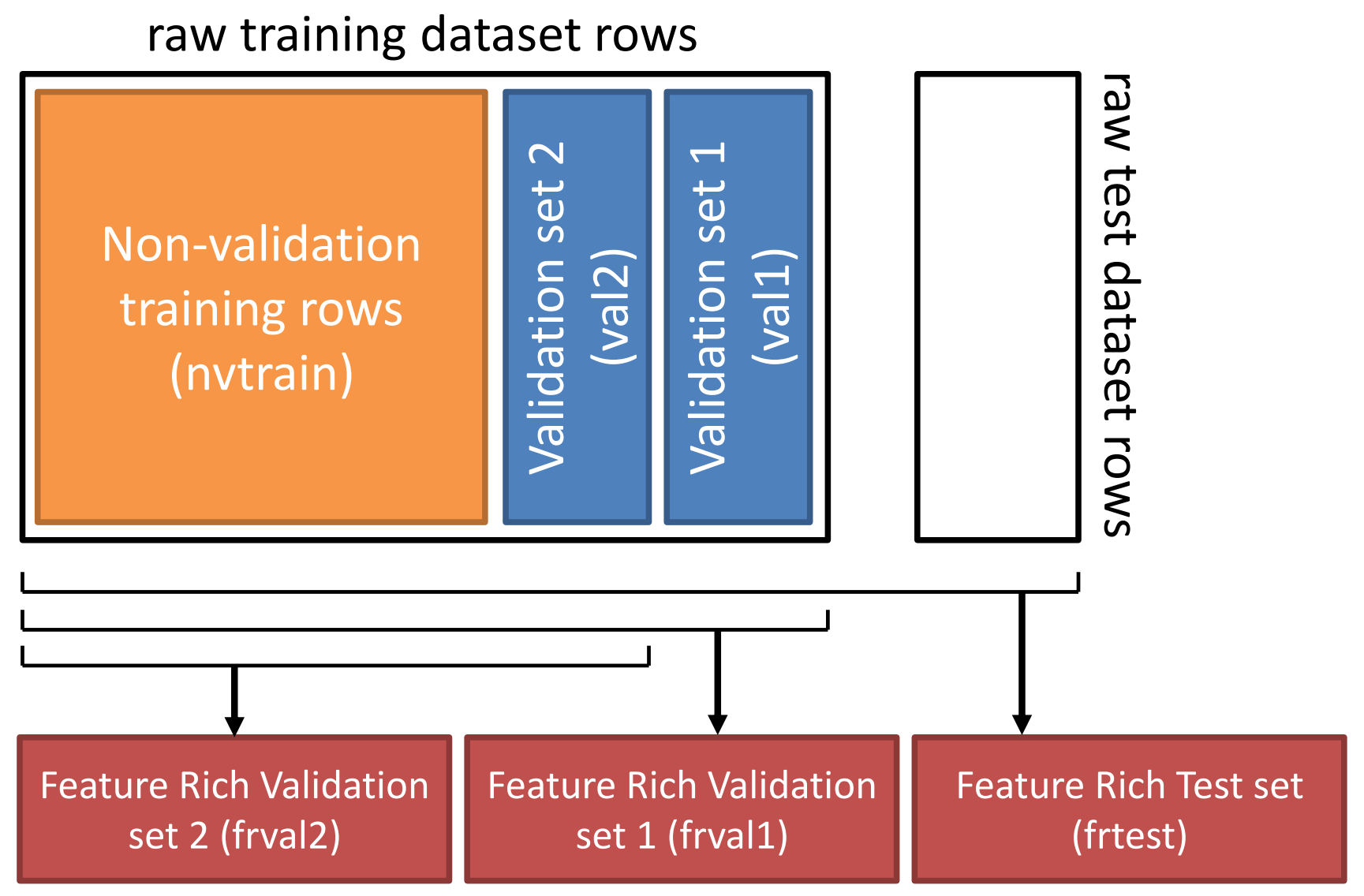
Machine Learning Appraoch

- While the student modeling approach was effective, it ignored much of the feature information included in the dataset. In order to utilize this information; a machine learning approach was also simultaneously pursued.
- After testing a variety of algorithms, Random Forests (Leo Breiman, ML 2001) was determined to be the most accurate at predicting this dataset
- The method trains T number of separate random decision trees. Each decision tree selects a random 1/P portion of the available features. The tree is grown until there are at least M observations in the leaf. When classifying unseen data, each tree votes on the binary class. The average of the votes is taken as the prediction.

Hardware: Two rocks clusters were used to train the Bayesian skill models (176 CPUs total). Two 32GB, 16 core machines were used to train the Random Forests classifiers. Software: MATLAB was used for all analysis. The Bayes Net Toolbox (Kevin Murphy), Statistics Toolbox and Parallel Computing Toolbox was used.

The first author is a National Science Foundation GK-12 Fellow

Feature extraction and engineering



raw training dataset rows
Non-validation training rows (nvtrain)
Validation set 2 (val2)
Validation set 1 (val1)
raw test dataset rows
Feature Rich Validation set 2 (frval2)
Feature Rich Validation set 1 (frval1)
Feature Rich Test set (frtest)

Feature extraction

- Two subsets of the training data were created to mimic the structure of the test set. Features were extracted from the previous data to generate features for these sets.

Feature engineering

Student progress features (avg. importance: 1.67)

- Number of data points [today, since the start of unit]
- Number of correct responses out of the last [3, 5, 10]
- Zscore sum for step duration, hint requests, incorrects
- Skill specific version of all these features

Percent correct features (avg. importance: 1.60)

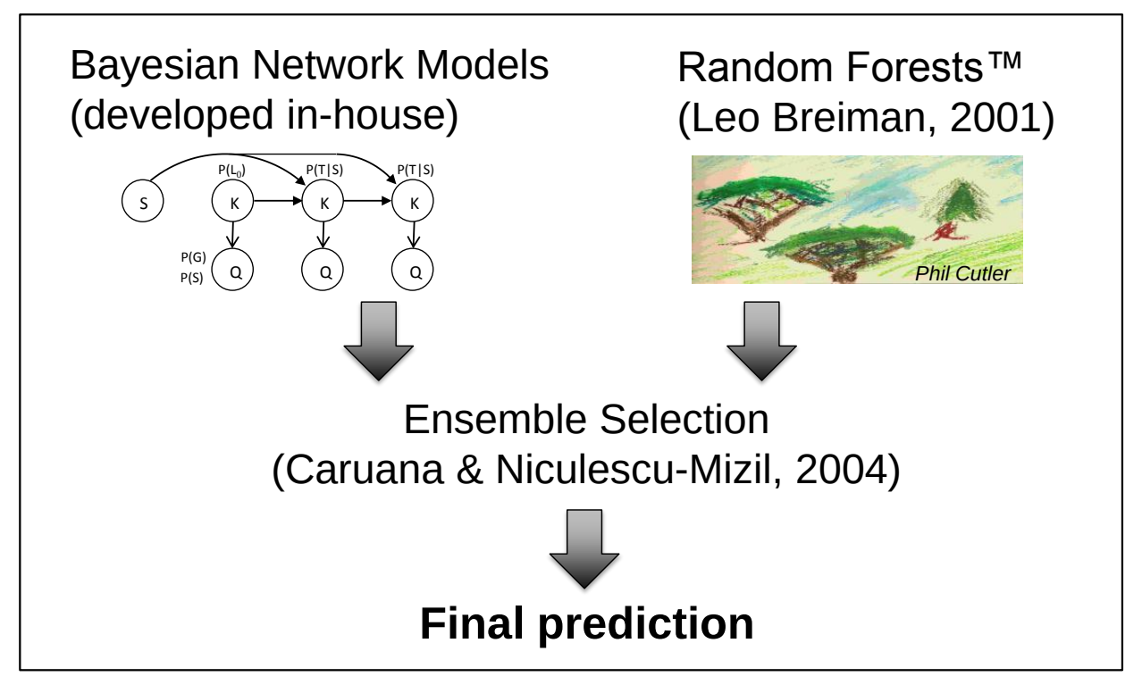
- % correct of unit, section, problem and step and total for each skill and also for each student (10 features)

Student Modeling Approach features (avg. importance: 1.32)

- The predicted probability of correct for the test row
- The number of data points used in training the parameters
- The final EM log likelihood fit of the parameters / data points

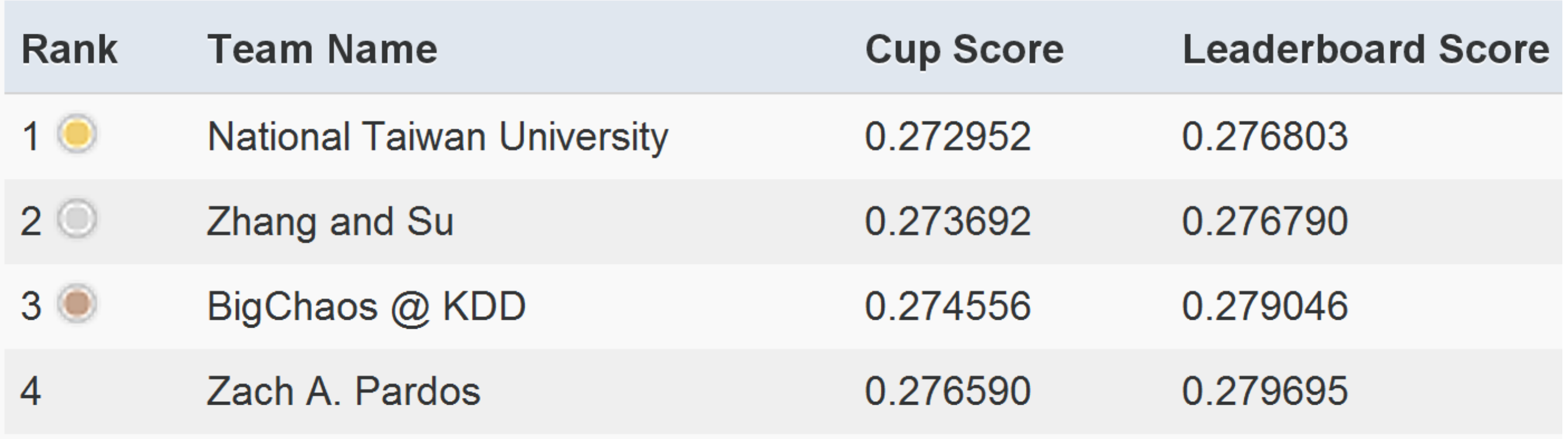
Competition Outcome

Bayesian Networks and Random Forests predictions were blended with Ensemble selection (Caruana et al., ICML 2004)



Bayesian Network Models (developed in-house)
P(L0)
P(T|S)
P(T|S)
S
K
K
K
P(G)
P(S)
Q
Q
Q
Random Forests™ (Leo Breiman, 2001)
Phil Cutler
Ensemble Selection (Caruana & Niculescu-Mizil, 2004)
Final prediction

This solution achieved 2nd place student prize and 4th place overall!



Rank	Team Name	Cup Score	Leaderboard Score
1 ●	National Taiwan University	0.272952	0.276803
2 ●	Zhang and Su	0.273692	0.276790
3 ●	BigChaos @ KDD	0.274556	0.279046
4	Zach A. Pardos	0.276590	0.279695

Conclusions

- Student Models developed from a 1995 Hidden Markov Model of learning and advanced here at WPI are formidable on the world stage.
- Random Forests that leverage feature of user and skill is a powerful tool for prediction student performance in Intelligent Tutoring Systems