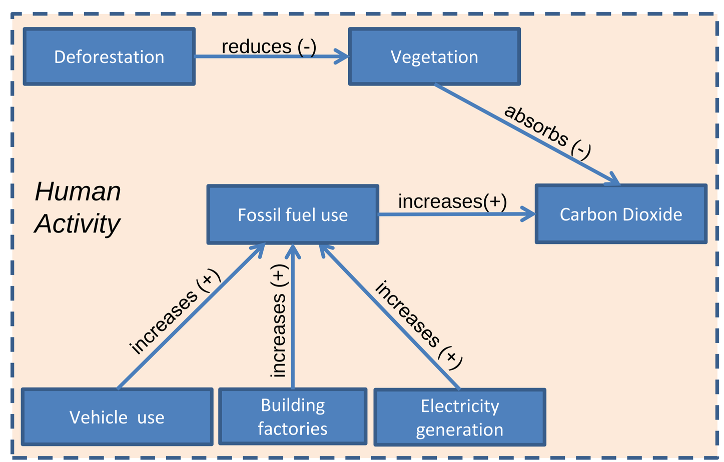
Comparison of Selection Criteria for Multi-Feature Hierarchical Activity Mining in Open Ended Learning Environments

Yi Dong, John S. Kinnebrew & Gautam Biswas

Supported by NSF CISE IIS EAGER GRANT # 1548499

Background and Motivation

- **Open-Ended Learning Environments (OELEs):** In OELEs, students utilize resources and tools provided to solve a challenge science problem
  - For example in Betty's Brain Learning Environment, students build a causal map for scientific phenomena to teach a virtual agent. They can read to learn, build, and check their models by interacting with the virtual agent, Betty



Partial Causal Map of climate change phenomena

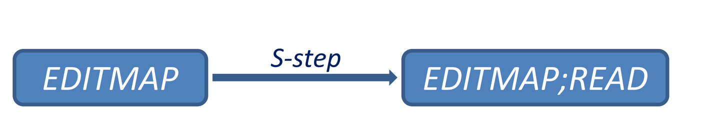
- **Motivations:**
  - OELEs have been used by researchers to track student's learning performance and many details of their learning interactions
  - The wealth of data collected provides new opportunities for developing algorithms to accurately model, understand, assess students' learning behaviors and strategies

Related Works

- **Sequential Pattern Mining (SPAM):** SPAM with bitmap representation for sequences performs a Depth-First Search traversal to find patterns that exceed a pre-defined frequency threshold (*Ayres*, 2002)
- **Differential Sequence Mining (DSM):** DSM finds frequent patterns that are good to differentiate between different groups of students (*Kinnebrew*, 2013)

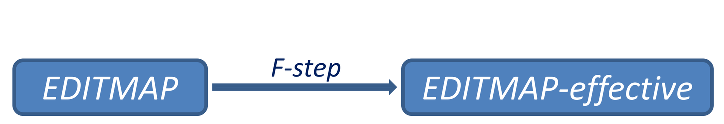
MFH-SPAM algorithm

- **Multi-Feature Hierarchical Sequential Pattern Mining:** MFH-SPAM extends SPAM algorithm by adding two new pattern extension steps and forms a multi-dimensional pattern extension schema:
  - *Sequence-extension* step (*S-step*), a flat extension originally in SPAM



  - *Hierarchical-extension* step (*H-step*) extends the last action hierarchically

  - *Feature-extension* step (*F-step*) extends the last action by associating with features



Pattern Selection:

- Even for moderate size of datasets, MFH-SPAM finds a huge number of frequent patterns
- We adopt the idea from DSM to find frequent patterns that are good differentiators between different students groups. For example:
  - High performers vs. Low Performers

Pattern Selection Metrics

Classifier Wrapper (CW): It uses the *F1-score* of a 5-fold cross validation for a Decision Tree Classifier as the metric to select the best set of frequent patterns

Information Gain (IG): The computational complexity of CW, which builds decision tree multiple times and results in redundant *Information Gain* computation is very high.
We directly apply *Information Gain* as a new selection metric

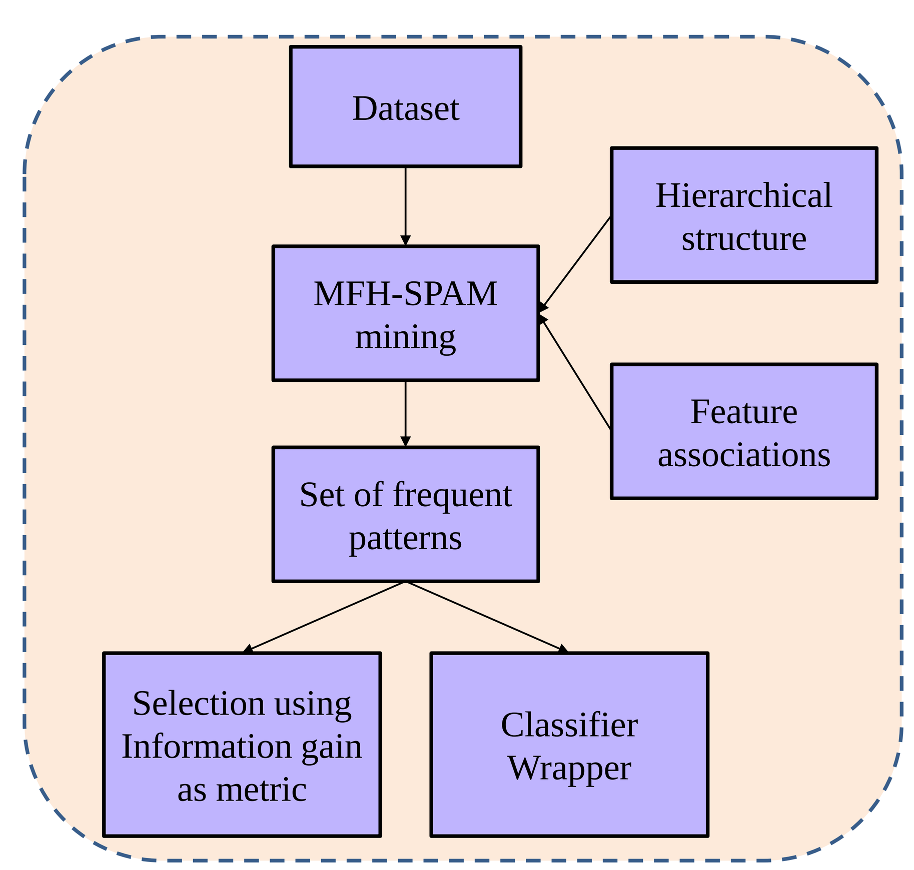
- The *Information Gain* for a given pattern P_1 is computed from the reduction of *Shannon Entropy* when P_1 becomes known, where *Shannon Entropy* measures homogeneity for a sample data and it is given by:

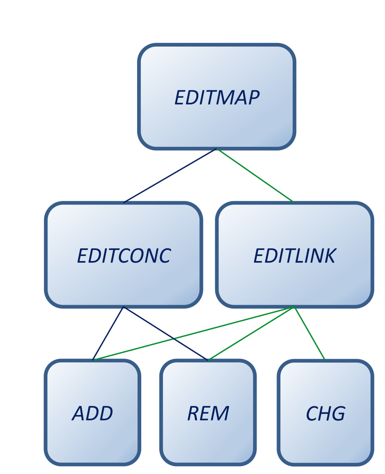
$$H(P_1) = -\sum_i^n p_i \log_2 p_i$$

where n is the number of student groups, p_i is the probability of P_1 been used by students in group i

- We use *instance frequency* (*i-frequency*) of patterns to compute the information gain. The *i-frequency* is a normalized measure of how often individual students uses a particular pattern
- For each patter, the information gain will be computed only once which improved the computational complexity over the classifier wrapper method

Algorithm Workflow





Partial Hierarchical Structure for Actions in Betty's Brain Learning Environment.

Two associated feature types:
1. *Effective vs. Ineffective*
2. *Supported vs. Unsupported*

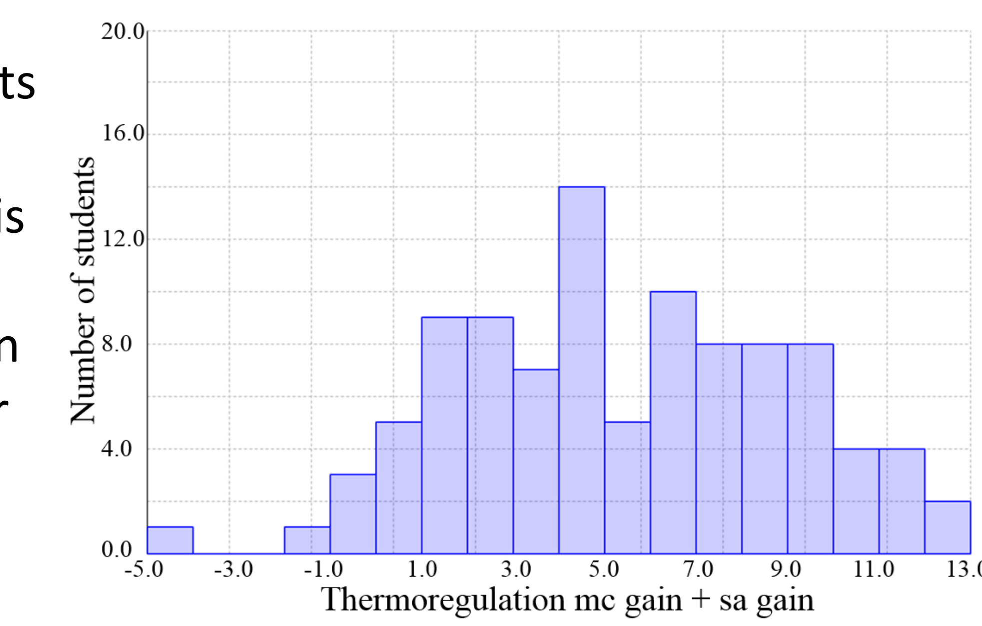
Case Study

Dataset:

- The dataset comes from an experiment we ran with 98 middle school students who used Betty's Brain system to learn *thermoregulation* phenomena
- We used the sum of the students' *thermoregulation* multiple choice and *thermoregulation* short answer score gains from pre- to post-test, S_{gain}, to divide them into:

  high performing group vs. low performing group

- The histogram of students distribution over S_{gain} (the figure on the right) is approximately bimodal. We used the median gain (5.0) as the boundary for the two groups



Results and Analysis

Classifier Wrapper High vs. Low Performers Mean usage(Stand deviation)	Information Gain as Metric High vs. Low Performers Mean usage(Stand deviation)
EDITLINK;QUIZTAKEN 25.9(21.9) vs. 10.6(13.3)	EDITLINK;QUIZTAKEN 25.9(21.9) vs. 10.6(13.3)
EDITMAP-eff-sup 24.1(17.6) vs. 12.3(11.5)	QUIZ;EDITLINK;READ 6.1(7.6) vs. 2.3(2.5)
READ-long 19.8(30.2) vs. 34.0(29.2)	NOTE 9.5(11.3) vs. 23.9(24.4)

Top 3 patterns as differentiators for both selection methods

Analysis:

- Both selection methods are able to find frequent patterns that are good differentiators as shown in the table
- The run time favors the Information Gain method compared to the Classifier Wrapper method (reduced from 28 seconds to 16 seconds), therefore, better computational efficiency for this metric

Future Work

- We will perform more systematic analysis of the differences between groups using hypothesis testing methods
- In addition, we will use correlational analysis to study in more depth the relations between behaviors and performance
- We will also work towards using the patterns derived to detect learner behaviors online, and develop scaffolding mechanisms