Data-based Student Modeling in Exploratory Learning Environments

Cristina Conati

Dept. of Computer Science,
University of British Columbia, Canada

Thanks to

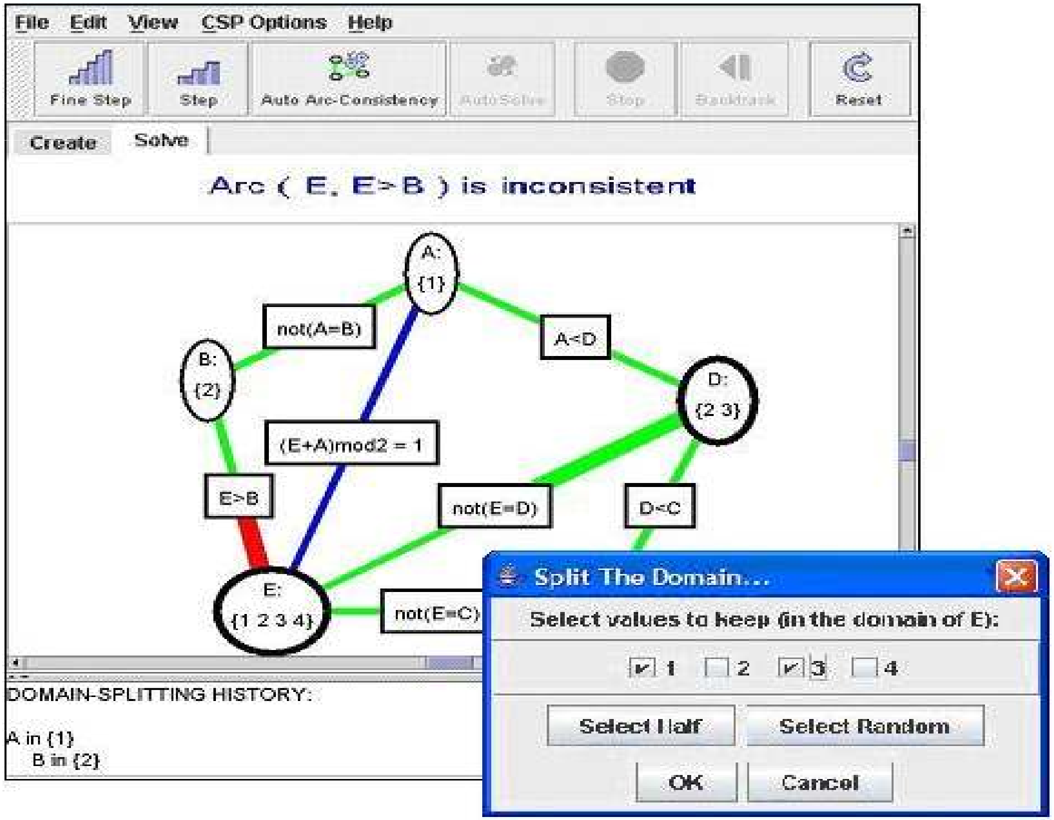
- Saleema Amershi
- Andrea Bernardini
- Samad Kardan
- David Ternes

Exploratory Learning Environments (ELE)

- Educational tools designed to support free exploration of target domain
  - Student initiated
  - No tutor
- Pros:
  - highly constructive and engaging
- Cons:
  - not all students explore effectively (e.g., Njoo and de Jong, 1993; Shute 1993), because of individual differences on the necessary skills

Example

- AISpace applet for Constraint Satisfaction Problems (CSP) – (Amershi et al 2006)



Overall Research Goal

- Create that provide *adaptive* support for students who don't explore effectively…
- …while maintaining as much as possible the sense of student freedom and control

⇓

- Adaptations are guided by the *student model*.
  - abstract representation of the student in terms of relevant traits (behaviors, knowledge, etc).

Challenge

- Necessary to model student exploratory behaviour and related skills
- Hard because
  - No clear definition of behavior correctness
  - No clear definition of effective behavior.
- Behavior effectiveness may depend on
  - Student pre-exiting knowledge
  - Student reasoning during interaction
  - Interface design

Previous Approaches

- Knowledge-based (e.g., Shute 1994, Conati et al 2002)
- Data-based
  - Supervised (e.g., Merten and Conati 2007)

Knowledge-based Approaches

- Require the knowledge and collaborative efforts of *experts* in:
  1. the instructional domain
  2. the interaction type and how it affects desired outcome
  3. the type of user model
- Time consuming, expensive and possibly unfeasible when there is limited expertise on 2

Data-based Approaches: Supervised

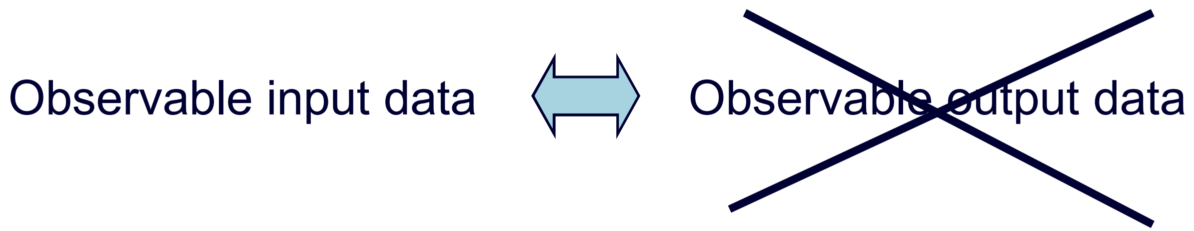
- Make use of supervised machine learning techniques.

Observable input data ⟺ ~~Observable output data~~

- If observable output data are not readily available, developers must supply them manually (e.g. Conati and Merten 2007)
  - Time consuming
  - Error prone

Proposed Approach

- Use *both* unsupervised and supervised machine learning to help build the student model
  - Light weight alternative to knowledge-based approaches
  - Alternative to supervised databased approaches when output labels are not readily available from the system



Contributions to EDM

- Focus on ELE as opposed to more structured problem solving or drill-and-practice activities (e.g., Zaiane 2002, Beck 2005, Backer et al 2008)

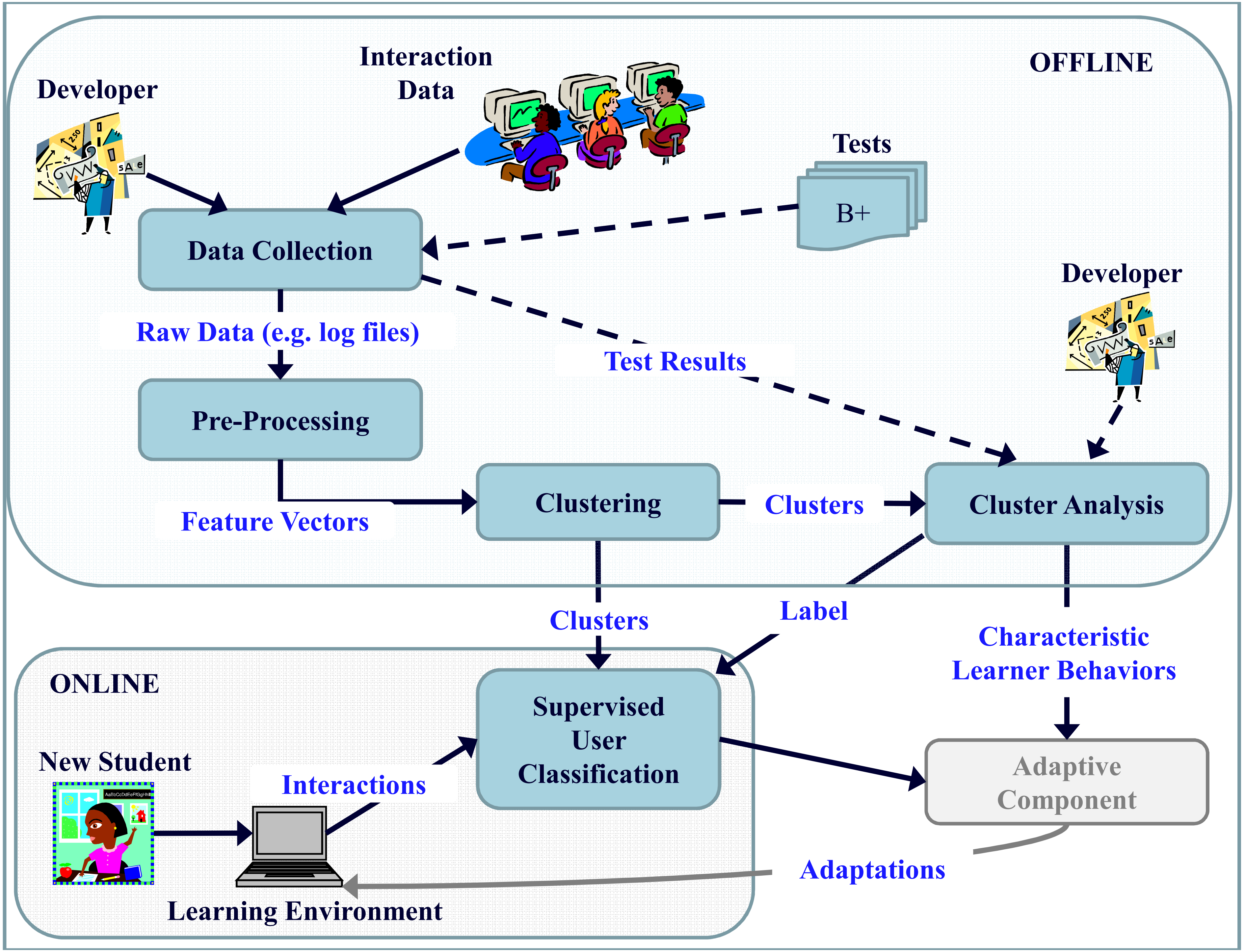
- Show effectiveness on
  - Two different types of ELE
    - CSP applet
    - ACE – Adaptive Coach for Exploration of mathematical functions
  - Different types of data (interface actions only, vs. inclusion of eye-tracking data)

Outline

- Approach Overview
- Results with the AISpace applet
- Results with ACE
- Extension with Association Rules
- Conclusions

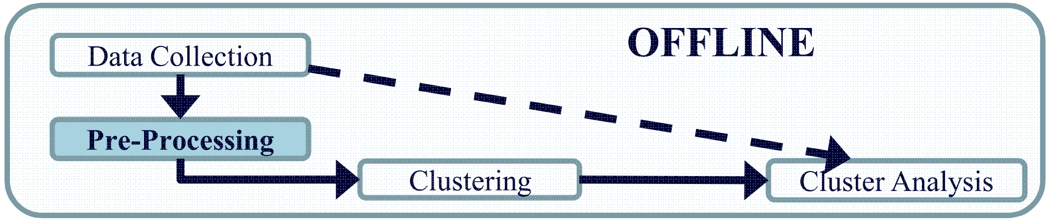
Outline

- **Approach Overview**
- Results with the AISpace applet
- Results with ACE
- Extension with Association Rules
- Conclusions



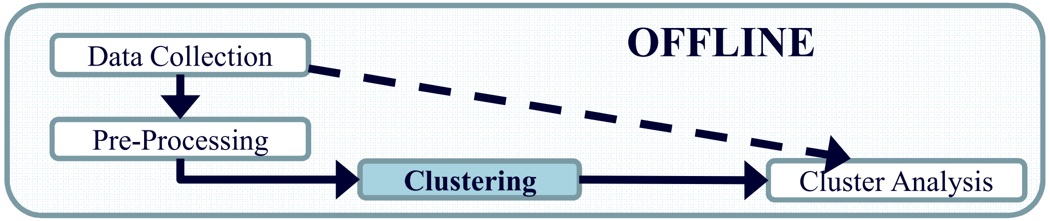
OFFLINE
Developer
Interaction Data
Tests
B+
Data Collection
Raw Data (e.g. log files)
Test Results
Developer
Pre-Processing
Clustering
Clusters
Cluster Analysis
Feature Vectors
Clusters
Label
Characteristic Learner Behaviors
ONLINE
Supervised User Classification
New Student
Interactions
Adaptive Component
Adaptations
Learning Environment

Preprocessing



- Each student must be represented by a multidimensional feature vector:
  - Frequency of actions
  - Mean and standard deviation of the latency between actions
  - Others (e.g., eye-tracking data)
- Feature selection to prune irrelevant features.

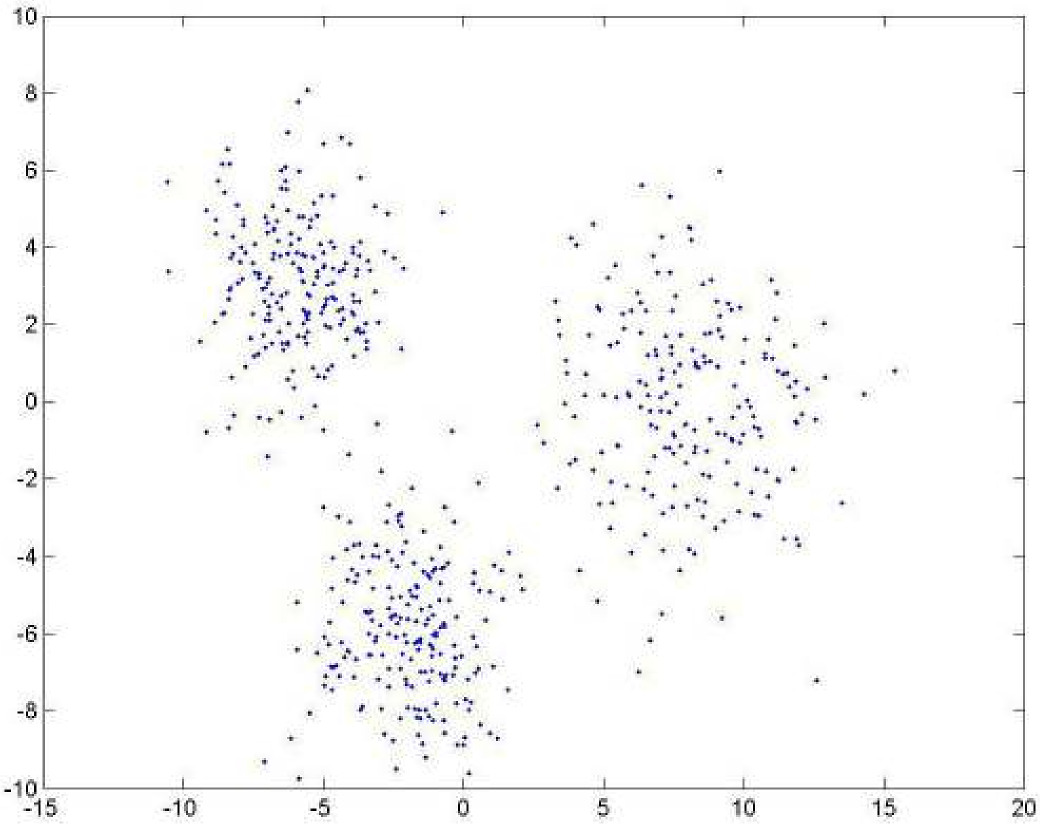
Clustering



- Automatically groups feature vectors by their similarity
- K-means clustering

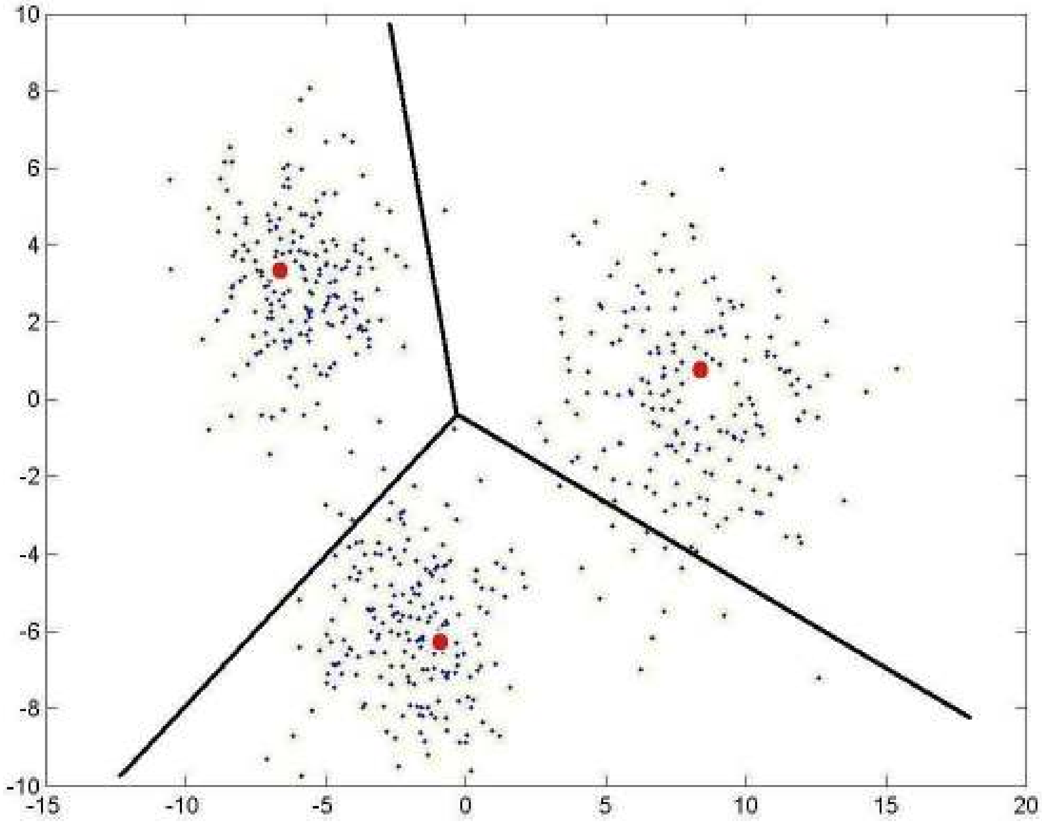
K-means Clustering

- Tell k-means how many clusters to find ($k = 3$)



K-means Clustering

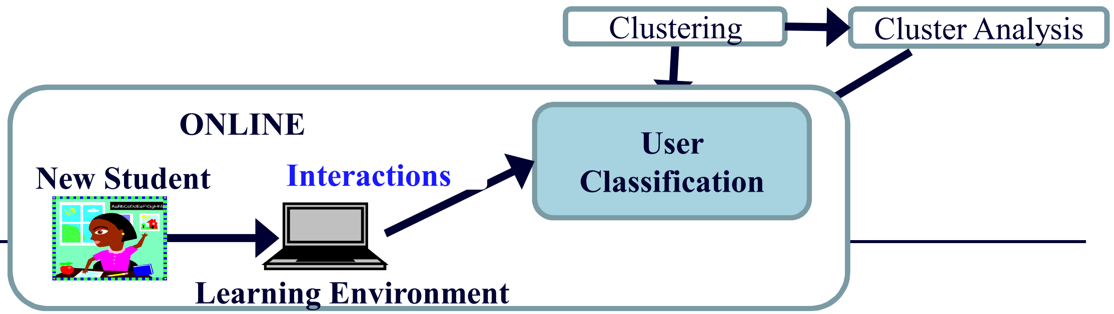
- Tell k-means how many clusters to find ($k = 3$)



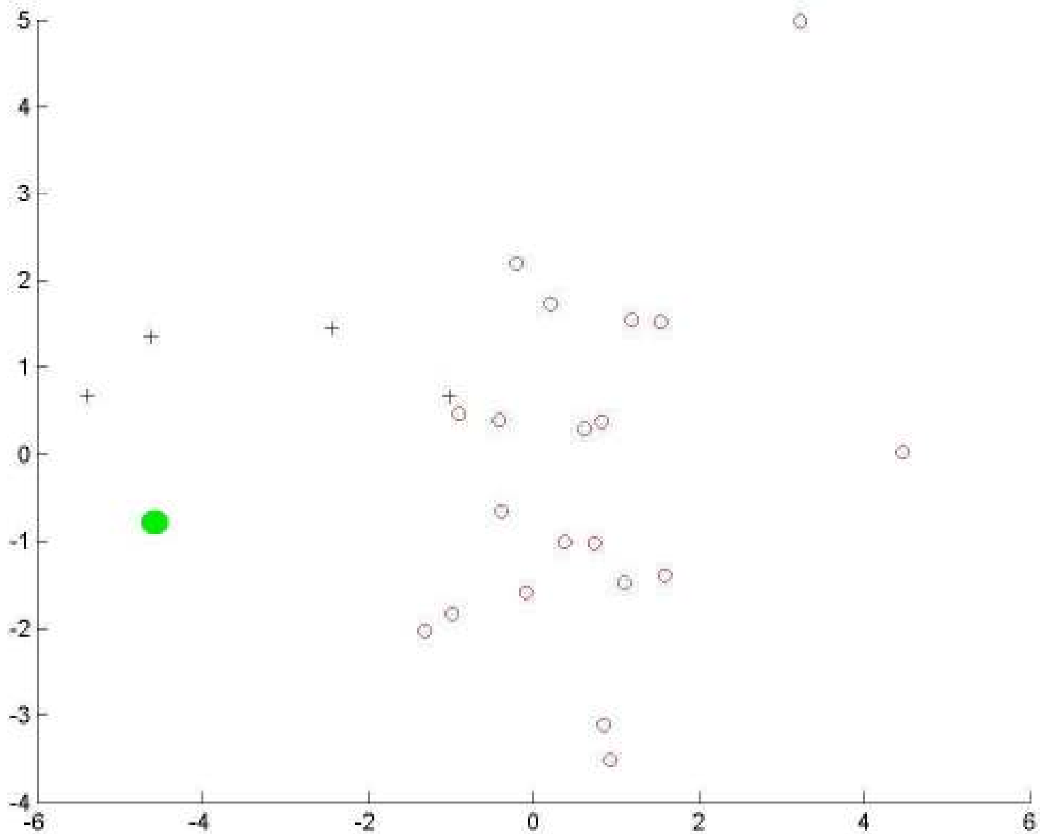
Cluster Analysis

- Characterize the clusters.
  - Isolate behaviors that distinguish clusters from each other
- Determine which clusters represent effective or ineffective learners.
  - Based on objective data (e.g., pre and posttests) if available
  - Based on experts subjective judgment
- Here we use formal statistical tests to compare clusters in terms of learning and feature similarity
  - Welch t-test, ANOVA (for statistical significance)
  - Cohens' *d* and partial η^2 (for practical significance)

User Model



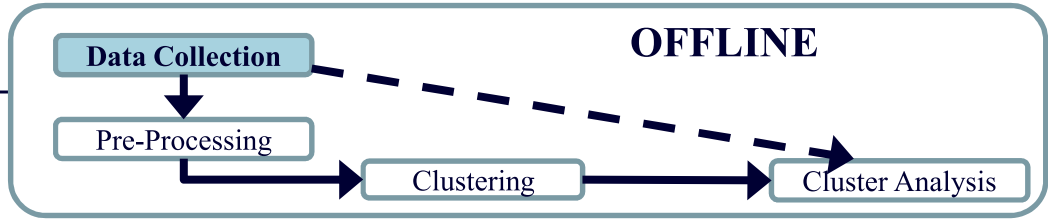
- Create data point for new student
- Update the data point as the student interacts
- Reassign to nearest cluster



Outline

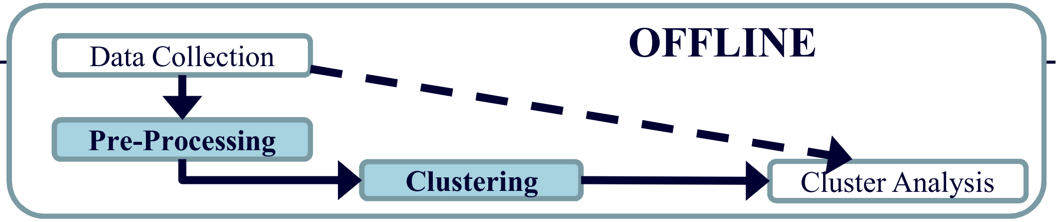
- Approach Overview
- **Results with the AISpace applet**
- Results with ACE
- Extension with Association Rules
- Conclusions

Data

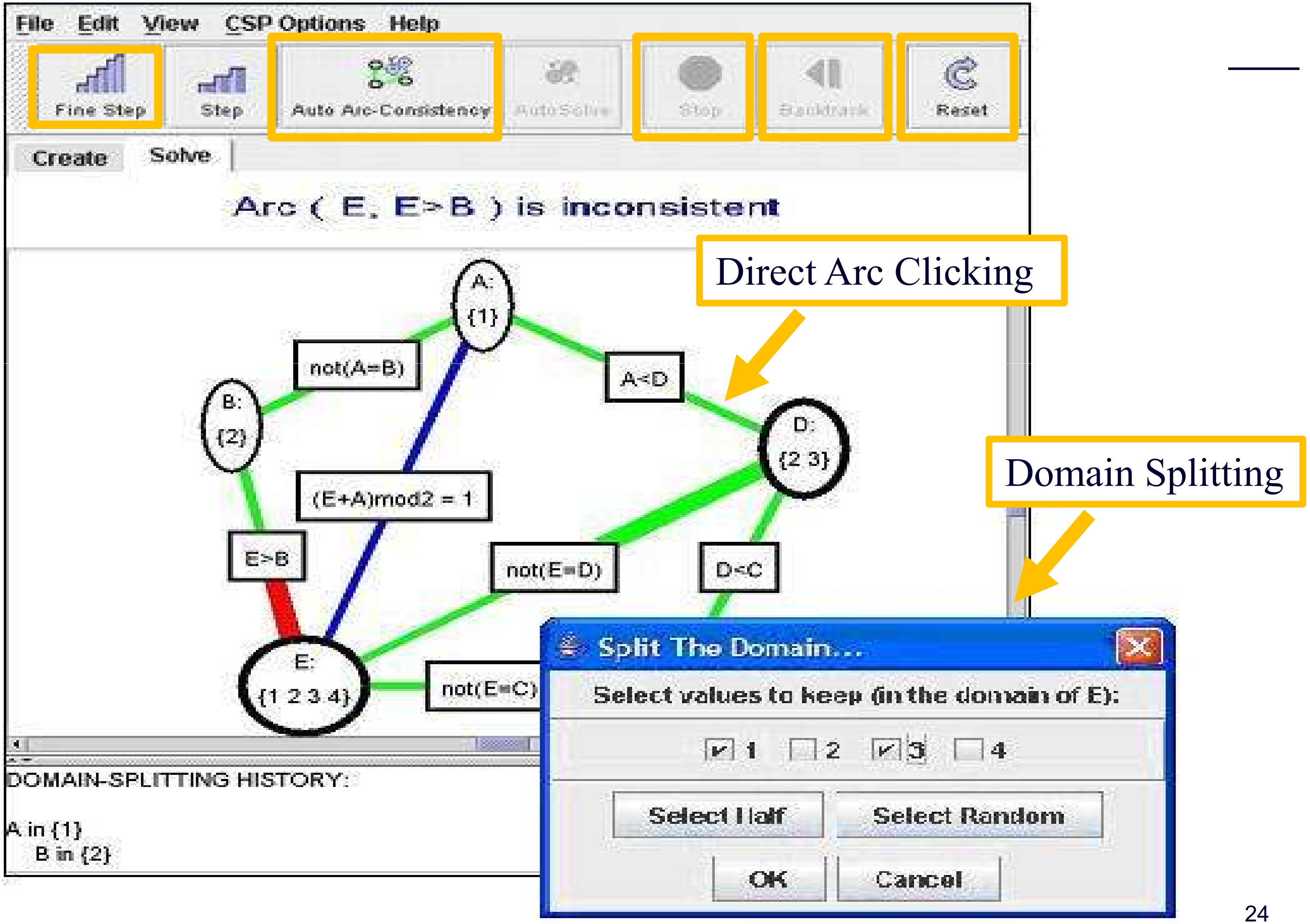


- Derived from a previous study (Amershi et al 2007)
  - 24 participants
  - Pre test
  - Logged data from the AIspace CSP applet
    - 1931 actions over 205.3 minutes
  - Post test

Preprocessing and Clustering



- One data point for each student
- 21 dimensional feature space
  - Seven interface actions
  - Average frequencies of use
  - Mean and standard deviation of pause durations
- *K-means* clustering with *k*=2 and *k*=3

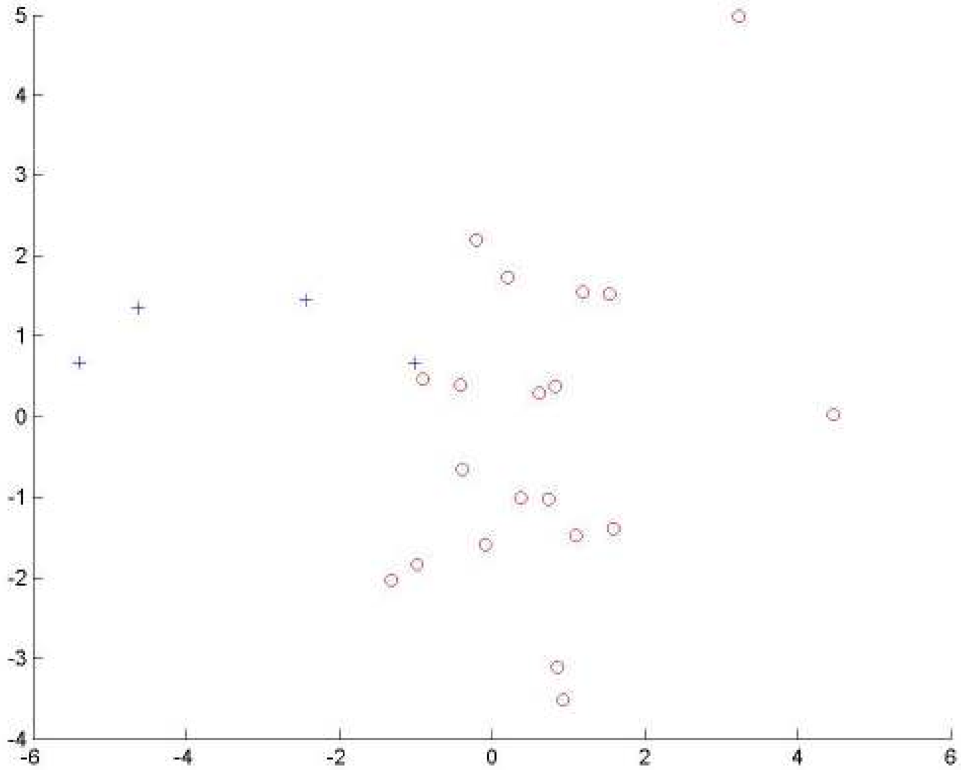

File Edit View CSP Options Help
Fine Step
Step
Auto Arc-Consistency
Auto Solve
Stop
Backtrack
Reset
Create
Solve
Arc (E, E>B) is inconsistent
Direct Arc Clicking
Domain Splitting
A: {1}
B: {2}
D: {2 3}
E: {1 2 3 4}
not(A=B)
A<D
(E+A)mod2 = 1
E>B
not(E=D)
D<C
not(E=C)
Split The Domain...
Select values to keep (in the domain of E):
1 2 3 4
Select Half
Select Random
OK
Cancel
DOMAIN-SPLITTING HISTORY:
A in {1}
B in {2}

CSP actions considered

- ***Fine Stepping***. Cycles through three detailed algorithm steps:
- ***Direct Arc Clicking***. User decides which arc to test
- ***Auto Arc Consistency*** *(Auto AC)*. Automatically *Fine Steps* through the network.
- ***Stop***. Stops *Auto AC*.
- ***Domain Splitting*** *(DS)*. User selects a variable domain to split, and specify a sub-network for further application of AC-3.
- ***Backtracking***. Recovers the alternative sub-network set aside by *DS*.
- ***Resetting***. Resets the CSP network to its initial state.

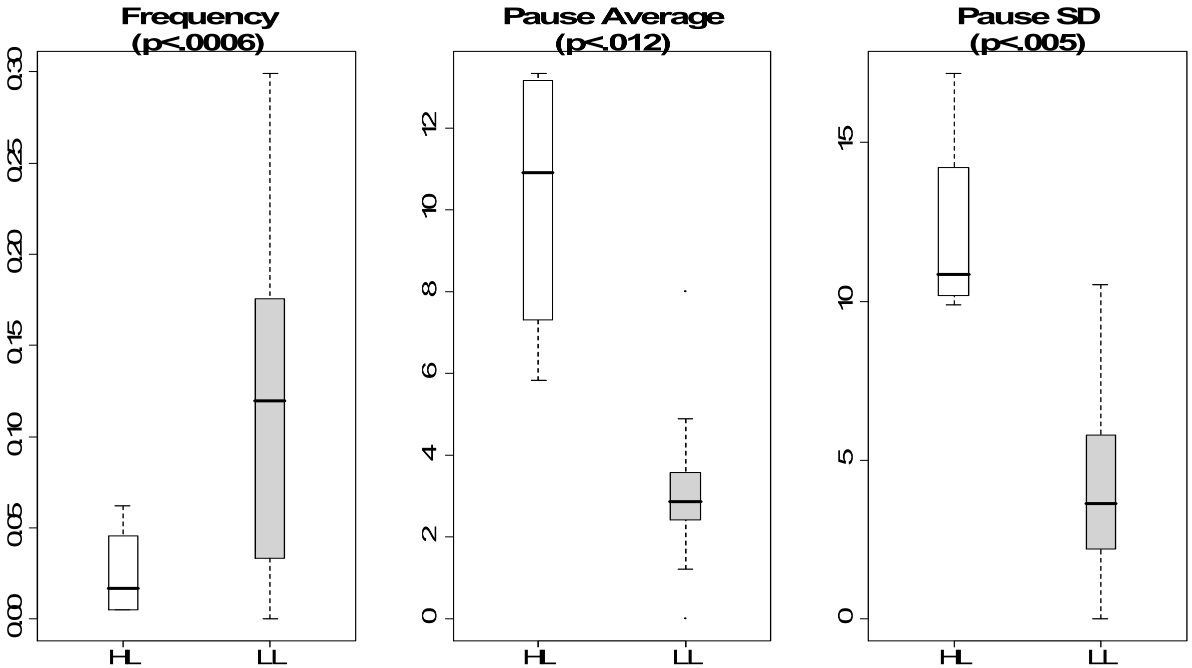
Cluster Analysis

- ***K* = 2**
- Significant differences in learning gains
  - p<.006
  - Large effect size ($d > 0.8$)
- Low Learning (LL) and High Learning (HL) Group



Cluster Analysis k=2

- Characteristics of LL Group
  - More frequent use of Fine Step feature
  - Consistently shorter pauses after using the Fine Step feature

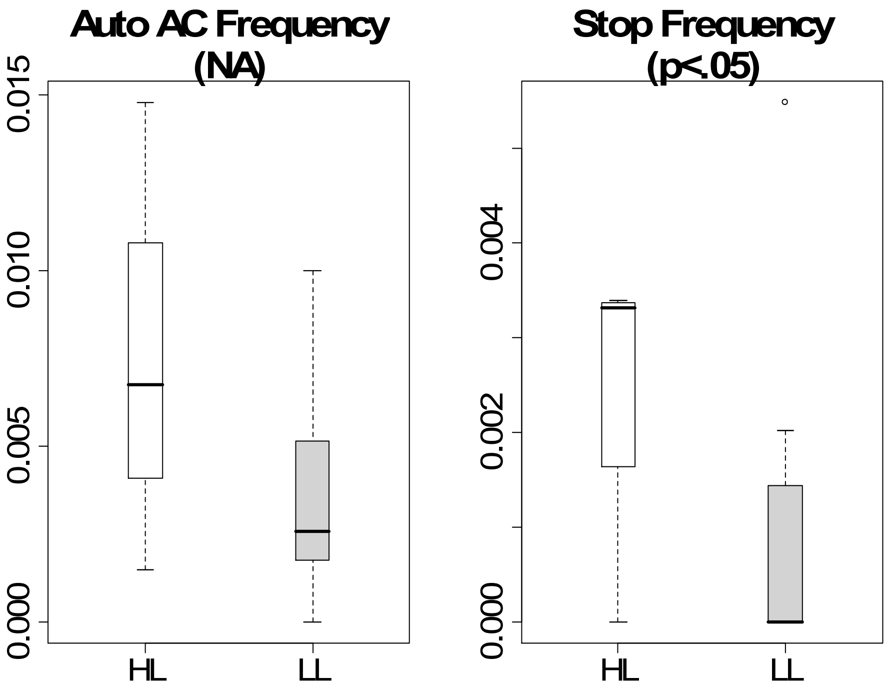


Cluster Analysis *k*=2

- Characteristics of LL Group
  - Consistently shorter pauses after using the Reset feature ($p<.001$)

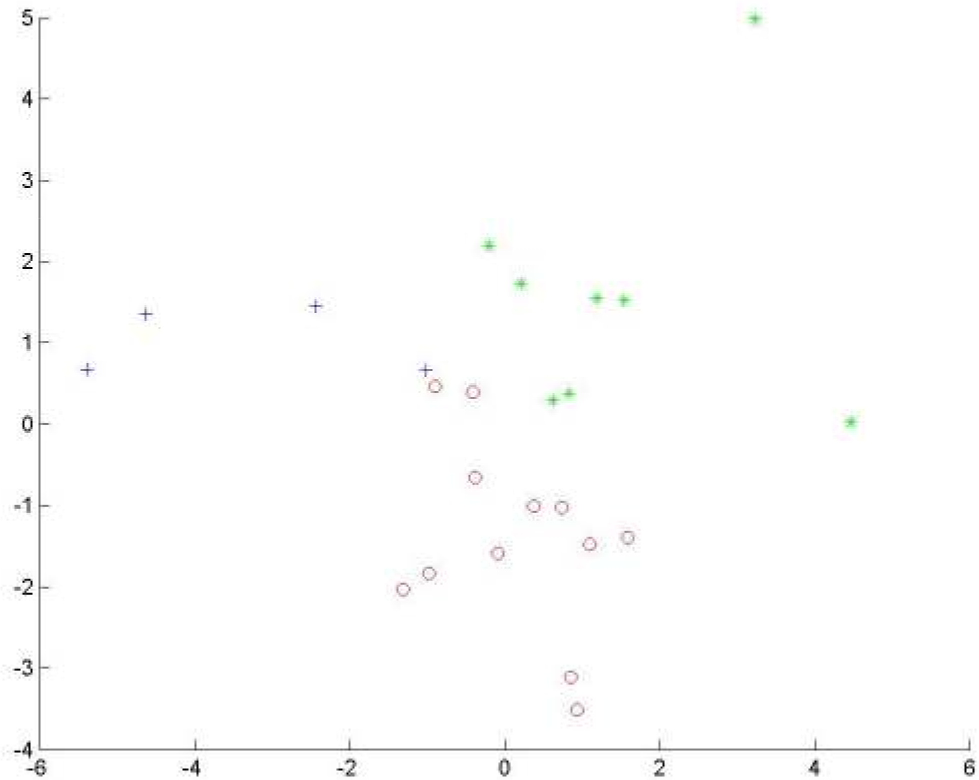
Cluster Analysis k=2

- Characteristics of LL Group
  - Less frequent use of the Auto AC feature (trend)
  - Less frequent use of the Stop feature (p<.01)



Cluster Analysis

- ***K* = 3**
- Significant differences in learning gains
  - F=3.58, p<.046
- One group had higher learning gains (HL) then the other groups (LL1, LL2)
  - p<.014
  - Large effect size (d > 0.8)

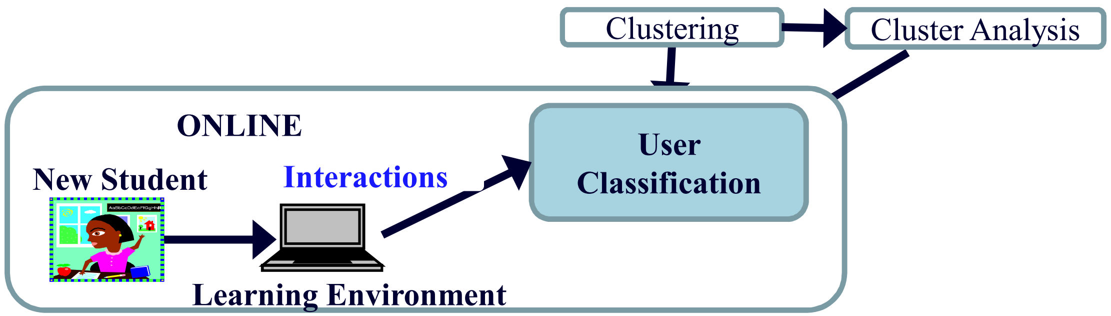


Cluster Analysis *k*=3

- Same distinctions between low and high learners as with k=2, plus additional behaviors

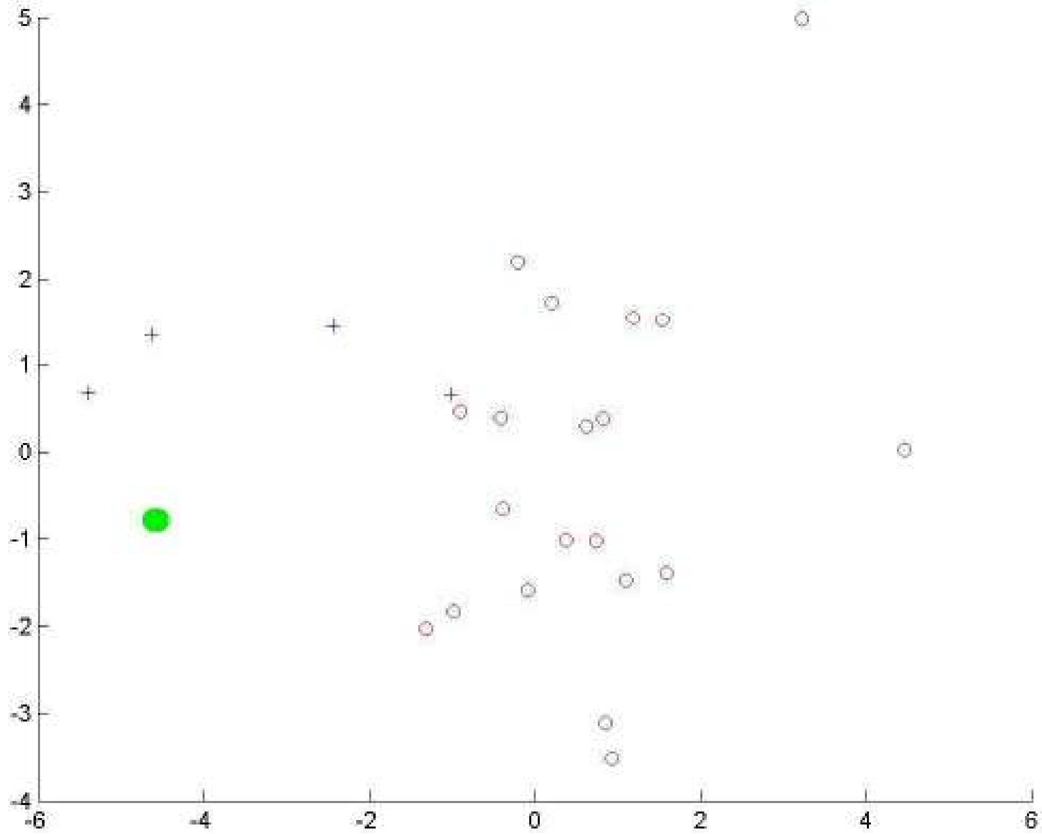
Key Point

- K-means clustering was able to automatically discover behavioral patterns detrimental to learning
- Pattern data could be used
  - As input in a knowledge-based approach
  - To train a classifier in a data-based approach



Online K-means Classification

- Find clusters offline
- Create data point for new student
- Update the data point as the student interacts
- Reassign to nearest cluster

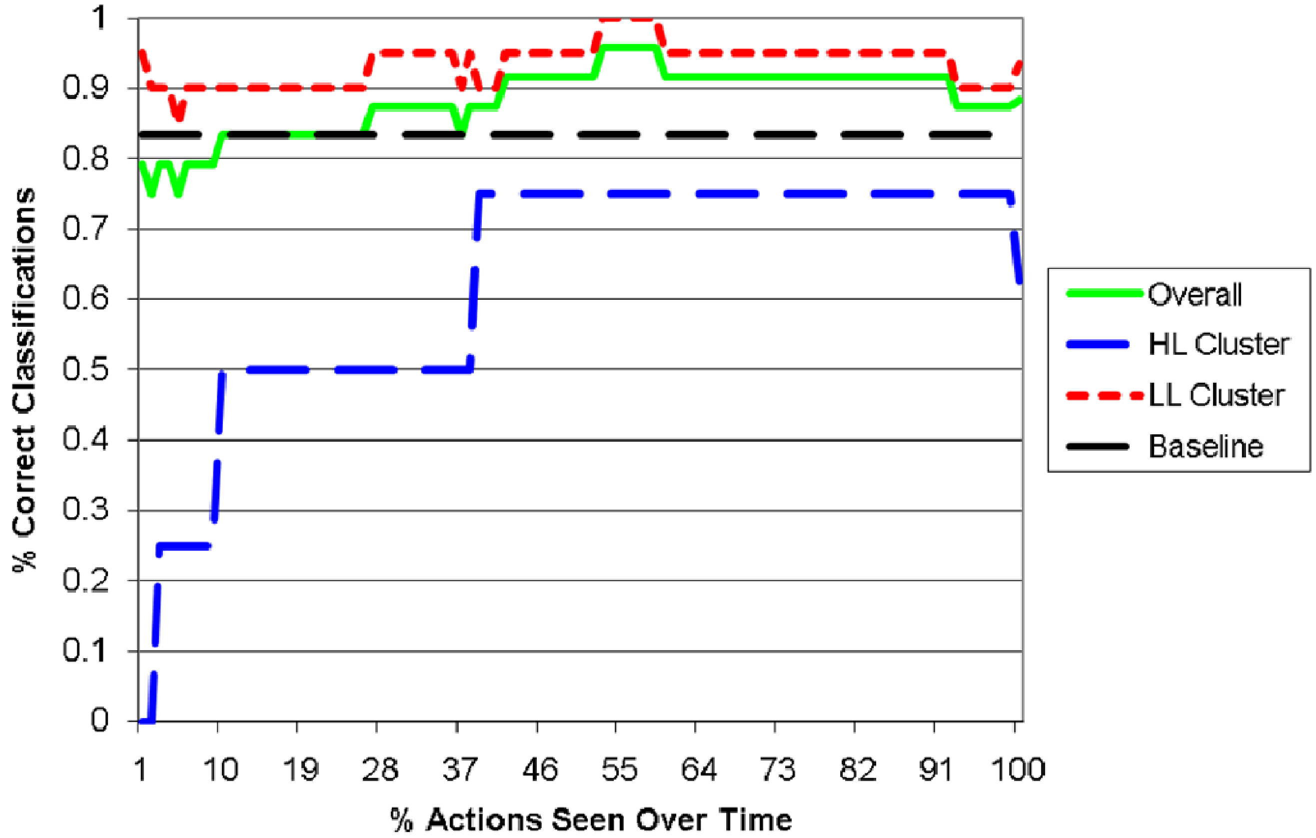


K-means Classifier Evaluation

- 24 fold, leave one out cross validation
  - Removed one student’s data at a time
  - Performed offline k-means clustering
  - Trained the classifier with the clusters found
  - Tested the classifier with the removed student’s logged data

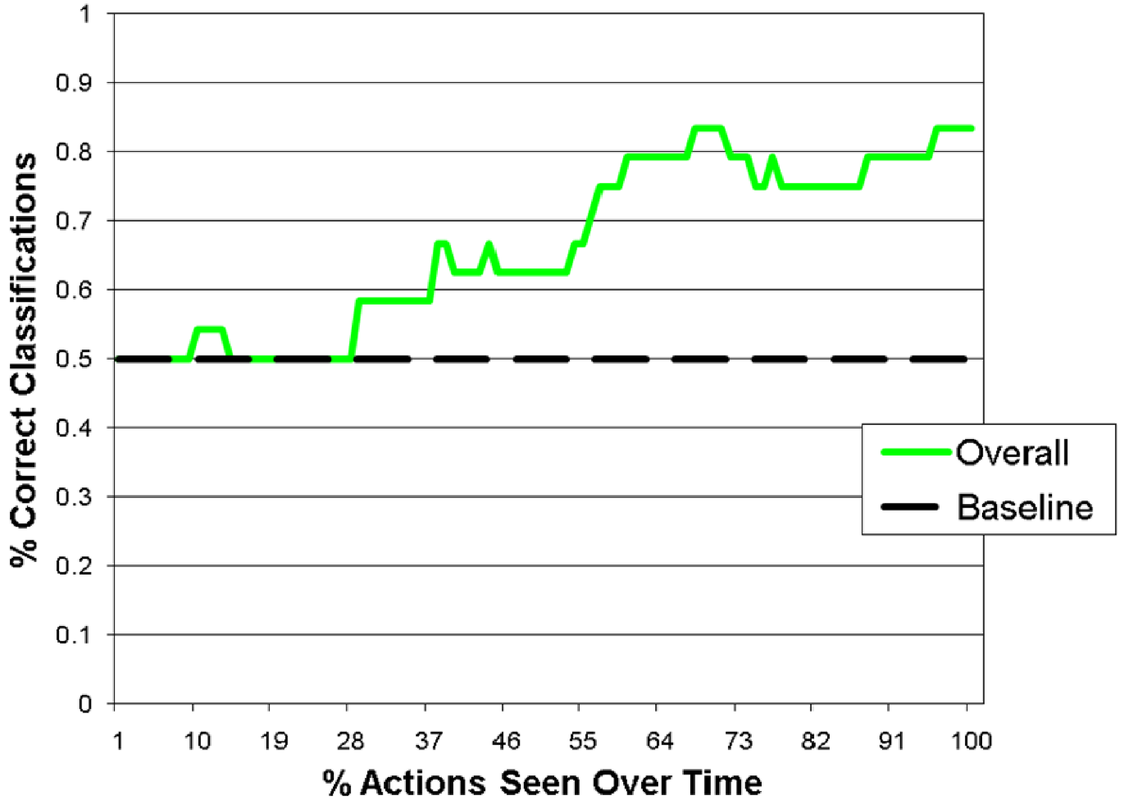
Classification Results *K*=2

- Overall accuracy for *k*=2 approximately 88%



Classification Results *K=3*

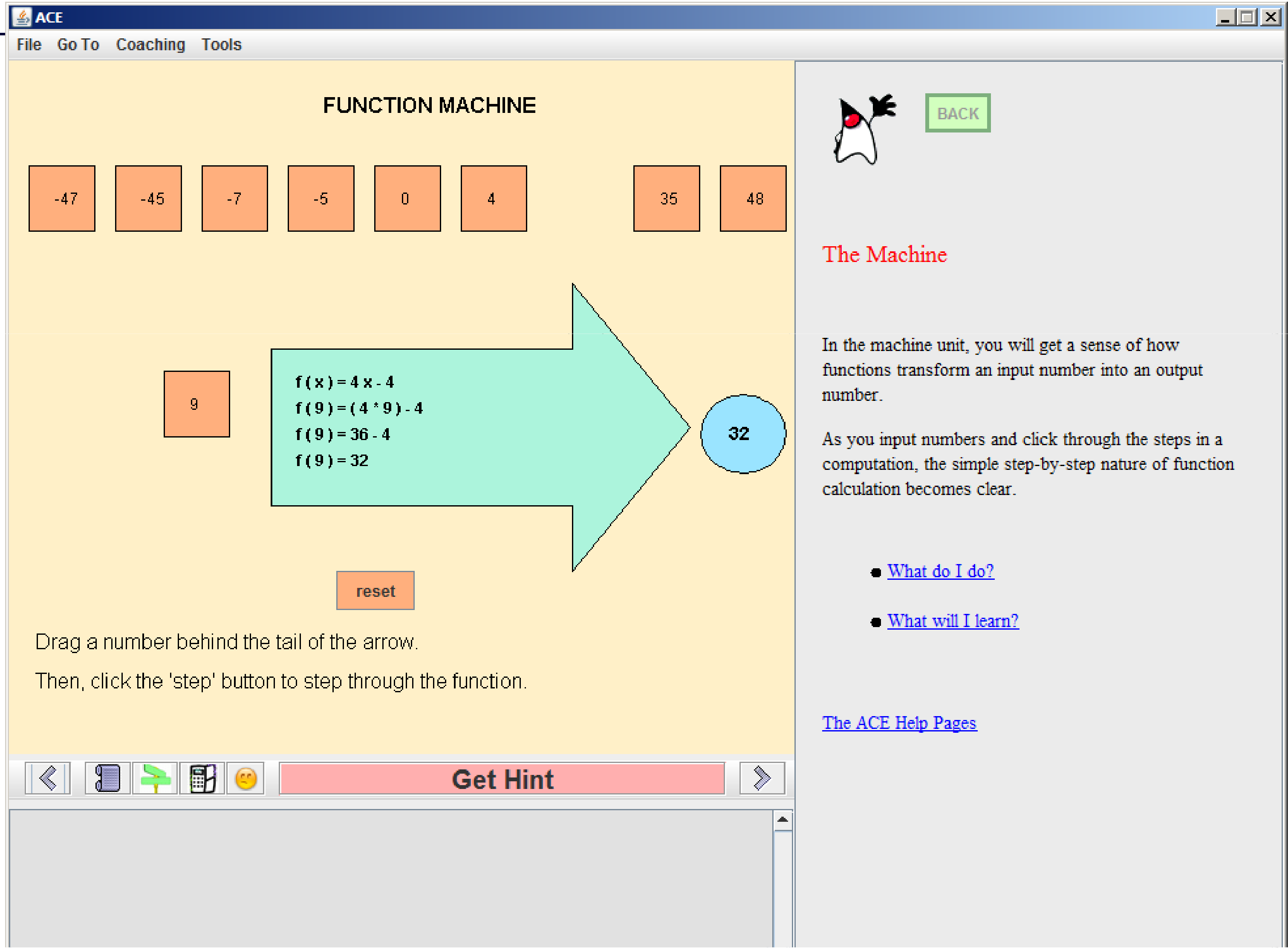
- Overall accuracy for *k*=3 approximately 74%
  - ~80% for LL2, ~66% for HL, ~45% for HL



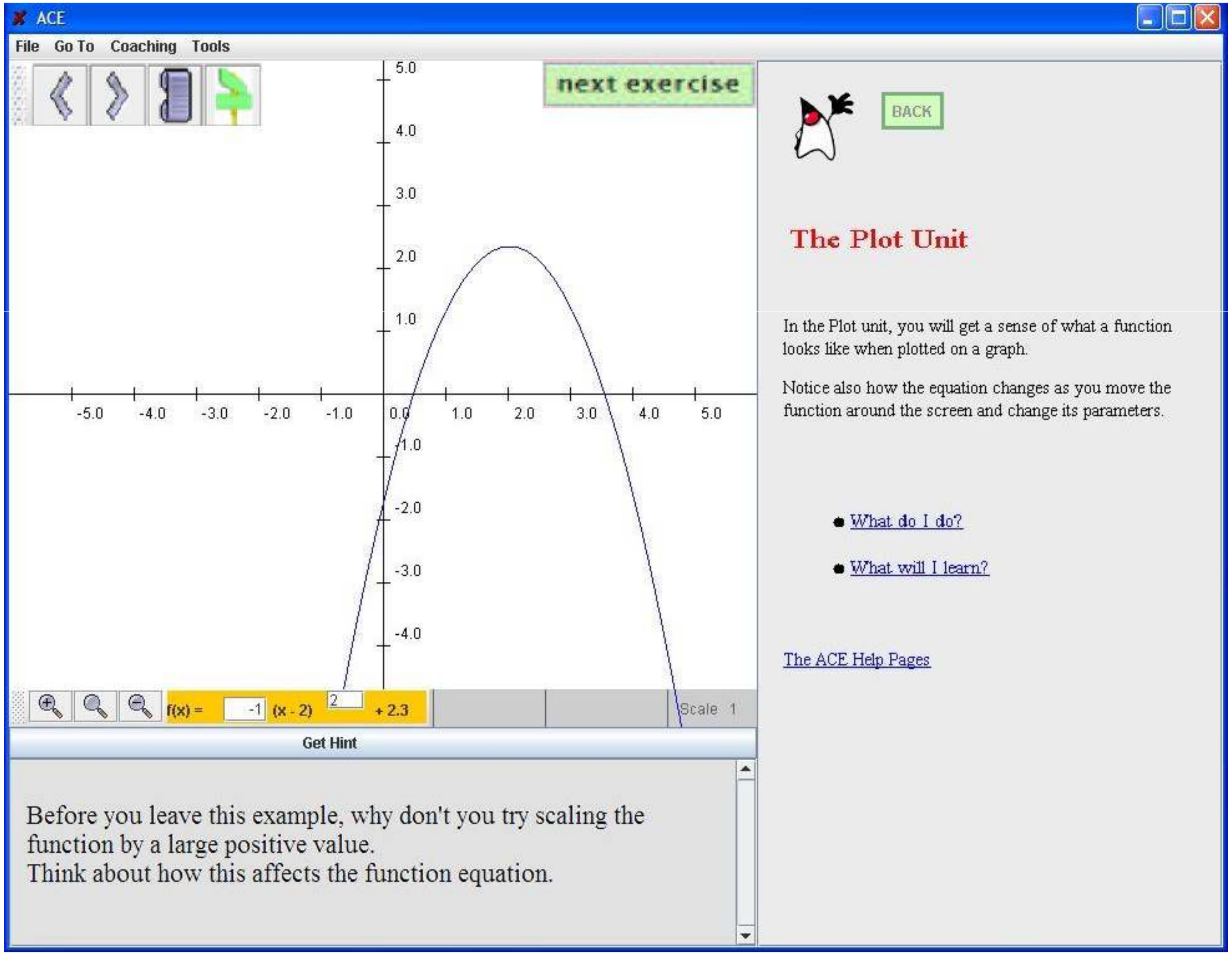
Outline

- Approach Overview
- Results with the AISpace applet
- **Results with ACE**
- Extension with Association Rules
- Conclusions

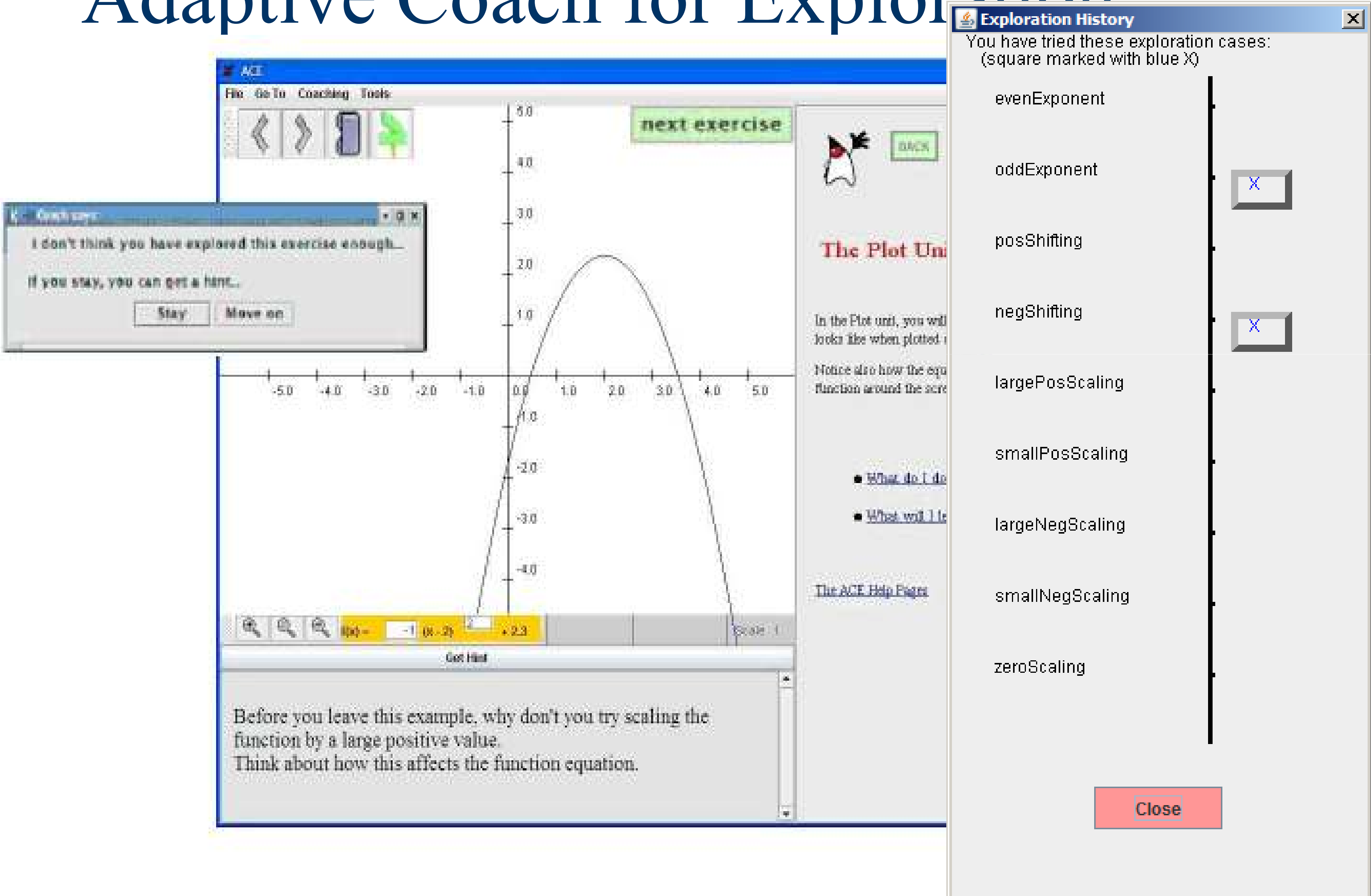
Adaptive Coach for Exploration



Adaptive Coach for Exploration



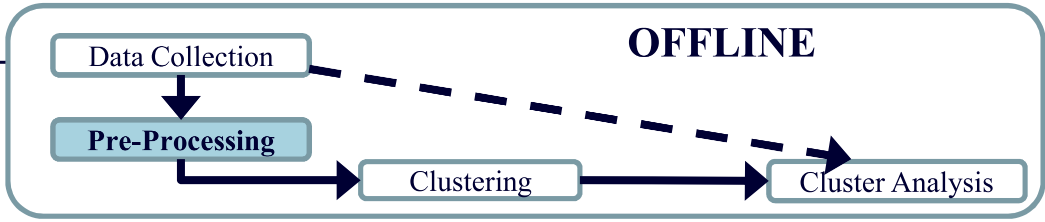
Adaptive Coach for Exploration



Data Collection for ACE

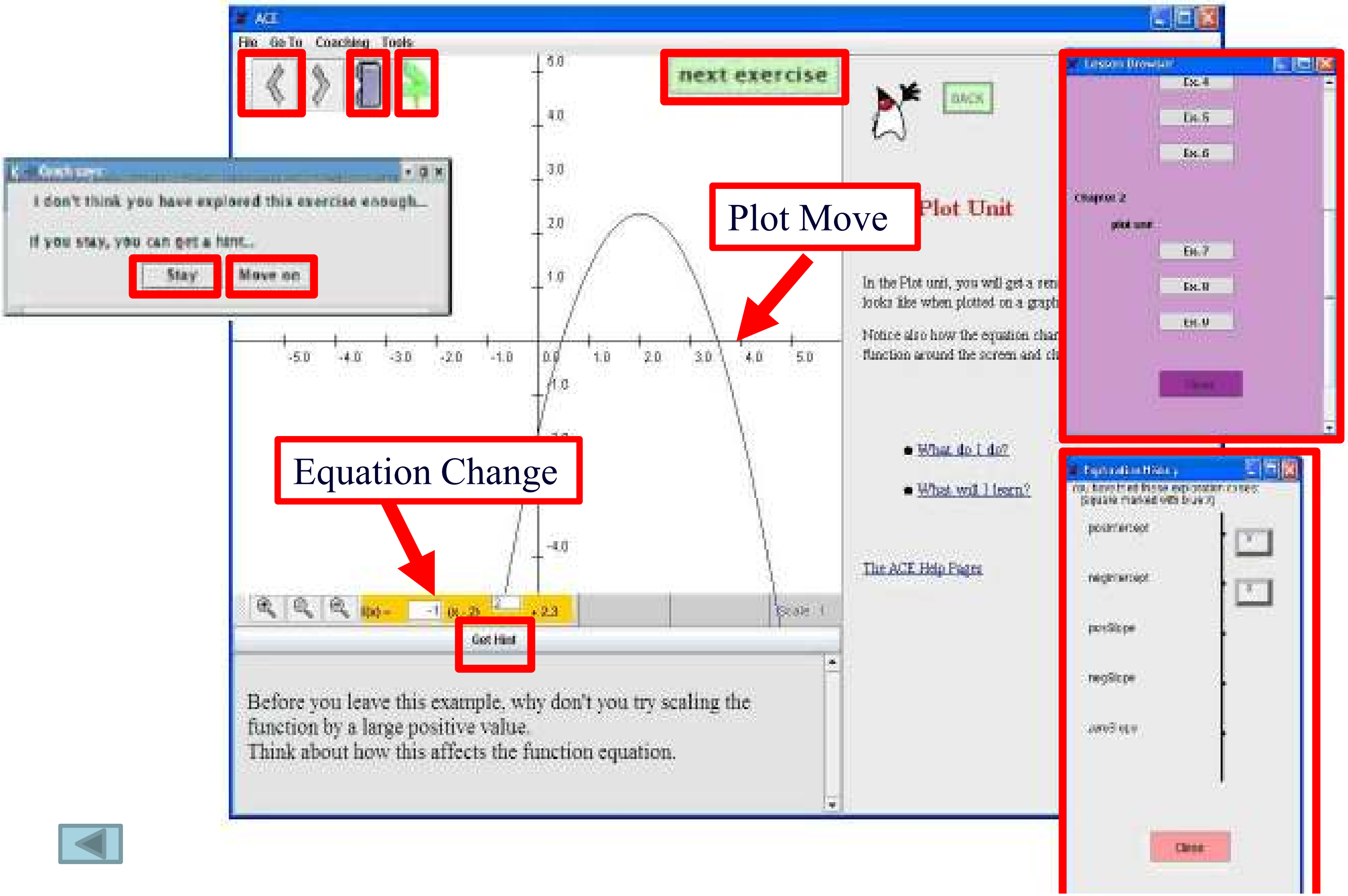
- 36 participants
- Pre test
- Interacted with ACE – here we focus on Plot Unit
  - Interactions were logged
  - Eye-gaze was tracked by a head mounted eye tracker
- Post test
- A total of 3783 interface actions were recorded, along with the accompanying gaze data.

Preprocessing for ACE

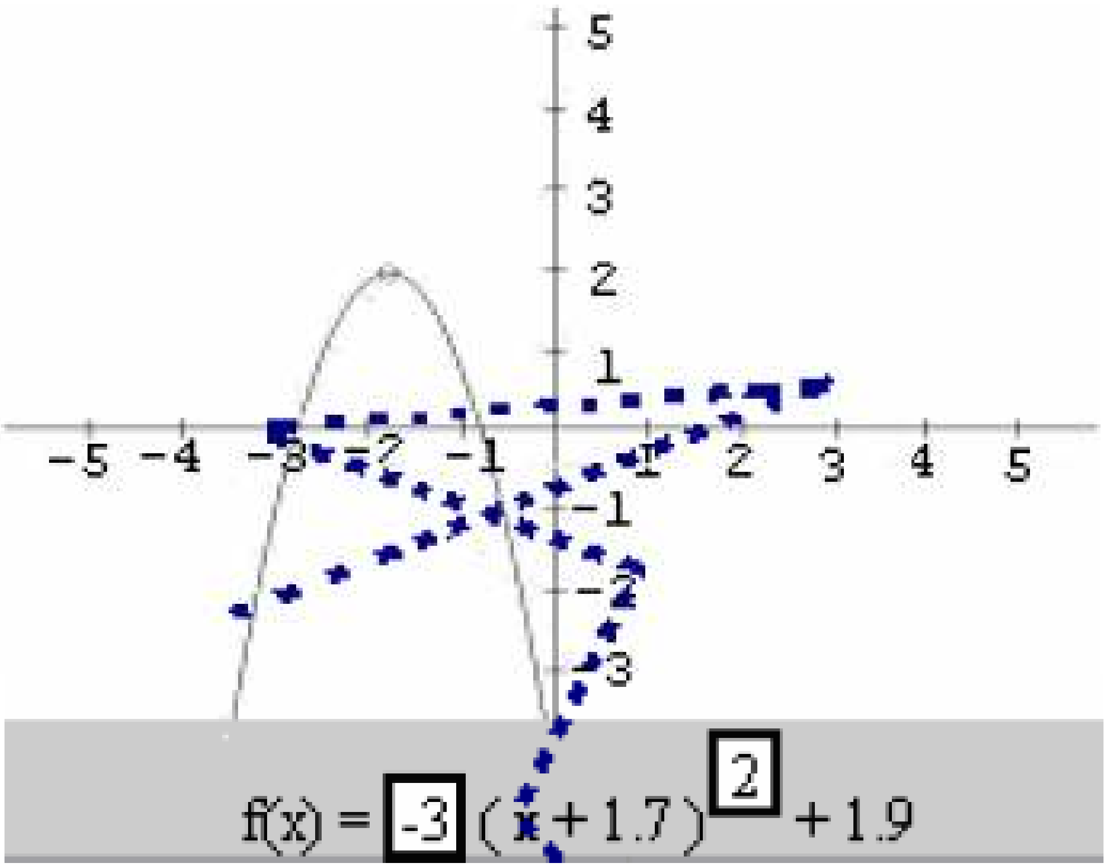


- *FeatureSet1*:
  - Only interface features (action frequencies and latencies).
  - 13 actions, 39 features in total.
- *FeatureSet2*:
  - Interface features
  - eye-gaze features (direct and indirect *gaze shifts*).
    - Mean and St. Dev. of gaze shifts per action type
  - 91 features in total.

Adaptive Coach for Exploration



Direct and Indirect Gaze Shifts



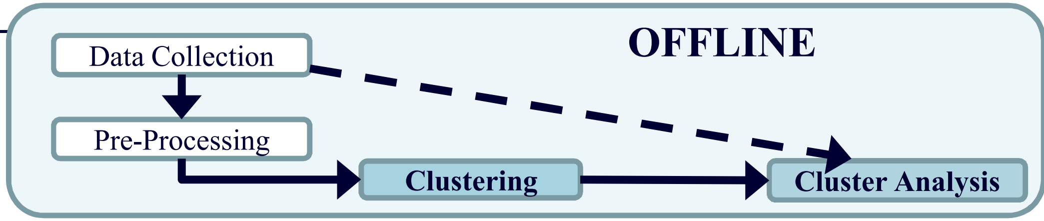
Goal

- See if our approach transfers to data beyond interface actions (gaze data here)
- Reproduce results showing that gaze data improves assessment of student exploration (Conati and Merten 2007)

Preprocessing for ACE (cont.)

- Feature selection (entropy-based, Dash and Liu 2000).
- *FeatureSet1* (without gaze features) reduced to 33 features on average
  - Only features removed relate to St.Dev. of latency
- *FeatureSet2* (with gaze features) reduced to 36 features
  - All action frequencies selected
  - Gaze dimensions selected only in conjunction with latency
    - After *equation change*, *next exercise*, *step back*, *stay* and *zoom*
    - Primarily important to discriminate between idle time and time spent thinking about action outcome (consistent with findings in Conati and Merten 2007)

Clustering and Cluster Analisys



- Applied k-means clustering to both feature sets.
- Meaningful results only for *k=2*

Cluster Analysis for ACE

- No significant differences in learning between clusters found in *FeatureSet1* (without gaze features).
- Marginally significant difference in learning between clusters found in *FeatureSet2* (with gaze features).
  - High Learners (***HL***) 11 students, Low Learners (***LL***), 25 students
  - (p=.069), d> 0.5

Cluster Analysis for ACE

- Findings that confirm results by (Conati and Merten 2007)
  - No significant differences in the frequency of *plot moves* and *equation changes* between HL and LL clusters
    - Sheer number of actions in not a good predictor of learning
  - HL students paused longer and made more eye-gaze movements after *equation changes* than LL students

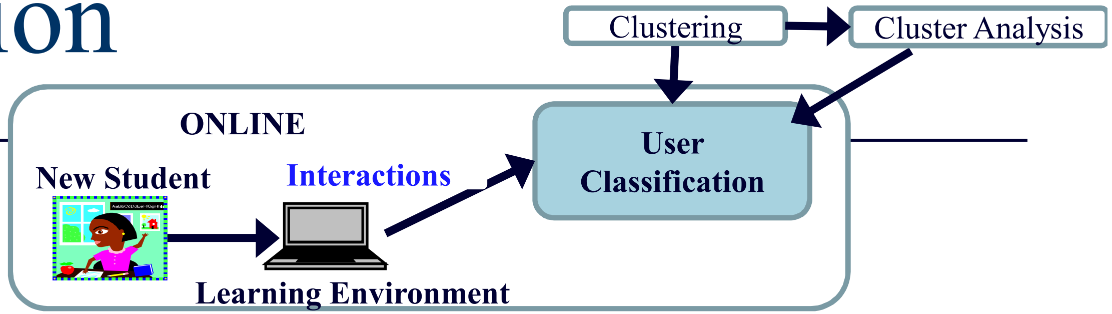
Cluster Analysis for ACE

- Other findings
  - HL students paused longer and made more eye-gaze movements after *next exercise* than LL students
  - LL students chose to ignore the Coach's suggestion to continue exploring the current exercise more frequently then HL students
  - HL students paused longer than LL students when the Coach suggested they continue exploring the current exercise

Cluster Analysis for ACE

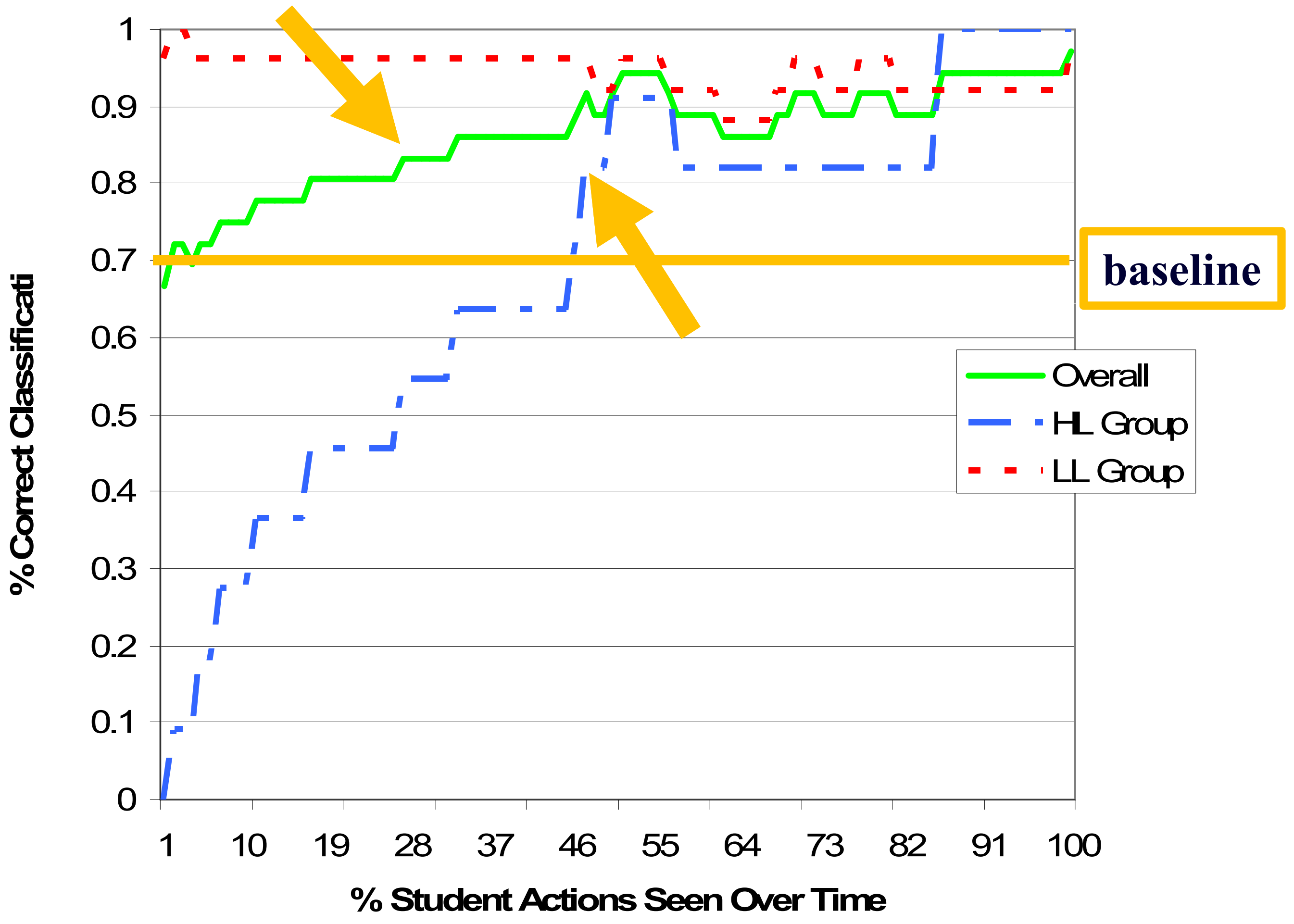
- LL students appeared to move impulsively back and forth through the curriculum, indicated by more frequent steps forward, backwards, and using other navigation tools provided by ACE .
- LL students paused longer after navigating to a help page.

User Classification



- Used the clusters to train a classifier user model.
- 36 fold Leave-one-out cross validation evaluation of the user model:
  - Removed one student's data at a time
  - Performed offline k-means clustering
  - Trained the classifier with the clusters found
  - Tested the classifier with the removed student's logged data

Evaluation



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
% Correct Classificati
1
10
19
28
37
46
55
64
73
82
91
100
% Student Actions Seen Over Time
baseline
Overall
HL Group
LL Group

Comparison Of Accuracy Results

	ACE	CSP (k =2)	CSP (k =3)
Overall	86.3	88.3	66.2
LL Group	94.2	93.5	66.1
HL Group	68.3	62.4	63.3
Baseline	69.4	83.3	50.0

Discussion

- Approach based on clustering gives good results for
  - Identifying general user types based on learning outcomes
  - Recognize users in real time based on these clusters
- Limitation: coarse classification granularity
  - Clusters consist of a variety of behaviors
  - Not possible to isolate which ones impact a new user's classification at any given time
  - Not possible to generate precise interventions on the relevant behaviors (e.g., discourage superficial browsing)

Solution

- Extend our approach by using
  - Class Association Rules (CAR)
  - A CAR Classifier

Summary

- Challenge:
  - build student models for environments in which it is hard to define effective interaction – e.g. exploratory learning environments
- Our solution
  - combines unsupervised and supervised learning to find and recognize relevant behaviors
- Tested on two different ELEs and on different types of data, including gaze patterns
- Encouraging results, although needs testing on larger datasets

Future Work

- Get more data!
  - Running subjects on CSP applet now and more will come when I teach Intro to AI this fall
- Experiment with other clustering and CAR algorithms
- Experiment with other ELE
  - We have a project to apply this techniques to ELE embedded in Virtual Worlds
- Build adaptive interventions based on this student modeling approach
- It would be great to compare them with interventions based on more precise modeling techniques

Thanks to

- Saleema Amershi
- Andrea Bernardini
- Samad Kardan
- David Ternes

- And to you for your kind attention!