US ARMY RDECOM

U.S. Army Research, Development and Engineering Command

Learning in Intelligent Tutoring Environments (LITE) Lab personnel at USMA, April 2011 (L-R):

- Dr. Robert Sottilare
- Dr. Heather Holden
- Mr. Keith Brawner
- Mr. Benjamin Goldberg

TECHNOLOGY DRIVEN. ***WARFIGHTER FOCUSED.***

Defense and Homeland Security Simulation

Realtime Clustering Of Unlabelled Sensory Data For Trainee State Assessment

September 2011

Trainee State Assessment for ITS

- Student Actions
- Sensor Data
- Assessment and Classification
- Instructional Strategy Decision?

Difficulties of Sensing

- People are not consistent
  - Day to day
  - Baseline to Baseline
- Unsupervised learning
- Real-time processing
  - Deterministic Algorithms
- Datastream problems
  - Infinite Length
  - Concept Detection
  - Concept Drift
  - Concept Evolution

Potential Solutions

- Incremental Clustering
  - K-means
  - Agglomerate
- Growing Neural Gas
- Adaptive Resonance Theory

Incremental *K*-Means Clustering

- Strengths
  - Benchmark approach
  - Well supported
- Weaknesses
  - Must know *K*
  - Responds to data frequency
  - Partitions poorly
  - NP-hard (general case)
  - Order sensitive (inc. case)

Algorithm:

For each point

Compare point to all known clusters

If no cluster is within vigilance

create new cluster here

else

move matched cluster up to <delta> in the direction of the recent point

Agglomerate Clustering

- Strengths
  - Modified inc. *k*-means
  - Accounts for cluster merging
  - Order insensitive
  - Do not have to know *k*

- Weaknesses
  - None (*a priori*)
  - Low coverage (*a posteriori)*

Algorithm:

Move the closest centroid towards the datapoint

Merge the two closest centroids, if appropriate

Creates one redundant centroid

Set redundant centroid equal to the datapoint

Adaptive Resonance Theory

- Strengths
  - Order insensitive
  - Merges
  - Responds to new concepts
- Weaknesses
  - Box-shapes
  - Parameterization issues
  - NN issues (trending)

Algorithm:

Apply new input pattern

Compute activation of all neurons

Select winning neuron

Vigilence test

If vigilence is relevant, add new pattern

Else not, test next best neuron

Else (no neurons), initialize new neuron

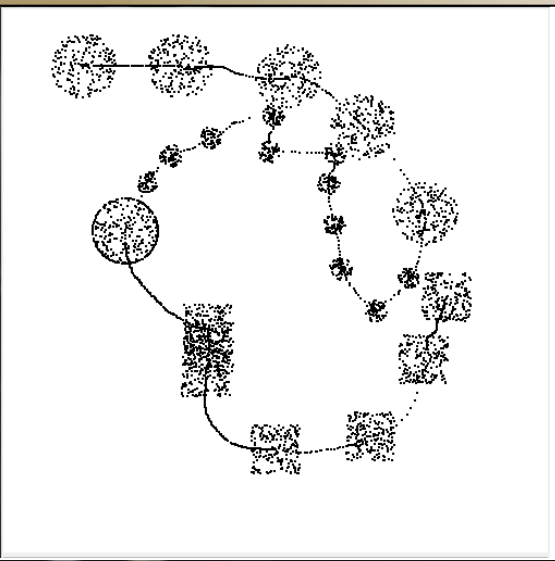
Growing Neural Gas

- Strengths
  - Responds to new concepts
  - Order sensitive
- Weaknesses
  - Order sensitive
  - Data frequency response
  - Gradient Descent
    - Slows with additional data

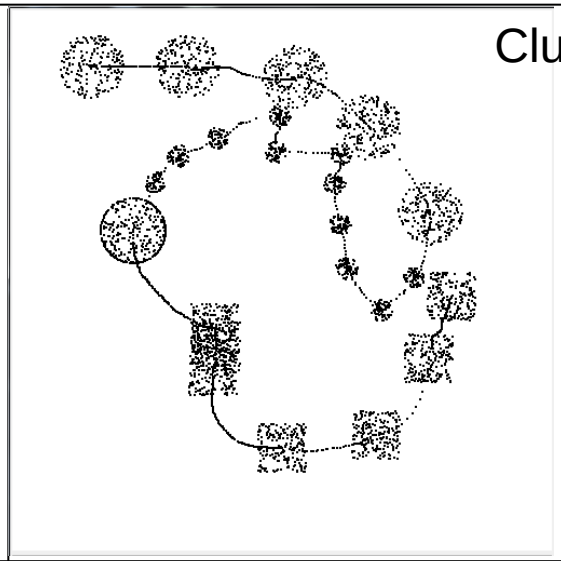
Algorithm:

If appropriate (current point does not correspond to known information)
create new reference arc
store error
Else, increment age of all arcs in this area
move existing arcs towards new data,
establish new ages for arcs
Remove Aged arcs
If any non-emanating arcs exist, remove
If it is the time to add a new point (timing)
Add a new reference point, halve the distances of the existing arcs to this point, scale the existing errors
Compute path of all arcs
For this point against each class:
If there are few related nodes, compute the probability of the point belonging to the lowest error class
Else determine the modified shape of the cluster it is most likely to belong to

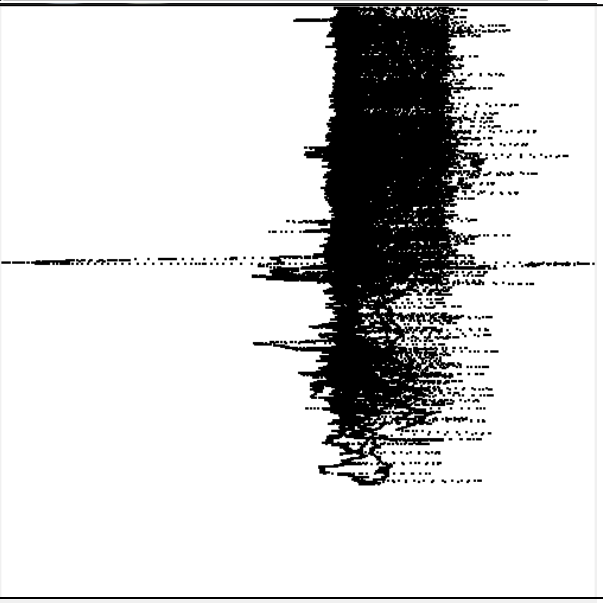
Raw Datasets



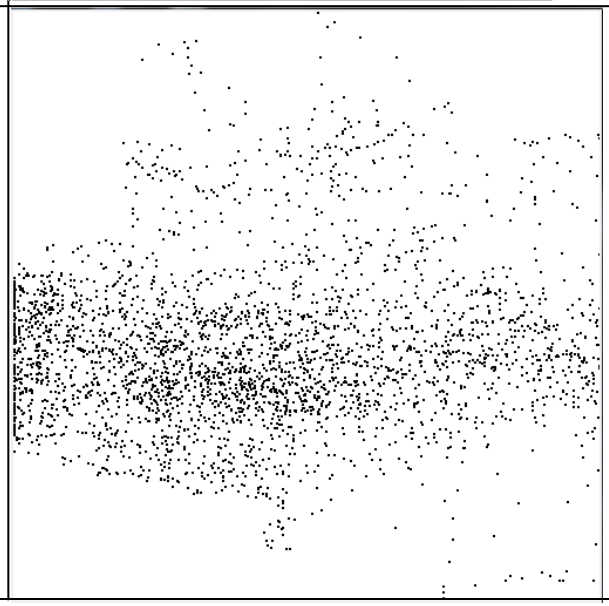
Clusters, Ordered



Clusters, Unordered

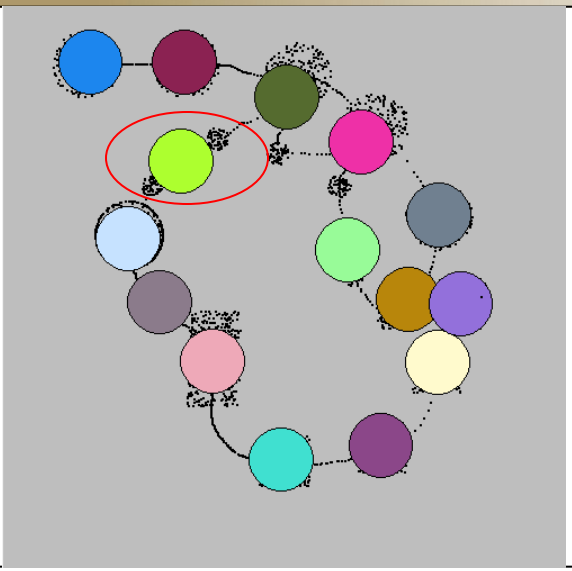


X: ECG
Y: GSR

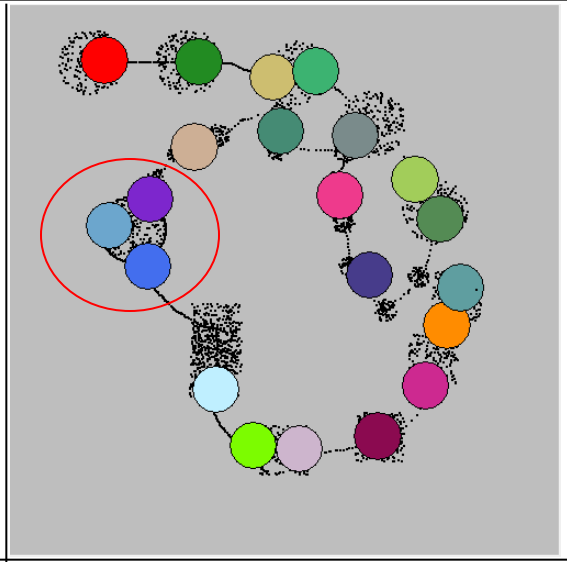


X: STE
Y: Engagement

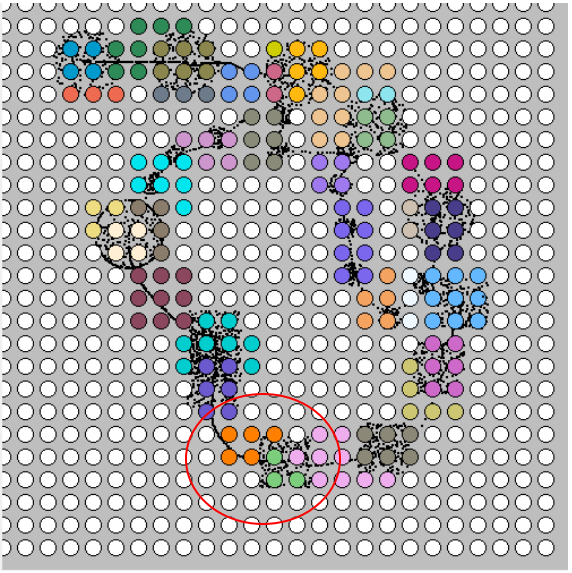
Performance – Ordered Shapes



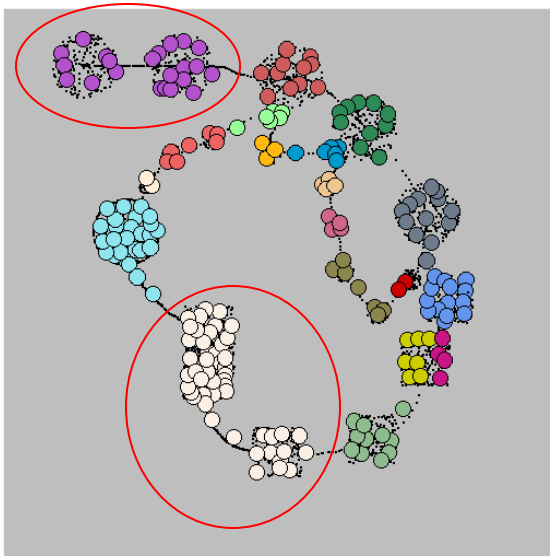
Agglomerate



Inc. K-means

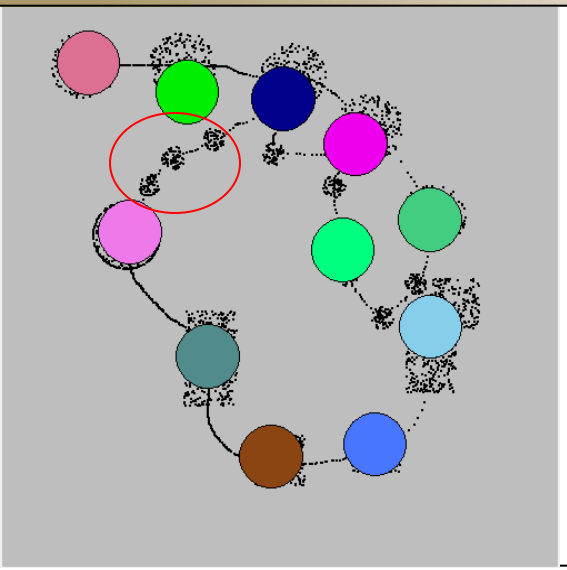


ART

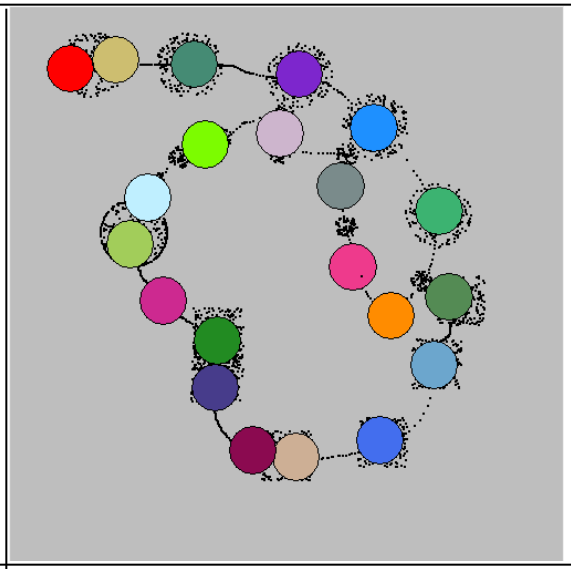


GNG

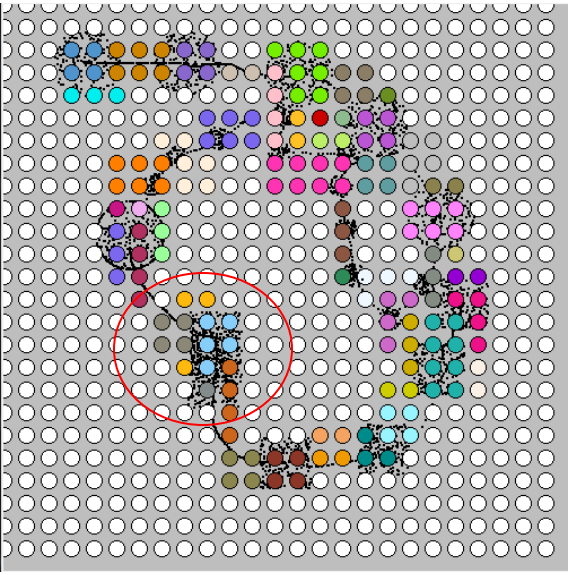
Performance – Unordered Shapes



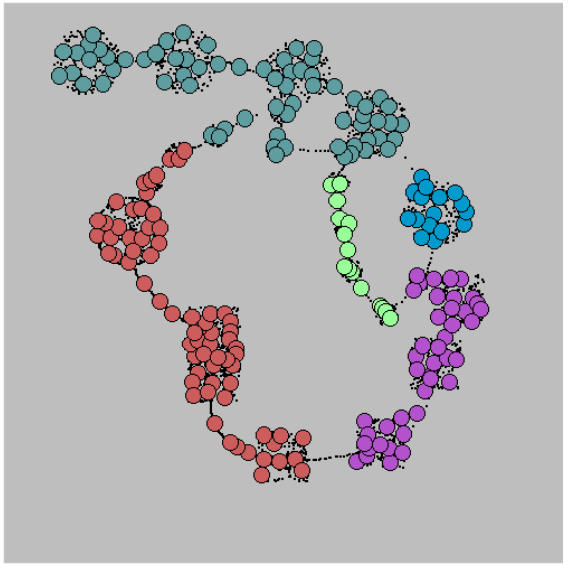
Agglomerate



Inc. K-means

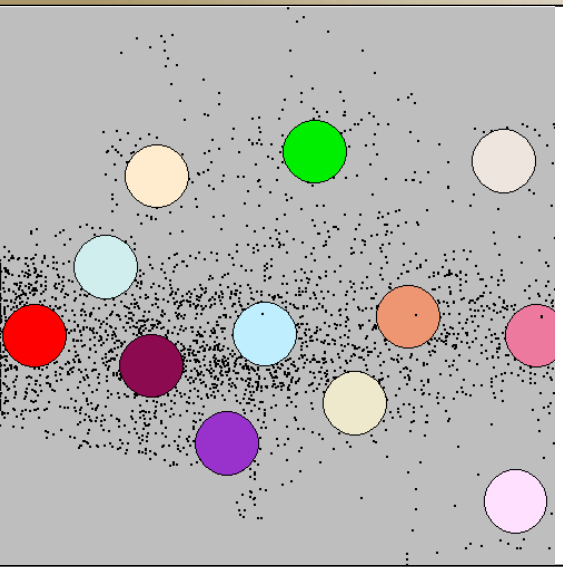


ART

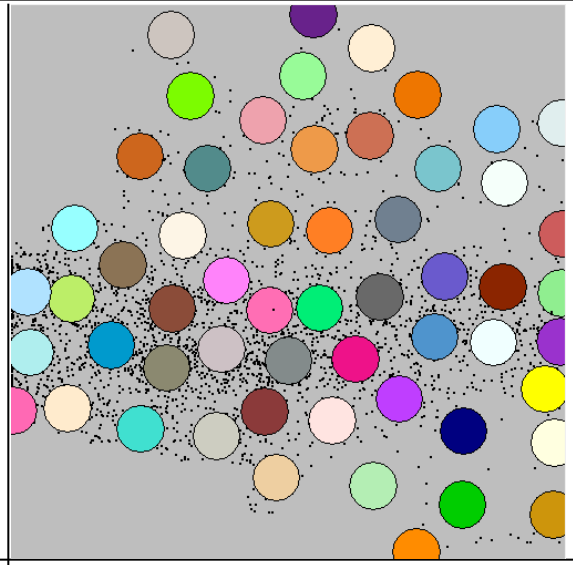


GNG

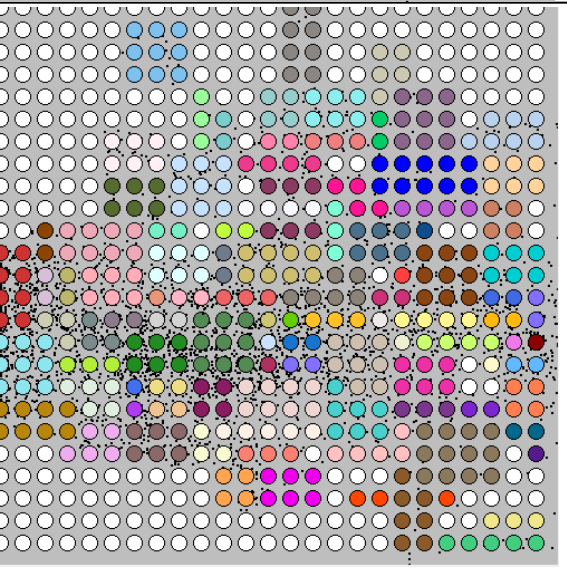
Performance – EEG (STE/Engagement)



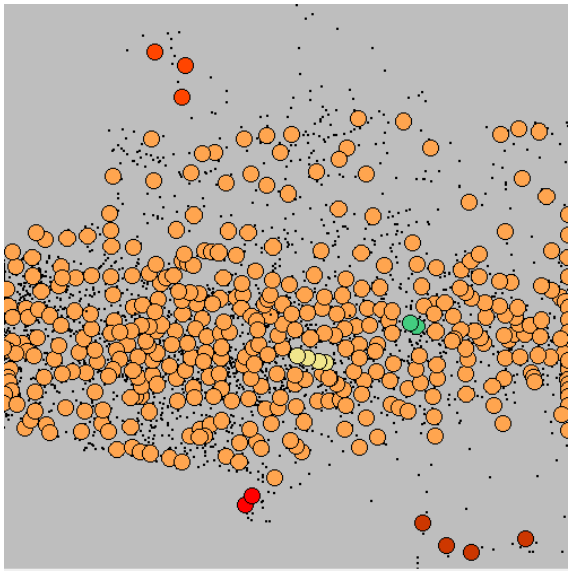
Agglomerate



Inc. K-means



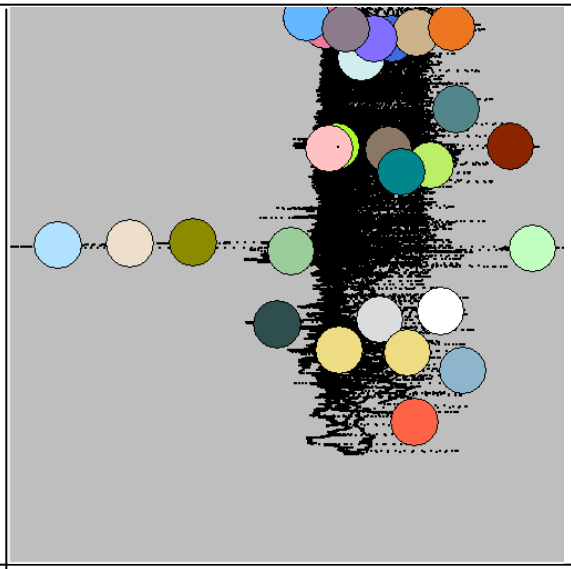
ART



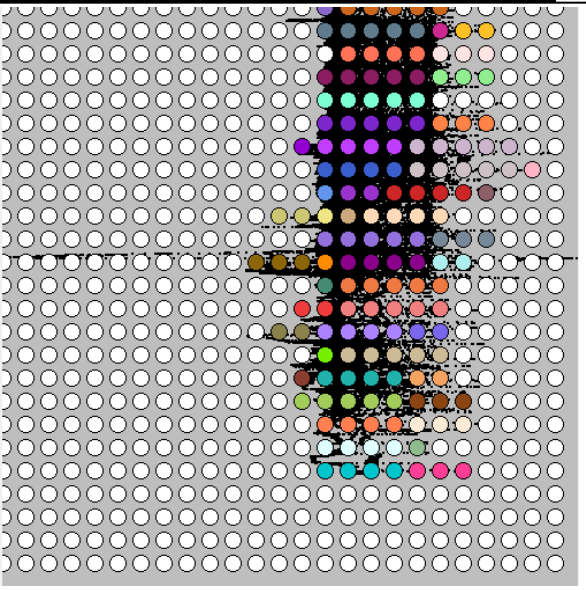
GNG

Performance – ECG/GSR

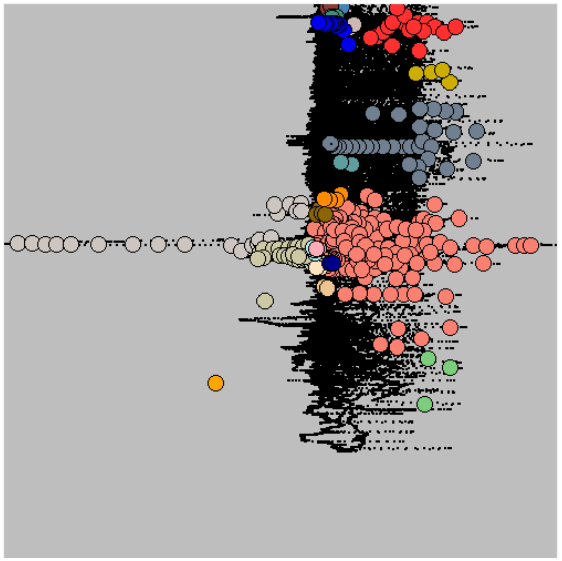
Agglomerate (Failed)



Inc. K-means

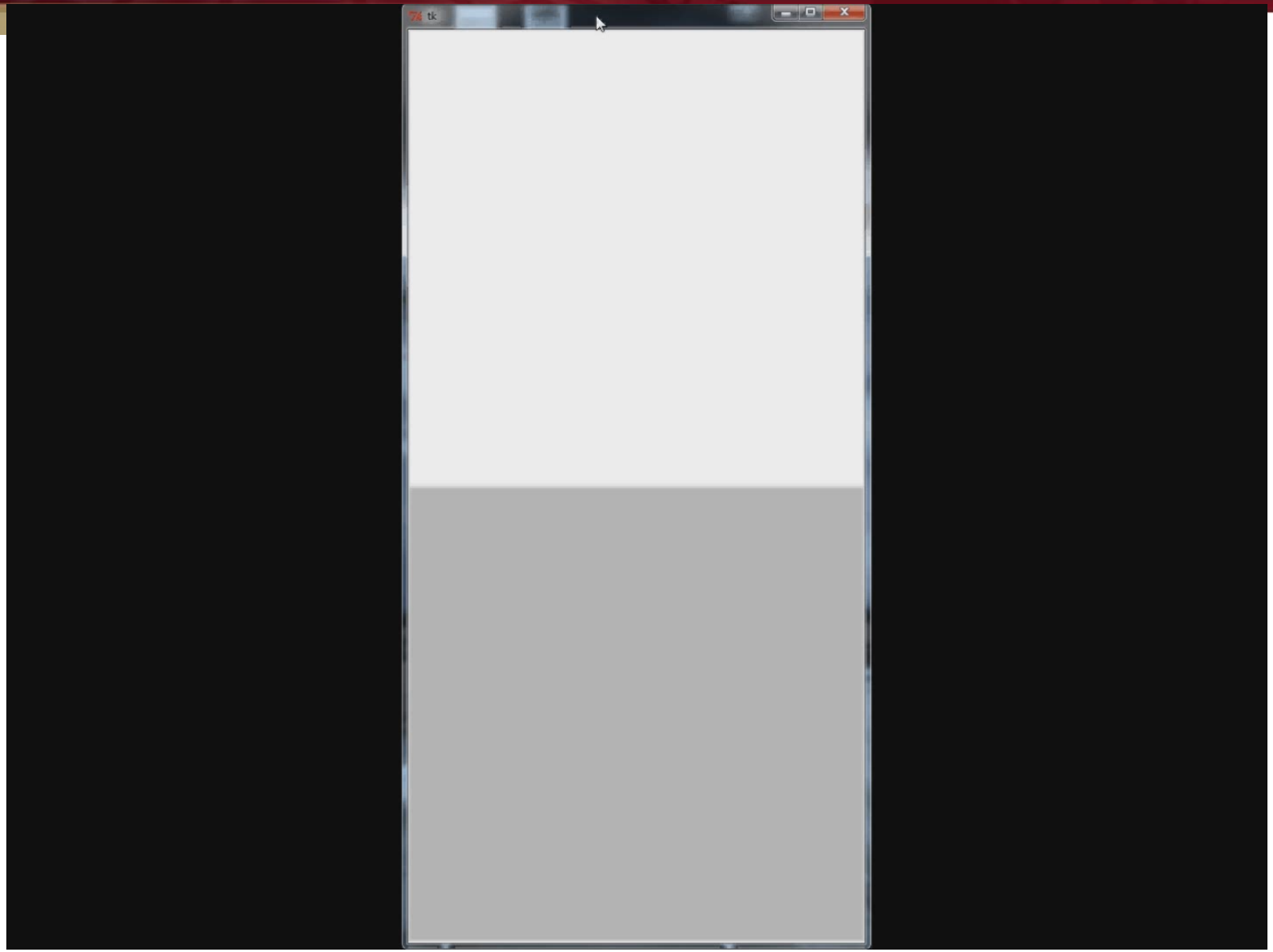


ART



GNG

Performance – 4x speed movie



Conclusions / Future Work

- Use constraint-based approaches
  - Semi-supervised clustering
  - Requires selection of initial algorithm
- Associate performance data with state data
  - More complete student picture
- Evaluate against validated dataset
  - Determine sensors to use
- Evaluate in an ITS system
  - Includes instructional strategy selection
- Use clusters as states, forecast movement between them