

Big Data with ADAMS

What the heck is ADAMS?





- Java, GPLv3
- Data mining: MOA, WEKA, MEKA, R
- Spreadsheets and databases
- Image and video processing
- Visualizations (plots, GIS)
- Scripting via Jython and Groovy

• ...



- Operators are called "actors"
- Actors arranged in tree, no connections
- Actor "handlers" nest other actors
 - e.g., sequence of actors
- Control actors control data flow
 - e.g., branch, tee, if-then-else, switch
- Input/output defines
 - standalone CA, source the transformer E, sink





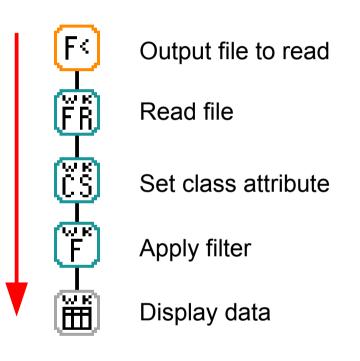
- Tree only supports 1-to-n connections
- Simulating n-to-m semantics
 - Containers
 - Variables
 - Internal storage
 - Callable actors







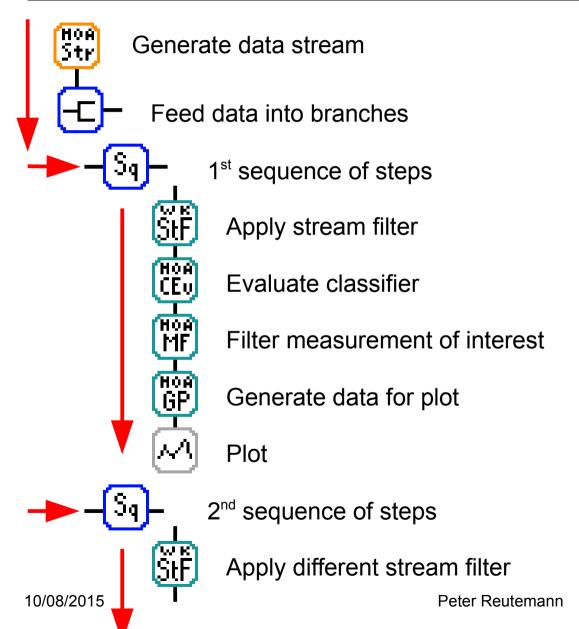
Execute nested actors one after the other



Load dataset, apply filter and display dataset







Filter data stream in two separate branches with different filters, evaluate classifier and plot metric





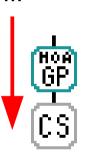


groups actors accessible via their name ("callable actors")



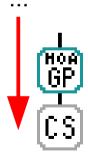
combined plot





1st evaluation: create plotting data

Pump data into referenced plot



2nd evaluation: create plotting data

Pump data into referenced plot

Generate combined plot of two evaluations by using "callable actors" functionality



Research (demos)

- Compare two MOA classifiers (drift)
- Compare MOA classifier on different streams
- MOA cluster visualization
- Track mouse in video

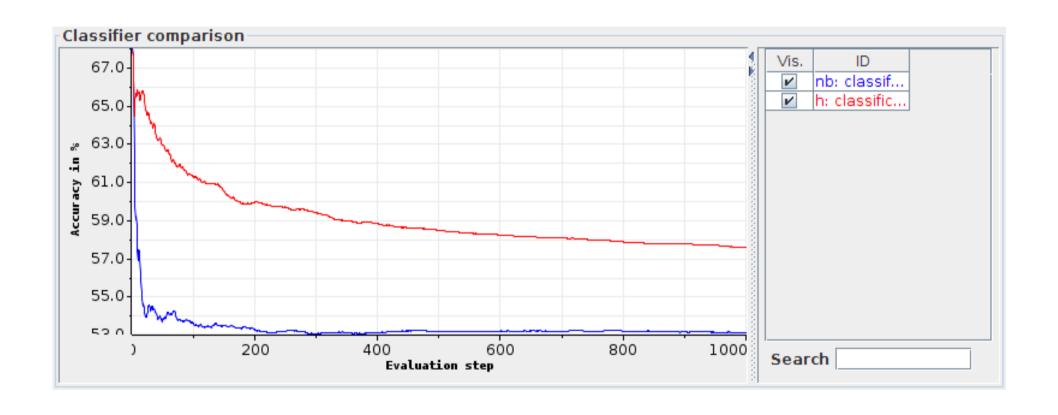
MOA - Drift



- FI Compares the accuracy of a Hoeffding tree classifier with NaiveBayes on a RandomRBFGenerator (with drift) data stream.
 - CA CallableActors
 - Hoeffding HoeffdingTree
 - NaiveBayes NaiveBayes
 - SequencePlotter x:left, Y:top, W:800, H:350
 - MOAStream RandomRBFGeneratorDrift/1000000, chunks of 1000
 - PC-Branch parallel, threads: #cores
 - ր -[Տգ]-hoeffding
 - MOAClassifierEvaluation Hoeffding, BasicClassificationPerformanceEvaluator/1000
 - MOAMeasurementsFilter .*classifications.*
 - MOAMeasurementPlotGenerator h:
 - CS CallableSink SequencePlotter
 - ণ জি-naivebayes
 - MOAClassifierEvaluation NaiveBayes, BasicClassificationPerformanceEvaluator/1000
 - MOAMeasurementsFilter .*classifications.*
 - MOAMeasurementPlotGenerator nb:
 - (S) CallableSink SequencePlotter

MOA - Drift







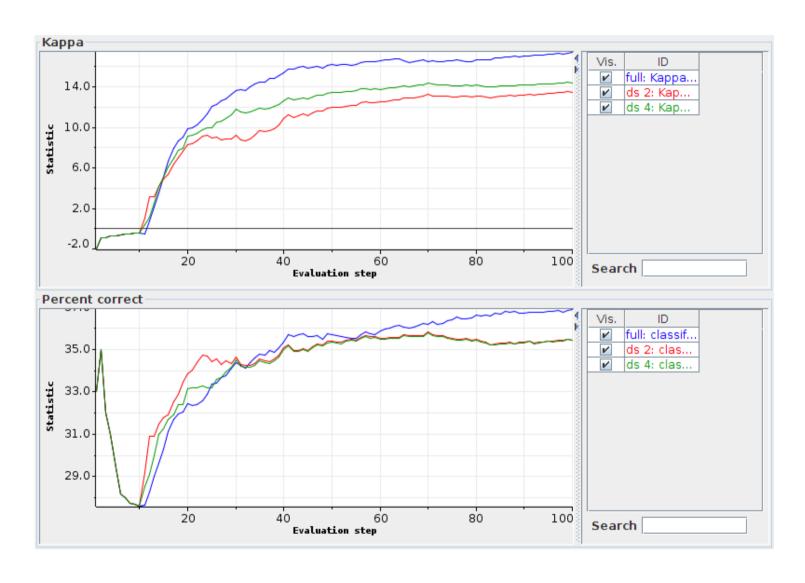
MOA - different streams

- Compares the performance of a classifier on different data streams.

 The DownSample stream filter is used to reduce the number of attributes for two of the three streams.
 - GridView X:left, Y:top, W:800, H:600, Rows: 2, Cols: 1
 - Kappa X:right, Y:top, W:800, H:350
 - Percent correct X:right, Y:top, W:800, H:350
 - CA CallableActors
 - MOAClassifier DecisionStump
 - MOAStream RandomRBFGenerator/10000, one-by-one
 - 9 -C-Branch sequential
 - 9 -(Sa)-full
 - MOAClassifierEvaluation MOAClassifier, BasicClassificationPerformanceEvaluator/100
 - P (-C)-Branch sequential
 - ր -Տգ-kappa
 - MOAMeasurementsFilter Kappa St.*
 - MOAMeasurementPlotGenerator full:
 - CS CallableSink Kappa
 - 9 -(Sq)-percent correct
 - MOAMeasurementsFilter .*correct.*
 - MOAMeasurementPlotGenerator full:
 - (S) CallableSink Percent correct
 - 9 Sq-downsample 2
 - WekaStreamFilter DownSample
 - MOAClassifierEvaluation MOAClassifier, BasicClassificationPerformanceEvaluator/100
 - - ۹ -{Sq}-kappa
 - MAMASSUrementsFilter Kanna St.*



MOA - different streams





MOA - Cluster visualization

- Visualizes the clustering of the data stream. One on the raw data, the other one with filtered data.
 - CA CallableActors
 - MOAClustererSetup ClusterGenerator
 - MOAStream RandomRBFGeneratorEvents/500000, chunks of 1000
 - - ף -S₄-raw
 - MOAClusterVisualization (raw) MOAClustererSetup
 - 9 -(Sq)-filtered
 - 🗱 WekaStreamFilter MultiFilter
 - MOAClusterVisualization (filtered) MOAClustererSetup





```
Flow

FI Let's the user select an initial object to be tracked.
In this case, select the head of the person.
             Frame X:left, Y:top, W:800, H:600, short title
                                    a background background
                                   Trail X:right, Y:top, W:1000, H:600, short title, paintlet: Circles, zoom: -1.0
                        SetVariable @{hm_factor} = 16
                         ▶ Start
             P (Tr) process video
                                    SelectFile directory: ${HOME}/documents/presentations/adams/bigmine-2015/flows/data
                          BaseName remove extension
                                                 ($v) SetVariable @{file} [REPLACE]
                                    MovielmageSequence 100ms
                                    SetStorageValue frame

    T-get timestamp

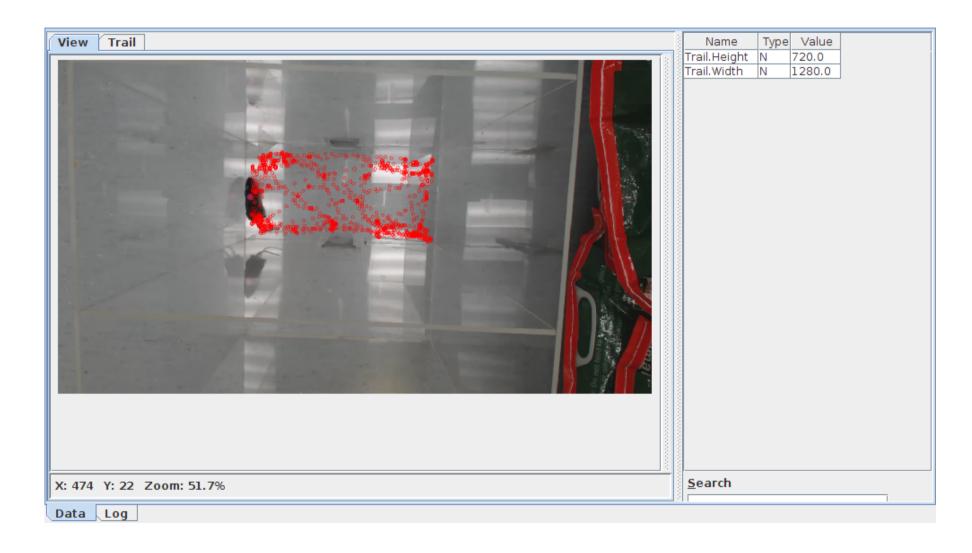
                                                 GetReportValue Timestamp[S]
                                                  (PB) Convert StringToDateTimeType
                                    P (T)-short
                                                            PB Convert-1 DateTimeTypeToString
                                                             $ SetVariable @{timestamp} [REPLACE]
                                                             RB Convert-1 DateTimeTypeToString
                                                             $ SetVariable @{timestamp_long} [REPLACE]
                          9 (1x)-select object @{variable}
                                                [66] Inspect X:left, Y:top, W:800, H:600, provider: ImageViewer (X:left, Y:top, W:640, H:480), skip: _Skip, accept: _Accept
                                                 SetStorageValue frame
                          oreate copy for map @{variable}
                                                 SetStorageValue map

    create copy for background @{variable}

                                                 SetStorageValue background
                          P (1x)-create heatmap/trail @{variable}
                                    ∘ Twidth
                                                             ImageInfo width
                                                             SetVariable @{width} [REPLACE]
                                    γ (T)-height
                                                            ImageInfo HEIGHT
                                                             $v SetVariable @{height} [REPLACE]
                                                             \begin{tabular}{l} $\P$ NewTrail ID: $$ $\mathbb{Q}_{\hat{a}} : \mathbb{Q}_{\hat{a}} : \mathbb
```







Industry





- BLGG environmental lab in NL
- Spectral analysis
 - XRF: 10,000, MIR: 2,000, NIR: 1,500
- In operation since 2006
- Predictive modelling: soil, plant (~250 models)
- 1,000 to 3,000 samples per day
- Savings due to less wet chemistry
 - USD 18 million to USD 33 million per year

Interested?



