Big Data with ADAMS

Machine learning

Peter Reutemann

What's on offer?

- WEKA *
  - mainly batch learning, some incremental schemes
  - classification, regression, clustering, association rules, data preprocessing
- MOA *
  - online learning (= data streams)
  - classification, regression, clustering
- MEKA
  - multi-label and multi-target classification
- R
  - depends on your installed packages

* *will be covered*

WEKA

- Actors have “Weka” prefix in name
- Icons have “WK”
- Examples
  -  WekaFileReader
  -  WekaFileWriter
  -  WekaCrossValidationEvaluator
  -  WekaTrainClassifier
  -  WekaClassifierSetup
- ADAMS contains additional algorithms

Cross-validate classifier

let user select dataset

load dataset and set class

create summary and display it

CV uses callable actor

Flow editor [weka_1_crossvalidate_classifier.flow -- /home/fracpete/documents/presentations/adam

File Edit Run View Window

weka_1_crossval..date_classifier

- Flow
  - CallableActors
    - WekaClassifierSetup J48 -C 0.25 -M 2
  - SelectFile directory: ${CWD} [stops flow if canceled]
  - WekaFileReader automatic (DATASET)
  - WekaClassSelector last, name: .*
  - WekaCrossValidationEvaluator WekaClassifierSetup, folds: 10, seed: 1, sequential
  - WekaEvaluationSummary
  - Display X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12

Parameters

Name

adams.flow.source.SelectFile

Parameters

- **stopFlowIfCanceled**
  enabled
- **customStopMessage**
  Select ARFF file
- **extensions**
  arff

Flow.Display

File Edit View

```
=== Summary ===

Correctly Classified Instances         884               98.441  %
Incorrectly Classified Instances        14                1.559  %
Kappa statistic                          0.9605
Mean absolute error                      0.0056
Root mean squared error                  0.0669
Relative absolute error                  4.1865 %
Root relative squared error             25.9118 %
Coverage of cases (0.95 level)          98.7751 %
Mean rel. region size (0.95 level)      16.7223 %
Total Number of Instances              898
```

Per fold statistics

- Output summary and plot accuracy per fold
- Actors to use
  - WekaCrossValidationSplit
  - WekaTrainTestSetEvaluator
  - WekaEvaluationSummary
  - WekaEvaluationValuePicker
  - SimplePlot
  - HistoryDisplay

Per fold statistics

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

split into train/test evaluate callable actor

create summary of fold pair

extract accuracy (= perc correct)

Flow editor [weka_2_per_fold_stats.flow -- /home/fracpete/documents/presentations/adams/kuching-2015/

File Edit Run View Window

weka_2_per_fold_stats

- Flow
  - CallableActors
    - WekaClassifierSetup J48 -C 0.25 -M 2
    - HistoryDisplay X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12, # toke
    - SimplePlot X:left, Y:top, W:800, H:350
  - SelectFile directory: ${CWD} [stops flow if canceled]
  - WekaFileReader automatic (DATASET)
  - WekaClassSelector last, name: .*
  - WekaCrossValidationSplit folds: 10, seed: 1, relation: @
  - WekaTrainTestSetEvaluator WekaClassifierSetup
  - Branch parallel, threads: #cores
    - Sequence
      - WekaEvaluationSummary
      - CallableSink HistoryDisplay
    - Sequence-1
      - WekaEvaluationValuePicker first: Percent correct (nominal)
      - MakePlotContainer name: Accuracy, type: PLOT
      - CallableSink SimplePlot

Parameters

Name

adams.flow.sink.SimplePlot

Parameters

plot lines

- **writer**
  adams.gui.print.NullWriter
- **paintlet**
  adams.gui.visualization.sequence.LinePaintlet
- **overlayPaintlet**
  adams.gui.visualization.sequence.NullPaintlet
- **mouseClickAction**
  adams.flow.sink.sequenceplotter.NullClickActio
- **colorProvider**
  adams.gui.visualization.core.DefaultColorProvic
- **axisX**
  adams.gui.visualization.core.AxisPanelOptions -
  -tick-generator
  \"adams.gui.visualization.core.axis.FancyTickGe
  -num-ticks 20\" -nth-value 2 -width 40 -custom-
  0.0

Flow.CallableActors.HistoryDisplay

File Edit View

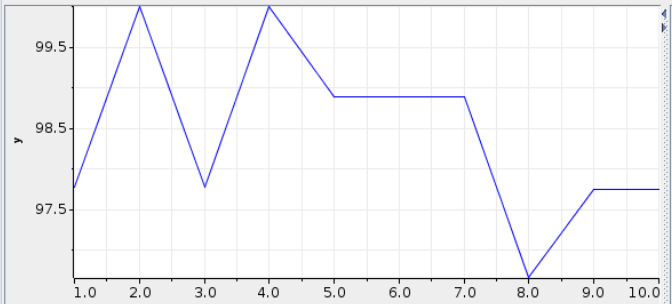
2015-11-23 12:59:31.934	
2015-11-23 12:59:32.004	
2015-11-23 12:59:32.022	
2015-11-23 12:59:32.043	
2015-11-23 12:59:32.057	
2015-11-23 12:59:32.078	
2015-11-23 12:59:32.096	
2015-11-23 12:59:32.122	
2015-11-23 12:59:32.140	
2015-11-23 12:59:32.164	

```
=== Summary ===

Correctly Classified Instances          88
Incorrectly Classified Instances         2
Kappa statistic                          0.9434
Mean absolute error                      0.0077
Root mean squared error                  0.0849
Relative absolute error                  5.7889 %
Root relative squared error             33.1576 %
Coverage of cases (0.95 level)          97.7778 %
Mean rel. region size (0.95 level)      16.6667 %
Total Number of Instances               90
```

Flow.CallableActors.SimplePlot

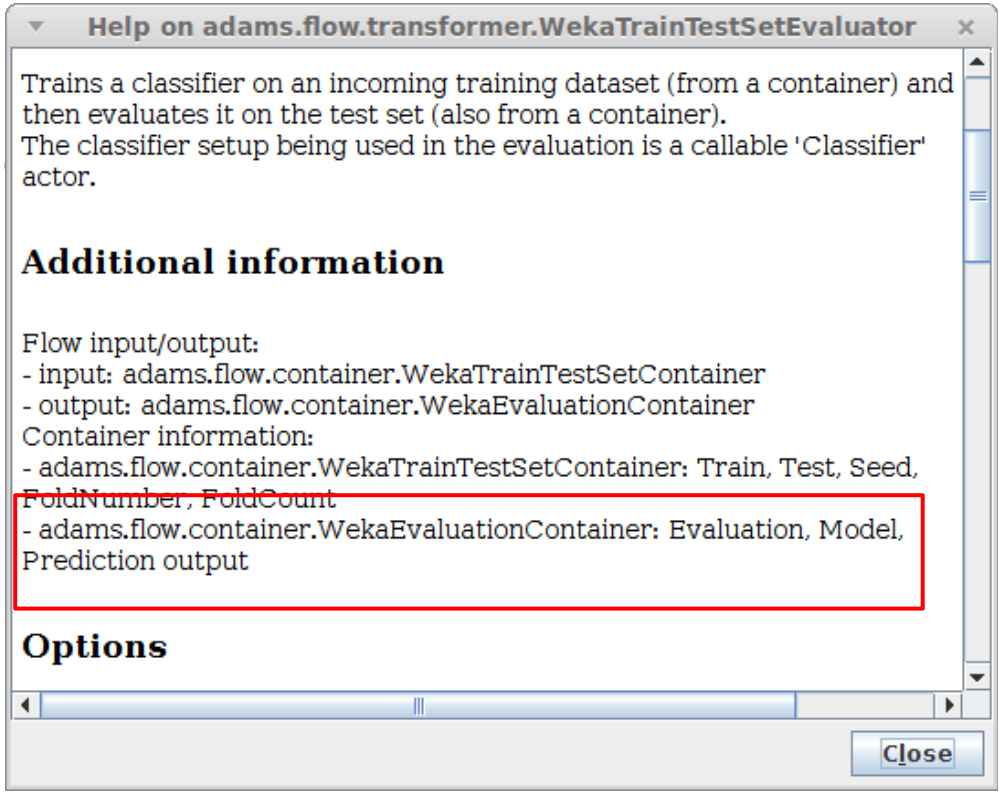
File Plot



Vis.

Per fold stats (numbered)

- output models too
  - WekaTrainTestSetEvaluator outputs container
  - ContainerValuePicker
- use variables to number
  - SetVariable
  - IncVariable
  - use *entryNameVariable* in HistoryDisplay



Per fold stats (numbered)

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

initialize var "fold" with 0

increment var "fold"

retrieve "model" from container

Flow editor [weka_3_per_fold_stats-numbered.flow -- /home/fracpete/documents/presentations/adams/kuc

File Edit Run View Window

weka_3_per_fold_stats-numbered

- Flow
  - CallableActors
    - WekaClassifierSetup J48 -C 0.25 -M 2
    - Statistics X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12, # tokens: 1,
    - Models X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12, # tokens: 1, en
    - SimplePlot X:left, Y:top, W:800, H:350
  - SelectFile directory: ${CWD} [stops flow if canceled]
  - WekaFileReader automatic (DATASET)
  - WekaClassSelector last, name: .*
  - SetVariable @{fold} = 0 [REPLACE]
  - WekaCrossValidationSplit folds: 10, seed: 1, relation: @
  - IncVariable @{fold}, INTEGER, inc: 1
  - WekaTrainTestSetEvaluator WekaClassifierSetup
  - Branch parallel, threads: #cores
    - Sequence
    - Sequence-2
      - ContainerValuePicker Model [outputs switched]
      - CallableSink Models
    - Sequence-1

Parameters

Name

adams.flow.control.ContainerValuePicker

Parameters

- **valueName**
  Model
- **switchOutputs**
  enabled

specify name of container value to retrieve

forward container value rather than container (and value in sub-flow)

Flow.CallableActors.Models

File Edit View

```
1  J48 pruned tree
2  ------------------
3
4  hardness <= 70
5  |   family = ?
6  |   |   strength <= 350
7  |   |   |   enamelability = ?
8  |   |   |   |   surface-quality = ?
9  |   |   |   |   |   condition = ?: 3 (59.0)
10 |   |   |   |   |   condition = S
11 |   |   |   |   |   |   steel = ?: 3 (2.0)
```

Flow.CallableActors.Statistics

File Edit View

```
1  === Summary ===
2
3  Correctly Classified Instances          88               97.7778 %
4  Incorrectly Classified Instances         2                2.2222 %
5  Kappa statistic                          0.9434
6  Mean absolute error                      0.0077
7  Root mean squared error                  0.0849
8  Relative absolute error                  5.7889 %
9  Root relative squared error             33.1576 %
10 Coverage of cases (0.95 level)          97.7778 %
11 Mean rel. region size (0.95 level)      16.6667 %
```

Train classifier

- Rather than evaluating, let's build a classifier!
- Also, save to disk and load it back in again
- Actors to use
  - WekaTrainClassifier
  - WekaModelWriter
  - WekaModelReader

Train classifier

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

re-use filename with a variable

train classifier and save it

load model (model + header)

Flow editor [weka_4_train_classifier_save_and_load.flow -- /home/fracpete/documents/presentations/adam

File Edit Run View Window

weka_4_train_cl..r_save_and_load

- Flow
  - CallableActors
    - WekaClassifierSetup J48 -C 0.25 -M 2
    - SetVariable @{model} = ${TMP}/test.model
    - Start
  - train
    - SelectFile directory: ${CWD} [stops flow if canceled]
    - WekaFileReader automatic (DATASET)
    - WekaClassSelector last, name: .*
    - WekaTrainClassifier WekaClassifierSetup
    - WekaModelWriter @{model}
  - load and display
    - Variable @{model}
    - WekaModelReader
    - Display X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12

Flow.SetVariable

Parameters

Name

adams.flow.standalone.SetVariable

Parameters

- **variableName**
  model
- **variableValue**
  ${TMP}/test.model

placeholder filenames (system + user defined)

Flow.load and display.Display

File Edit View

```
|   |   |   strength > 0: 1 (2.0)
|   strength > 350
|   |   steel = ?: 2 (8.0)
|   |   steel = R: 3 (5.0)
|   |   steel = A: 3 (5.0)
|   |   steel = U: 2 (0.0)
|   |   steel = K: 2 (0.0)
|   |   steel = M: 2 (11.0)
|   |   steel = S: 1 (5.0)
|   |   steel = W: 2 (0.0)
|   |   steel = V: 2 (0.0)
hardness > 70
|   hardness <= 80
|   |   cbond = ?: U (3.0)
|   |   cbond = Y: 3 (2.0)
|   hardness > 80: U (35.0)

Number of Leaves  :     35

Size of the tree :      47
```

Train and use classifier

- Let's make some predictions!
- Rather than loading serialized model, we'll store it in *internal storage* (key-value pairs)
- Actors to use
  - SetStorageValue (transformer)
  - StorageValue (source)
  - SequenceSource
  - WekaInstanceBuffer
  - WekaClassifying

Train and use classifier

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

retrieve model container, extract model

store model container in storage

turn dataset into rows

use model from callable actor "model"

Flow editor [weka_5_train_and_use_classifier.flow -- /home/fracpete/documents/presentations/adams/kuchi

File Edit Run View Window

weka_5_train_an.._use_classifier

- Flow
  - CallableActors
    - WekaClassifierSetup J48 -C 0.25 -M 2
    - model
      - StorageValue model
      - ContainerValuePicker Model [outputs switched]
  - Start
  - train
    - SelectFile directory: ${CWD} [stops flow if canceled]
    - WekaFileReader automatic (DATASET)
    - WekaClassSelector last, name: .*
    - WekaTrainClassifier WekaClassifierSetup
    - SetStorageValue model
  - use model
    - SelectFile directory: ${CWD} [stops flow if canceled]
    - WekaFileReader automatic (DATASET)
    - WekaClassSelector last, name: .*
    - WekaInstanceBuffer INSTANCES_TO_INSTANCE,
    - WekaClassifying model
    - Display X:left, Y:top, W:640, H:480, font: Monospaced

Parameters

Name

adams.flow.transformer.WekaClassifying

Parameters

- **modelActor**
  model
- **onTheFly**
  enabled

model is built "on the fly"

Flow.use model.Display

File Edit View

```
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.021739130434782608,0.0,0.9782608
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=1.0, Classification label=2, Distribution=0.0,0.981132075471698 1,0.018867924
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.021739130434782608,0.0,0.9782608
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.0,0.0,1.0,0.0,0.0,0.0, Instance=
Classification=2.0, Classification label=3, Distribution=0.021739130434782608,0.0,0.9782608
Classification=4.0, Classification label=5, Distribution=0.0,0.0,0.0,0.0,1.0,0.0, Instance=
Classification=4.0, Classification label=5, Distribution=0.0,0.0,0.0,0.0,1.0,0.0, Instance=
```

Use Filters

- Filters can be applied using
  - WekaFilter (batch + stream)
  - WekaStreamFilter (only stream)
- **Caution:** if information could leak, use FilteredClassifier approach!

Use Filters

using same variable, must use Branch in sequential mode

variable for HistoryDisplay

remove numeric attributes

remove nominal attributes

Flow editor [weka_5b_use_filters.flow -- /home/fracpete/documents/presentations/adams/kuching-2015/2_basic_flows/

File Edit Run View Window

weka_5b_use_filters

- Flow
  - CallableActors
    - WekaClassifierSetup J48 -C 0.25 -M 2
    - HistoryDisplay X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12, # tokens: 1, entry nam
  - SelectFile directory: ${CWD} [stops flow if canceled] load anneal
  - WekaFileReader automatic (DATASET)
  - WekaClassSelector last, name: .*
  - Branch sequential
    - full
      - WekaCrossValidationEvaluator WekaClassifierSetup, folds: 10, seed: 1, sequential
      - WekaEvaluationSummary
      - SetVariable @{entry} = full [REPLACE]
      - CallableSink HistoryDisplay
    - nominal
      - WekaFilter RemoveType -T numeric
      - WekaCrossValidationEvaluator WekaClassifierSetup, folds: 10, seed: 1, sequential
      - WekaEvaluationSummary
      - SetVariable @{entry} = nominal [REPLACE]
      - CallableSink HistoryDisplay
    - numeric
      - WekaFilter RemoveType -T nominal
      - WekaCrossValidationEvaluator WekaClassifierSetup, folds: 10, seed: 1, sequential
      - WekaEvaluationSummary
      - SetVariable @{entry} = numeric [REPLACE]
      - CallableSink HistoryDisplay

Flow.Branch.nominal.WekaFilter

Parameters

Name

adams.flow.transformer.WekaFilter

Parameters

- **filter**
  weka.filters.unsupervised.attribute.RemoveType -T numeric

removing only numeric attributes

Flow.CallableActors.HistoryDisplay

File Edit View

full
nominal
numeric

```
=== Summary ===

Correctly Classified Instances          790               87.9733 %
Incorrectly Classified Instances        108               12.0267 %
Kappa statistic                           0.6891
Mean absolute error                       0.0462
Root mean squared error                   0.1686
Relative absolute error                  34.3515 %
Root relative squared error              65.2781 %
Coverage of cases (0.95 level)           97.3274 %
Mean rel. region size (0.95 level)       22.6615 %
Total Number of Instances               898
```

numeric

Change classifier options

- variables can be attached to options
- doesn't work for frameworks other than ADAMS
- solution

  manipulate Java object via property path
- available actors
  - SetProperty (of callable actor)
  - GetProperty (of object passing through)
  - UpdateProperty (of object passing through)
  - UpdateProperties (of sub-actor)

Change classifier options

the classifier to modify

load dataset and put in storage

update property of callable actor

evaluate on data from storage

Flow editor [weka_6_change_classifier_options.flow -- /home/fracpete/documents/presentations/ad

File Edit Run View Window

weka_6_change_c..ssifier_options

- Flow
  - CallableActors
    - WekaClassifierSetup J48 -C 0.25 -M 2
    - HistoryDisplay X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12, # tokens: 1, entry n
  - Start
  - load training data
    - SelectFile directory: ${CWD} [stops flow if canceled] use labor.arff
    - WekaFileReader automatic (DATASET)
    - WekaClassSelector last, name: .*
    - SetStorageValue data
  - iterate confidence
    - StringConstants 4 values, one by one
    - SetVariable @{confidence} [REPLACE]
    - SetProperty property: classifier.confidenceFactor, name: WekaClassifierSetup
    - evaluate
      - StorageValue data
      - WekaCrossValidationEvaluator WekaClassifierSetup, folds: 10, seed: 1, sequential
      - WekaEvaluationSummary
      - CallableSink HistoryDisplay

Flow.iterate confidence.SetProperty

Parameters

Name

adams.flow.transformer.SetProperty

Parameters

- **property** classifier.confidenceFactor
- **actorName** WekaClassifierSetup

the property path of the confidence factor of J48

Flow.CallableActors.HistoryDisplay

File Edit View

0.2
0.25
0.3
0.35

```
=== Summary ===

Correctly Classified Instances          43               75.4386 %
Incorrectly Classified Instances        14               24.5614 %
Kappa statistic                          0.4845
Mean absolute error                      0.2972
Root mean squared error                  0.4552
Relative absolute error                 64.9755 %
Root relative squared error             95.3445 %
Coverage of cases (0.95 level)          89.4737 %
Mean rel. region size (0.95 level)      83.3333 %
Total Number of Instances               57
```

Regression

- Instead of classification use regression
- Display classifier errors as graph
- Actors to use
  - WekaClassifierErrors

    actual vs. predicted
  - WekaAccumulatedError

    sorts the error values and creates plot containers

Regression

Weka plot of act vs pred

errors sorted by size for plot

various ways of displaying data

Flow editor [weka_7_evaluate_regressor.flow -- /home/fracpete/documents/presentations/adams/ku

File Edit Run View Window

weka_7_evaluate_regressor

- Flow
  - CallableActors
    - WekaClassifierSetup LinearRegression -S 0 -R 1.0E-8
    - SelectFile directory: ${CWD} [stops flow if canceled] use bodyfat.arff
    - WekaFileReader automatic (DATASET)
    - WekaClassSelector last, name: .*
    - WekaCrossValidationEvaluator WekaClassifierSetup, folds: 10, seed: 1
  - Branch parallel, threads: #cores
    - Sequence
      - WekaClassifierErrors X:left, Y:top, W:640, H:480, error scaler: Au
    - Sequence-1
      - WekaAccumulatedError Plot
      - SimplePlot X:left, Y:top, W:800, H:350

Parameters

Name

adams.flow.sink.WekaClassifierErrors

Parameters

- writer
  adams.gui.print.NullWriter
- errorScaler
  adams.data.weka.predictions.AutoScaler -scale
  adams.data.weka.predictions.RelativeNumericl

Flow.Branch.Sequence.WekaClassifierErrors

File

X: class (Num) | Y: predictedclass (Num)

Colour: class (Num) | Select Instance

Reset | Clear | Open | Save | Jitter

Plot: bodyfat.names-classifier_errors

46.085 / 20.997 / -4.09 — 0 / 23.75 / 47.5

Class colour

0 / 23.75 / 47.5

Flow.Branch.Sequence-1.SimplePlot

File

Plot

8.0 / 6.0 / 4.0 / 2.0 / 0.0 — y

0 / 20.0 / 40.0 / 60.0 / 80.0 / 100.0 / 120.0 / 140.0 / 160.0 / 180.0 / 200.0 / 220.0 / 240.0 — x

Vis.	ID
☑	Plot

Search

Clustering

- Data handling like classifiers, but no class
- ADAMS also offers some post-processors
  - cluster centers
  - cluster statistics (min, max, mean, ...)
- Example actors
  - WekaClustererSetup
  - WekaTrainClusterer
  - WekaCrossValidationClustererEvaluator
  - WekaClustering

Clustering

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

remove class attribute

train clusterer and post-process

post-processing the clusterer/data

Flow editor [weka_7b_clustering.flow -- /home/fracpete/documents/presentations/adams/kuching-20

File Edit Run View Window

weka_7b_clustering

- Flow
  - CallableActors
    - WekaClustererSetup SimpleKMeans -init 0 -max-candidates 10
    - centers X:left, Y:top, W:640, H:480, short title
    - statistics X:right, Y:top, W:640, H:480, short title
  - FileSupplier ${HOME}/documents/presentations/adams/kuching-2015
  - WekaFileReader automatic (DATASET)
  - WekaFilter Remove
  - Branch parallel, threads: #cores
    - centers
      - WekaTrainClusterer WekaClustererSetup
      - ContainerValuePicker Clustered dataset [outputs switched
      - CallableSink centers
    - statistics
      - WekaTrainClusterer WekaClustererSetup
      - ContainerValuePicker Clustered dataset [outputs switched
      - CallableSink statistics

Parameters

Name

adams.flow.transformer.WekaTrainClusterer

Parameters

- **postProcessor**
  adams.flow.transformer.wekaclusterer.ClusterCente

centers

File

Relation: bodyfat.names-weka.filters.unsupervised.attribute.Remove-weka.filters.uns...

No.	1: Cluster index Numeric	2: Density Numeric	3: Age Numeric	4: Weight Numeric	5: Height Numeric	6: Neck Numeric	7: Chest Numeric	8: Abdomen Numeric	9: Hip Numeric	1
1	0.0	1.04160...	46.2...	204.15...	70.672...	39.85...	107.95...	101.9373...	105....	
2	1.0	1.06588...	43.9...	160.30...	69.762...	36.61...	95.564...	85.63310...	95.7...	

statistics

File

Relation: bodyfat.names-weka.filters.unsupervised.attribute.Remove-weka.filters.uns...

No.	1: Statistic String	2: Density Numeric	3: Age Numeric	4: Weight Numeric	5: Height Numeric	6: Neck Numeric	7: Chest Numeric	8: Abdomen Numeric	9: Hip Numeric	10: Th Nume
1	0-Min	0.995	23.0	163.75	29.5	34.4	95.4	88.6	94.2	5
2	0-Max	1.0991	74.0	363.15	77.75	51.2	136.2	148.1	147.7	8
3	0-Median	1.04160...	46.2...	204.15...	70.672...	39.85...	107.95...	101.9373...	105....	63.41
4	0-Mean	1.04160...	46.2...	204.15...	70.672...	39.85...	107.95...	101.9373...	105....	63.41
5	0-StdDev	0.01514...	12.8...	24.497...	4.8351...	1.993...	6.8528...	8.495287...	6.64...	4.665
6	1-Min	1.0378	22.0	118.5	64.0	31.1	79.3	69.4	85.0	4
7	1-Max	1.1089	81.0	191.0	77.5	41.1	106.9	99.8	103.9	6
8	1-Median	1.06588...	43.9...	160.30...	69.762...	36.61...	95.564...	85.63310...	95.7...	56.44
9	1-Mean	1.06588...	43.9...	160.30...	69.762...	36.61...	95.564...	85.63310...	95.7...	56.44
10	1-StdDev	0.01450...	12.3...	15.556...	2.4127...	1.696...	4.8682...	5.991024...	4.00...	3.354

Quiz: Weka

- Evaluate each of these classifier setups
  - RandomForest with 250 trees
  - SMO with RBF kernel and logistic models
- On these datasets
  - labor.arff
  - anneal.arff
- Display evaluation summaries

Quiz: Weka

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

classifier setup has variable attached

user selects datasets to evaluate on

extract relation name from current dataset

classifier setups are stored as string (copy/paste from Weka Explorer)

cross-validate current classifier/dataset combination and display

Flow editor [weka_8_quiz.flow -- /home/fracpete/documents/presentations/adams/kuching-2015/2

File Edit Run View Window

weka_8_quiz

- Flow
  - CallableActors
    - WekaClassifierSetup @{classifier}
    - HistoryDisplay X:left, Y:top, W:640, H:480, font: Monospaced-PLAIN-12, # toker
    - SelectFile directory: ${CWD} [stops flow if canceled]
    - WekaFileReader automatic (DATASET)
  - Tee
    - WekaInstancesInfo RELATION_NAME
    - SetVariable @{relation} [REPLACE]
    - WekaClassSelector last, name: .*
    - SetStorageValue data
  - iterate classifiers
    - StringConstants 2 values, one by one
    - SetVariable @{classifier} [REPLACE]
    - Tee
      - StringReplace replace '([a-zA-Z\\.]+)(\\.)([a-zA-Z\\.]+)(.*)' with '$3'
      - SetVariable @{cls_short} [REPLACE]
    - combination
      - CombineVariables @{relation}/@{cls_short}
      - SetVariable @{combination} [REPLACE]
    - evaluate
      - StorageValue data
      - WekaCrossValidationEvaluator WekaClassifierSetup, folds: 10, seed: 1,
      - WekaEvaluationSummary
      - CallableSink HistoryDisplay

Flow finished.

Parameters

Name

adams.flow.transformer.StringReplace

Parameters

- **find**
  ([a-zA-Z\\\\.]+)(\\\\.)([a-zA-Z\\\\.]+)(.*)
- **replace**
  $3

extract classifier name using regular expression

Flow.CallableActors.HistoryDisplay

File Edit View

anneal/RandomForest
anneal/SMO
labor/RandomForest
labor/SMO

```
=== Summary ===

Correctly Classified Instances          50               87.7193 %
Incorrectly Classified Instances         7               12.2807 %
Kappa statistic                          0.7394
Mean absolute error                      0.1289
Root mean squared error                  0.304
Relative absolute error                 28.1838 %
Root relative squared error             63.6643 %
Coverage of cases (0.95 level)          94.7368 %
Mean rel. region size (0.95 level)      62.2807 %
Total Number of Instances               57
```

MOA

- Actors have “MOA” prefix in name
- Icons have “MOA”
- Examples
  - MOAClassifierEvaluation
  - MOAMeasurementsFilter
  - MOAMeasurementsPlotGenerator
  - MOAClassifierSetup
  - MOAStream

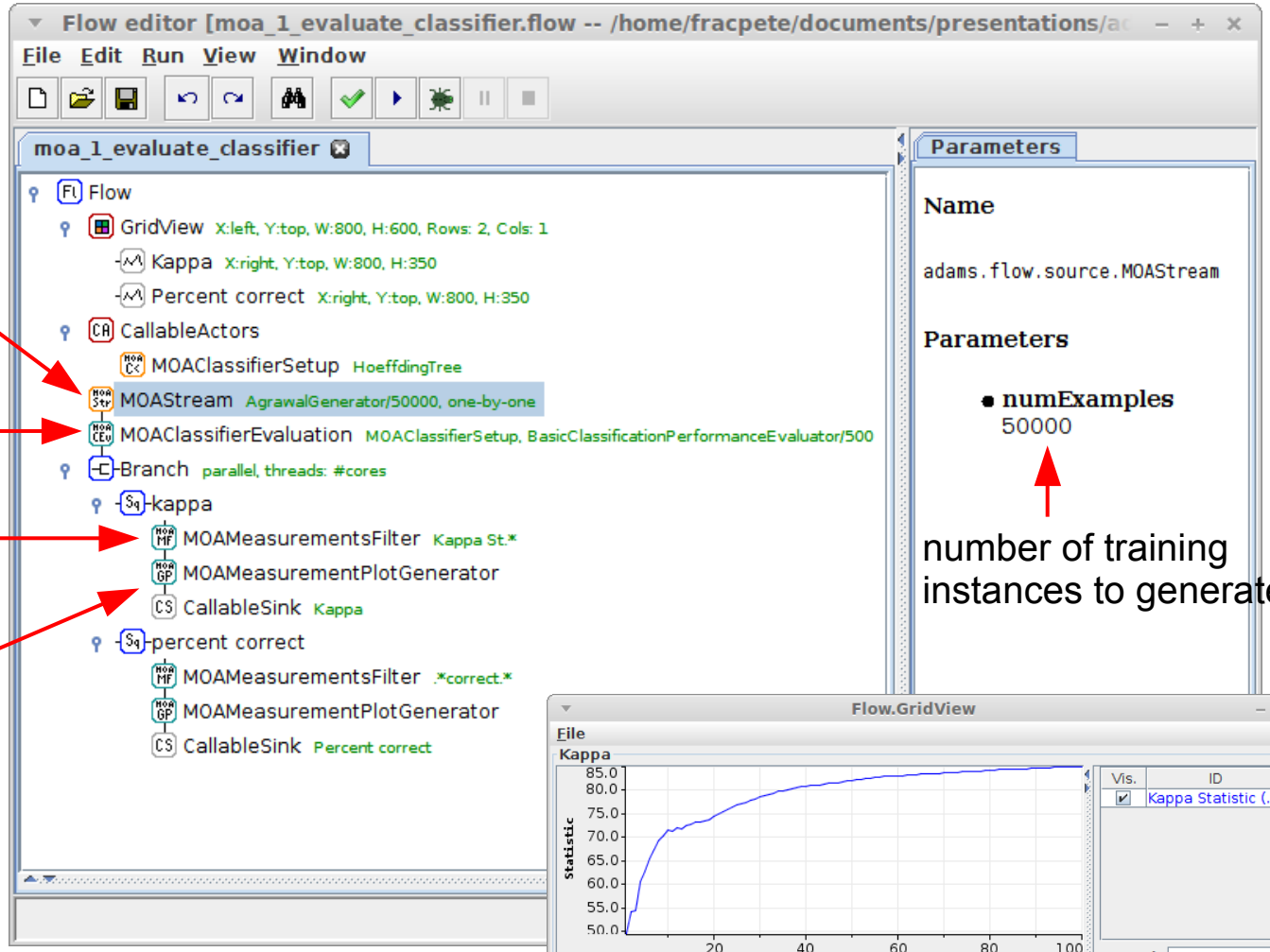
Evaluate classifier

artificial data stream generator

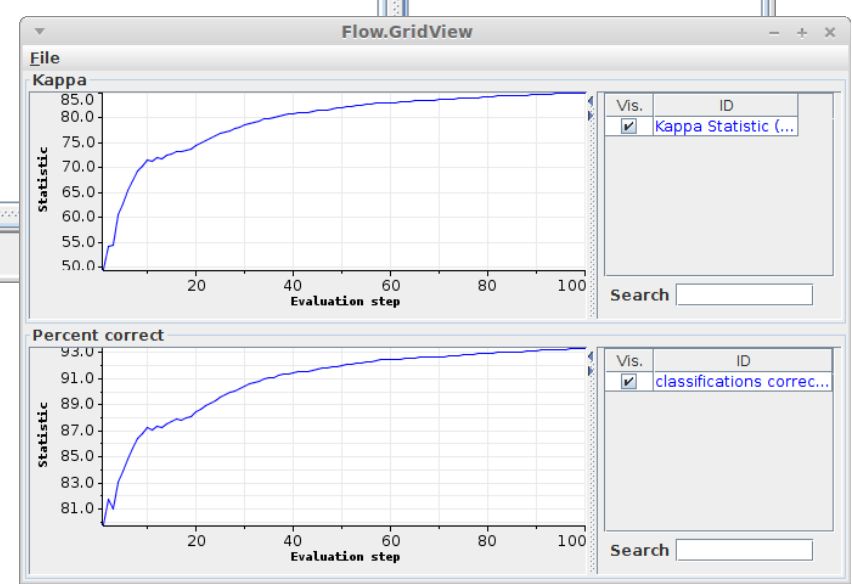
evaluate classifier output evaluation every 500 instances

get measurement of interest

create plot container and forward it to plot



number of training instances to generate



Filter streams

- Number and type of attributes can have impact on classifier performance
- Stream filters can be used to filter data streams
  - WekaStreamFilter

Filter streams

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

data stream with
40 numeric attributes

remove every
2nd attribute

remove every
4th attribute

Flow editor [moa_2_filter_streams.flow -- /home/fracpete/documents/presentations/adams/kuching-2

File Edit Run View Window

moa_2_filter_streams

- Flow
  - CallableActors
    - MOAClassifier HoeffdingTree
    - Kappa X:left, Y:top, W:800, H:350
  - MOAStream RandomRBFGenerator/10000, one-by-one
  - Branch parallel, threads: #cores
    - full
      - MOAClassifierEvaluation MOAClassifier, BasicClassificationPerf
      - MOAMeasurementsFilter Kappa St:*
      - MOAMeasurementPlotGenerator full:
      - CallableSink Kappa
    - downsample 2
      - WekaStreamFilter DownSample
      - MOAClassifierEvaluation MOAClassifier, BasicClassificationPerf
      - MOAMeasurementsFilter Kappa St:*
      - MOAMeasurementPlotGenerator ds 2:
      - CallableSink Kappa
    - downsample 4
      - WekaStreamFilter DownSample
      - MOAClassifierEvaluation MOAClassifier, BasicClassificationPerf
      - MOAMeasurementsFilter Kappa
      - MOAMeasurementPlotGenerat
      - CallableSink Kappa

Parameters

Name

adams.flow.transformer.WekaStreamFilter

Parameters

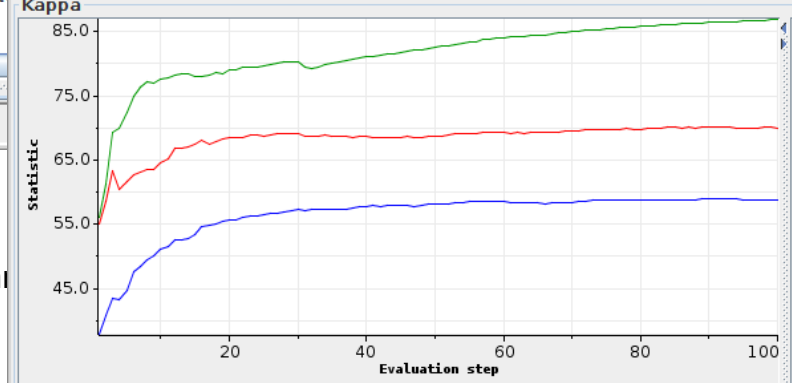
- **filter**
  weka.filters.unsupervised.attribute.DownSample
  -nth 2

remove every
*n*th attribute

Flow.CallableActors.Kappa

File

Kappa



Vis.	ID
✓	ds 4: Kappa
✓	ds 2: Kappa
✓	full: Kappa S

Search

Drift analysis

- Select a drift stream generator as source

  e.g., RandomRBFGeneratorDrift
- Generate lots of examples
- Output examples in chunks to avoid flow overheads (e.g., provenance)
- Compare several algorithms

  adaptive and non-adaptive

Drift analysis

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

adaptive and non-adaptive algorithms

evaluation in parallel, prefixing plots output every 1000 examples

*Flow editor [moa_2b_drift_analysis.flow -- /home/fracpete/documents/presentations/adams/kuching-2015/3_machi

File Edit Run View Window

*moa_2b_drift_analysis

- Flow
  - CallableActors
    - Hoeffding HoeffdingTree
    - Hoeffding Adaptive HoeffdingAdaptiveTree
    - OzaBagAdwin OzaBagAdwin
    - SequencePlotter X:left, Y:top, W:800, H:350
  - MOAStream RandomRBFGeneratorDrift/500000, chunks of 1000
  - Branch parallel, threads: #cores
    - hoeffding
      - MOAClassifierEvaluation Hoeffding, BasicClassificationPerformanceEvaluator/
      - MOAMeasurementsFilter .*classifications.*
      - MOAMeasurementPlotGenerator h:
      - CallableSink SequencePlotter
    - hoeffding adaptive
      - MOAClassifierEvaluation Hoeffding Adaptive, BasicClassificationPerformanceE
      - MOAMeasurementsFilter .*classifications.*
      - MOAMeasurementPlotGenerator ha:
      - CallableSink SequencePlotter
    - ozabag adwin
      - MOAClassifierEvaluation OzaBagAdwin, BasicClassificationPerformanceEvalua
      - MOAMeasurementsFilter .*classifications.*
      - MOAMeasurementPlotGenerator oba:
      - CallableSink SequencePlotter

Flow.MOAStream

Parameters

Name

adams.flow.source.MOAStream

Parameters

- **streamGenerator**
  moa.streams.generators.RandomRBFGeneratorDrift -s 0.001
- **numExamples**
  500000
- **chunkSize**
  1000

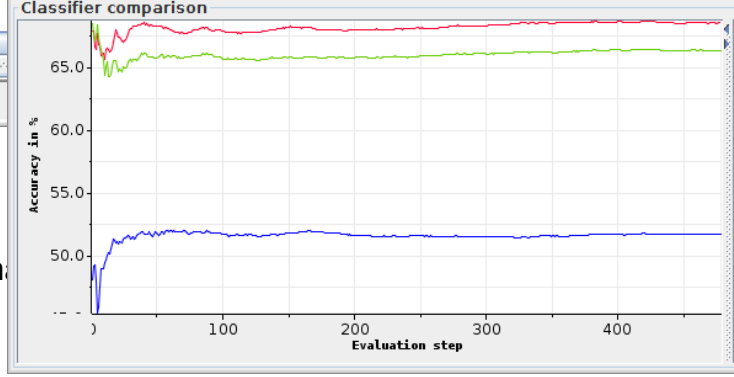
drift generator

speed up transfer of data in flow

Flow.CallableActors.SequencePlotter

File

Classifier comparison



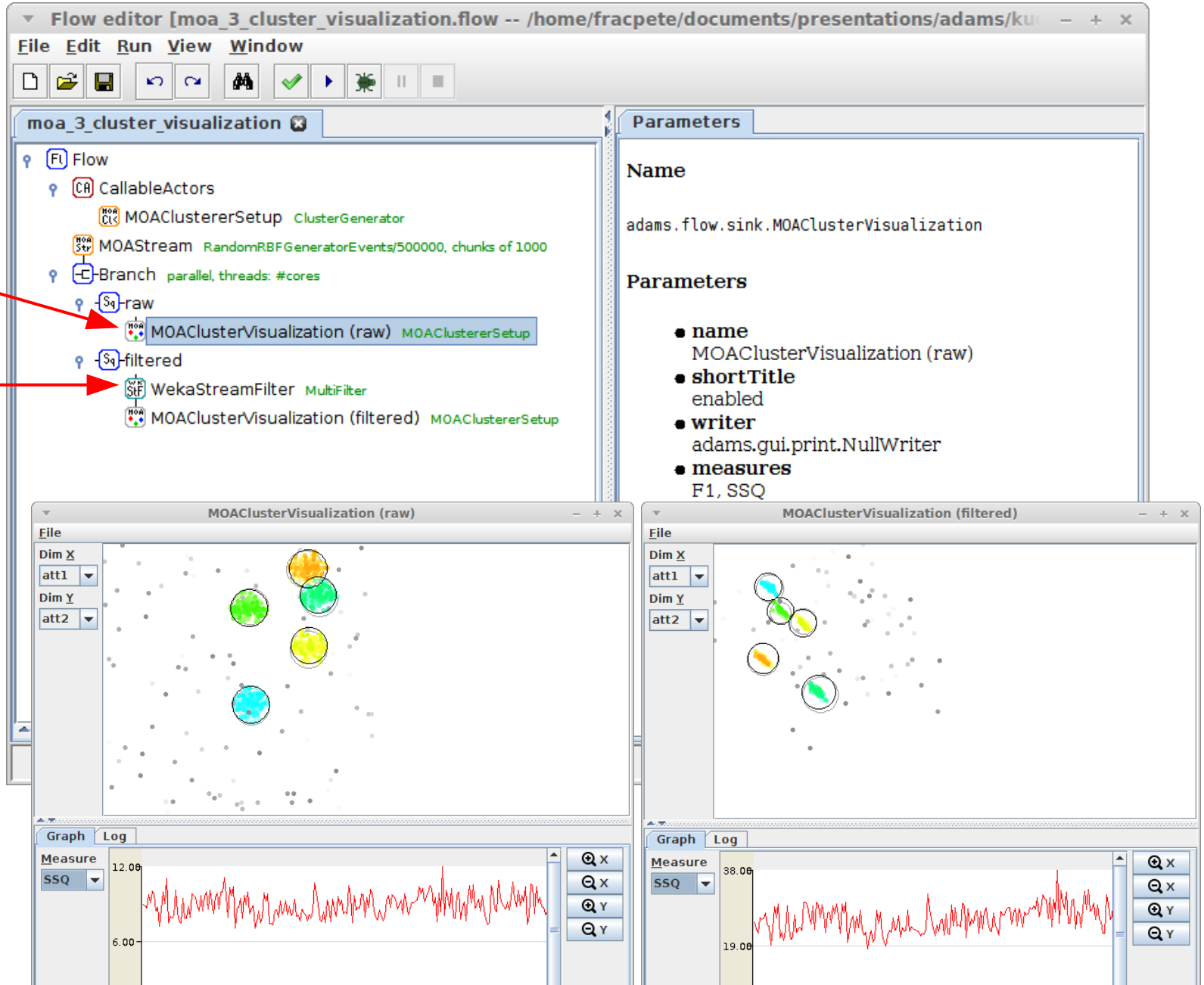
Clustering

- MOA, like WEKA, supports clustering
- MOA also has live clustering visualization
- Different preprocessing changes clustering
- Actors to use
  - WekaStreamFilter
  - WekaClusterVisualization

Clustering

visualize clusters

filter data stream



Questions?

https://adams.cms.waikato.ac.nz/

@TheAdamsFlow