Te Ipu o te Mahara
Artificial Intelligence
Institute
THE UNIVERSITY OF WAIKATO

Taking Weka to the next level with ADAMS

No need to write code

Peter Reutemann

Outline

- What is ADAMS?
- Weka Investigator
- Flow editor
- Other useful tools
- Demo

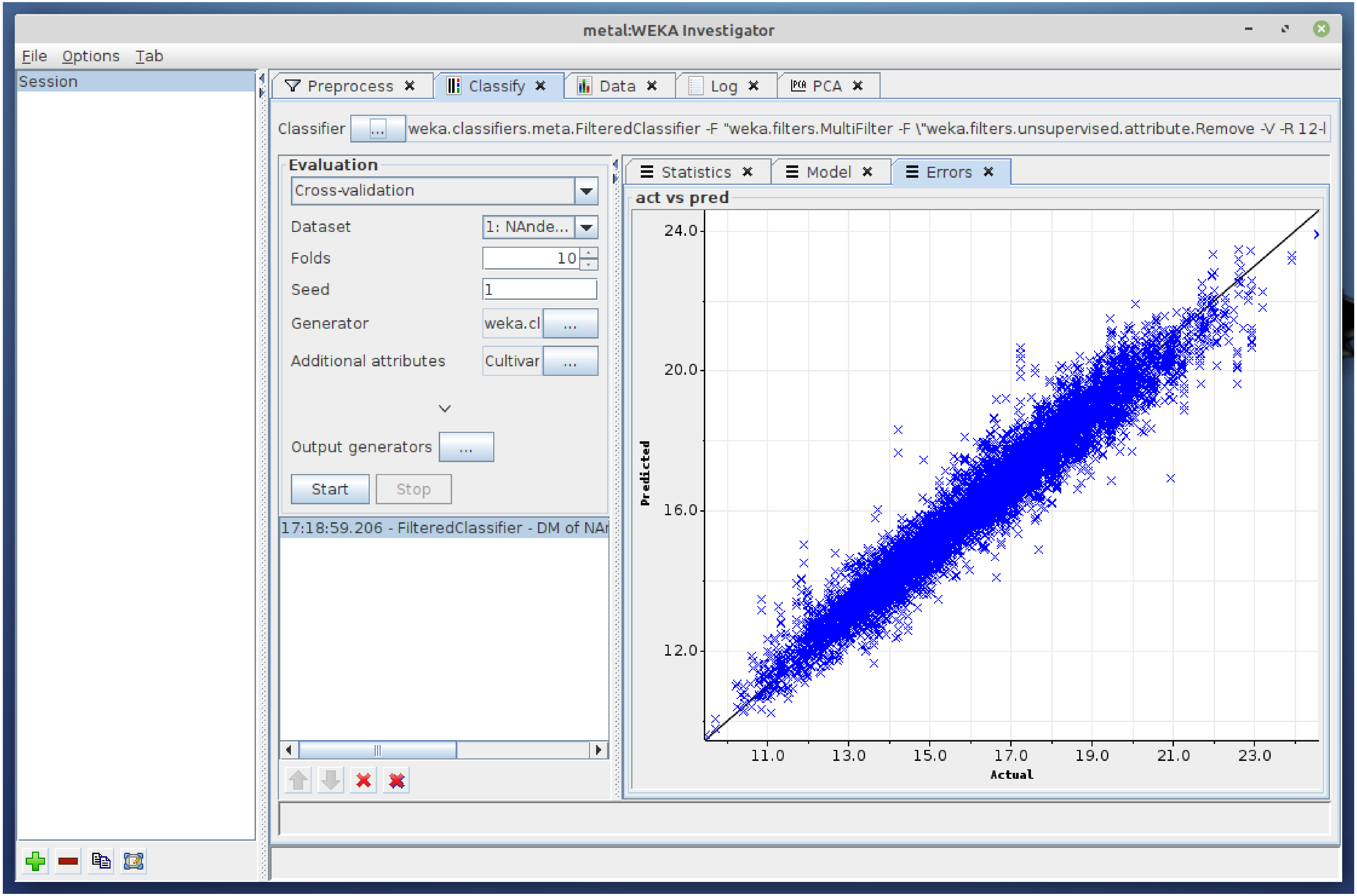
What is ADAMS?

- Java plugin framework (Linux, OSX, Windows)
- Modules with various functionality
  - eg modules for Weka, Meka, MOA
- Interfaces for various tasks
- Workflow engine
- Started in 2009 to work with GC-MS data

Weka Investigator

- Exploration tool like Weka Explorer
- But it supports
  - multiple sessions
  - multiple datasets loaded at the same time
  - batch filtering
  - arbitrary number of tabs
  - predefined output generators
  - different/additional visualizations
  - easy exports of results

Weka Investigator



Classifier ... weka.classifiers.meta.FilteredClassifier -F "weka.filters.MultiFilter -F \"weka.filters.unsupervised.attribute.Remove -V -R 12-|

Evaluation: Cross-validation

Dataset: 1: NAnde...

Folds: 10

Seed: 1

Generator: weka.cl ...

Additional attributes: Cultivar ...

17:18:59.206 - FilteredClassifier - DM of NAr

Flow editor

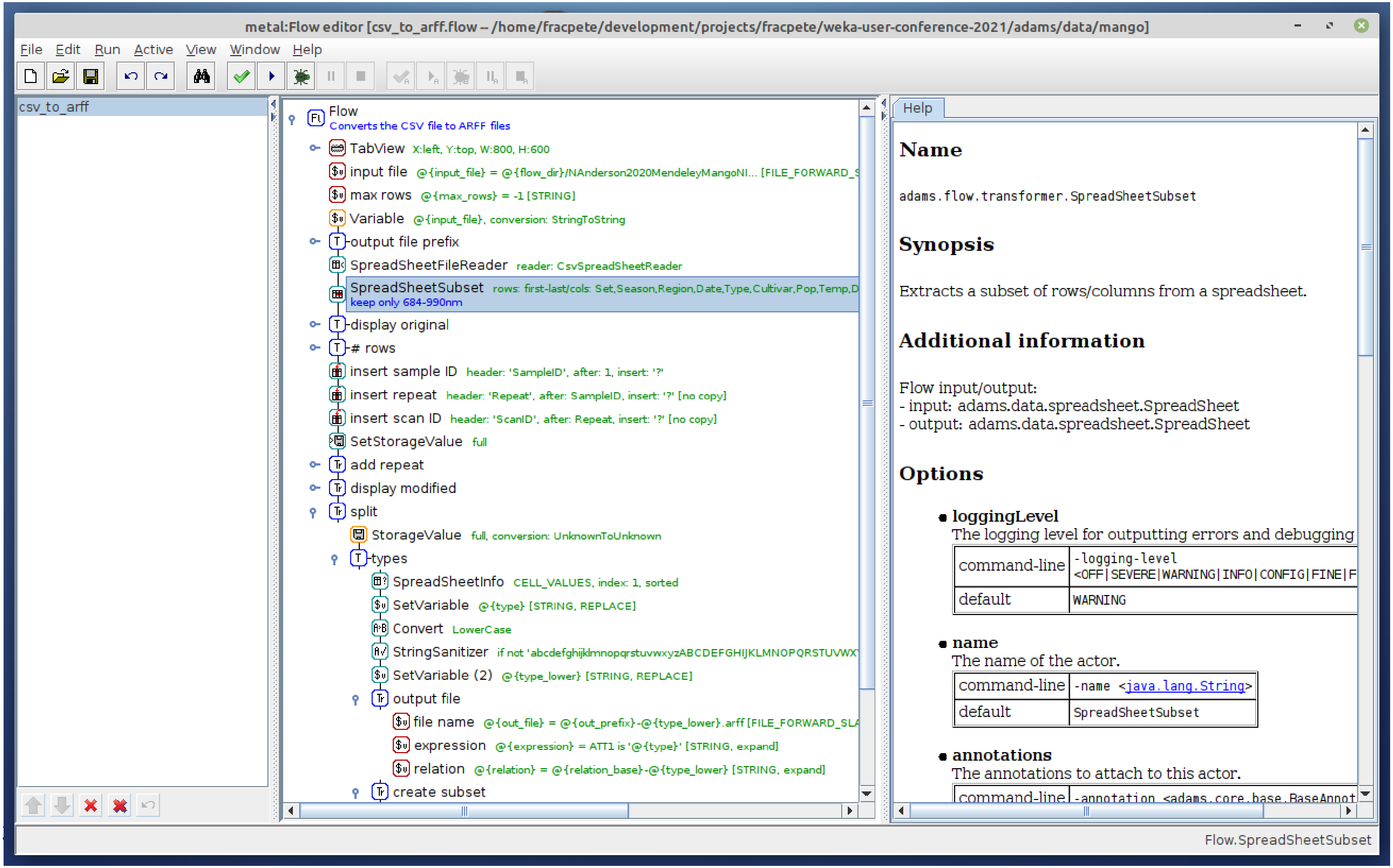
- Why another workflow system?
  - Canvas-based set up is tedious
  - Doesn't scale well (when using 100s or 1000s of operators)
  - KnowledgeFlow too Weka-centric
- How is it different/better?
  - Actors (= operators) arranged in tree (collapsible)
  - Actor handlers nest other actors (eg sequence)
  - Control actors control data flow (eg tee, branch, switch)
  - Input/output defines:

    standalone (no I/O), source (only O), transformer (I/O), sink (only I)
  - Offers debugging (incl. breakpoints and data inspection)

Flow editor

- How does it work?
  - Data-driven system (actor output triggers next one)
  - Event-based triggers available (eg cron, web-services)
- But: Tree only supports 1-to-n connections
- Simulating n-to-m semantics
  - Containers (combine multiple outputs)
  - Variables (change actor options on-the-fly)
  - Internal storage (key-value storage)
  - Callable actors (feed data into named actor)

Flow editor


metal:Flow editor [csv_to_arff.flow -- /home/fracpete/development/projects/fracpete/weka-user-conference-2021/adams/data/mango]
File
Edit
Run
Active
View
Window
Help
csv_to_arff
Help
Flow.SpreadSheetSubset

```
Flow
Converts the CSV file to ARFF files
    TabView  X:left, Y:top, W:800, H:600
    input file  @{input_file} = @{flow_dir}/NAnderson2020MendeleyMangoNI... [FILE_FORWARD_S
    max rows  @{max_rows} = -1 [STRING]
    Variable  @{input_file}, conversion: StringToString
    output file prefix
    SpreadSheetFileReader  reader: CsvSpreadSheetReader
    SpreadSheetSubset  rows: first-last/cols: Set,Season,Region,Date,Type,Cultivar,Pop,Temp,D
    keep only 684-990nm
    display original
    # rows
    insert sample ID  header: 'SampleID', after: 1, insert: '?'
    insert repeat  header: 'Repeat', after: SampleID, insert: '?' [no copy]
    insert scan ID  header: 'ScanID', after: Repeat, insert: '?' [no copy]
    SetStorageValue  full
    add repeat
    display modified
    split
        StorageValue  full, conversion: UnknownToUnknown
        types
            SpreadSheetInfo  CELL_VALUES, index: 1, sorted
            SetVariable  @{type} [STRING, REPLACE]
            Convert  LowerCase
            StringSanitizer  if not 'abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWX
            SetVariable (2)  @{type_lower} [STRING, REPLACE]
            output file
                file name  @{out_file} = @{out_prefix}-@{type_lower}.arff [FILE_FORWARD_SLA
                expression  @{expression} = ATT1 is '@{type}' [STRING, expand]
                relation  @{relation} = @{relation_base}-@{type_lower} [STRING, expand]
            create subset
```

Name

`adams.flow.transformer.SpreadSheetSubset`

Synopsis

Extracts a subset of rows/columns from a spreadsheet.

Additional information

Flow input/output:
- input: adams.data.spreadsheet.SpreadSheet
- output: adams.data.spreadsheet.SpreadSheet

Options

- **loggingLevel**
The logging level for outputting errors and debugging

command-line	`-logging-level`
	`<OFF|SEVERE|WARNING|INFO|CONFIG|FINE|F`

default	`WARNING`

- **name**
The name of the actor.

command-line	`-name <java.lang.String>`
default	`SpreadSheetSubset`

- **annotations**
The annotations to attach to this actor.

command-line	`-annotation <adams.core.base.BaseAnnot`

Flow editor

eurofins | Agro

- Any real world applications?
  - S3000 - commercial tool based on ADAMS to analyze soil/plant samples (NIR, MIR, XRF)
  - User only manages configuration flows, generators create low-level worker flows
  - Eurofins Agro: ~450 models, ~3000 samples/day
  - Largest (production) worker flow generated so far:
    - ~44,000 actors

Other useful tools

- Multi-Experimenter
- Spreadsheet file viewer
- SQL Workbench
- File commander (ftp, smb, sftp)
- Preview browser
- Append/merge datasets
- Check dataset compatibility

Available functionality

- Databases (MySQL, SQLite, PostgreSQL, HSQL, MSSQL, Sybase, MS Access)
- XML/XSLT/XPath, HTML, JSON
- LaTeX (generate/execute), PDF (read/extract/generate)
- spreadsheets (CSV, TSV, Gnumeric, ODF, MS Excel, fixed column, ...)
- webservices (SOAP/REST)
- scripting (Groovy/Jython/Python)
- Weka, Meka, MOA, R-Project, parameter optimization
- Natural language processing (parsing, word clouds)
- OCR, barcodes
- spectral data (AniML, CAL, CML, DPT, EEM, JCampDX, MPS, NIR, Opus, Relab, SPA, SPC, SpecLib, spreadsheet-based)
- scatter/line plots, gnuplot, control charts
- images, heatmaps, audio, video, webcams, ffmpeg
- de-/compression (tar, zip, bzip2, gzip, lzma, xz, zstd)
- remote access (sftp, ssh, scp, rsync)
- Java code generation
- Generate custom ADAMS application
- Docker images and image generation
- frontend (user interaction), backend (Linux daemon/Windows Service)

Demo

Enough talking, let's see ADAMS in action!

Questions?

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

https://github.com/fracpete/weka-user-conference-2021