



MLOSS Workshop - NIPS 2008

Experiment Databases for Machine Learning

Joaquin Vanschoren
Hendrik Blockeel

Machine learning experiments

- Summarized in papers
- Individual experiments and experiment details lost

Share experiments

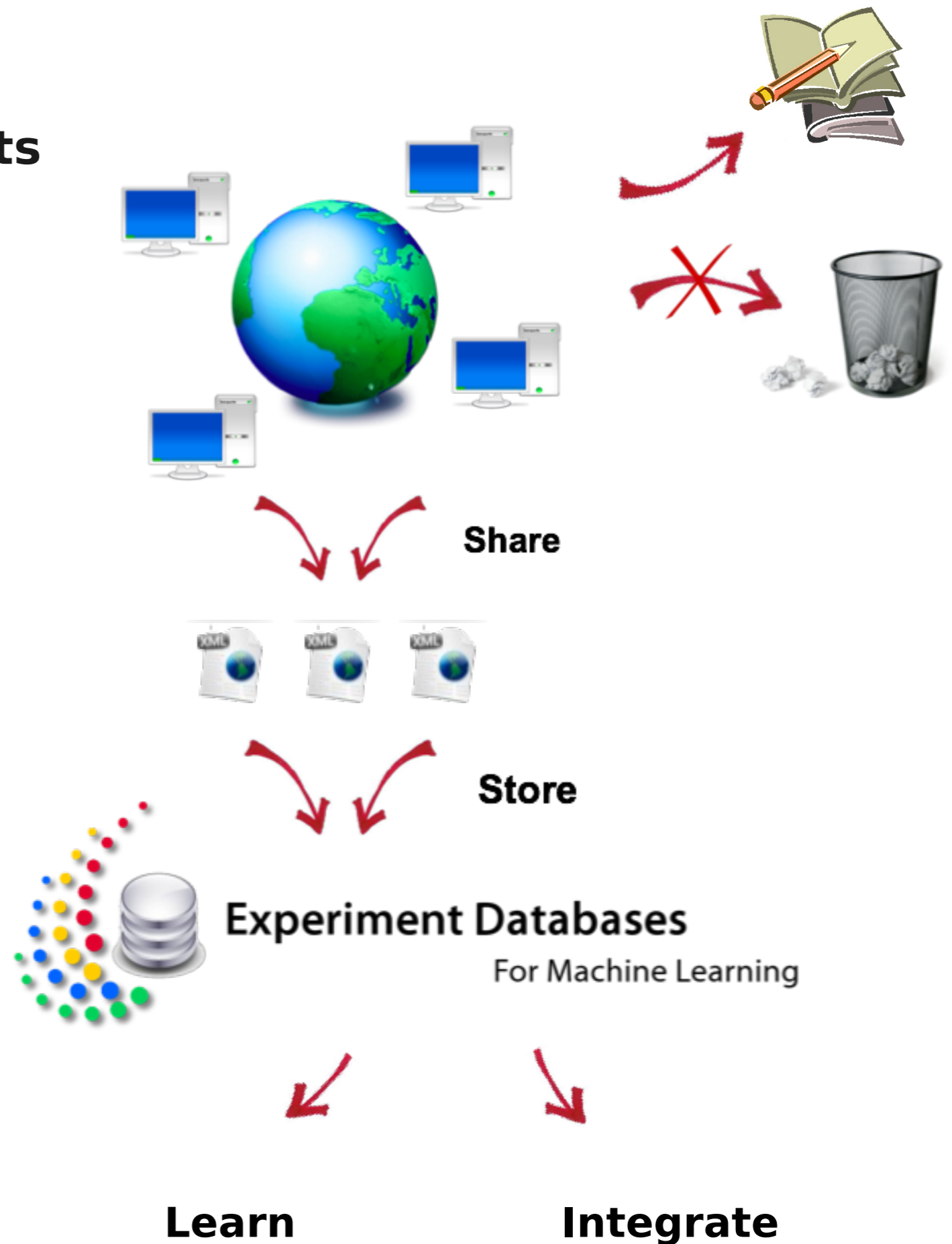
- Common description language
- All details to ensure repeatability
- ExpML: a first attempt

Store experiments

- All information organized
- Ask any question by writing query (SQL)

Reuse information

- Learn from the past
- Save time/resources



Why share experiments?

- In machine learning research:

Good science
Reproducibility

Organization

'Map' of known approaches
Also: negative results

Visibility
Algorithms pop up
in searches



Save time & energy:
Re-use experiments
(e.g., benchmarking)



New possibilities:

Larger, more
generalizable studies



ExpML: An Example

- Definitions:
 - Enforce repeatability
 - Theoretical information

```
<definition >
  <algorithm name="Bagging" version="1.31.2.2" libname="weka"
    libversion="3.4.8" url="http://www.cs.waikato.ac.nz/ml/weka/"
    classpath="weka.classifiers.meta.Bagging">
    <parameter name="P">
      <description>Size of each bag as percentage of data set size </description>
      <default>100</default>
      <property name="suggested_min" value="20"/>
      ...
    </parameter>
    ...
    <property name="class" value="ensemble">
    <property name="handles_classification" value="true">
    ...
  </algorithm >
</definition >
```

ExpML: An Example (2)

- Complex experiment setups:
 - Algorithms + (meta)parameters
 - Datasets + preprocessors

```
<experiment>
  <setting>
    <algorithm name="Bagging" version="1.31.2.2" libname="weka">
      <parameter name="P" value="90"/>
      <parameter name="O" value="false"/>
      <parameter name="I" value="40"/>
      <parameter name="W" value="algorithm">
        <algorithm name="NaiveBayes" version="1.16" libname="weka"/>
      </parameter>
    </algorithm>
    <dataset name="pendigits-90%">
      <preprocessor name="RemovePercentage" version="1.3" libname="weka">
        <parameter name="P" value="10"/>
        <dataset name="pendigits" url="http://archive.ics.uci.edu/ml/">
          <classIndex>-1</classIndex>
        </dataset>
      </preprocessor>
      <classIndex>-1</classIndex>
    </dataset>
    <evalmethod name="CrossValidation" version="1.53" libname="weka"
      libversion="3.4.8">
      <parameter name="nbfolds" value="10"/>
      <parameter name="randomseed" value="1"/>
    </evalmethod>
    <environment>machine14 </environment>
  </setting>
```

ExpML: An Example (3)

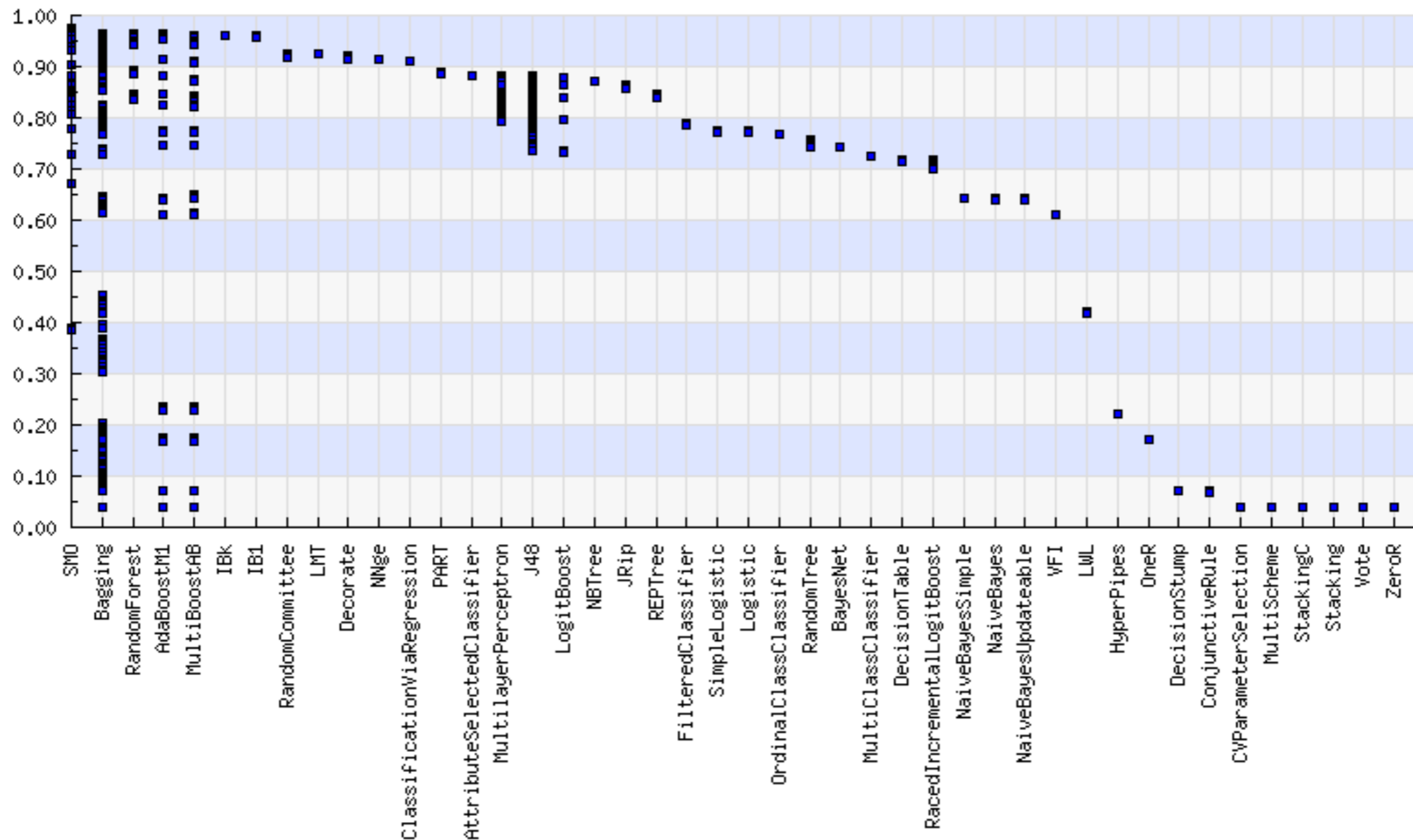
- Experiment results:
 - Evaluations
 - Predictions

```
<evaluation>
  <metric name="build_cputime" value="5.67"/>
  <metric name="build_memory" value="17929416"/>
  <metric name="mean_absolute_error" value="0.030570337062541805"/>
  <metric name="root_mean_squared_error" value="0.15960607792291556"/>
  <metric name="predictive_accuracy" value="0.8570778748180494"/>
  <metric name="kappa" value="0.8411692914743762"/>
  <metric name="confusion_matrix" value="
    [[0,1,2,3,4,5,6,7,8,9],[1021,0,0,0,2,0,3,0,51,4],[1,883,84,58,...],...]"/>
  <metric name="precision_array" value="
    [[0,1,2,3,4,5,6,7,8,9],[0.94449586,0.7132472,0.8305752,0.86625874,...]]"/>
  ...
</evaluation>
<predictions target="0">
  <instance nr="00000" prediction="8">
    <prob prediction="0" value="1.8761967426234115E-5"/>
    ...
    <prob prediction="8" value="0.9991914442703987"/>
    <prob prediction="9" value="3.2190267582597184E-31"/>
  </instance>
  ...

```

“Learning” from stored Experiments

- UCI dataset “letter”
- Accuracy of all stored classifiers



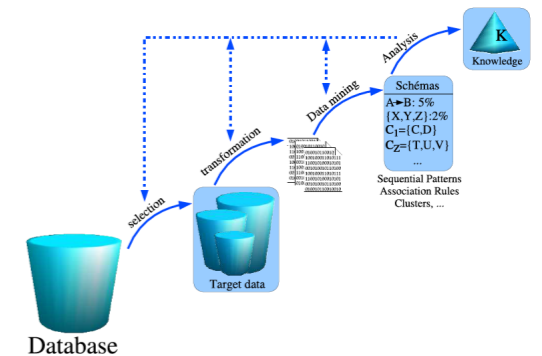
Integration in tools



Experimentation tools
skip known experiments



DM/ML toolbenches
share experiments



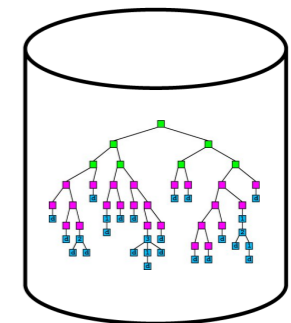
DM assistance tools
advise, e.g., workflows



Parallelization
ML experiment grids



Molecule databases
organize modelling efforts



Inductive databases
query model properties

Efharisto

Danke

Thanks y'all

Xie Xie

Dank U

Toda

Merci

Grazie

Spasiba

Thanks

Gracias

Arigato

Köszönöm

Tesekkurler

Obrigado

Dhanyavaad

Hvala



<http://expdb.cs.kuleuven.be>