Execution of R Built Predictive Solutions

Alex Guazzelli, PhD
VP, Analytics - Zementis, Inc.

useR! 2010
Exporting Models from R

Why?

- Memory
- Speed
- Freedom
- Transparency
- Interoperability
- Accessibility
- Because you can!
Exporting Models from R

How?

PMML
PMML

Predictive Model Markup Language (PMML)

- PMML is an XML-based language to
  - Define data mining models
  - Share models between compliant applications
- Standard for exchange of models to
  - Avoid proprietary issues and incompatibilities
  - Easily put models to work
- Clear separation of tasks
  - Model development vs. model execution
  - Scientists focus on building the best model
  - Eliminates need for custom model deployment
PMML defines a standard not only to represent data-mining models, but also data handling and data transformations (pre- and post-processing).

- A **Data Dictionary** defines all the raw data fields (including missing value strategy and outlier treatment).

- Several **Data Transformations** strategies allow for intelligent extraction of feature detectors from raw data (“data massaging”).

- A comprehensive list of **Data-Mining Models** offers power and flexibility.

- Post-processing of results allow for tailored decisions.

- Model Explanation allows for performance evaluation.
Matured and Supported by Industry

- Data Mining Group [http://www.dmg.org](http://www.dmg.org)
- Vendor independent consortium
- Mature standard
  - Current version 4.0
  - Active group and constant enhancements
- Industry supporters
  - Major Players: IBM/SPSS, Oracle, SAP, Microsoft
  - Analytics: KXEN, SAS, Salford, Togaware, Zementis
  - BI: Microstrategy, Teradata, Tibco, Pentaho
  - Open Source: R, KNIME, Rapid-I
  - Others: Equifax, FICO, Open Data Group, Visa, Pervasive, NASA
Using the PMML package to export a Neural Network model.

```r
> library(nnet)
> Iris <- read.csv("Iris.csv")
> IrisNet <- nnet(CLASS~., data=Iris, size=4)
# weights:  35
initial  value 171.675942
iter  10 value 43.796331
iter  20 value 0.062310
iter  30 value 0.001175
final  value 0.000038
converged
> library(pmml)
Loading required package: XML
> pmml(IrisNet)
```
Model is readily exported in PMML and ready to be used.

```xml
<PMML version="3.2" xmlns="http://www.dmg.org/PMML-3_2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.dmg.org/PMML-3_2 http://www.dmg.org/PMML-3_2/PMML-3_2.xsd">
  <Header copyright="Copyright (c) 2008 Alex Guazzelli" description="Neural Network PMML Model">
    <Extension name="timestamp" value="2008-10-23 17:45:45" extender="Rattle"/>
    <Extension name="description" value="Alex Guazzelli" extender="Rattle"/>
    <Application name="Rattle/PMML" version="1.1.9"/>
  </Header>
  <DataDictionary numberOfFields="5">
    <DataField name="CLASS" optype="categorical" dataType="string">
      <Value value="Iris-setosa"/>
      <Value value="Iris-versicolor"/>
      <Value value="Iris-virginica"/>
    </DataField>
    <DataField name="SEPAL_LE" optype="continuous" dataType="double"/>
    <DataField name="SEPAL_WI" optype="continuous" dataType="double"/>
    <DataField name="PETAL_LE" optype="continuous" dataType="double"/>
    <DataField name="PETAL_WI" optype="continuous" dataType="double"/>
  </DataDictionary>
  <NeuralNetwork modelName="NeuralNet_model" functionName="classification" numberOfLayers="2">
    <MiningSchema>
      <MiningField name="CLASS" usageType="predicted"/>
      <MiningField name="SEPAL_LE" usageType="active"/>
      <MiningField name="SEPAL_WI" usageType="active"/>
      <MiningField name="PETAL_LE" usageType="active"/>
      <MiningField name="PETAL_WI" usageType="active"/>
    </MiningSchema>
    <NeuralInputs numberOfInputs="4">
      <NeuralInput id="1">
        <DerivedField name="derivedNI_SEPAL_LE" optype="continuous" dataType="double">
          <FieldRef field="SEPAL_LE"/>
        </DerivedField>
      </NeuralInput>
    </NeuralInputs>
  </NeuralNetwork>
</PMML>
```
From R to PMML
Supported Packages/Objects

- nnet: Neural Networks
- hclust: Clustering
- rpart: Decision Trees
- arules: Association Rules
- lm/glm: Regression
- kmeans: Clustering
- ksvm: SVMs
- randomForest
Got Models…

- Data Analysis
- Statistical Model
- PMML Export

What Now?
ADAPA

- **ADAPA by Zementis**
  - Predictive Decisioning Platform
  - PMML-based
  - Drools to integrate business logic
  - Scalable execution platform
  - Real-time integration into business processes
  - Accessible from anywhere
  - Not a model development environment

- **ADAPA on Amazon Elastic Compute Cloud**
  - Software as a service
  - Up/Down scaling as needed
  - Pay-as-you-go
    - Amazon Payments ($0.99 per hour)
  - Amazon experience & reliability
From Model Building to Model Deployment

Model Building

- Pervasive
- IBM
- KNIME
- R
- Salford Systems
- SAS
- TIBCO
- RapidMiner
- SPSS
- Microsoft SQL Server
- KXEN
- MicroStrategy
- Statistica

Model Deployment

- PMML (Predictive Model Markup Language)
- ADAPA
- Amazon Web Services
Model Execution

ADAPA

Web Console

Web Services

Excel Add-in
Model Execution via iPhone
Zementis Contributions

• **ADAPA**: A decision engine that deploys models expressed in PMML and executes them in real-time. Available for on-site and cloud deployments.

• **Excel Add-in**: Allows for scoring in ADAPA directly from within Excel.

• **PMML Converter**: Validates, converts, and corrects old and new PMML code. Available at the DMG website and at [http://www.zementis.com/pmml.htm](http://www.zementis.com/pmml.htm).

• **Contributing Member of the DMG**: Submitted several proposals for PMML 4.0 and already working with other members on PMML 4.1.

• Code contributor for the **R PMML package** (available on CRAN).


• **PMML Book**: Available on Amazon.com.

• **PMML Blogs**: Several blogs on PMML topics ([http://adapasupport.zementis.com](http://adapasupport.zementis.com) and [http://www.predictive-analytics.info](http://www.predictive-analytics.info)).