PMML Execution of R Built Predictive Solutions

Alex Guazzelli[1*] , Kostantinos Stathatos[1] , Michael Zeller[1]

1. Zementis, Inc.
* Contact author: Alex.Guazzelli@zementis.com

Keywords: PMML, Model Deployment, ADAPA, Predictive Solutions, Predictive Applications

The rule in the past was that whenever a model was built in a particular development environment, it remained in that environment forever, unless it was manually recoded to work somewhere else. This rule has been shattered with the advent of *PMML* (Predictive Modeling Markup Language). Defined as an XML-based language used to represent predictive data mining models, it was specified by the Data Mining Group, an independent group of leading technology companies. By providing a uniform standard to represent predictive models, *PMML* allows for the exchange of predictive solutions between different applications and various vendors. The *R PMML* package, which is currently available through CRAN (the Comprehensive *R* Archive Network), exports *PMML* for a variety of modeling techniques which include: neural network models, support vector machines, decision trees, regression models, association rules and clustering models. Besides *R*, many statistical tools also support the standard; these include, for example, tools from *KNIME, SAS, IBM/SPSS*, and *TIBCO*.

Once exported as *PMML* files, models are readily available for deployment into an execution engine for scoring or classification. *ADAPA* is one example of such an engine. It takes in models expressed in *PMML* and transforms them into web-services. Models can be executed either remotely by using web-services calls, or via a web console. Users can also use an Excel add-in to score data from inside Excel using models built in *R*.

R models have been exported into *PMML* and uploaded in *ADAPA* for many different purposes. Use cases where clients have used the flexibility of *R* to develop and the *PMML* standard combined with *ADAPA* to deploy range from financial applications (e.g., risk, compliance, fraud) to energy applications for the smart grid. The ability to easily transition solutions developed in *R* to the operational IT production environment helps eliminate the traditional limitations of *R*, e.g. performance for high volume or real-time transactional systems and memory constraints associated with large data sets.

References

A. Guazzelli, K. Stathatos, and M. Zeller (2009). Efficient Deployment of Predictive Analytics through Open Standards and Cloud Computing. *SIGKDD Explorations Newsletter*, 11/1, 32-38.

A. Guazzelli, M. Zeller, W. C. Lin, G. Williams (2009). PMML: An Open Standard for Sharing Models. *The R Journal*, 1/1, 60-65.

Data Mining Group (2009). *PMML version 3.2*, http://www.dmg.org/pmml-v3-2.html.

G. Williams, M. Harshler, A. Guazzelli, M. Zeller, W. Lin, H. Ishwaran, U. B. Kogalur, and R. Guha. (2009). *PMML: Generate PMML for various models*. http://rattle.togaware.com/. R package version 1.2.7.

R. Pechter (2009). What's PMML and What's New in PMML 4.0? *SIGKDD Explorations Newsletter*, 11/1, 19-25.