Massive Predictive Modeling

Mark F. Hornick

Oracle
mark.hornick@oracle.com

Keywords: Advanced Analytics, Predictive Modeling, Oracle, Big Data

As enterprises continue to amass data at ever increasing rates and with greater variety – what is being called *big data* – the ability to extract value from that data demands high performance and scalable tools – both in hardware and software. In various industries, enterprise take on *massive predictive modeling* projects, where the goal is to build models, one per customer, to understand behavior and tailor predictions at the customer level. These predictions can then be aggregated to assess future demand. When there are millions of customers, each with their own accumulated data, such as frequent utility meter readings or retail sales, the scale of such projects takes on a new dimension. Massive predictive modeling comes with challenges: effectively partitioning data, storing and managing resulting models, associating models with customers during prediction, as well as backup, recovery, and security.

While *R* has parallel capabilities to facilitate taking advantage of clusters of computers, significant coding is usually required to meet the challenges noted above. In this talk, we present the business problem and illustrate how *Oracle R Enterprise*, one of Oracle's *R* technologies [1], facilitates massive predictive modeling in a pair of succinct *R* scripts. With Oracle R Enterprise, the data, *R* scripts, and models all reside in Oracle Database, which simplifies and speeds production deployment.

References

[1] Oracle R Technologies, http://www.oracle.com/technetwork/database/database-technologies/r/r-technologies/overview/index.html.