

Using DBMS_DATA_MINING With Oracle Database 10g

One of the new features in Oracle 9i was Oracle Data Mining, a data mining engine which allowed data analysts and application developers to perform a range of data mining algorithms on data held in the Oracle database. Oracle 9i came with a number of mining algorithms, such as Adaptive Bayes Networks, Clustering and Association Rules, together with a Java API to allow ODM functions to be included in Java applications.

Whilst this was useful for Java programmers, it wasn't all that relevant for PL/SQL programmers and to remedy this, Oracle Database 10g comes with a new package called DBMS_DATA_MINING that provides PL/SQL access to the data mining engine.

Like the Java API, DBMS_DATA_MINING allows you to build a data mining model, test it and then apply the model to provide scores or predictive information for an application. One of the key differentiators for Oracle Data Mining is that mining models can be applied directly to data in the database - there's no need to extract the data and then separately load it into the mining engine, meaning that data mining can be now carried out in 'real time'.

The Oracle Data Mining engine can be pointed at any schema in the database, and if the data needs processing beforehand (to place continuous and discrete values into range 'bins') there's also a new accompanying package, DBMS_DATA_MINING_TRANSFORM to carry this out automatically.

In addition to the new PL/SQL interface, Oracle Data Mining in Oracle Database 10g has the following additional enhancements;

- Features extraction using Non-Negative Matrix Factorization Algorithm
- Enhanced Data Preprocessing
- Enhanced Adaptive Bayes Network
- Extension of standard Oracle database security to Oracle Data Mining user data and mining results
- DM4J, the add-in to JDeveloper that allows the graphical building of mining models
- Support for Support Vector Machines