Implementing Data Mining Algorithms With Microsoft SQL Server

Abstract

Implementing data mining algorithms with Microsoft SQL Server C. L. Curotto1 & N. F. F. Ebecken2 1 CESEC/UFPR–Civil Engineering Graduate Program, Brazil 2 COPPE/UFRJ–Civil Engineering Graduate Program, Brazil Abstract The OLE DB for DM (Microsoft's object-based technology for sharing information and services across process and machine boundaries focused on database mining applications) specification provides an industry standard for implementation of data mining algorithms aggregated with Microsoft SQL Server 2000. The Simple Naive Bayes classifier is implemented using the OLE DB for DM Resource Kit. Numeric input attributes, multiple prediction trees and incremental classification are considered. All necessary steps to implement this algorithm are explained and discussed. Some results are shown to illustrate the capabilities of the implementation. 1 Introduction Nowadays database system managers like MS (Microsoft) SQL (Standard Query Language) Server [1] are available, with resources for manipulation of terabytes of data with parallel processing of queries (with multiprocessor servers) using microcomputers [2]. This situation suggests the integration of DM technology by using database managers to enlarge the scope of this technology at a low cost. This approach of integration, achieved by tightly coupling DM and OLAP (On-Line Analytical Processing) techniques in database application development environments, is matter of current interest. It has been discussed in conferences such as ICDM’98 (First International Conference on DM), happened on September 1998 in Rio de Janeiro – Brazil, ICDM’00 (Second International Conference on DM), happened on July 2000, Cambridge – UK and more recent ones. Agrawal [3] presented a methodology for tightly coupling of DM application to relational database system - IBM DB2/CS – based on utilization of user

Keywords
SQL Server data mining offers Data Mining Add-ins for office 2007 that allows discovering the patterns and relationships of the data..... - All of the Microsoft data mining algorithms can be customized and are fully programmable using the provided APIs, or by using the data mining components in SQL Server Integration Services. - Third-party algorithms can also be used to comply with the OLE DB for Data Mining specification, or develop custom algorithms that can be registered as services and then used within the SQL Server Data Mining framework. Next Page ». SQL server - Explain how to use DMX the data mining query language. Data mining extension is based on the syntax of SQL. Data Mining Algorithms for Directed/Supervised Data Mining Tasks—linear regression models are the most common data mining algorithms for estimation data mining tasks. Of course, linear regression is a very well known and familiar technique. A number of data mining algorithms can be used for classification data mining tasks including logistic regression, decision trees, neural networks, memory based reasoning (k-nearest neighbor), and Naïve Bayes. 2. Data Mining Tasks without a Target or Dependent Variable. SQL Server Data Tools uses Microsoft Visual Studio (VS) as the Integrated Development Environment (IDE) which will be familiar to VB.NET or C# users. When VS opens, most likely the top will include the menu and tool bar with the Start Page tab active. The next hues show the syntax of SQL for DM command to populate the mining model using data from MS SQL Server. 80 Data Mining III For debugging purposes and for use in client applications DM SQL queries, can be made to retrieve the rowset contents of the DM model. Figure 3 brings out Kim’s [24] utility DMSamp showing part of the results of the query that retrieves the entire contents of the test model.
Why is this happening to me?