

Data Mining with SQL Server Analysis Services and Excel 2007

Andrew J. Brust
Chief, New Technology
twentysix New York

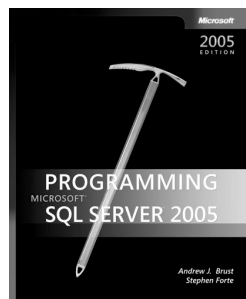
Pre-requisites for this presentation:

- 1) Basic Understanding of Relational Databases
 - 2) Some C# knowledge
 - 3) Familiarity with Excel
- Level: *Intermediate*

Who Am I?

- Microsoft Regional Director, NY and NJ, MVP for VB, INETA Speaker's Bureau member
- Member, Microsoft BI Partner Advisory Council
- Co-moderator, NYC .NET Developers Group
– <http://www.nycdotnetdev.com>
- Regular Speaker @ VSLive!
- Contributing Editor, *Visual Studio Magazine*
- Vice Chair, NY Software Industry Association
- Chief, New Technology 
 - Microsoft Gold Certified Partner
 - BI, Integration, Portals, Custom .NET development

Read All About it in Chapter 20 (by Elsie Pan)...



What is Data Mining?

- SQL Server Analysis Services consists of:
 - OLAP: for dimensional/drill-down analysis
 - Data Mining: for discovery of trends, patterns in data
- So data mining involves active observation, not mere aggregation
- Employs "machine learning"
 - Statistics, e.g. regression
 - Artificial Intelligence, e.g. neural networks

Vocabulary and Concepts

- Training data, test data, new data
- Input variables
- Predictable variables
- Continuous vs. discrete data
- Mining Structures
- Mining Models
- Prediction queries
- DMX and XMLA

Algorithms

	Cross-Selling	Targeted Marketing	Fraud Detection	Customer Profiling	Forecasting	Data Validation
Association Rules	✓					
Clustering		✓	✓	✓		✓
Decision Trees		✓		✓		
Linear Regression				✓		
Logistic Regression				✓		
Naïve Bayes		✓		✓		
Neural Nets				✓		
Sequence Clustering				✓		
Time Series					✓	

Data Mining: Preparatory Steps

- Identify data source
 - Can be any relational data source
 - View(s) are often required
 - Data exploration usually helpful
- Determine predictable, input variables
- Partition data
 - Training data (required; used to build model)
 - Test data (optional; used to verify model and determine accuracy)

Data Mining and OLAP

- Mining models can be built on cubes too
- Can then create a dimension from model
 - Clustering and association algorithms work best
- Typically OLAP cubes *not* used as source data for mining models
- When processing a model, you will see that pseudo-cubes, dimensions and measure groups are created for them. Ignore this.

Data Mining with BI Dev Studio (VS)

- Create Integration Services project to partition data
- Create Analysis Services project
 - Or use existing (OLAP) project
- Add data source and data source view
 - Point to training data
 - Explore data with OWC PivotTable and charts
 - Add calculated columns and named queries
- Add mining structure
- Add mining model(s)
- Process/train model(s)
- Browse and test model(s)
- Design prediction queries

Data Mining with SQL Server Management Studio

- Use same viewers/designers available in BIDS/VS
 - Browse
 - Test: lift & profit charts; classification matrices
 - Query
- Process/Train
- Script
 - DMX scripts for structure/model creation, training and (prediction) querying
 - Use Template Explorer to learn DMX
 - Use Script xxx As... to generate XMLA scripts

Data Mining with Reporting Services

- Since DMX is just SQL, RS can be used for a variety of data mining tasks:
 - Classification matrices
 - Prediction queries

Data Mining with .NET

- Embed model browser controls
- Use AdomdClient provider to:
 - Discover structures and models in a database
 - Bring back prediction query data as DataReaders/DataSets; use data binding to display
 - Execute model content queries
 - Execute any DMX query
- Use AdomdServer provider to:
 - Create DMX stored procedures
- Use AMO to:
 - Create and maintain structures and models
- Use ASP.NET viewer control sample code

Control Freak

- WinForms Model Viewer Controls
 - Right click Toolbox and select Choose Items...
 - Assembly *not* in GAC, but included with BIDS:

C:\Program Files\Microsoft Visual Studio 8\Common7\IDE\PrivateAssemblies\Microsoft.AnalysisServices.Viewers.DLL
- Can also get controls from Feature Pack for SQL Server 2005. Get the right version:
 - RTM: November 2005; SP1: April 2006
 - SP2: February 2007
 - <http://tinyurl.com/yumpyp>
 - Must download and install AS 9.0 OLE DB provider and ADO MD.NET (also part of Feature Pack) before installing DM viewer controls

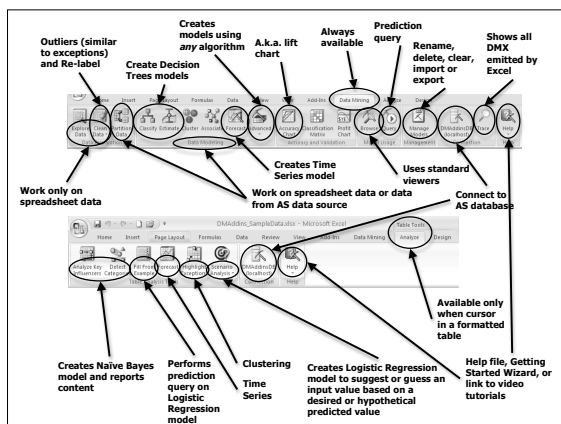
Frame of Reference

- AdomdClient (a.k.a. ADO MD.NET)
 - Microsoft.AnalysisServices.AdomdClient
 - Typically added to WinForms or ASP.NET project
- AdomdServer
 - Microsoft.AnalysisServices.AdomdServer
 - Typically added to Class Library project
- Analysis Management Objects (AMO)
 - Microsoft.AnalysisServices
 - Typically added to WinForms, ASP.NET or Class Library project

Data Mining Add-Ins for Excel

- Released concurrently with SQL 2005 SP2
 - But work with SP1 and RTM servers
- Provide full Data Mining client
 - Create, Browse, Test, Query and Manage structures and models
 - Data Mining ribbon
- Provide DM-powered tools for table analysis
 - Use "disposable" SQL 2005 AS models to analyze and clean data, add forecasts and predictions, and perform what-if and goal-seek analysis
 - Table Tools/Analyze ribbon
- Get it at <http://tinyurl.com/28xpne>

Ribbons



DM Add-Ins: Fast Facts

- Wizard-driven model creation
 - Allow input/predictable variable selection
 - Optionally enable drill-through
 - Optionally browse model and copy diagrams to spreadsheet as bitmaps
- Produce great-looking reports
- Integrate with Excel filtering, sorting, charting
- Almost completely replace BIDS and SSMS
 - Cannot create AS databases
- Offer option of "session models"

Prediction Queries

- Also Wizard-driven
- Select model
- Specify column mappings (model to sheet or data source columns)
- DMX query editor
 - Includes templates
 - Hyperlinks and dialogs for each query clause
- Specify outputs
- Output to appended column, new sheet or existing sheet

Other Items Worth of Mention

- Excel 2007 row & column limits improved
- Nested table issues
- Visio templates

Resources

- *Programming Microsoft SQL Server 2005*
 - (Microsoft Press, 2006) Brust and Forte; Chapter 20 by Elsie Pan
 - <http://tinyurl.com/ytmfpc>
- *Data Mining with SQL Server 2005*
 - (Wiley, 2005) Tang and MacLennan
 - <http://tinyurl.com/2d8zp6>
- SQL Server Data Mining site:
 - <http://www.sqlserverdatamining.com>
- Excel 2007 DM Video tutorials, Webcast:
 - <http://tinyurl.com/yvzfta>

Questions?

- andrew.brust@26ny.com
- <http://www.brustblog.com>