Targeted Drug Discovery through Data Mining

John Brocklebank, PhD
Development Director for Enterprise Miner

Wayne Thompson, PhD
Product Strategist
Data Mining
Introduction

- This presentation will provide an overview of how Enterprise Miner can be used to explore and analyze pharmaceutical data

- Process flow examples include:
  - Diagnostic testing through gene expression profiling
  - Detecting drop outs in a clinical trial
  - Adverse events prediction in a clinical trial
  - Association discovery to determine symptoms that commonly occur together
SAS Defines Data Mining

As the *process* of *selecting, exploring, and modeling* large amounts of data to uncover previously unknown information for a *business advantage*.
Data Mining Tools

Visualization

Clustering

Market Basket Analysis

Neural Networks

Decision Trees

Regression
Two Passages into the Data Mine

Find Trends, Patterns and Relationships
Enterprise Miner Process: SEMMA

1. **Sample**
   - Input Data Source
   - Sampling
   - Data Partition

2. **Explore**
   - Distribution Explorer
   - Multiplot
   - Insight
   - Association
   - Variable Selection

3. **Modify**
   - Data Set Attributes
   - Transform Variables
   - Filter Outliers
   - Replacement
   - Clustering
   - SOM/Kohonen

4. **Model**
   - Regression
   - Tree
   - Neural Network
   - User Defined Mode
   - Ensemble

5. **Assess**
   - Assessment

---

- create one or more representative data tables
- search for anticipated relationships, unanticipated trends, and anomalies in order to gain understanding and ideas.
- create, select, and transform the variables to focus the model selection process.
- use the analytical tools to search for a combination of the data that reliably predicts a desired outcome.
- evaluate the usefulness and reliability of the findings from the data mining process.
Worldwide Statistical and Data Mining Revenue Share for 1999

SAS statistical and data mining revenues for 1999 were $209.7 million, a 21.2% growth over 1998.

Source: 2000 IDC Market Share report for Statistic and Data Mining Software Vendors
“the most successful data mining tool vendor in the world…”
*Data Mining News, March 29, 1999*

SAS Institute, with Enterprise Miner, identified as data mining “Winner”
*Forrester Research, February 1999*

SAS Institute leads in installation share and mind share
*Meta Group, 1999 Data Warehouse Marketing Trends/Opportunities*
The 1999 Magic Quadrant on Data Mining Workbenches

The resulting Magic Quadrant

Challengers

Leaders

Ability to Execute

Niche Players

Visionaries

Completeness of Vision

Silicon Graphics

SAS

IBM

SPSS

Angoss

Oracle

As of 8/99

Source: GartnerGroup
“Mine Your Business”

- Credit Scoring
- Preventive Maintenance
- Capacity Planning
- Customer Retention
- Churn Analysis
- Profiling/Segmentation
- Cross-Selling
- Customer Acquisition
- Fraud Detection
- Network Intrusion Detection

How can data mining be applied in pharmaceutical research?
Enterprise Miner is Used to Help Better Market Drugs

• A pharmaceutical company can analyze its recent sales force activity and their results to improve targeting of high-value physicians and determine which marketing activities will have the greatest impact in the next few months.

• The data needs to include competitor market activity as well as information about the local health care systems.

• The results can be distributed to the sales force via a wide-area network that enables the representatives to review the recommendations from the perspective of the key attributes in the decision process.

• The ongoing, dynamic analysis of the data warehouse allows best practices from throughout the organization to be applied in specific sales situations.
Customer Retention

IMS America

*Large Pharmaceutical Research Company*

Collects 65 gigabytes per month from 175,000 sites.

Detects Brand Loyalty and Brand Product Trends.
R&D Spending

Source PhRMA Annual Survey ’98
Total Number of New Drugs Approved

Source: FDA
Faster Drug Discovery

Tremendous volumes of chemical compounds and biological data are being collected.

Resources are often lacking but computing powering is increasing tremendously.

Chemical and Biological Data

healthcare products
Let's look at some areas where EM can help speed the drug pipeline.
Genome Mining

Gene Expression Profiling

5'-AATTAGCCTTAGCC
AATTGAGCCTTAGCC

SNP Linkage Analysis

SAS Enterprise Miner

Self-Organizing Map Results

Linkage Disequilibrium Plot

Population-Targeted Clinical Trials
SNP Linkage Analysis

- Due to the vast number of SNPs that can be identified, data mining may prove to be advantageous in “sifting” through the SNP data to map human diseases.

- Enterprise Miner has already been used to analyze SNP data in the APOE region to pinpoint the markers associated with Alzheimer's.
Recent advances in microarray technologies allow for the investigation of 1000’s of genes simultaneously.

Through data mining, we can identify those genes that have similar patterns of expression relative to a disease status.
Classification of Acute Leukemia using Gene Expression Profiling

- Acute lymphoblastic leukemia and acute myeloid leukemia tumors are nearly identical.
- Distinguishing the two cancers is critical for successful treatment.
- We used Enterprise Miner to identify the genes that clearly differentiate the two cancers -- ensuring successful treatment of patients, and saving money.
Proliferation of Chemoinformatic Data

- Huge investments are being made to collect large chemical compound libraries
- Automated robotic assay, material storage, and retrieval systems are enabling assay rates of 10,000 to 20,000 per week, which may soon surpass 100,000 per day.
- Dredging for Pharmacophores - preliminary compounds on which chemical modifications can be made to develop a potential "lead" compound or drug candidate
- Data Mining can be applied to these large volumes of data to discover interesting classes of biological compounds
Clustering and classification techniques to find groups of molecules that exhibit similar structure-to-activity, structure-to-reactivity, and structure-to-pharmacological behavior relationships.

Clustering compounds into chemical families, e.g., does the compound have a benzene ring, a double bond between atoms of this type.

Determining what particular set of compounds show activity against a particular disease target.

Determining what structural characteristics the compounds share.
Using Enterprise Miner to Explore and

Exploit Drug Discovery Data

Michael S. Lajiness
Pharmacia Upjohn
Kalamazoo, MI
Other Applicable Mining Areas

- Pharmaco Economics - comparing the cost of different drug treatment programs with similar efficacy rates
- Pharamaco Vigilance - finding the adverse events that occur more frequently with specific drugs
- Quality Control - ensuring that our production processes are running in a statistical state of quality control
- Manufacturing - capacity planning to insure that the right quantities of drugs are available at the right times
Excellent Papers from the PharmaSUG 2000 Conference Proceedings

- **Pharmaceutical Application of Enterprise Miner**
  - Thomas Burger, Eli Lilly and Company, Indianapolis, IN
  - Jerry Plaatje, SAS Institute
  - John LaBore, Eli Lilly and Company, Indianapolis, IN
  - James K. Monroe, Eli Lilly and Company, Indianapolis, IN

- **Data Mining and Statistics in a Pharmaceutical Environment**
  - Franky De Cooman, SPS (Europe), Mechelen, Belgium
Conclusion

- Through mining, we can make the discovery process more efficient and effective
- Enterprise Miner 4.0 already provides an excellent framework for drug discovery and general pharmaceutical research
- A genomics warehouse and mining solutions is under development
Questions?

Thank you for your time.
John.Brocklebank@sas.com
Wayne.Thompson@sas.com