



SAS® Enterprise Miner™ Software

Data Mining auch für den Endanwender
Stoßen Sie zum Kern der Informationen vor



Data Mining eröffnet bisher

ungeahnte Möglichkeiten, der Datenflut

in Unternehmen Herr zu werden.

Bislang fehlte es aber an einer Software,

die auf der einen Seite über ein

mächtiges Spektrum

an leistungsfähigen statistischen

Methoden und Algorithmen verfügte

und diese Leistungsfähigkeit auf

der anderen Seite mit

einer intuitiv zu bedienenden

Oberfläche verband. Diese Lücke schließt

der SAS Enterprise Miner.

Mit dem SAS Enterprise Miner können komplexe Analysen schnell und einfach durchgeführt werden. Alle wichtigen Analysemodelle sind bereits vordefiniert und lassen sich für den High-End-Anwender problemlos an seine speziellen Frage- und Problemstellungen anpassen.

Direkter Zugriff auf den Datenbestand

Die Einsatzgebiete des SAS Enterprise Miner sind damit praktisch unbegrenzt, zum Beispiel für die Analyse von Bondaten im Handel, Kundensegmentierungen im Rahmen von Customer Relationship Management oder Portfolio-Analysen im Finanzdienstleistungsbereich. Ebenso groß wie die Zahl der denkbaren Anwendungen sind die statistischen Methoden und Verfahren, die die Software den Anwendern zur Verfügung stellt. Wichtigstes Merkmal des SAS Enterprise Miner ist sein direkter Zugriff auf den Datenbestand im Data Warehouse. Damit profitiert die Software von den Vorteilen der SAS Software: Plattformunabhängigkeit, Flexibilität, leistungsfähige Anwendungen, Skalierbarkeit und durchgehendes Metadatenkonzept. Die Bewertungsmodule sämtlicher Data Mining-Analysen können durch einen automatisch erzeugten Code direkt im gesamten Data Warehouse angewendet und umgesetzt werden.

Data Warehouse als Basis für Data Mining

Das Data Mining-Konzept von SAS versteht sich damit auch als Teil und Erweiterung von Data Warehousing. Ein Data Warehouse als Zusammenführung themenorientierter, konsolidierter und gereinigter Daten stellt dabei die optimale

Basis für Data Mining dar. Denn Data Mining ohne ein Data Warehouse erfordert einen erheblichen Aufwand für das Zusammentragen und Vorbereiten der Daten.

Definierte Arbeitsschritte

Der SAS Enterprise Miner führt alle Methoden und Elemente der Data Mining-Technologie „SEMMA“ in einer homogenen Arbeitsoberfläche zusammen. In dieser grafischen Oberfläche sind alle Data Mining-Projekte frei konfigurierbar und mit der SEMMA-Technologie realisierbar.

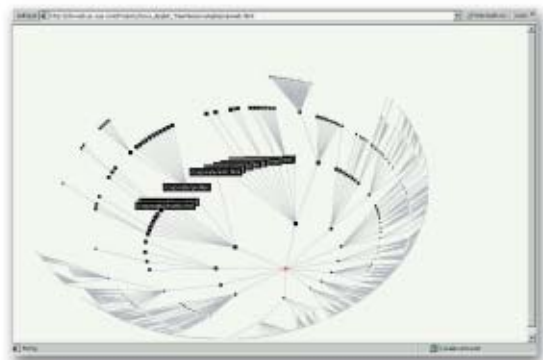
Die SEMMA-Technologie umfasst die von SAS definierten Arbeitsschritte im Data Mining:

Sample – repräsentative Stichproben (zufällig und einfach geschichtet, jede Nte und erste N-Beobachtungen, clusterweise).

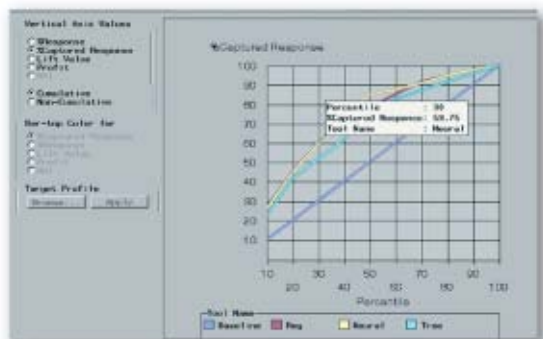
Explore – Exploration (Suche nach Mustern zwischen den Daten). Die Visualisierung ermöglicht die interaktive, grafische Erkundung der Daten, Clusteranalyse für algorithmenbasierte Vorgruppierung der Daten in möglichst signifikant unterschiedliche Objekt- bzw. Subjektsegmente).

Modify – Modifikation und Transformation der Daten durch Definieren, Auswählen, Verändern der Variablen (Ausreißer-Filter, Bildung transformierter Variablen, Berechnung statistischer Grundkennziffern).

Model – Modellbildung (Neuronale Netze, Zeitreihenanalysen, Statistische Verfahren, Entscheidungsbäume, Kombination verschiedener Modelle)



Das neue Treeview erlaubt es komplexe Zusammenhänge in einer Exponentialprojektion zu visualisieren.



Der SAS Enterprise Miner verbindet leistungsfähige statistische Methoden und Algorithmen mit einer intuitiven Oberfläche. Dabei stehen vielfältige Visualisierungsmöglichkeiten der Daten und Ergebnisse zur Verfügung.

Assess – Auswertung (Vergleich der Ergebnisse unterschiedlicher Verfahren, interaktive Auswertegrafiken für Güte, Gewinn, Aufwand und Nutzen)

Automatisierung des Data Mining-Prozess

In seiner anwenderfreundlichen Oberfläche kombiniert der SAS Enterprise Miner Bedienungskomfort und analytische Tiefe. Das Prozessfluss-Diagramm, eine grafische Benutzeroberfläche mit Drag-and-Drop-Funktionalität, automatisiert den gesamten Data Mining-Prozess. Die Software bietet jedoch nicht nur die Möglichkeit differenzierter Verfahren: Sie unterstützt den Anwender im Modellierungsteil auch bei der Auswahl der effektivsten Methode und beim Ergebnisvergleich. Dabei werden die Anwender Schritt für Schritt durch die Datenanalyse geführt. Als einziges Softwarewerkzeug umfasst der SAS Enterprise Miner dabei alle modernen Data Mining-Methoden wie Cluster-, Assoziations- und Regressionsanalyse, Entscheidungsbäume und Neuronale Netze.

Applikation für Neuronale Netze

Welche Data Mining-Methode der Anwender für die Modellierung einsetzt, hängt von seinen speziellen Fragestellungen und den zugrundeliegenden Daten ab. Der SAS Enterprise Miner unterstützt den Anwender bei dieser Entscheidung.

Bestandteil der Software ist auch eine Applikation für Neuronale Netze, die über folgende Features verfügt:

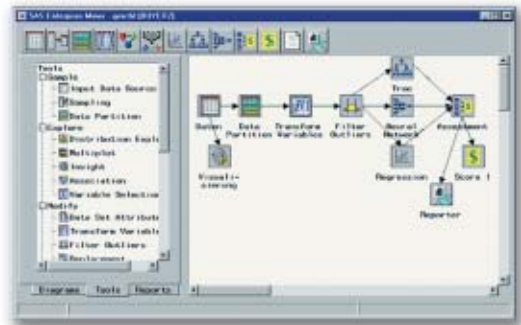
- Höchste numerische Effizienz durch Einbindung moderner numerischer Algorithmen für die nichtlineare Optimierung.

- Gleichzeitiger Datenzugriff auf Daten aus relationalen und hierarchischen Datenbanken, sequentiellen Dateien und SAS Dateien.
- Unternehmensweiter Datenzugriff über die anerkannte Data Warehouse-Technologie der SAS Software.
- Identische Funktionalität auf allen Hardware-Plattformen mit der Möglichkeit verteilter Verarbeitung bei rechenintensiven Arbeitsschritten.
- Offenlegung der den Trainingsverfahren zugrunde liegenden numerischen Algorithmen.

Der SAS Enterprise Miner ist als Client/Server-Anwendung entwickelt. Der Client läuft auf Windows, die Server-Komponenten auf Windows, MVS und verschiedenen Unix-Derivaten: Sun, Solaris, HP-UX, AIX, Compaq True 64 Unix, Intel ABI+ etc.

Um mehr über die SAS Enterprise Miner Software zu erfahren, besuchen Sie uns unter www.sas.com/products/miner

Der gesamte SEMMA-Prozess:
Stichprobe, Exploration, Modifikation,
Modellbildung, Assessment



Einsatzmöglichkeiten des SAS Enterprise Miner

Projektziel	Data Mining Technik	Softwarelösungen
Zielgruppen-Marketing - Kundenklassifikation aus aktuellen Verkaufs- und sozio-ökonomischen Daten	Klassifikation	Chi-Square Automatic Interaction Detection (Chaid), Neuronale Netze, Cluster-Analyse; Memory Based Reasoning
Database Marketing - Kaufwahrscheinlichkeit für bestimmte Produkte	Klassifikation und Analyse von Zusammenhängen	Chaid-Menu-Systeme, Regression, Conjoint-Analyse, Datenexploration und -Visualisierung
Medienanalyse - welche Kunden bevorzugen welche Medien?	Channel Management	Neuronale Netze, Regression
Analyse der Kundenzufriedenheit	Analyse von Umfrageergebnissen	Hauptkomponentenanalyse, Neuronale Netze, Regression, Diskriminanzanalyse, Datenexploration und -visualisierung
Absatzprognose	Zeitreihen	Neuronale Netze, Zeitreihenverfahren
Preisanalyse - Kaufverhalten bei verschiedenen Preisen	Schätzung von Preiselastizitäten	Spezialisierte Zeitreihenverfahren und Regressionsanalysen
Bonitätsprüfung - Analyse der Kreditwürdigkeit aus Kundenmerkmalen	Klassifikation und Analyse von Zusammenhängen	Logistische Regression, Diskriminanzanalyse, Neuronale Netze
Betrugserkennung bei Versicherungen und Kreditkartenfirmen aus Transaktions- und Kundenmerkmalen	Klassifikation und Analyse von Zusammenhängen	Assoziationsanalysen und Link-Analysen
Klassifikation elektronisch abgespeicherter Daten	Text Mining	Text Parsing, Singular Value Decomposition, Expectation Maximization Clustering
Prognose von Rechner- und Netzbelastungen	Zeitreihenmuster	Neuronale Netze, Zeitreihenverfahren

Tools im Enterprise Miner

Sample

- Input Data Source
- Sampling
- Data Partition

Explore

- Distribution Explorer
- Multiplot
- Insight
- Association
- Variable Selection

Modify

- Data Set Attributes
- Transform Variables
- Filter Outliers
- Replacement
- Clustering
- SOM/Kohonen

Model

- Regression
- Decision Trees
- Neural Networks
- Ensemble
- User defined

Assess

- Assessment
- Scoring
- Reporting

Utility

- Group Processing
- Code Note
- Control Point
- Subdiagram

Accessing Data

- Warehouse/data mart
- access to over 50 different
- file structures
- SQL front-end generator

Sampling

- Simple Random
- Stratified
- Systematic
- First N
- Cluster

Variable Selection

- Unrelated to target
- In hierarchies
- Many missing values
- Large number of levels

Transformations

- Log
- Square root
- Inverse
- Square
- Exponential
- Standardized
- Bucketed, Quantile
- Binning
- Best Power Transformation
- Maximize Normality
- User defined

Filtering Outliers

- Statistically
- rare values
- extreme values
- Manually

Missing value imputation

- Mean, median
- Distribution based
- Tree imputation
- Default constant

Clustering

- K-means
- Kohonen
- SOM

Data Partition

- Trainings, Validation and Testdata
- Good generalization
- Oversampling

Regression

- Linear and Logistic
- Nonlinear
- Stepwise, forward and
- backward selection

Decision Trees

- CHAID
- CART
- C 4.5
- Hybrid
- Interactive growing/pruning of trees

Neural Nets

- Multiple network architectures
- Several Training Techniques
- Preliminary optimization
- Automatic standardization of inputs

Ensemble

- Combination of several models
- boosting
- bagging

Assessment

- Lift charts
- ROC curves
- Profit charts (based on user-
- specified profit/cost matrix)
- Confusion matrix

Scoring

- Automated score code generation
- Captures modeling, clustering,
- transformations, and missing value
- imputation code
- Export code to mainframe

Reporting

- Documentation of analysis
- HTML or text format

Environment

- Client/server architecture
- Client: Windows 95 and NT
- Server: Windows NT and UNIX
- (SUN, HP, IBM, and DEC)



SAS Institute GmbH
In der Neckarhelle 162
D-69118 Heidelberg
Tel: 06221/415-123
Fax: 06221/415-145

www.sas.de

SAS World Headquarters
SAS Campus Drive
Cary, NC 27513 USA
Tel: (919) 677 8000
Fax: (919) 677 4444
Web: www.sas.com