

Search and Data Mining: Techniques

Applications

Anya Yarygina

Boris Novikov

Introduction

- Data mining applications
- Data mining system products and research prototypes
- Additional themes on data mining
- Social impacts of data mining
- Trends in data mining

Reference

Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques, 2nd Edition. The Morgan Kaufmann Series in Data Management Systems, 2006

Data Mining Applications

- Data Mining for financial data analysis
- Data Mining for retail industry
- Data Mining for telecommunication industry
- Data Mining for biological data analysis
- Data Mining for other scientific applications
- Data Mining for intrusion detection

Data Mining for financial data analysis

- Banking, credit, investment, insurance services
- Data: relatively complete, reliable, high quality
- Systematic data analysis and data mining
 - Design and construction of data warehouses for multidimensional data analysis and data mining
 - Loan payment prediction and customer credit policy analysis
 - Classification and clustering of customers for target marketing
 - Detection of money laundering and other financial crimes

Data Mining for retail industry

- Huge amounts of data on sales, customer shopping history, goods transportation
- Data mining for retail industry
 - Design and construction of data warehouses based on the benefits of data mining
 - Multidimensional analysis of sales, customers, products, time, and region
 - Analysis of the effectiveness of sales campaigns
 - Analysis of customer loyalty
 - Product recommendation and cross-referencing of items

Data Mining for telecommunication industry

- Multidimensional analysis of telecommunication data
- Fraudulent pattern analysis and the identification of unusual patterns
- Multidimensional association and sequential pattern analysis
- Mobile telecommunication services
- Use of visualization tools in telecommunication data analysis

Data Mining for biological data analysis

- Semantic integration of heterogeneous, distributed genomic and proteomic databases
- Alignment, indexing, similarity search, and comparative analysis of multiple nucleotide/protein sequences
- Discovery of structural patterns and analysis of genetic networks and protein pathways
- Identifying co-occurring gene sequences and linking genes to different stages of disease development
- Visualization tools in genetic data analysis

Data Mining for other scientific applications

- Data warehouses and data preprocessing
- Mining complex data types
- Graph-based mining
- Visualization tools and domain-specific knowledge

Data Mining for intrusion detection

- Development of data mining algorithms for intrusion detection
- Association and correlation analysis, and aggregation to help select and build discriminating attributes
- Analysis of stream data
- Distributed data mining
- Visualization and querying tools

How to choose a data mining system?

- Data types
- System issues
- Data sources
- Data mining functions and methodologies
- Coupling data mining with database and/or data warehouse systems
- Scalability
- Visualization tools
- Data mining query language and graphical user interface

Additional Themes on Data Mining

- Theoretical foundations of data mining
- Statistical data mining
- Visual and audio data mining
- Data mining and collaborative filtering

Theoretical foundations of data mining

- Data reduction
- Data compression
- Pattern discovery
- Probability theory
- Microeconomic view
- Inductive databases

Statistical data mining

- Regression
- Generalized linear models
- Analysis of variance
- Mixed-effect models
- Factor analysis
- Discriminant analysis
- Time series analysis
- Survival analysis
- Quality control

Visual data mining

- Data visualization
- Data mining result visualization
- Data mining process visualization
- Interactive visual data mining

Social Impacts of Data Mining

- Ubiquitous and invisible data mining
- Data mining, privacy, and data security

Trend in Data Mining

- Application exploration
- Scalable interactive data mining methods
- Integration of data mining with database systems
- Standardization of data mining language
- Visual data mining
- New methods for mining complex types of data
- Biological data mining
- Data mining and software engineering
- Web mining
- Distributed data mining
- Real-time and time-critical data mining
- Graph mining, link analysis, and social network analysis
- Privacy protection and information security

Outline

- Data mining applications
- Data mining system products and research prototypes
- Additional themes on data mining
- Social impacts of data mining
- Trends in data mining