Pertemuan ke 11

Aplikasi dan Trend dalam Data Mining

- Aplikasi Data mining
- Produk Sistem dan Penelitian Data mining
- Tema Tambahan pada Data Mining
- Pengaruh Sosial dari Data Mining
- Trend dalam Data Mining

Aplikasi Data Mining

- Data mining adalah disiplin ilmu yang masih baru dengan aplikasi yang luas dan beragam
  - Masih ada satu nontrivial gap antara prinsip umum dari data mining dan domain-specific, effective data mining tools untuk aplikasi tertentu.
- Beberarap domain aplikasi, antara lain:
  - Biomedical and DNA data analysis
  - Financial data analysis
  - Retail industry
  - Telecommunication industry

Biomedical and DNA Data Analysis

- Urutan DNA: 4 blok dasar yang membangun DNA: (nucleotides): adenine (A), cytosine (C), guanine (G), and thymine (T).
- Gene: satu urutan/barisan dari ratusan individual nucleotides tersusun dalam urutan tertentu.
- Manusia mempunyai sekitar 30,000 genes
- Sangat banyak cara sehingga nucleotides dapat diurutkan dan dibariskan untuk membentuk genes yang berbeda.
- Integrasi semantik dari keberagaman, database genome yang terdistribusi
  - Current: highly distributed, uncontrolled generation dan menggunakan data DNA yang sangat luas kebergamannya
  - Metode Data cleaning dan data integration dikembangkan dalam data mining akan membantu

Analisis DNA : Contoh

- Pencarian keserupaan dan perbandingan diantara barisan DNA
  - Bandingkan pola yang sering muncul dari setiap kelas (misal, penyakit dan kesehatan)
  - Identifikasi pola barisan gene yang berpengaruh dalam berbagai penyakit.
- Analisis Association : Pengidentifikasian dari kemunculan barisan gen
  - Sebagian penyakit tidak di triger melalui satu gen tunggal tetapi oleh kombinasi gen yang berlaku bersama.
  - Analysis Association dapat membantu menentukan macam macam dari gen yang kelihatannya akan muncul secara bersamaan dalam contoh target.
- Analisis Path : menghubungkan gen ke tingkatan pengembangan penyakit yang berbeda.
  - Gen yang berbeda dapat menjadi aktif pada tingkatan berbeda dari penyakit
  - Mengembangkan intervensi pharmaceutical yang mentargetkan tingkatan yang berbeda secara terpisah.
- Tool Visualisasi dan analisis data genetika

Data Mining untuk Analisis Data Keuangan

- Data keuangan terkumpul di bank dan intstitusi keuangan yang pada umumnya adalah lengkap, handal dan tinggi kualitasnya.
- Desain dan konstruksi dari data warehouse untuk analisis data multidimensi dan data mining.
  - View perubahan debet dan pendapatan/keuntungan berdasarkan bulan, daerah, sektor dan faktor.
  - Akses informasi statistik seperti max, min, total, average, trend, dll.
- Peramalan/prediksi pembayaran pinjaman / analisis kebijaksanaan kredit konsumen.
  - Pemeringkatan pemilihan fitur dan keterhubungan atribut
  - Kinerja pembayaran pinjaman
  - Rating kredit konsumen

Data Mining Keuangan

- Classification dan clustering dari konsumen untuk sasaran pemasaran.
  - multidimensional segmentation melalui nearest-neighbor, classification, decision trees, dll. untuk mengidentifikasi kelompok konsumen atau mengasosiasi satu konsumen baru ke satu kelompok konsumen yang tepat/sesuai.
- Detection of money laundering dan kejahatan keuangan lainnya
  - integration of from multiple DBs (e.g., bank transactions, federal/state crime history DBs)
  - Tools: data visualization, linkage analysis, classification, clustering tools, outlier analysis, and sequential pattern analysis tools (find unusual access sequences)

Data Mining untuk Retail Industry

- Retail industry: jumlah data yang sangat besar pada sales, customer shopping history, dll.
- Aplikasi dari retail data mining
  - Identify customer buying behaviors
  - Discover customer shopping patterns and trends
  - Improve the quality of customer service
  - Achieve better customer retention and satisfaction
  - Enhance goods consumption ratios
  - Design more effective goods transportation and distribution policies

Data Mining dalam Retail Industry: Contoh

- 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
- Customer retention: Analysis of customer loyalty
  - Use customer loyalty card information to register sequences of purchases of particular customers
  - Use sequential pattern mining to investigate changes in customer consumption or loyalty
  - Suggest adjustments on the pricing and variety of goods
- Purchase recommendation and cross-reference of items

Data Mining untuk Industri Telekomunikasi (1)

- A rapidly expanding and highly competitive industry and a great demand for data mining
  - Understand the business involved
  - Identify telecommunication patterns
  - Catch fraudulent activities
  - Make better use of resources
  - Improve the quality of service
- Multidimensional analysis of telecommunication data
  - Intrinsically multidimensional: calling-time, duration, location of caller, location of callee, type of call, etc.

Data Mining untuk Industri Telekomunikasi (1)

- Fraudulent pattern analysis and the identification of unusual patterns
  - Identify potentially fraudulent users and their atypical usage patterns
  - Detect attempts to gain fraudulent entry to customer accounts
  - Discover unusual patterns which may need special attention
- Multidimensional association and sequential pattern analysis
  - Find usage patterns for a set of communication services by customer group, by month, etc.
  - Promote the sales of specific services
  - Improve the availability of particular services in a region
- Use of visualization tools in telecommunication data analysis

Bagaimana memilih satu Sistem Data Mining?

- Commercial data mining systems have little in common
  - Different data mining functionality or methodology
  - May even work with completely different kinds of data sets
- Need multiple dimensional view in selection
- Data types: relational, transactional, text, time sequence, spatial?
- System issues
  - running on only one or on several operating systems?
  - a client/server architecture?
  - Provide Web-based interfaces and allow XML data as input and/or output?

Bagaimana memilih satu Sistem Data Mining? (2)

- Data sources
  - ASCII text files, multiple relational data sources
  - support ODBC connections (OLE DB, JDBC)?
- Data mining functions and methodologies
  - One vs. multiple data mining functions
  - One vs. variety of methods per function
    - More data mining functions and methods per function provide the user with greater flexibility and analysis power
- Coupling with DB and/or data warehouse systems
  - Four forms of coupling: no coupling, loose coupling, semitight coupling, and tight coupling
    - Ideally, a data mining system should be tightly coupled with a database system

Bagaimana memilih satu Sistem Data Mining? (3)

- Scalability
  - Row (or database size) scalability
  - Column (or dimension) scalability
  - Curse of dimensionality: it is much more challenging to make a system column scalable that row scalable
- Visualization tools
  - "A picture is worth a thousand words"
  - Visualization categories: data visualization, mining result visualization, mining process visualization, and visual data mining
- Data mining query language and graphical user interface
  - Easy-to-use and high-quality graphical user interface
  - Essential for user-guided, highly interactive data mining

Contoh Sistem Data Mining (1)

- IBM Intelligent Miner
  - A wide range of data mining algorithms
  - Scalable mining algorithms
  - Toolkits: neural network algorithms, statistical methods, data preparation, and data visualization tools
  - Tight integration with IBM's DB2 relational database system
- SAS Enterprise Miner
  - A variety of statistical analysis tools
  - Data warehouse tools and multiple data mining algorithms
- Mirosoft SQLServer 2000
  - Integrate DB and OLAP with mining
  - Support OLEDB for DM standard

Contoh Sistem Data Mining (2)

- SGI MineSet
  - Multiple data mining algorithms and advanced statistics
  - Advanced visualization tools
- Clementine (SPSS)
  - An integrated data mining development environment for end-users and developers
  - Multiple data mining algorithms and visualization tools
- DBMiner (DBMiner Technology Inc.)
  - Multiple data mining modules: discovery-driven OLAP analysis, association, classification, and clustering
  - Efficient, association and sequential-pattern mining functions, and visual classification tool
  - Mining both relational databases and data warehouses

Data Mining dan Intelligent Query Answering

- A general framework for the integration of data mining and intelligent query answering
  - Data query: finds concrete data stored in a database; returns exactly what is being asked
  - Knowledge query: finds rules, patterns, and other kinds of knowledge in a database
    - Intelligent (or cooperative) query answering: analyzes the intent of the query and provides generalized, neighborhood or associated information relevant to the query

Trends dalam Data Mining (1)

- Application exploration
  - development of application-specific data mining system
  - Invisible data mining (mining as built-in function)
- Scalable data mining methods
  - Constraint-based mining: use of constraints to guide data mining systems in their search for interesting patterns
- Integration of data mining with database systems, data warehouse systems, and Web database systems
- Invisible data mining

Trends dalam Data Mining (2)

- Standardization of data mining language
  - A standard will facilitate systematic development, improve interoperability, and promote the education and use of data mining systems in industry and society
- Visual data mining
- New methods for mining complex types of data
  - More research is required towards the integration of data mining methods with existing data analysis techniques for the complex types of data
- Web mining
- Privacy protection and information security in data mining