

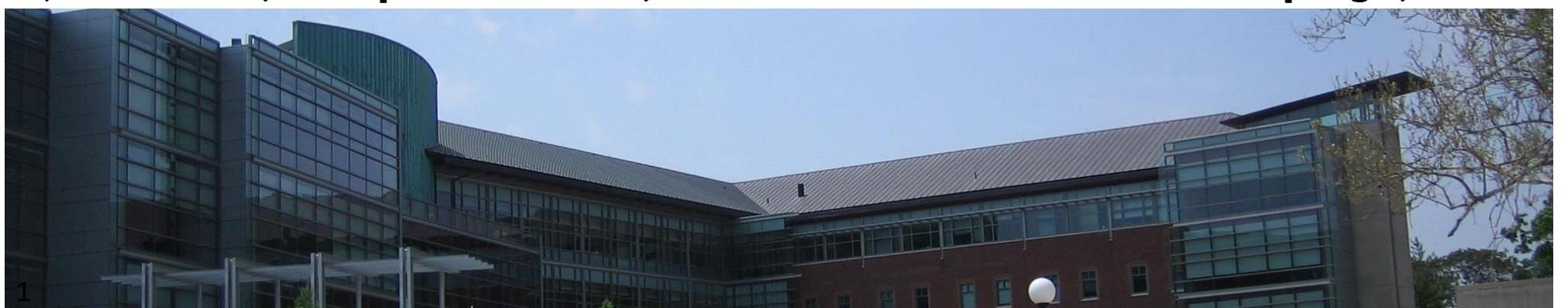


CS 412 Intro. to Data Mining

Data Warehouse & Data mining

Chapter 1. Introduction

Jiawei Han, Computer Science, Univ. Illinois at Urbana-Champaign, 2017





Data and Information Systems (DAIS)

❑ Database Systems

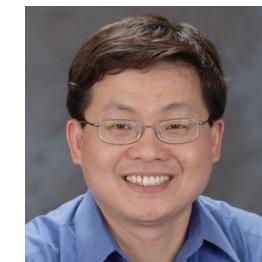


Jiawei Han

❑ Data Mining



Aditya
Parameswaran



Kevin Chang

❑ Text Information Systems



Hari
Sundaram



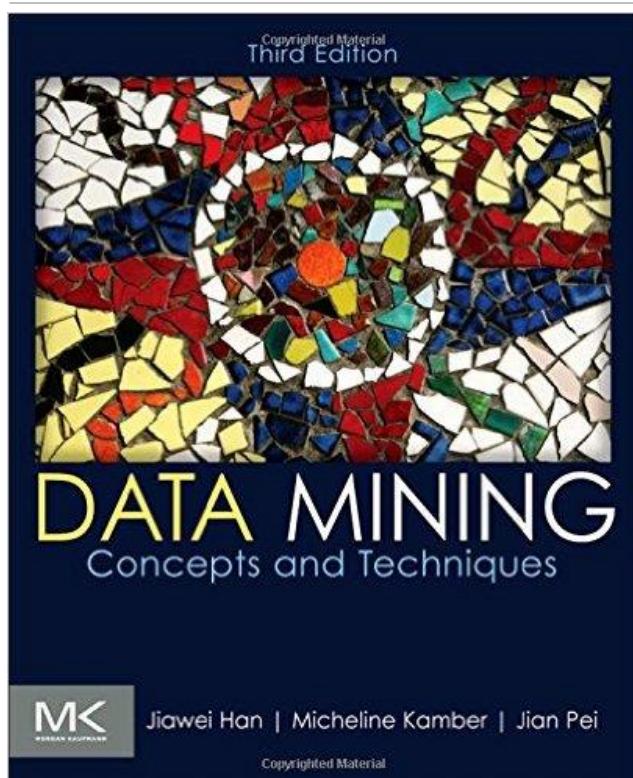
ChengXiang
Zhai

Data and Information Systems (DAIS:) Course Structures at CS/UIUC

- Coverage: Database, data mining, text information systems, Web and bioinformatics
- Data mining
 - [Intro. to data warehousing and mining \(CS412\)](#)
 - [Data mining: Principles and algorithms \(CS512\)](#)
- Database Systems:
 - Intro. to database systems (CS411)
 - Advanced database systems (CS511)
- Text information systems
 - Text information system (CS410)
 - Advanced text information systems (CS510)

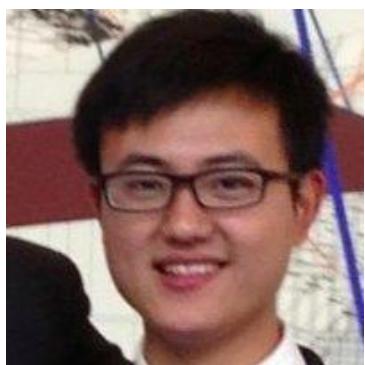
ໜັງສືອກ ຖະນາຍານ

CS 412. Course Page & Class Schedule



- Textbook
 - Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining: Concepts and Techniques* (3rd ed), Morgan Kaufmann, 2011
- Class Homepage:
<https://wiki.engr.illinois.edu/display/cs412>
- Bookmark on course schedule page
- **Class Schedule: 9:30-10:45 am Tues./Thurs. @1404 SC**
- Office hours: 10:45-11:30am Tues./Thurs. @2132 SC
- Lecture media: recorded; but class attendance is critical

CS 412. Fall 2017. Teach Assistants



Dongming Lei



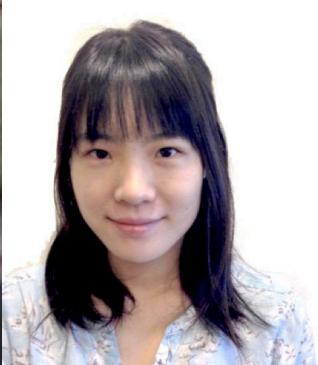
Carl Yang
(Online Session)



Yu Shi



Chao Zhang



Shi Zhi

- TA office hours: **4-5pm (Mon.), 11-12pm (Wed.)@0207SC**. Additional hours before due date will be announced at Piazza
- Wait list (No wait list at this time, keep attending class, see if there is space available or there is overflow section opening)
- If you cannot register but still desperately want to get in, please sign on when there is “potential opening”: Explain why you have to take the course This Fall!

CS 412. Course Work and Grading

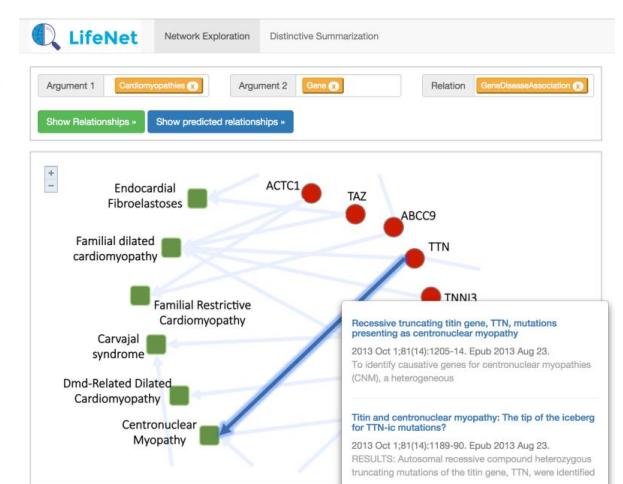
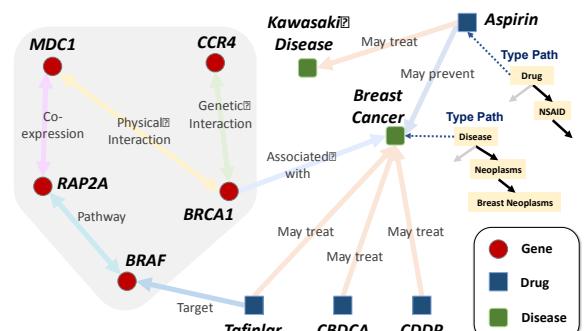
- ❑ Assignments, Programming Assignments, and Exams
 - ❑ Written Assignments: ~~15%~~ (three homework assignments expected)
 - ❑ Programming assignments: ~~20%~~ (two programming assignments expected)
 - ❑ Midterm exam: ~~30%~~ ~~0%~~ ~~40%~~
 - ❑ Final exam: ~~35%~~ ~~30%~~
- ❑ For students taking 4th credit (TA will provide concrete instructions on the 4th credit project)
- ❑ For students registering 4 credits: 25%. The overall scores will be scaled proportionally
- ❑ Need help and/or discussions?
 - ❑ Sign on: [Piazza](https://piazza.com/illinois/cs412) (<https://piazza.com/illinois/cs412>)
- ❑ Check your homework/exam scores:
 - ❑ Compass

ເພື່ອຄານຫຼັກຄາວຂອງຕະຫຼາດ

ສົມບັດລະບົບຂອງເປົ້າຫວີ

Help Needed: LifeNet—A Structured Network-Based Knowledge Exploration and Analytics System for Life Sciences

- What we are doing?
 - A scalable system that transforms biomedical papers into a knowledge graph & supports various search/analytics functions
ກາງຄັ້ງນຸ້າ / Analysis ນຳມານ
- What we already have? → ສິ້ງທີ່ຈົບ
ຮັບອັນດັບ
 - A working prototype system & an ACL demo paper
- What we are looking for?
 - Students with expertise on HTML/CSS & JavaScript
 - Experiences on web frameworks and databases
 - System design experience will be a big plus
- What you will gain?
 - Hourly pay (\$12-\$15 per hour, 6-20 hours per week)
 - Possible research publications & a good thesis topic



Send us your resume if interested: Jiaming Shen (mickeysjm@gmail.com)

Chapter 1. Introduction

- Why Data Mining? 
- What Is Data Mining?
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined?
- What Kinds of Patterns Can Be Mined?
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining
- A Brief History of Data Mining and Data Mining Society
- Summary

ក្រុងព័ត៌មានបានបញ្ជាក់ឡើង

Why Data Mining?

ទីនាំទៅ

- The Explosive Growth of Data: from terabytes to petabytes
 - ▣ Data collection and data availability
 - ▣ Automated data collection tools, database systems, Web, computerized society
 - ▣ Major sources of abundant data
 - ❑ Business: Web, e-commerce, transactions, stocks, ...
 - ❑ Science: Remote sensing, bioinformatics, scientific simulation, ...
 - ❑ Society and everyone: news, digital cameras, YouTube
- We are drowning in data, but starving for knowledge!
 - ❑ គោរពចាប់ផ្តើម ដែលតាមការណែនាំសាខាពេទ្យ: តិចខ្សែ
 - ❑ ការវិភាគ: ប្រើប្រាស់បច្ចុប្បន្នសម្រាប់ការអនុវត្តន៍
- “Necessity is the mother of invention”—Data mining—Automated analysis of massive data sets

Chapter 1. Introduction

- Why Data Mining?
- What Is Data Mining? 
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined?
- What Kinds of Patterns Can Be Mined?
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining
- A Brief History of Data Mining and Data Mining Society
- Summary

What Is Data Mining?



- Data mining (knowledge discovery from data)
 - Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data
 - Data mining: a misnomer?
- Alternative names
 - Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc.
- Watch out: Is everything “data mining”?
 - Simple search and query processing
 - (Deductive) expert systems

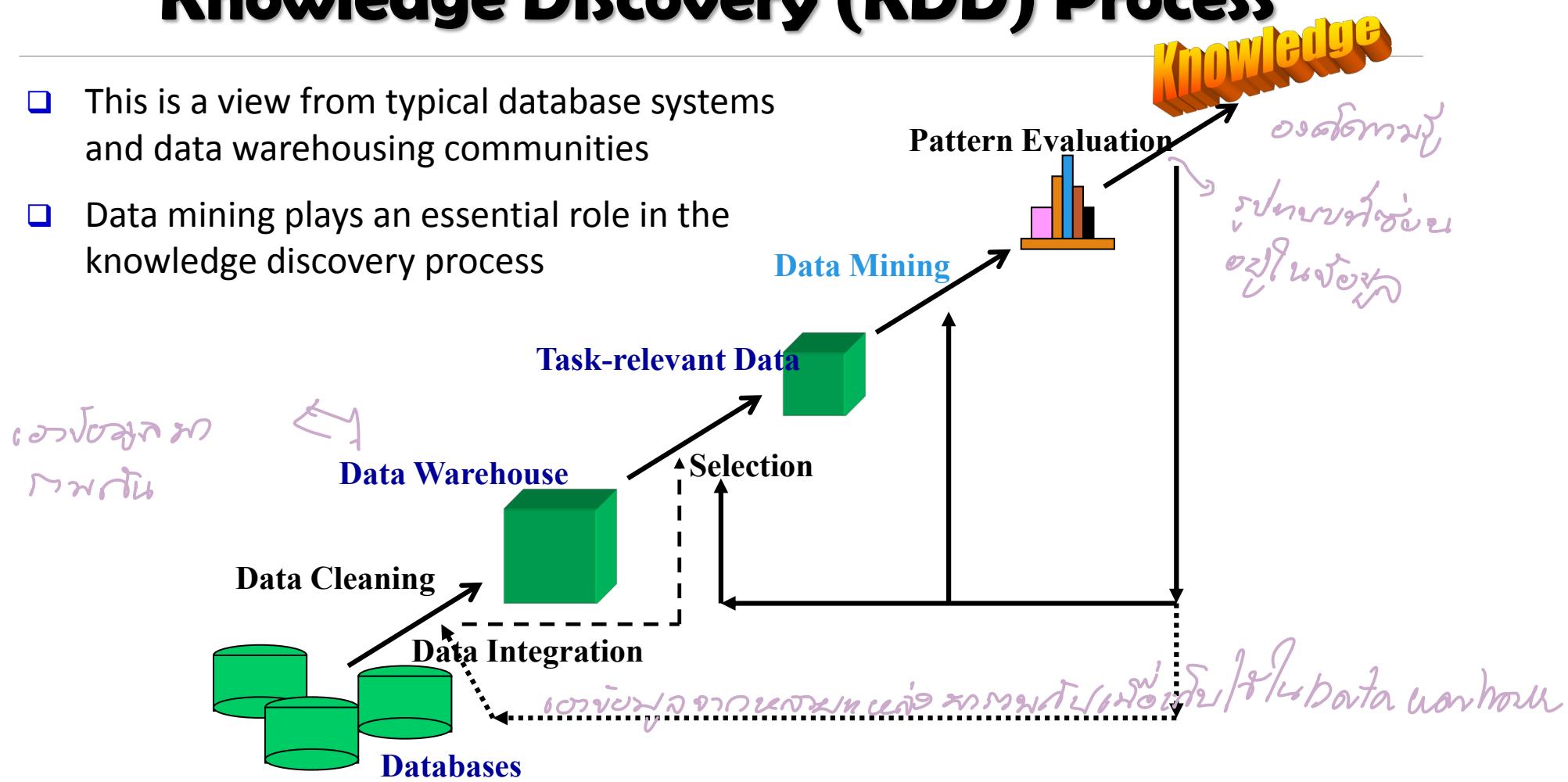
☞ នាគត់គកអ្នកពីការ Data រាយការ
ដើម្បីបង្ហាញពាណិជ្ជកម្ម
នៃក្នុងវា



ការវិភាគរបស់ការពាក្យទូទៅ

Knowledge Discovery (KDD) Process

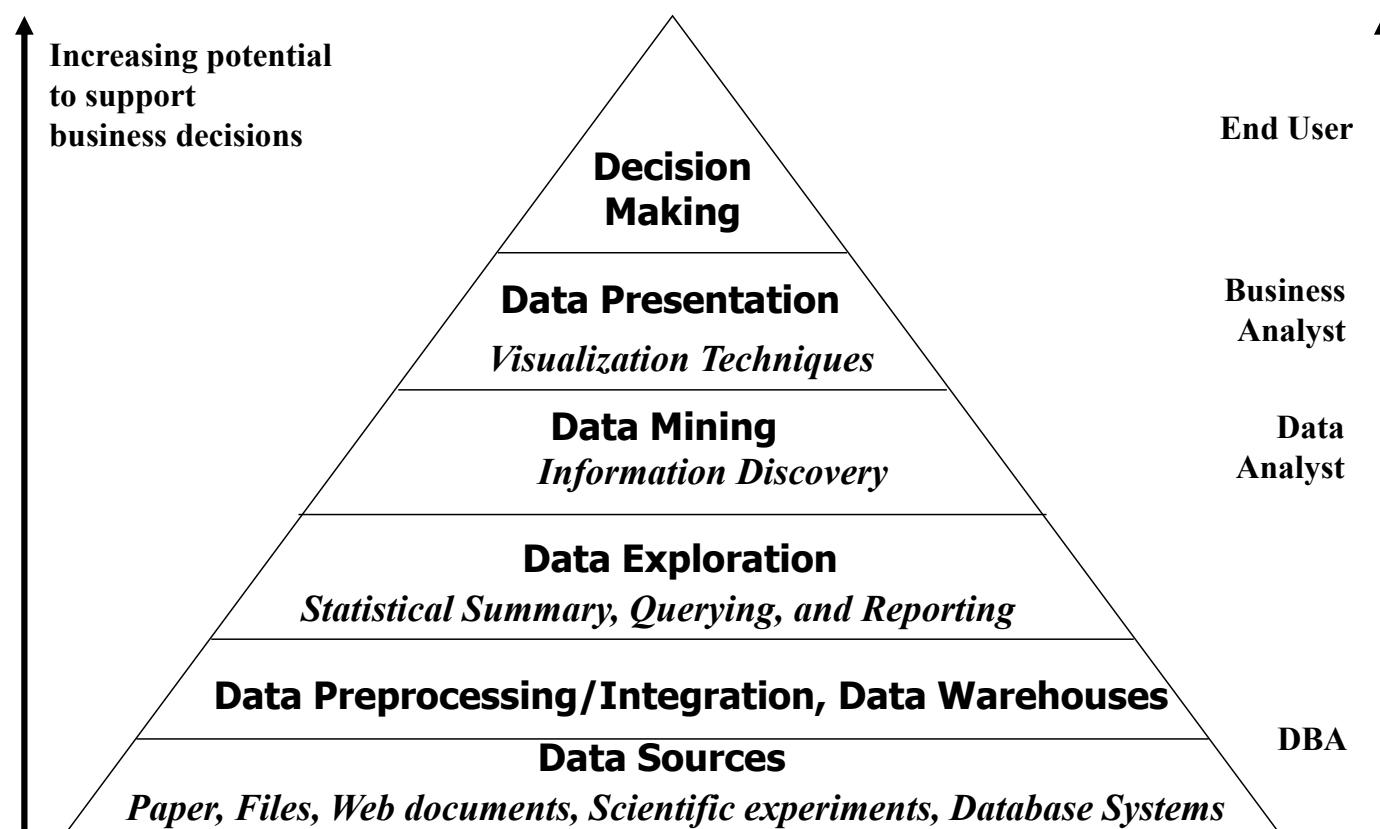
- This is a view from typical database systems and data warehousing communities
- Data mining plays an essential role in the knowledge discovery process



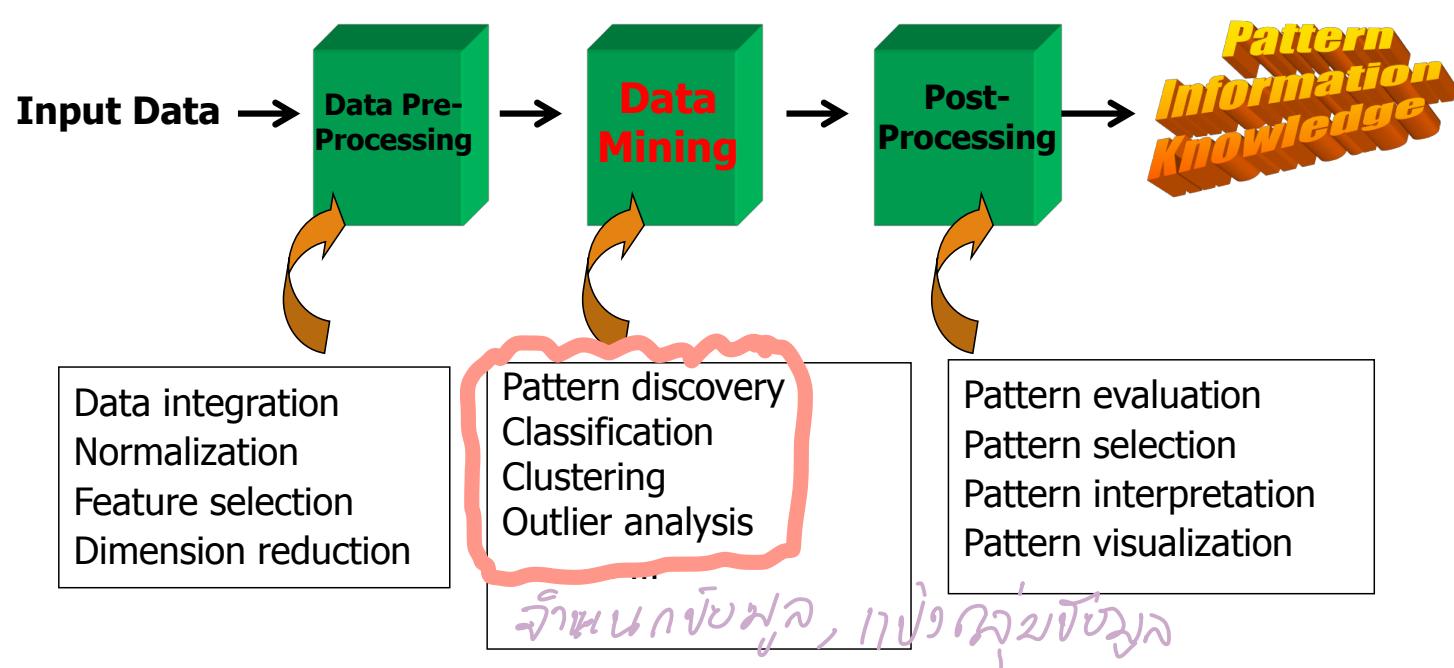
Example: A Web Mining Framework

- Web mining usually involves
 - Data cleaning
 - Data integration from multiple sources
 - Warehousing the data
 - Data cube construction
 - Data selection for data mining
 - Data mining
 - Presentation of the mining results *ເຕັມຄົດນິ້ງບົກລາກທຳອະໄຫາວິດ*
 - Patterns and knowledge to be used or stored into knowledge-base

Data Mining in Business Intelligence



KDD Process: A View from ML and Statistics



- This is a view from typical machine learning and statistics communities

Data Mining vs. Data Exploration

- ❑ Which view do you prefer?
 - ❑ KDD vs. ML/Stat. vs. Business Intelligence
 - ❑ Depending on the data, applications, and your focus

- ❑ Data Mining vs. Data Exploration
 - ❑ Business intelligence view
 - ❑ Warehouse, data cube, reporting but not much mining
 - ❑ Business objects vs. data mining tools
 - ❑ Supply chain example: mining vs. OLAP vs. presentation tools
 - ❑ Data presentation vs. data exploration

Chapter 1. Introduction

- Why Data Mining?
- What Is Data Mining?
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined?
- What Kinds of Patterns Can Be Mined?
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining
- A Brief History of Data Mining and Data Mining Society
- Summary



ມູນຄວາມສະເໜີ

Multi-Dimensional View of Data Mining

- **Data to be mined**
 - Database data (extended-relational, object-oriented, heterogeneous), data warehouse, transactional data, stream, spatiotemporal, time-series, sequence, text and web, multi-media, graphs & social and information networks
- **Knowledge to be mined (or: Data mining functions)**
 - ກາງກຳນົດລົກທຸນ: ເລືດປັບປຸງ ຕັ້ງອິນໄຕ ກາງຄ້າແນກປະເກດ ຈົດກຸ່ວ Characterization, discrimination, association, classification, clustering, trend/deviation, outlier analysis, ...
 - Descriptive vs. predictive data mining
 - Multiple/integrated functions and mining at multiple levels
- **Techniques utilized**
 - Data-intensive, data warehouse (OLAP), machine learning, statistics, pattern recognition, visualization, high-performance, etc.
- **Applications adapted** ການລົ້າປລວ
 - Retail, telecommunication, banking, fraud analysis, bio-data mining, stock market analysis, text mining, Web mining, etc.

Chapter 1. Introduction

- Why Data Mining?
- What Is Data Mining?
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined? 
- What Kinds of Patterns Can Be Mined?
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining
- A Brief History of Data Mining and Data Mining Society
- Summary

ຈຸດບັນຈິດຂອງລົງປະເກດໄດ

Data Mining: On What Kinds of Data?

ຖາວອນດາ

- Database-oriented data sets and applications
ສ້າງນັ້ນທີ່ມີຄະນະສົມຜົນຫຼຸດ
- Relational database, data warehouse, transactional database
ຮຽກຮ່ວມ
ຕ່າງກົນ
- Object-relational databases, Heterogeneous databases and legacy databases
ຮູ້ເຂົ້າມີລົງປະເກດ
- Advanced data sets and advanced applications
- Data streams and sensor data
ອານຸກວດວາລາ
ຮັງຄຽກ
ກົດໆປ
- Time-series data, temporal data, sequence data (incl. bio-sequences)
- Structure data, graphs, social networks and information networks
ເຮືອພື້ນຖານ
- Spatial data and spatiotemporal data
- Multimedia database
- Text databases
- The World-Wide Web

Chapter 1. Introduction

- Why Data Mining?
- What Is Data Mining?
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined?
- What Kinds of Patterns Can Be Mined? 
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining
- A Brief History of Data Mining and Data Mining Society
- Summary

ផែមិន: ព័ត៌ម្យ

Data Mining Functions: (1) Generalization

ការរបរាយនៃសម្រាប់គ្រប់គ្រងទីផ្សារ

- Information integration and data warehouse construction

ស្តាំ ពេលវេលា របាយ

- Data cleaning, transformation, integration, and multidimensional data model ជាមធានភាពីតិ៍

- Data cube technology

ការប្រើប្រាស់បច្ចេកទេសការគាំទ្រនាំ

- Scalable methods for computing (i.e., materializing) multidimensional aggregates

ការរបបនាយកិត្តិវត្ថុ

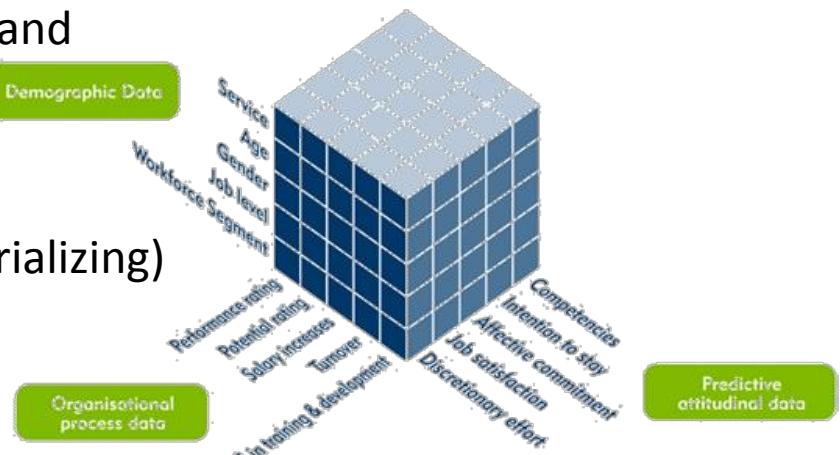
- OLAP (online analytical processing)

ខ្សោយការណ៍

- Multidimensional concept description: Characterization and discrimination ការអណៈតកម្មិនីដែលរាយការលេងក្នុងប៉ែត់

- Generalize, summarize, and contrast data characteristics, e.g., dry vs. wet region

ស្តីពី ប្រើប្រាស់ព័ត៌ម្យភាពទីផ្សារ

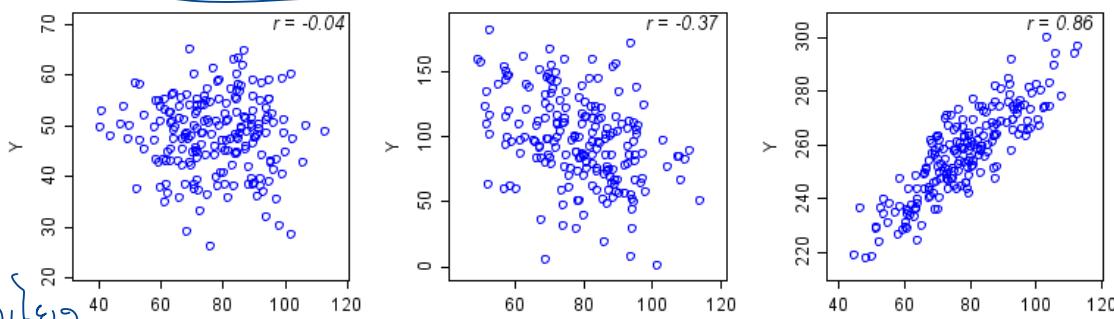


ກຳສາວຸດຈັດຂ່າຍ
ການຄົ້ນພບຮູງປະບຽບ

Data Mining Functions: (2) Pattern Discovery

ໃຊ້ບໍ່ຢ່າງ
ຮູບແບບ
ຫຼຸດສິ້ນຄ້າທີ່ໃຊ້ບໍ່ຢ່າຍ

- Frequent patterns (or frequent itemsets)
 - What items are frequently purchased together in your Walmart? → ຕາວຄໍາ ສາງ
- Association and Correlation Analysis ວິເຄາະ: ນີ້ສະໝັກພື້ນຖານ

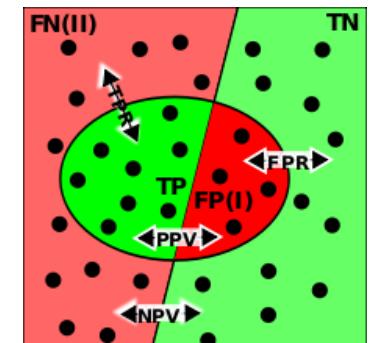


- A typical association rule Ex. ດັ່ງຕົວໄຕ້ສິ້ນຄ້າເບື້ອງໄປກາວໂກລັ້ມັດວັນຈະຫາຍ
- Diaper → Beer [0.5%, 75%] (support, confidence) ໄດ້ອັກສົ່ນ
- Are strongly associated items also strongly correlated?
- How to mine such patterns and rules efficiently in large datasets?
- How to use such patterns for classification, clustering, and other applications?

Data Mining Functions: (3) Classification

ການຫຼັບຫາຍຸສາກ

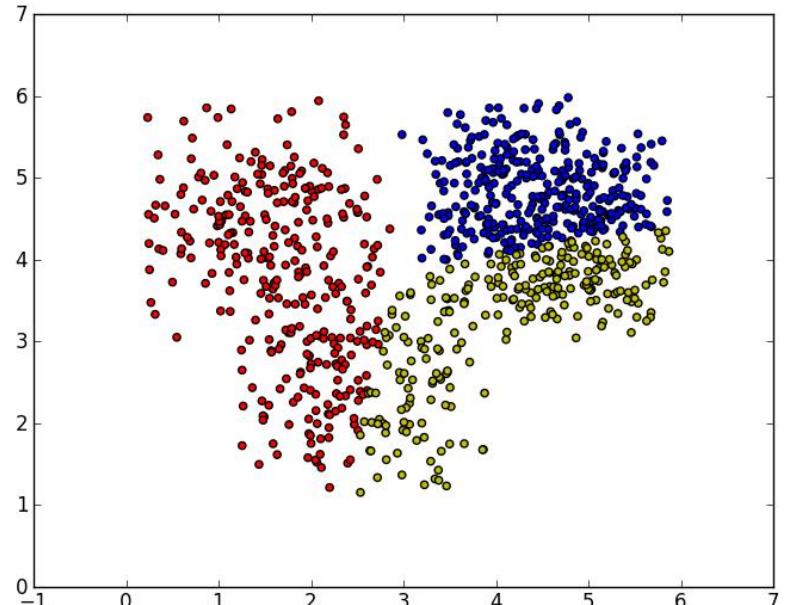
- Classification and label prediction
 - ❑ Construct models (functions) based on some training examples
ສ້າງແນບຈຳລັດ
 - ❑ Describe and distinguish classes or concepts for future prediction
ຕອບຫາຍ ແລະ ພະຍາຍ
ແນວ ກີໂຕ ກຳເຫດອນນາຄົມ
ຈຳນວນການອອນດູວັດການ
 - ❑ Ex. 1. Classify countries based on (climate)
ຮັດການ: ຍິກາ
 - ❑ Ex. 2. Classify cars based on (gas mileage)
 - ❑ Predict some unknown class labels
ວິຊີກ່ຽວຂ່າຍ
- Typical methods
 - ❑ Decision trees, naïve Bayesian classification, support vector machines, neural networks, rule-based classification, pattern-based classification, logistic regression, ...
- Typical applications:
 - ❑ Credit card fraud detection, direct marketing, classifying stars, diseases, web-pages, ...



ການວິເຄາະ; ແລ້ວສອຫາວັດ

Data Mining Functions: (4) Cluster Analysis

- ການຮັບຮືນວ່າຂບໍ່ໄມ່ມີສູງແລ້ວ ເຊິ່ນ ໄມ່ຈຳປັບປຸງກຳທົບ
- Unsupervised learning (i.e., Class label is unknown)
 - Group data to form new categories (i.e., clusters), e.g., cluster houses to find distribution patterns
 - Principle: Maximizing intra-class similarity & minimizing interclass similarity
 - Many methods and applications



การวิเคราะห์ค่าผิดปกติ

Data Mining Functions: (5) Outlier Analysis

□ Outlier analysis

ไม่สอดคล้องกับพฤติกรรม

- Outlier: A data object that does not comply with the general behavior of the data

ขึ้นๆ ลงๆ บ่อยๆ

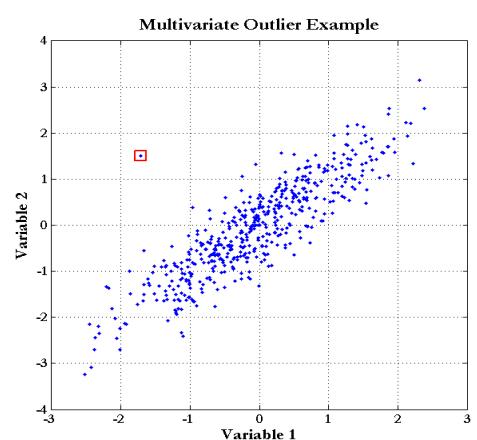
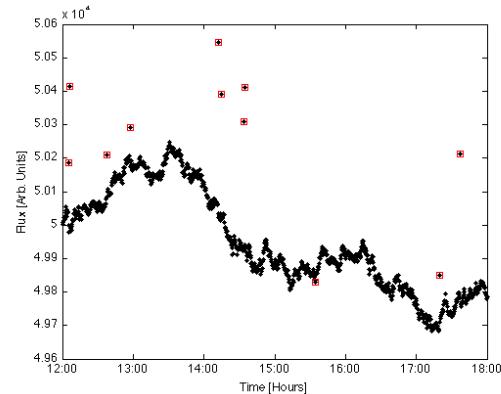
- Noise or exception?—One person's garbage could be another person's treasure

โดยผู้คนที่มองว่าเป็นการวิเคราะห์การจัดกลุ่มหรือการวิเคราะห์ความถี่

- Methods: by product of clustering or regression analysis, ...

เช่น การ聚类 หรือ การ回歸 分析

- Useful in fraud detection, rare events analysis,

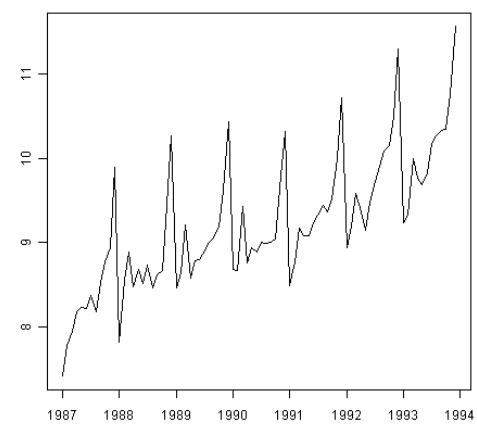


ເວລາ

ການຈົດລໍາດັບ

Data Mining Functions: (6) Time and Ordering: Sequential Pattern, Trend and Evolution Analysis

- **ລໍາດັບ** ກາງວິຄຣາ; ນັ້ນແນວໃນໜີ່ມ ແລະ ວິສ້ອມນາກຮ
Trend, time-series, and deviation analysis
- **ກາງເປົ່າຍະເປົນ** ອຸດຮາຍ ທຳນາຍຈຸດຄ່າ
e.g., regression and value prediction
- Sequential pattern mining
 - e.g., buy digital camera, then buy large memory cards
- Periodicity analysis
- Motifs and biological sequence analysis
 - **ລວມສາຍ** ກາງວິຄຣາ; ນັ້ນເປັນຮະບບ:
Approximate and consecutive motifs
 - **ໂຄຍປະການ** Similarity-based analysis
- Mining data streams
 - Ordered, time-varying, potentially infinite, data streams

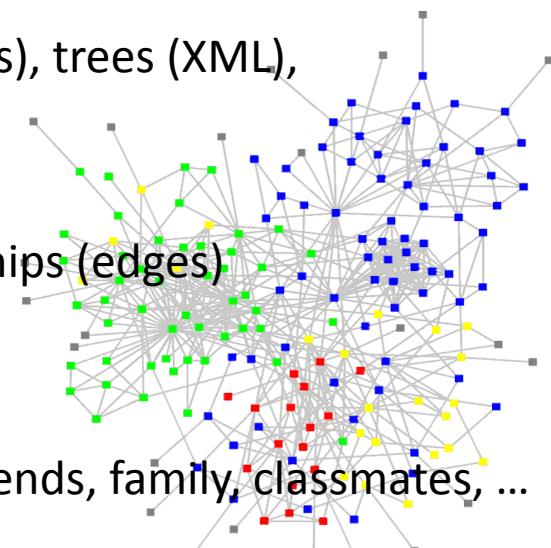


การวิเคราะห์โครงสร้าง|||
Data Mining Functions: (7) Structure and Network Analysis | ครึ่งท้าย

การทําเนื่องงาน

- Graph mining
 - Finding frequent subgraphs (e.g., chemical compounds), trees (XML), substructures (web fragments)
- Information network analysis
 - Social networks: actors (objects, nodes) and relationships (edges)
 - e.g., author networks in CS, terrorist networks
 - Multiple heterogeneous networks
 - A person could be multiple information networks: friends, family, classmates, ...
 - Links carry a lot of semantic information: Link mining
- Web mining
 - Web is a big information network: from PageRank to Google
 - Analysis of Web information networks
 - Web community discovery, opinion mining, usage mining, ...

สารปรัชญา



การปั่น: เว็บความรู้ **Evaluation of Knowledge**

ทั้งหมดที่คุณได้รับในเรื่อง

- Are all mined knowledge interesting?

สามารถถูกใช้สำหรับการค้นหา

- One can mine tremendous amount of "patterns"

บางส่วนอาจมีอยู่ในที่ที่ไม่สามารถเข้าถึง

- Some may fit only certain dimension space (time, location, ...)

บางคราว

- Some may not be representative, may be transient, ...

การปั่น: เว็บความรู้ที่ถูกต้อง

ข้อมูลที่สนใจโดยตรง

- Evaluation of mined knowledge → directly mine only interesting knowledge?

ต. ควรบุคคลิก เช่น พฤติกรรม

- Descriptive vs. predictive

การคาดการณ์

□ Coverage

ครอบคลุม

□ Typicality vs. novelty

ค. ความน่าจะเป็น

□ Accuracy

ค. แม่นยำ

□ Timeliness

ค. เวลา, ค. 时效性

□ ...

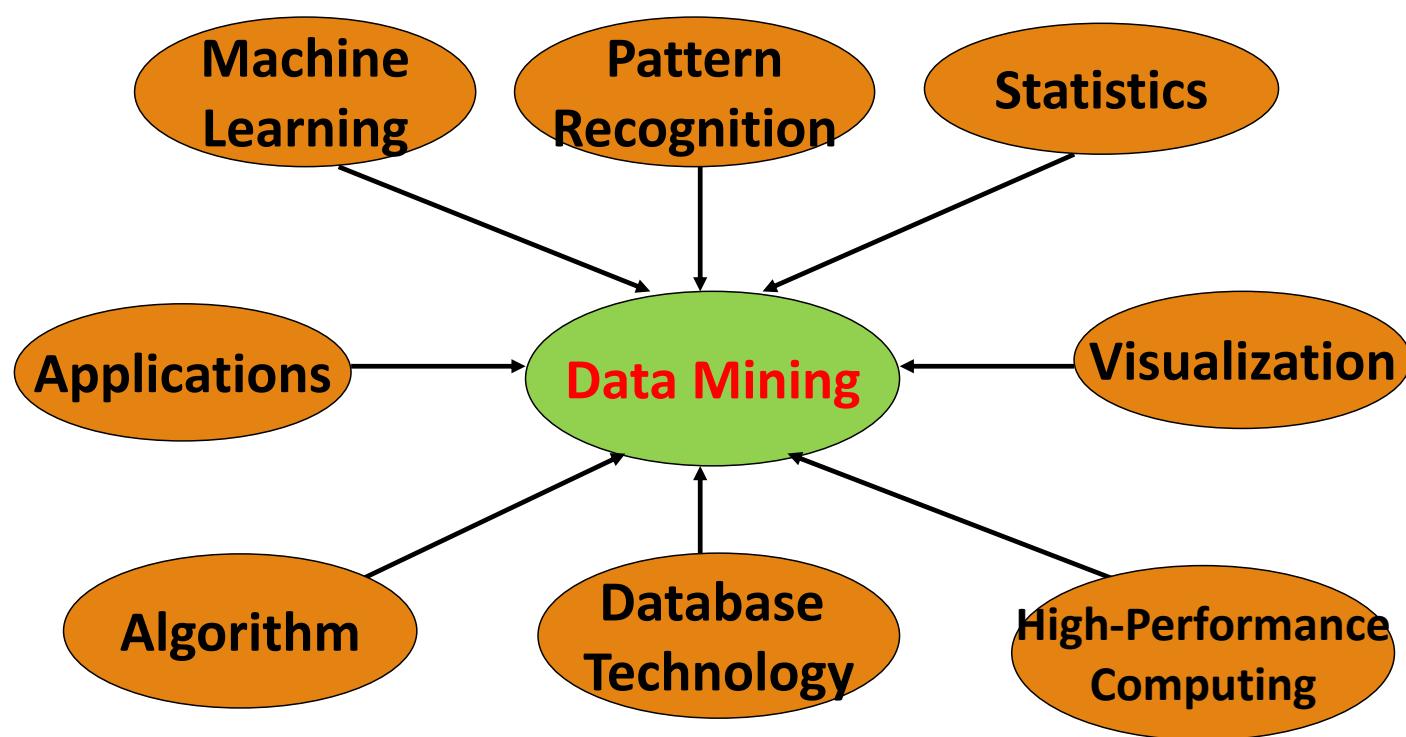


Chapter 1. Introduction

- Why Data Mining?
- What Is Data Mining?
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined?
- What Kinds of Patterns Can Be Mined?
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining
- A Brief History of Data Mining and Data Mining Society
- Summary



Data Mining: Confluence of Multiple Disciplines



ଓଡ଼ିଆ

ឧវាយន៍ា ទាំងឡាតាំង

Why Confluence of Multiple Disciplines?

ຈົດນູ້ມປ່ຽນມານາຄາ

- ❑ Tremendous amount of data
 - ❑ Algorithms must be scalable to handle big data
 - ❑ High-dimensionality of data
 - ❑ Micro-array may have tens of thousands of dimensions
 - ❑ High complexity of data
 - ❑ Data streams and sensor data
 - ❑ Time-series data, temporal data, sequence data
 - ❑ Structure data, graphs, social and information networks
 - ❑ Spatial, spatiotemporal, multimedia, text and Web data
 - ❑ Software programs, scientific simulations
 - ❑ New and sophisticated applications

Chapter 1. Introduction

- ❑ Why Data Mining?
- ❑ What Is Data Mining?
- ❑ A Multi-Dimensional View of Data Mining
- ❑ What Kinds of Data Can Be Mined?
- ❑ What Kinds of Patterns Can Be Mined?
- ❑ What Kinds of Technologies Are Used?
- ❑ What Kinds of Applications Are Targeted? 
- ❑ Major Issues in Data Mining
- ❑ A Brief History of Data Mining and Data Mining Society
- ❑ Summary

ການປະຢຸກຕ່າງໆ

Applications of Data Mining

ການວິຄາສີເລັດນ້ຳເວັບ ຈຳເນັກ ຈົດກຸລ ຈົດໜັດງູ

- Web page analysis: classification, clustering, ranking
ວິຄາສີນໍ້ຮູມກົດ
- Collaborative analysis & recommender systems
ຜູ້ການທຳອຳ
- Basket data analysis to targeted marketing
ເພື່ອການຕະຫຼາດປິ່ງຂອງຫາຍ່າງ
- Biological and medical data analysis
- Data mining and software engineering
- Data mining and text analysis
- Data mining and social and information network analysis
ຝຶກຂົນໃຈຕ່າງໆ ການຊັບວິມກົມຕ່າງໆໄວ້ນີ້ນີ້
- Built-in (invisible data mining) functions in Google, MS, Yahoo!, LinkedIn, Facebook, ...
ເຄື່ອງນີ້ ດັກ້ານເຂົ້າຕ່າງໆ
- Major dedicated data mining systems/tools
ເຄື່ອງນີ້ ດັກ້ານເຂົ້າຕ່າງໆ
- SAS, MS SQL-Server Analysis Manager, Oracle Data Mining Tools)



Chapter 1. Introduction

- Why Data Mining?
- What Is Data Mining?
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined?
- What Kinds of Patterns Can Be Mined?
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining 
- A Brief History of Data Mining and Data Mining Society
- Summary

ប្រចាំទីនៃលក្ខណៈការទូទាត់ទិន្នន័យ

Major Issues in Data Mining (1)

វិធានការទូទាត់

- Mining Methodology
ទូទាត់ការងារ & ពិនិត្យ
- Mining various and new kinds of knowledge
ទូទាត់រឿង & ថ្មី
- Mining knowledge in multi-dimensional space
គម្រោងប្រឈមស្ថាបនុយការ
- Data mining: An interdisciplinary effort
ការដំឡើងស្ថាបនុយការ
- Boosting the power of discovery in a networked environment
ការពេញលេញការស្ថាបនុយការ
- Handling noise, uncertainty, and incompleteness of data
- Pattern evaluation and pattern- or constraint-guided mining
ការត្រួតពិនិត្យការងារ
- User Interaction
ការរួមចុះហើយ
- Interactive mining
ការរួមចុះហើយ
- Incorporation of background knowledge
ការសម្រេចការងារទូទាត់ទិន្នន័យ
- Presentation and visualization of data mining results
ការបង្ហាញនូវការងារទូទាត់ទិន្នន័យ

Major Issues in Data Mining (2)

ປະສົງກົດລາຍ & ຂ.ສາຂາການ

- Efficiency and Scalability **ປະຈຸບັດ**
 - Efficiency and scalability of data mining algorithms
ກາງກຳເມືອງຕາງໆທີ່ມີລາຍແບບນໍາໃນ ກາຍ ສຕຣັນ ໄວ່ ສ້າງພົວ
 - Parallel, distributed, stream, and incremental mining methods
ຄ.ຊີລາກະສົດ
- Diversity of data types
 - Handling complex types of data
 - Mining dynamic, networked, and global data repositories
ຜລກອະການສິ່ງຄູນ
- Data mining and society
 - Social impacts of data mining
 - Privacy-preserving data mining
ສ້າງຕົວ
 - Invisible data mining
ກຳເນົາມືຍ້າທີ່ມີກຳສະຕິໄວ້ເໜີນ

Chapter 1. Introduction

- ❑ Why Data Mining?
- ❑ What Is Data Mining?
- ❑ A Multi-Dimensional View of Data Mining
- ❑ What Kinds of Data Can Be Mined?
- ❑ What Kinds of Patterns Can Be Mined?
- ❑ What Kinds of Technologies Are Used?
- ❑ What Kinds of Applications Are Targeted?
- ❑ Major Issues in Data Mining
- ❑ A Brief History of Data Mining and Data Mining Society
- ❑ Summary



A Brief History of Data Mining Society

- 1989 IJCAI Workshop on Knowledge Discovery in Databases
 - Knowledge Discovery in Databases (G. Piatetsky-Shapiro and W. Frawley, 1991)
- 1991-1994 Workshops on Knowledge Discovery in Databases
 - Advances in Knowledge Discovery and Data Mining (U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, 1996)
- 1995-1998 International Conferences on Knowledge Discovery in Databases and Data Mining (KDD'95-98)
 - Journal of Data Mining and Knowledge Discovery (1997)
- ACM SIGKDD conferences since 1998 and SIGKDD Explorations
- More conferences on data mining
 - PAKDD (1997), PKDD (1997), SIAM-Data Mining (2001), (IEEE) ICDM (2001), WSDM (2008), etc.
- ACM Transactions on KDD (2007)

Conferences and Journals on Data Mining

- KDD Conferences
 - ACM SIGKDD Int. Conf. on Knowledge Discovery in Databases and Data Mining ([KDD](#))
 - SIAM Data Mining Conf. ([SDM](#))
 - (IEEE) Int. Conf. on Data Mining ([ICDM](#))
 - European Conf. on Machine Learning and Principles and practices of Knowledge Discovery and Data Mining ([ECML-PKDD](#))
 - Pacific-Asia Conf. on Knowledge Discovery and Data Mining ([PAKDD](#))
 - Int. Conf. on Web Search and Data Mining ([WSDM](#))
- Other related conferences
 - DB conferences: ACM SIGMOD, VLDB, ICDE, EDBT, ICDT, ...
 - Web and IR conferences: WWW, SIGIR, WSDM
 - ML conferences: ICML, NIPS
 - PR conferences: CVPR,
- Journals
 - Data Mining and Knowledge Discovery (DAMI or DMKD)
 - IEEE Trans. On Knowledge and Data Eng. (TKDE)
 - KDD Explorations
 - ACM Trans. on KDD

Where to Find References? DBLP, CiteSeer, Google

-
- ❑ Data mining and KDD (SIGKDD)
 - ❑ Conferences: ACM-SIGKDD, IEEE-ICDM, SIAM-DM, PKDD, PAKDD, etc.
 - ❑ Journal: Data Mining and Knowledge Discovery, KDD Explorations, ACM TKDD
 - ❑ Database systems (SIGMOD)
 - ❑ Conferences: ACM-SIGMOD, ACM-PODS, VLDB, IEEE-ICDE, EDBT, ICDT, DASFAA
 - ❑ Journals: IEEE-TKDE, ACM-TODS/TOIS, JIIS, J. ACM, VLDB J., Info. Sys., etc.
 - ❑ AI & Machine Learning
 - ❑ Conferences: Machine learning (ML), AAAI, IJCAI, COLT (Learning Theory), CVPR, NIPS, etc.
 - ❑ Journals: Machine Learning, Artificial Intelligence, Knowledge and Information Systems, IEEE-PAMI, etc.
 - ❑ Web and IR
 - ❑ Conferences: SIGIR, WWW, CIKM, etc.
 - ❑ Journals: WWW: Internet and Web Information Systems,
 - ❑ Statistics
 - ❑ Conferences: Joint Stat. Meeting, etc.
 - ❑ Journals: Annals of statistics, etc.
 - ❑ Visualization
 - ❑ Conference proceedings: CHI, ACM-SIGGraph, etc.
 - ❑ Journals: IEEE Trans. visualization and computer graphics, etc.

Summary

ການຊັ້ນຕົວນີ້ ການຄັ້ງພຽງປະເລີດຄາມຮູ່ທີ່ສຳເນົາຈາກຫຼັມນິ້ນຂອງມາດໂລກ

- Data mining: Discovering interesting patterns and knowledge from massive amount of data
- ວິວທີ່ການຫຼັມຮູ່ມາດໂລກ ຕົ້ນຕົວນີ້ ເປັນກໍຕ້ອງການດໍາທຳ ພາກ
- A natural evolution of science and information technology, in great demand, with wide applications
- A KDD process includes data cleaning, data integration, data selection, transformation, data mining, pattern evaluation, and knowledge presentation
- Mining can be performed in a variety of data
- Data mining functionalities: characterization, discrimination, association, classification, clustering, trend and outlier analysis, etc.
- Data mining technologies and applications
- Major issues in data mining

Recommended Reference Books

- Charu C. Aggarwal, Data Mining: The Textbook, Springer, 2015
- E. Alpaydin. Introduction to Machine Learning, 2nd ed., MIT Press, 2011
- R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2ed., Wiley-Interscience, 2000
- U. Fayyad, G. Grinstein, and A. Wierse, Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann, 2001
- J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd ed. , 2011
- T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, 2009
- T. M. Mitchell, Machine Learning, McGraw Hill, 1997
- P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Wiley, 2005 (2nd ed. 2016)
- I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 2nd ed. 2005
- Mohammed J. Zaki and Wagner Meira Jr., Data Mining and Analysis: Fundamental Concepts and Algorithms 2014

