

CS 412 Intro. to Data Mining

Chapter 1. Introduction

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





Data and Information Systems (DAIS)

- Database Systems

- Data Mining

- Text Information Systems

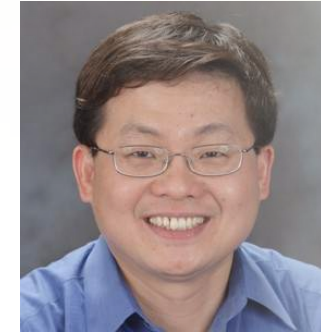
- Networks



Jiawei Han



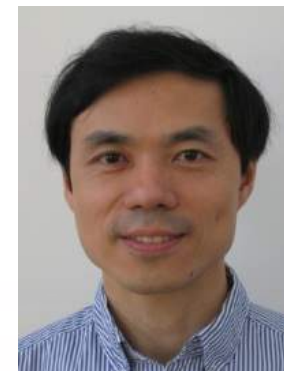
Aditya
Parameswaran



Kevin Chang



Hari
Sundaram



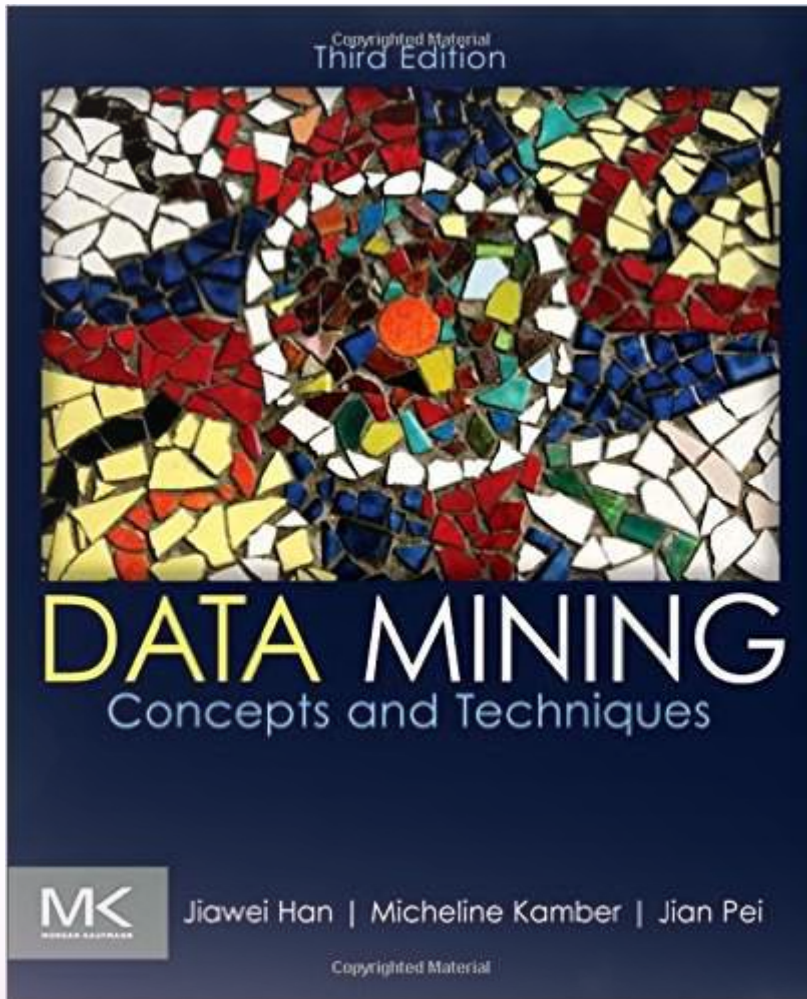
ChengXiang
Zhai

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)

CS 412. Course Page & Class Schedule



- Textbook

หนังสือ

- 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

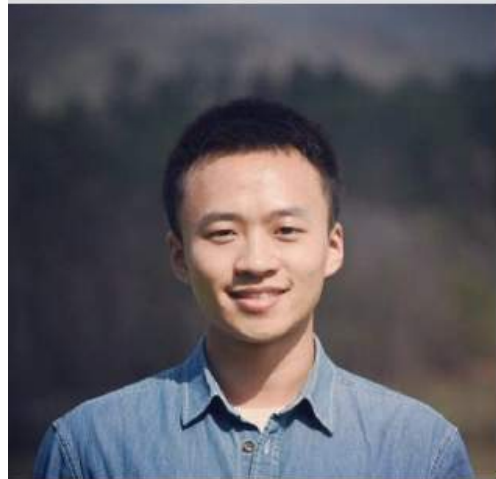


Dongming Lei



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!

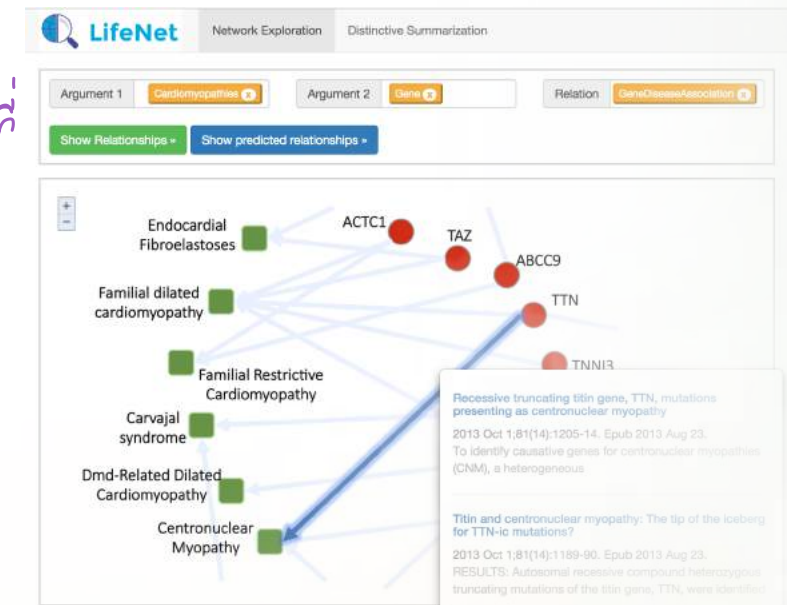
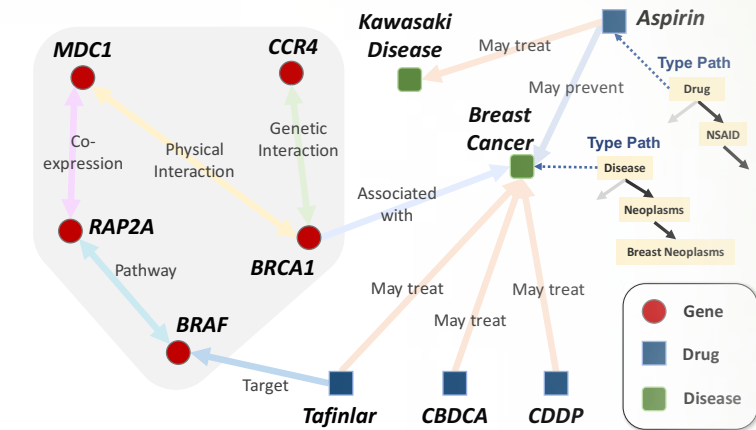
CS 412. Course Work and Grading

- Assignments, Programming Assignments, and Exams
 - Written Assignments: 15% (three homework assignments expected)
 - Programming assignments: 20% (two programming assignments expected)
 - Midterm exam: 30%
 - Final exam: 35%
- 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

ระบบสำรวจและวิเคราะห์ข้อมูลงานเคสช่วยที่มีโครงสร้าง


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
- What we already have?
 - 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

Why Data Mining? ทำไมต้องทำเหมืองข้อมูล

การเติบโต

- The Explosive Growth of Data: from terabytes to petabytes

- Data collection and data availability รวบรวม พร้อมใช้งาน

- Automated data collection tools, database systems, Web, computerized society

- Major sources of abundant data แหล่งข้อมูลที่สำคัญมากมาย

- 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!

- “Necessity is the mother of invention” — Data mining — Automated analysis of massive data sets ความจำเป็น เป็นต้นกำเนิดของการประดิษฐ์ ชุดข้อมูล การวิเคราะห์ชุดข้อมูลขนาดใหญ่



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

อะไรคือเหมืองข้อมูล What Is Data Mining?

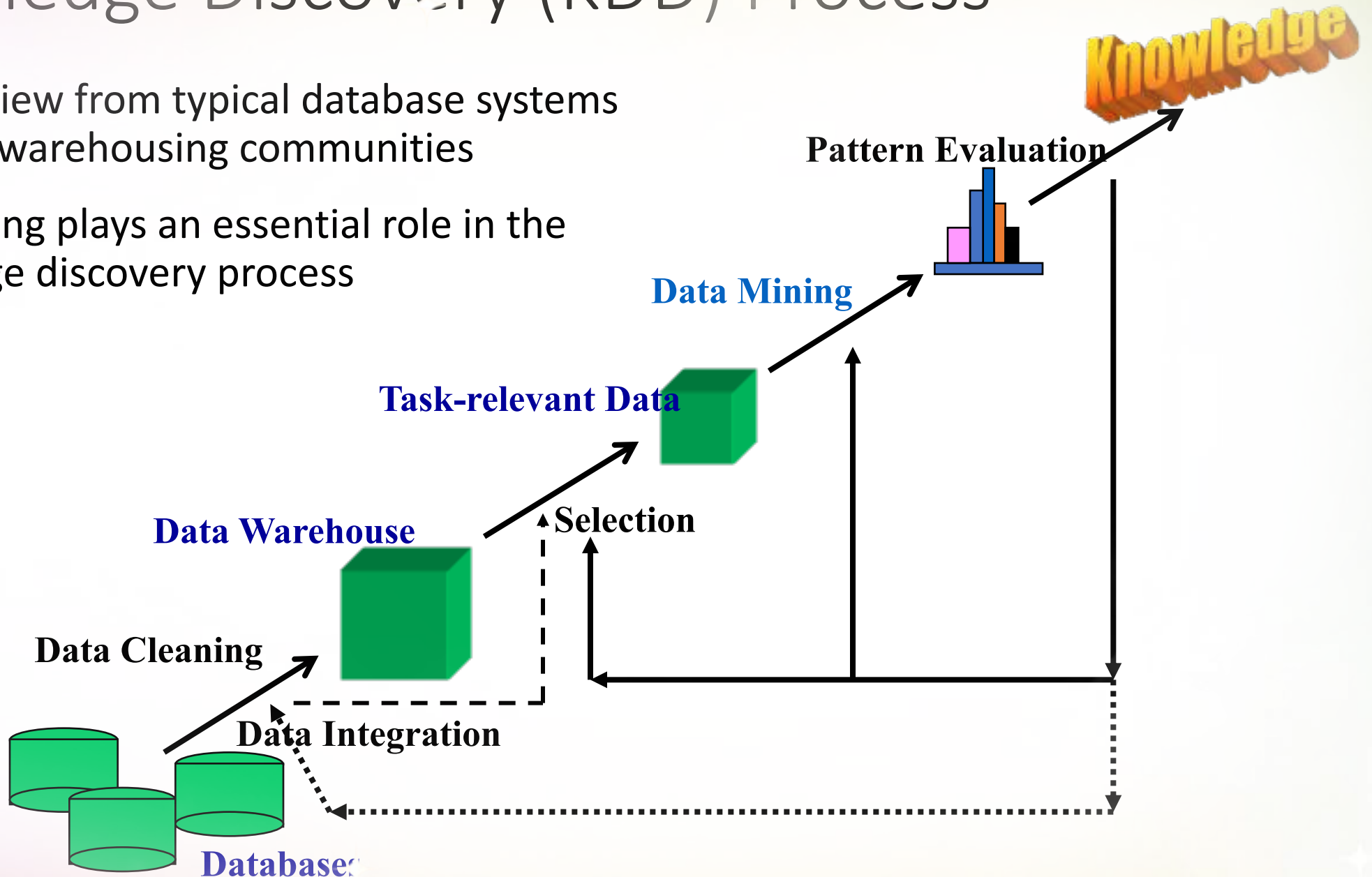


- Data mining (knowledge discovery from data)
 - การค้นพบความรู้จากข้อมูล
 - การค้นพบแบบข้อมูลที่น่าสนใจ
 - Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data
 - เรียกชื่อผิด
 - Data mining: a misnomer?
- Alternative names: ชื่อทางเลือก
 - การค้นพบความรู้
 - 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”?
 - การประมวลผล การค้นหา และแบบสอบถามอย่างง่าย
 - Simple search and query processing
 - ระบบผู้เชี่ยวชาญ
 - (Deductive) expert systems



Knowledge Discovery (KDD) Process

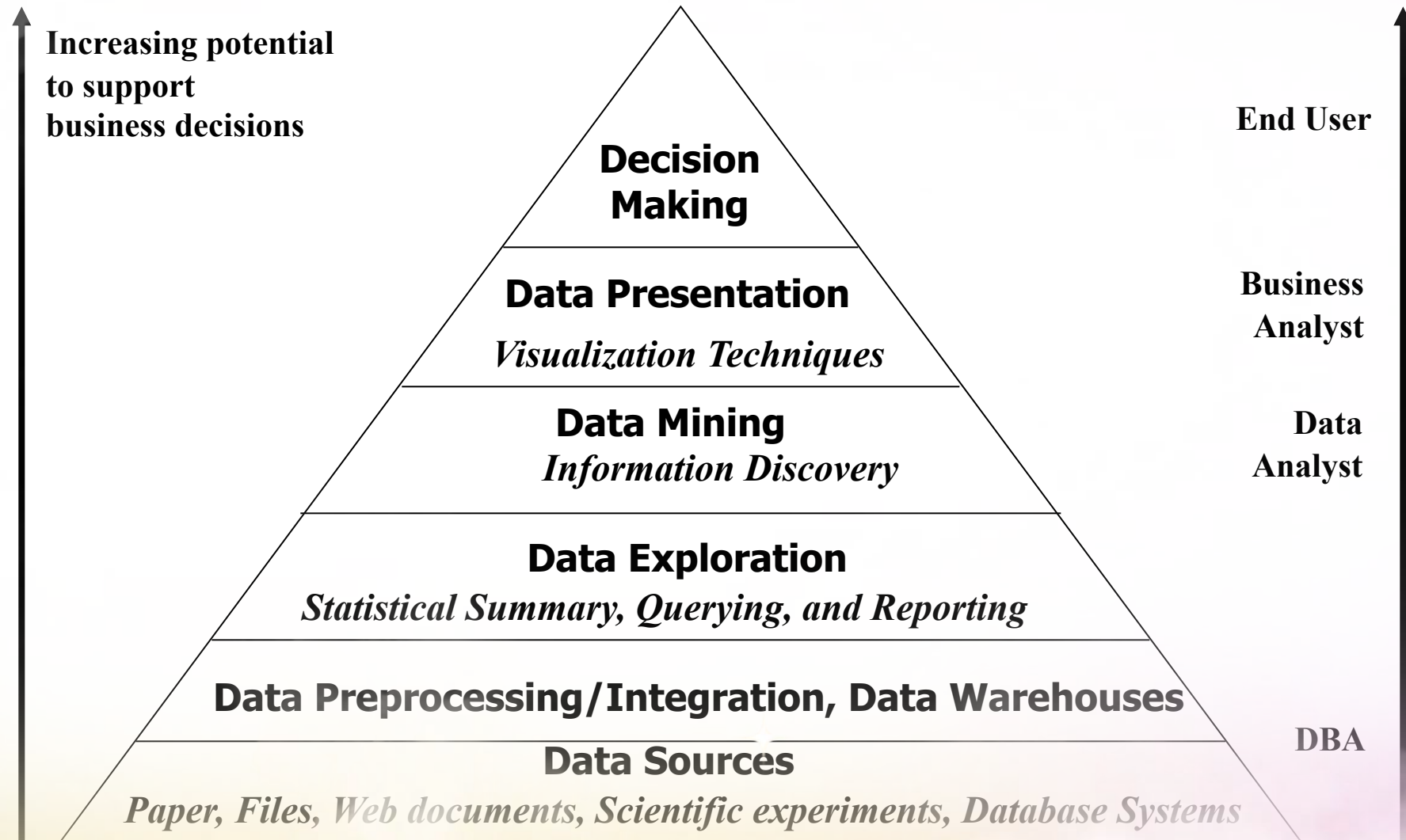
- This is a view from typical database systems and data warehousing communities
- Data mining plays an essential role in the knowledge discovery process



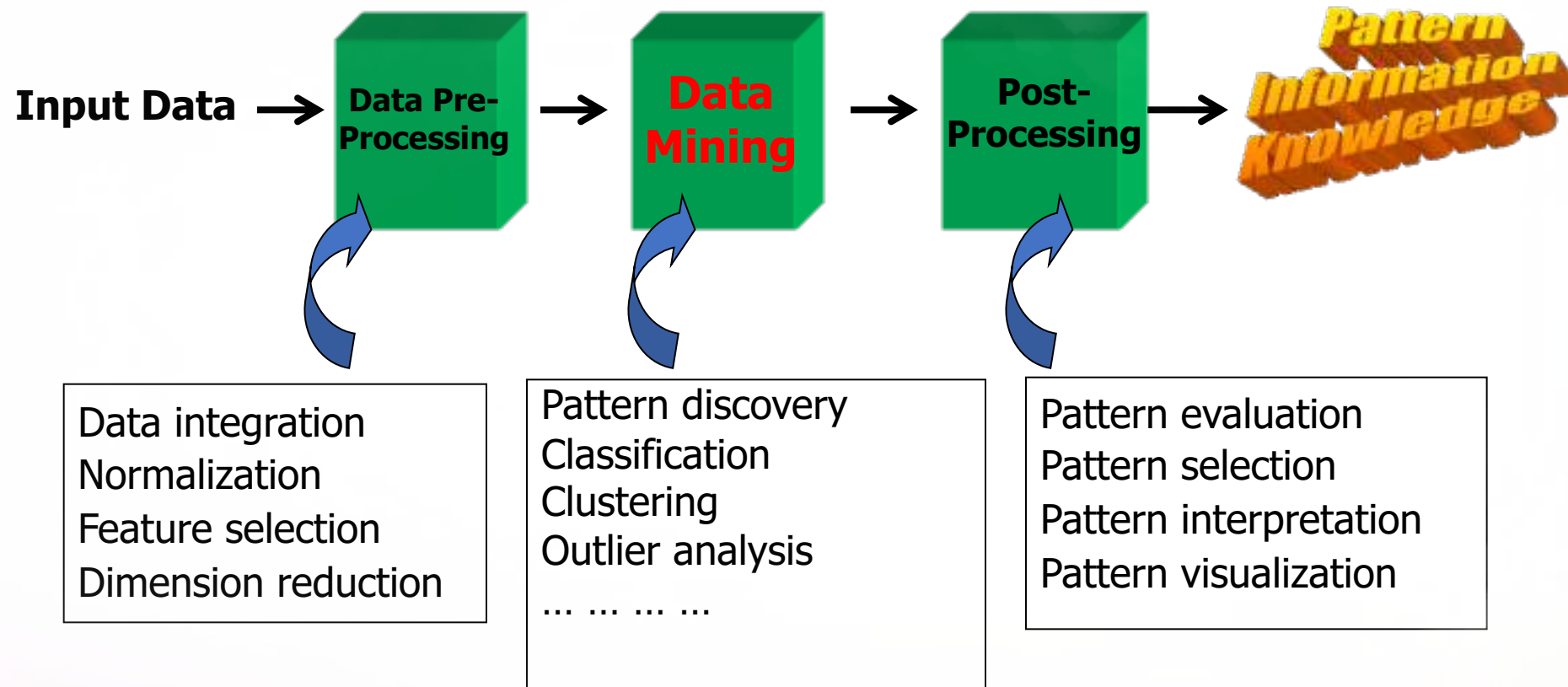
Example: A Web Mining Framework

- 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
 - 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

Multi-Dimensional View of Data Mining

- **Data to be mined** ^{ข้อมูลที่จะขุด}
 - Database ^{ฐานข้อมูล} data (extended-relational, object-oriented, heterogeneous), data warehouse, ^{ส่วนขยาย} 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)** ^{ฟังก์ชันการทำเหมืองข้อมูล}
 - Characterization, discrimination, association, classification, clustering, trend/deviation, outlier analysis, ...
^{การค้นหาลักษณะ: เลือกประเภท เชื่อมโยง การจำแนกประเภท จัดกลุ่ม}
 - 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** ^{การดัดแปลง}
 - Retail, telecommunication, banking, fraud analysis, bio-data mining, stock market analysis, text mining. Web mining. etc.

Chapter 1. Introduction

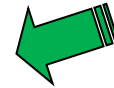
- Why Data Mining?
- What Is Data Mining?
- A Multi-Dimensional View of Data Mining
- What Kinds of Data Can Be Mined? 
- What Kinds of Patterns Can Be Mined?
- What Kinds of Technologies Are Used?
- What Kinds of Applications Are Targeted?
- Major Issues in Data Mining
- A Brief History of Data Mining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

Data Mining: On What Kinds of Data? ^{ชุดข้อมูลประเภทใด?}

- Database-oriented data sets and applications ^{ชุดข้อมูล}
 - Relational database, data warehouse, transactional database ^{ฐานข้อมูลเชิงสัมพันธ์ ระบบธุรกรรม}
 - Object-relational databases, Heterogeneous databases and legacy databases ^{ต่างกัน}
- Advanced data sets and advanced applications ^{ชุดข้อมูลขั้นสูง}
 - Data streams and sensor data
 - Time-series data, temporal data, sequence data (incl. bio-sequences) ^{อนุกรมเวลา ช่วงเวลา ลำดับ}
 - Structure data, graphs, social networks and information networks
 - Spatial data and spatiotemporal data ^{เชิงพื้นที่}
 - 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



ลักษณะทั่วไป

Data Mining Functions: (1) Generalization

การรวบรวมและสร้างคลังข้อมูล

- Information integration and data warehouse construction

- Data ^{ล้าง}cleaning, ^{แปลง}transformation, ^{รวม}integration, and multidimensional data model ^{จำลองลักษณะที่}

- Data cube technology

การนำข้อมูลมาวิเคราะห์ตามความต้องการ

- Scalable methods for computing (i.e., materializing) multidimensional aggregates ^{การรวบรวมข้อมูลและ}

- OLAP (online analytical processing) ^{เชิงวิเคราะห์}

- Multidimensional concept description: Characterization and discrimination ^{การนิยามลักษณะ: เฉพาะ และ การเลือกปฏิบัติ}

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

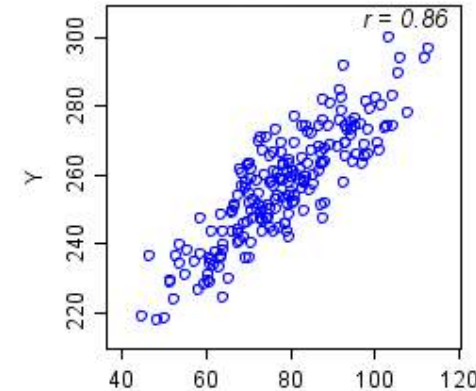
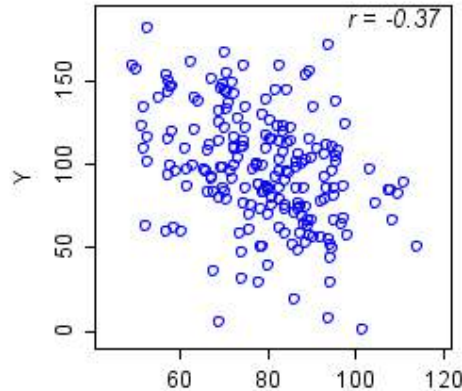
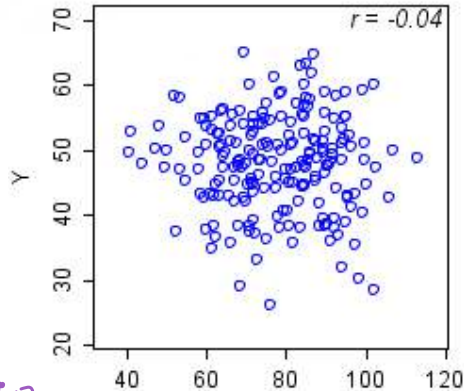
^{สรุป} ^{เปรียบเทียบคุณลักษณะข้อมูล}



Data Mining Functions: (2) Pattern Discovery

การค้นพบรูปแบบ → ที่ต้องการของข้อมูล

- Frequent patterns (or frequent itemsets)
 - What items are frequently purchased together in your Walmart?
- Association and Correlation Analysis



- A typical association rule
 - Diaper → Beer [0.5%, 75%] (support, confidence)
 - 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?

การเชื่อมโยง

Ex. ถ้าคุณซื้อสินค้าไปทางไปทางใกล้ๆ มักจะพบได้เหมือนกัน

Data Mining Functions: (3) Classification

การจำแนกประเภท: 1 ทศ

การทำแบบจำลอง

- Classification and label prediction

สร้างแบบจำลอง

- Construct models (functions) based on some training examples

อธิบาย: 1. 2. 3.

- Describe and distinguish classes or concepts for future prediction

แนวคิด

ทำนายอนาคต

- Ex. 1. Classify countries based on (climate)

จำแนกตามภูมิอากาศ

- Ex. 2. Classify cars based on (gas mileage)

รถตามระยะ: ย. กม

- Predict some unknown class labels

รู้ล่วงหน้า

- Typical methods

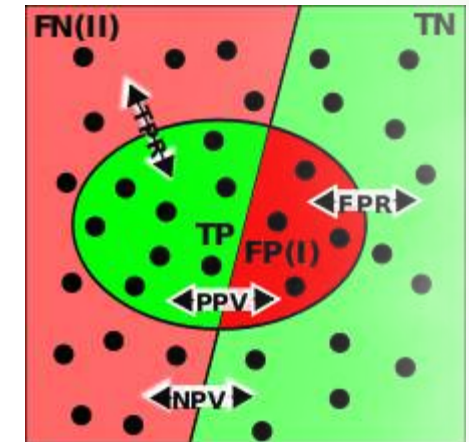
- Decision trees, naïve Bayesian classification, support vector machines, neural networks, rule-based classification, pattern-based classification, logistic regression, ...

- Typical applications:

- Credit card fraud detection, direct marketing, classifying stars, diseases, web-pages, ...

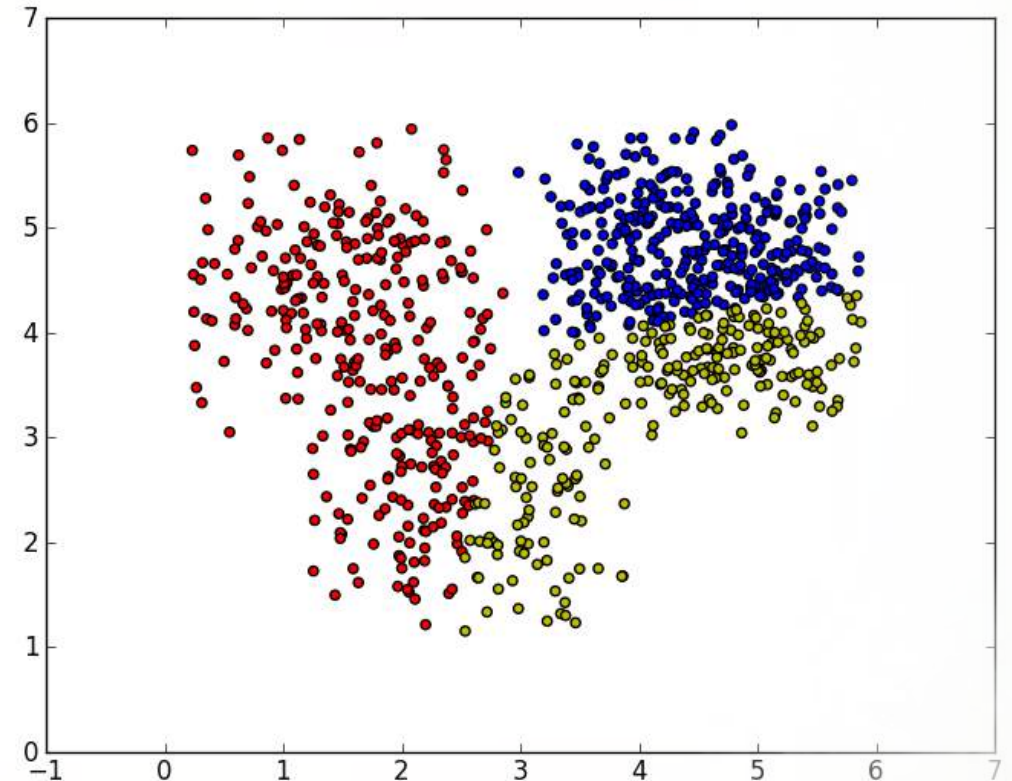
การตรวจจับการฉ้อโกงบัตรเครดิต

การตลาดทางตรง



Data Mining Functions: (4) Cluster Analysis

- Unsupervised learning (i.e., Class label is unknown)
การเรียนรู้แบบไม่มีผู้ดูแล เช่น ไม่รู้จัดประเภทกับ
- Group data to form new categories (i.e., clusters), e.g., cluster houses to find distribution patterns
เพื่อสร้างหมวดหมู่ใหม่
บ้านคลัสเตอร์เพื่อค้นหารูปแบบการกระจาย
- Principle: Maximizing intra-class similarity & minimizing interclass similarity
เพิ่มการคล้ายคลึงในคลาส ลดความแตกต่างระหว่างคลัสเตอร์
- Many methods and applications



Data Mining Functions: (5) Outlier Analysis

- Outlier analysis

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

ไม่สอดคล้องกับพฤติกรรมทั่วไปของข้อมูล

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

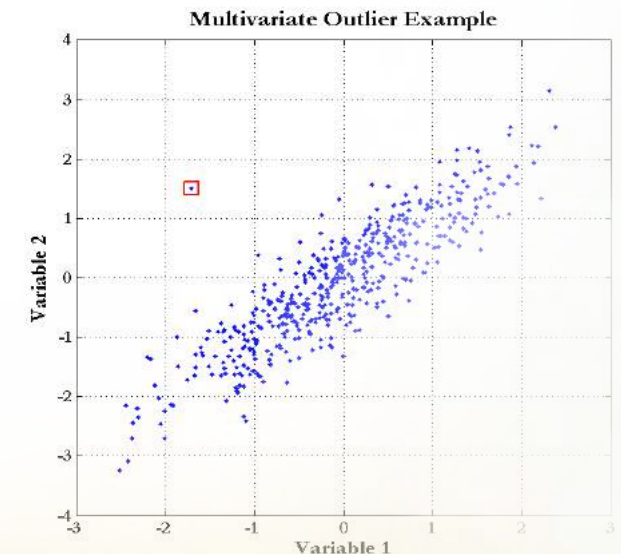
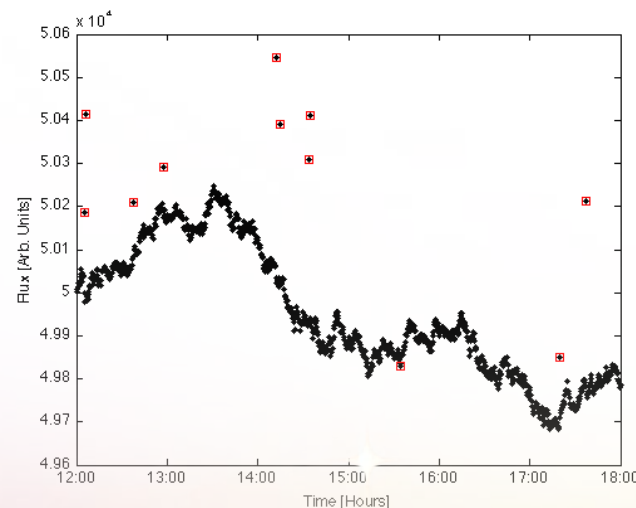
ของใครบางคนเสีย

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

โดยผลคูณของการจัดกลุ่ม และการวิเคราะห์ถดถอย

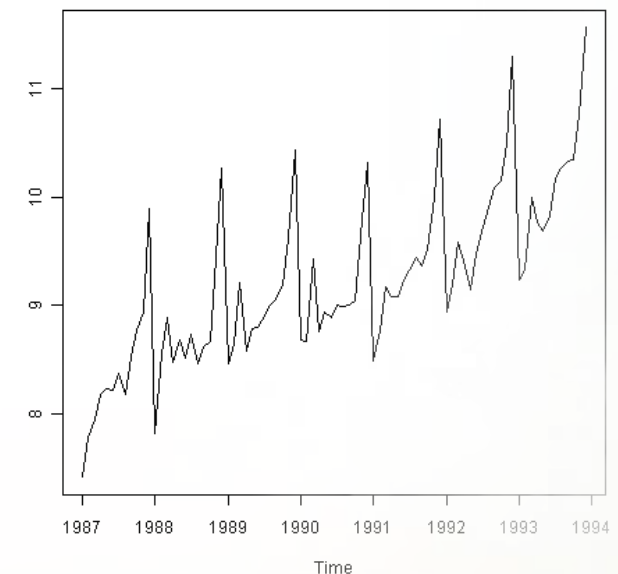
สเปกโตรัมในการตรวจการดัดแปลง , การวิเคราะห์พันธุกรรม

- Useful in fraud detection, rare events analysis



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

- ลำดับ การวิเคราะห์แนวโน้ม และวัดผลการ
• Sequence, trend and evolution analysis
 - แนวโน้ม การเปลี่ยนแปลง
• Trend, time-series, and deviation analysis
 - การตอบ ทำนายผลค่า
• e.g., regression and value prediction
 - Sequential pattern mining
 - e.g., buy digital camera, then buy large memory cards
- การวิเคราะห์เชิงตรรกะ
• Periodicity analysis
- Motifs and biological sequence analysis
 - ลำดับ การวิเคราะห์ลำดับการตั้งภาพ
• Approximate and consecutive motifs
- ตามรูปแบบ
• Similarity-based analysis
- Mining data streams
 - Ordered, time-varying, potentially infinite, data streams



Data Mining Functions: (7) Structure and Network Analysis

การวิเคราะห์โครงสร้างและเครือข่าย

- Graph mining

การขุดค้นกราฟ

 - Finding frequent subgraphs (e.g., chemical compounds), trees (XML), substructures (web fragments)

การค้นหาโครงสร้างย่อยที่พบบ่อย
- Information network analysis

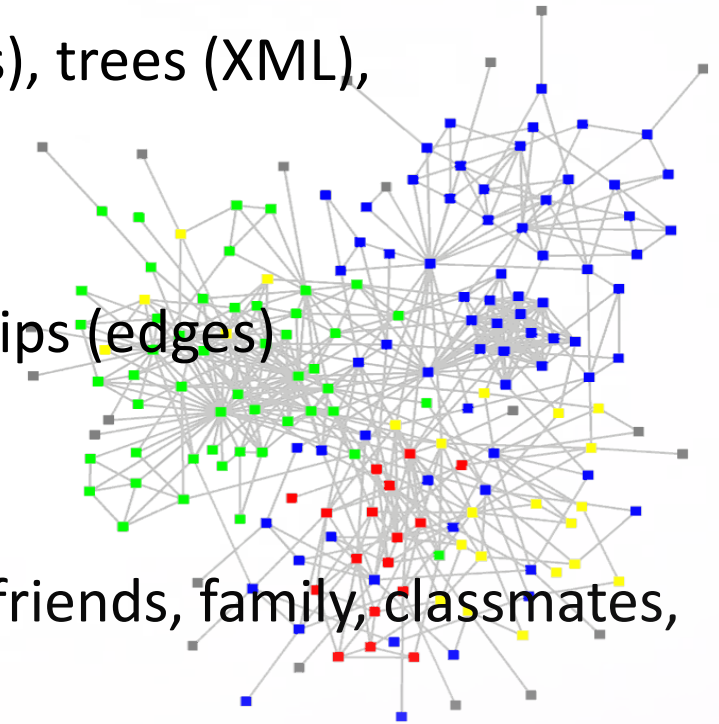
การวิเคราะห์เครือข่ายข้อมูล

 - Social networks: actors (objects, nodes) and relationships (edges)

เครือข่ายสังคม: ผู้แสดง (วัตถุ, โหนด) และความสัมพันธ์ (เส้นเชื่อม)

 - e.g., author networks in CS, terrorist networks

เช่น เครือข่ายผู้เขียนในวิทยาการคอมพิวเตอร์, เครือข่ายก่อการร้าย
 - 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



การประเมินความรู้

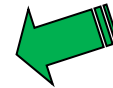
Evaluation of Knowledge

- Are all mined knowledge interesting? ข้อมูลที่ขุดได้มีความน่าสนใจหรือไม่
 - One can mine tremendous amount of “patterns” สามารถขุดได้จำนวนมาก
 - Some may fit only certain dimension space (time, location, ...) บางส่วนอาจสอดคล้องกับพื้นที่บางมิติเท่านั้น
 - Some may not be representative, may be transient, ... ชั่วคราว
- Evaluation of mined knowledge → directly mine only interesting knowledge? การประเมินความรู้ที่ถูกต้อง
 - Descriptive vs. predictive ความครอบคลุมเชิงพรรณนา
 - Coverage การคาดการณ์
 - Typicality vs. novelty แปลกใหม่
 - Accuracy ค.แม่นยำ
 - Timeliness ทันเวลา, ทันสมัย
 - ...

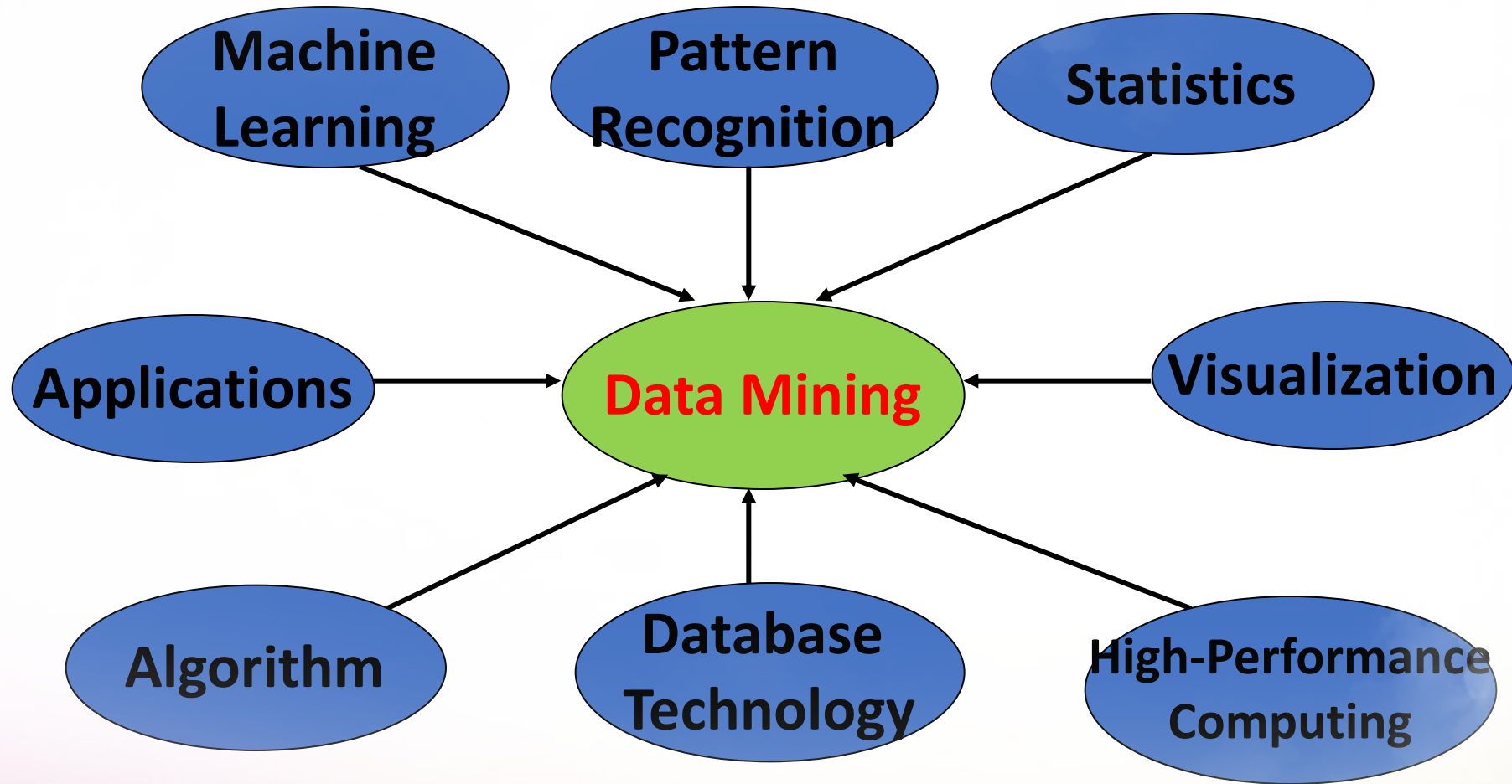


Chapter 1. Introduction

- 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



ဟာဘ

အသံပညာစာ

Why Confluence of Multiple Disciplines?

အလွန်အမင်းများပြားလာသော

- Tremendous amount of data

အလွန်အမင်းများပြားလာသော

- Algorithms must be scalable to handle big data

အလွန်အမင်းများပြားလာသော

- High-dimensionality of data

- Micro-array may have tens of thousands of dimensions

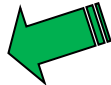
အလွန်အမင်းများပြားလာသော

- 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


การประยุกต์ใช้

Applications of Data Mining

- Web page analysis: classification, clustering, ranking
การวิเคราะห์หน้าเว็บ จำแนก จัดกลุ่ม จัดอันดับ
- Collaborative analysis & recommender systems
วิเคราะห์ร่วมกัน แนะนำ
- Basket data analysis to targeted marketing
เพื่อการตลาดเป้าหมาย
- Biological and medical data analysis
- Data mining and software engineering
- Data mining and text analysis
- Data mining and social and information network analysis
- Built-in (invisible data mining) functions in Google, MS, Yahoo!, Linked, Facebook, ...
ฝังไว้ในตัว การขุดข้อมูลที่มีอยู่แล้ว
- Major dedicated data mining systems/tools
เครื่องมือที่เน้นเฉพาะขุด
 - 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

Major Issues in Data Mining (1) ประเด็นหลักในการขุดข้อมูล

- Mining Methodology วิธีการขุด
 - Mining various and new kinds of knowledge หลากหลายและใหม่
 - Mining knowledge in multi-dimensional space หลายมิติ
 - Data mining: An interdisciplinary effort การผสมผสานแบบสหวิทยาการ
 - Boosting the power of discovery in a networked environment เพิ่มพลังการค้นพบในสภาพแวดล้อมเครือข่าย
 - Handling noise, uncertainty, and incompleteness of data
 - Pattern evaluation and pattern- or constraint-guided mining
- User Interaction การโต้ตอบผู้ใช้
 - Interactive mining
 - Incorporation of background knowledge การรู้พื้นฐาน
 - Presentation and visualization of data mining results แสดงผลจากการขุดข้อมูล

Major Issues in Data Mining (2)

ปรี: ประสิทธิภาพ และ: ความสามารถในการ

- Efficiency and Scalability

ปรับขนาด

- Efficiency and scalability of data mining algorithms
- Parallel, distributed, stream, and incremental mining methods

ความหลากหลาย

- Diversity of data types

- Handling complex types of data
- Mining dynamic, networked, and global data repositories

- Data mining and society

ผลกระทบต่อสังคม

- Social impacts of data mining
- Privacy-preserving data mining

- Invisible data mining

การแอบขโมยข้อมูลโดยไม่เห็น



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.

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



Summary

การขุดข้อมูล: การค้นพบรูปแบบและความรู้ที่น่าสนใจจากข้อมูลจำนวนมาก

- Data mining: Discovering interesting patterns and knowledge from massive amount of data
- A natural evolution of science and information technology, in great demand, with wide applications
วิวัฒนาการทางวิทยาศาสตร์และเทคโนโลยีที่มีความต้องการอย่างมาก
- 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

