

Data Warehouse & Data Mining

Chapter 1

Introduction

• Why Data Mining

Data Mining = ศาสตร์เพื่อการค้นหา / ศาสตร์ของข้อมูล
 ในปัจจุบัน Data ที่มีอยู่มาก แหล่งมาจากการเก็บมาได้
 เช่น เทคนิคการจดจำรูปแบบ, เทคนิคการจำแนก簇 หรือวิธีการที่มี MS-
 เก็บข้อมูลที่มีอยู่แล้ว ทำให้ดัน
 รวมไปถึง website ที่มี ลักษณะที่ต้องการที่ๆ จึง website ที่ต้อง^{ห้าม}
Website ที่มี = ถูกจัดทำมาเพื่อปะปนไปในเดือนสุด
ก็คือ ภารกิจที่มุ่งหวังนี้ หลักๆ เพื่อประโยชน์ของบุคลากรที่จะ = ควบ
คุมชุมชนไว้ในเมืองเดียว ดังเช่นความสามารถของเพื่อนมนต์
จะต้องเป็นคนของชาติที่มีประวัติ良好

What is Data Mining

Data Mining คือ การวิเคราะห์ข้อมูลจากข้อมูลจำนวนมาก (big data) เพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ โดยทำการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และหาความน่าจะเป็นที่จะเกิดขึ้น เพื่อให้ไดองค์ความรู้ใหม่ ที่สามารถนำไปใช้ประกอบการตัดสินใจในด้านต่างๆ เช่น ตลาดหลักทรัพย์, ทางธุรกิจ, ทางด้านการแพทย์, ยุทธศาสตร์ทหาร

กิจกรรม

knowledge discovery in databases "มรรคกิจกรรมค้นพบข้อมูลใน Databases"

knowledge extraction "มรรคกิจกรรมค้นพบ"

data/pattern analysis "มรรคกิจกรรมค้นพบรูปแบบ"

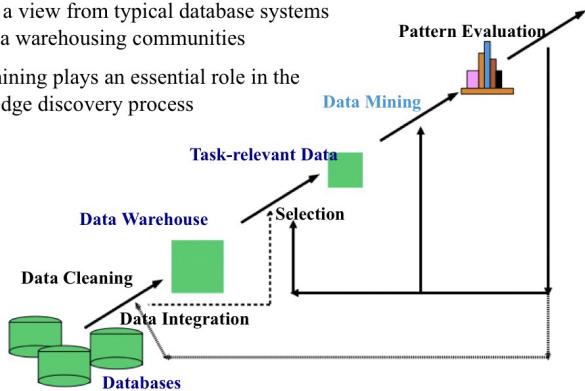
data archaeology "กลับไปค้นหากองโบราณ"

Business Intelligence "สร้างองค์ความรู้กับ data แล้วนำไปใช้"

Knowledge Discovery (KDD) Process

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

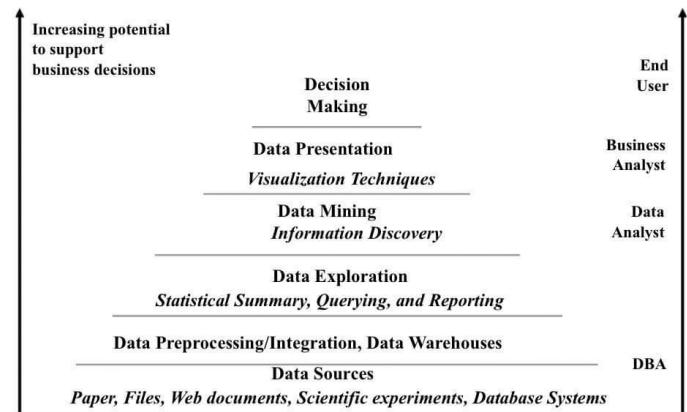


ક્રિયાનુંદિત ડેટા

ખાંડન કરી રહેલી ડેટા

- Data Cleaning એવી ખાંડન કરી રહેલી કાર્યક્રમ હૈ જે અસ્થિર ડેટાનું પોતું કરી શકે હોય.
- Data Integration એવી ખાંડન કરી રહેલી કાર્યક્રમ હૈ જે વિવિધ સ્તરીય ડેટાનું એક સાથે પોતું કરી શકે હોય.
- Data Selection એવી ખાંડન કરી રહેલી કાર્યક્રમ હૈ જે વિવિધ ડેટાનું એક સાથે પોતું કરી શકે હોય.
- Data Transformation એવી ખાંડન કરી રહેલી કાર્યક્રમ હૈ જે વિવિધ ડેટાનું એક સાથે પોતું કરી શકે હોય.
- Data Mining એવી ખાંડન કરી રહેલી કાર્યક્રમ હૈ જે વિવિધ ડેટાનું એક સાથે પોતું કરી શકે હોય.
- Pattern Evaluation એવી ખાંડન કરી રહેલી કાર્યક્રમ હૈ જે વિવિધ ડેટાનું એક સાથે પોતું કરી શકે હોય.
- Knowledge Representation એવી ખાંડન કરી રહેલી કાર્યક્રમ હૈ જે વિવિધ ડેટાનું એક સાથે પોતું કરી શકે હોય.

Data Mining in Business Intelligence



Data Mining ມີຕົວນັ້ນຍິ່ງ 3 ຊູ່

1. Pattern discovery = ມູຈຸລາວືອິຫຼາຕົວ
2. Classification = ຄົນຄ່ານິມີນ
3. Clustering = ລົກລົມຕົວ

How the data suppose to look like

columns (ລາຍລະອຽດ) = Attributes, Fields, Features = ດ້ວຍບັນດາຮັບສິນທີ່ມີກົດໝາຍ

	id	name	domain_id	closed	city_name	zipcode	geohash	new_open	weighted_average_rating	number_of_chains	...	good_for_groups
0	2	ພະຍົກສັນຕິພາບ	2	0	Samut Songkhram	75000	w4rh7g3	0	5.000000	NaN	...	NaN
1	4	Coffee House	1	0	Bangkok Metropolitan Region	12150	w4rx73h	0	2.000000	NaN	...	NaN
2	5	ທະນາຄານຫຼາຍ	4	0	Phra Nakhon Si Ayutthaya	13000	w4x98jk	0	4.000000	NaN	...	NaN
3	6	ເນັ້ນເຄີຍກໍາໄກ	1	0	Bangkok Metropolitan Region	107000	w4rqw9q	0	0.000000	NaN	...	NaN
4	7	Buono Caffe	1	0	Bangkok Metropolitan	10220	w4rx4gd	0	3.738462	NaN	...	NaN

Row (ແളນ)

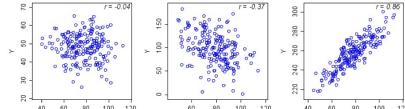
Records, Data point
= ຖະໜາຍອັນດັບ

ໃໝ່ ດາວໂຫຼວງ Data Mining

①

Data Mining Functions: (2) Pattern Discovery

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



- A typical association rule → *Association Data Mining*
 - Diaper \square 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?

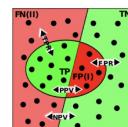
Association rule

ມີອອນກໍາໄລນັ້ນຕົວ \Rightarrow Diaper & beer ໃຫດວ່າເວັບໄຊນີ້ມີຄວາມ ຢຳຮັດຂຶ້ນທີ່ມີຄວາມ
ອະນຸຍາວມີມີຄວາມຮັດກົງ

②

Data Mining Functions: (3) Classification ມຽນແຈງຂອງມີ

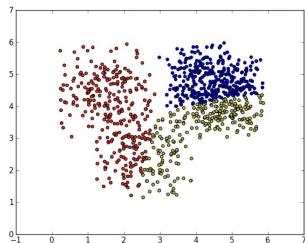
- Classification and label prediction
 - Construct models (functions) based on some training examples
 - Describe and distinguish classes or concepts for future prediction
 - Ex. 1. Classify countries based on (climate)
 - Ex. 2. Classify cars based on (gas mileage)
 - Predict some unknown class labels
- 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, ...



①

Data Mining Functions: (4) Cluster Analysis

- 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



ກົດລົມມູນຈຸດກົມ
ນໍາຈຸດກົມ

ຮັບມືວ່າມີຄວາມຮັດກົງ