Introduction to Data Mining
資料探勘導論

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系

http://mail.tku.edu.tw/myday/
2013-02-21
淡江大學101學年度第2學期
課程教學計畫表
(2013.02 - 2013.06)

• 課程名稱：資料探勘 (Data Mining)
• 授課教師：戴敏育 (Min-Yuh Day)
• 開課系級：資管四P (TLMXB4P)
• 開課資料：選修 單學期 2 學分 (2 Credits, Elective)
• 上課時間：週四 9,10 (Thu 16:10-18:00)
• 上課教室：B216
Data Mining at the Intersection of Many Disciplines

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Knowledge Discovery (KDD) Process

Data mining: core of knowledge discovery process

Data Cleaning

Data Integration

Data Warehouse

Task-relevant Data

Selection

Data Mining

Pattern Evaluation

Source: Han & Kamber (2006)
Increasing potential to support business decisions

- Decision Making
- Data Presentation
- Visualization Techniques
- Data Mining
- Information Discovery
- Data Exploration
- Statistical Summary, Querying, and Reporting
- Data Preprocessing/Integration, Data Warehouses
- Data Sources

Data Sources:
- Paper, Files, Web documents, Scientific experiments, Database Systems

Source: Han & Kamber (2006)
Business Pressures–Responses–Support Model

**Business Environmental Factors**
- Globalization
- Customer demand
- Government regulations
- Market conditions
- Competition
- Etc.

**Organization’s Responses**
- Strategy
- Partners’ collaboration
- Real-time response
- Agility
- Increased productivity
- New vendors
- New business models
- Etc.

**Decisions and Support**
- Analyses
- Predictions
- Decisions
  - Integrated computerized decision support
  - Business intelligence

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
課程簡介

• 本課程介紹資料探勘 (Data Mining) 的基礎概念及應用技術。

• 課程內容包括
  - 資料探勘導論
  - 關連分析
  - 分類與預測
  - 分群分析
  - SAS企業資料採礦實務與認證
  - 資料探勘個案分析與實作
Course Introduction

• This course introduces the fundamental concepts and applications technology of data mining.

• Topics include
  – Introduction to Data Mining
  – Association Analysis
  – Classification and Prediction
  – Cluster Analysis
  – Data Mining Using SAS Enterprise Miner
  – Case Study and Implementation of Data Mining
課程目標
(Objective)

• 瞭解及應用資料探勘基本概念與技術。

• Understand and apply the fundamental concepts and technology of data mining.
<table>
<thead>
<tr>
<th>週次</th>
<th>日期</th>
<th>內容 (Subject/Topics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102/02/21</td>
<td>資料探勘導論 (Introduction to Data Mining)</td>
</tr>
<tr>
<td>2</td>
<td>102/02/28</td>
<td>和平紀念日 (放假一天) (Peace Memorial Day) (No Classes)</td>
</tr>
<tr>
<td>3</td>
<td>102/03/07</td>
<td>關連分析 (Association Analysis)</td>
</tr>
<tr>
<td>4</td>
<td>102/03/14</td>
<td>分類與預測 (Classification and Prediction)</td>
</tr>
<tr>
<td>5</td>
<td>102/03/21</td>
<td>分群分析 (Cluster Analysis)</td>
</tr>
<tr>
<td>6</td>
<td>102/03/28</td>
<td>SAS企業資料採礦實務 (Data Mining Using SAS Enterprise Miner)</td>
</tr>
<tr>
<td>7</td>
<td>102/04/04</td>
<td>清明節、兒童節 (放假一天) (Children's Day, Tomb Sweeping Day) (No Classes)</td>
</tr>
<tr>
<td>8</td>
<td>102/04/11</td>
<td>個案分析與實作一 (SAS EM 分群分析)：Banking Segmentation (Cluster Analysis – K-Means using SAS EM)</td>
</tr>
<tr>
<td>週次</td>
<td>日期</td>
<td>內容 (Subject/Topics)</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>9</td>
<td>102/04/18</td>
<td>期中報告 (Midterm Presentation)</td>
</tr>
<tr>
<td>10</td>
<td>102/04/25</td>
<td>期中考試週</td>
</tr>
<tr>
<td>11</td>
<td>102/05/02</td>
<td>個案分析與實作二 (SAS EM 關連分析)：Web Site Usage Associations (Association Analysis using SAS EM)</td>
</tr>
<tr>
<td>12</td>
<td>102/05/09</td>
<td>個案分析與實作三 (SAS EM 決策樹、模型評估)：Enrollment Management Case Study (Decision Tree, Model Evaluation using SAS EM)</td>
</tr>
<tr>
<td>13</td>
<td>102/05/16</td>
<td>個案分析與實作四 (SAS EM 迴歸分析、類神經網路)：Credit Risk Case Study (Regression Analysis, Artificial Neural Network using SAS EM)</td>
</tr>
<tr>
<td>14</td>
<td>102/05/23</td>
<td>期末專題報告 (Term Project Presentation)</td>
</tr>
<tr>
<td>15</td>
<td>102/05/30</td>
<td>畢業考試週</td>
</tr>
</tbody>
</table>
教學方法與評量方法

• 教學方法
  - 講述、討論、實作

• 評量方法
  - 實作、報告、上課表現
教材課本

• 講義 (Slides)
• 參考書籍
  – 決策支援與企業智慧系統，九版，Efraim Turban 等著，李昇暾審定，2011，華泰
作業與學期成績計算方式

• 批改作業篇數
  – 2篇（Team Term Project）

• 學期成績計算方式
  – 期中評量：30 ％ (期中報告)
  – 期末評量：30 ％ (期末報告)
  – 其他（課堂參與及報告討論表現）：40 ％
Team Term Project

• Term Project Topics
  – Data mining
  – Web mining
  – Business Intelligence

• 3-5 人为一组
  – 分组名单于 2013.03.07 (四) 老师下课时缴纳
  – 由班代统一收集协调分组名单
# A Taxonomy for Data Mining Tasks

<table>
<thead>
<tr>
<th>Data Mining</th>
<th>Learning Method</th>
<th>Popular Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prediction</strong></td>
<td>Supervised</td>
<td>Classification and Regression Trees, ANN, SVM, Genetic Algorithms</td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td>Supervised</td>
<td>Decision trees, ANN/MLP, SVM, Rough sets, Genetic Algorithms</td>
</tr>
<tr>
<td><strong>Regression</strong></td>
<td>Supervised</td>
<td>Linear/Nonlinear Regression, Regression trees, ANN/MLP, SVM</td>
</tr>
<tr>
<td><strong>Association</strong></td>
<td>Unsupervised</td>
<td>Apriory, OneR, ZeroR, Eclat</td>
</tr>
<tr>
<td><strong>Link analysis</strong></td>
<td>Unsupervised</td>
<td>Expectation Maximization, Apriory Algorithm, Graph-based Matching</td>
</tr>
<tr>
<td><strong>Sequence analysis</strong></td>
<td>Unsupervised</td>
<td>Apriory Algorithm, FP-Growth technique</td>
</tr>
<tr>
<td><strong>Clustering</strong></td>
<td>Unsupervised</td>
<td>K-means, ANN/SOM</td>
</tr>
<tr>
<td><strong>Outlier analysis</strong></td>
<td>Unsupervised</td>
<td>K-means, Expectation Maximization (EM)</td>
</tr>
</tbody>
</table>

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
The Evolution of BI Capabilities

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
A High-Level Architecture of BI

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Mining the Social Web:
Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

Source: http://www.amazon.com/Mining-Social-Web-Analyzing-Facebook/dp/1449388345
Web Mining Success Stories

• Amazon.com, Ask.com, Scholastic.com, …
• Website Optimization Ecosystem

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
## Top 10 CIO Technology Priorities in 2013

<table>
<thead>
<tr>
<th>Top 10 Technology Priorities</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytics and business intelligence</td>
<td>1</td>
</tr>
<tr>
<td>Mobile technologies</td>
<td>2</td>
</tr>
<tr>
<td>Cloud computing (SaaS, IaaS, PaaS)</td>
<td>3</td>
</tr>
<tr>
<td>Collaboration technologies (workflow)</td>
<td>4</td>
</tr>
<tr>
<td>Legacy modernization</td>
<td>5</td>
</tr>
<tr>
<td>IT management</td>
<td>6</td>
</tr>
<tr>
<td>CRM</td>
<td>7</td>
</tr>
<tr>
<td>Virtualization</td>
<td>8</td>
</tr>
<tr>
<td>Security</td>
<td>9</td>
</tr>
<tr>
<td>ERP Applications</td>
<td>10</td>
</tr>
</tbody>
</table>

Source: Gartner Executive Programs (January 2013)
[http://www.gartner.com/newsroom/id/2304615](http://www.gartner.com/newsroom/id/2304615)
## Top 10 CIO Business Priorities in 2013

The following table lists the top 10 business priorities in 2013, along with their rankings:

<table>
<thead>
<tr>
<th>Top 10 Business Priorities</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasing enterprise growth</td>
<td>1</td>
</tr>
<tr>
<td>Delivering operational results</td>
<td>2</td>
</tr>
<tr>
<td>Reducing enterprise costs</td>
<td>3</td>
</tr>
<tr>
<td>Attracting and retaining new customers</td>
<td>4</td>
</tr>
<tr>
<td>Improving IT applications and infrastructure</td>
<td>5</td>
</tr>
<tr>
<td>Creating new products and services (innovation)</td>
<td>6</td>
</tr>
<tr>
<td>Improving efficiency</td>
<td>7</td>
</tr>
<tr>
<td>Attracting and retaining the workforce</td>
<td>8</td>
</tr>
<tr>
<td>Implementing analytics and big data</td>
<td>9</td>
</tr>
<tr>
<td>Expanding into new markets and geographies</td>
<td>10</td>
</tr>
</tbody>
</table>

Source: Gartner Executive Programs (January 2013)
http://www.gartner.com/newsroom/id/2304615
Summary

- This course introduces the fundamental concepts and applications technology of data mining.

- Topics include
  - Introduction to Data Mining
  - Association Analysis
  - Classification and Prediction
  - Cluster Analysis
  - Data Mining Using SAS Enterprise Miner
  - Case Study and Implementation of Data Mining
Contact Information

戴敏育博士 (Min-Yuh Day, Ph.D.)

專任助理教授
淡江大學 資訊管理學系

電話：02-26215656 #2347
傳真：02-26209737
研究室：i716 (覺生綜合大樓)
地址：25137 新北市淡水區英專路151號
Email：myday@mail.tku.edu.tw
網址：http://mail.tku.edu.tw/myday/