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/
2014-02-19
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：週三 6,7 (Wed 13:10-15:00)
上課教室：B216
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
  – 文字探勘與網頁探勘
  – 海量資料分析
  – SAS企業資料採礦實務與認證 (SAS EM)
  – 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
  – Text and Web Mining
  – Big Data Analytics
  – Data Mining Using SAS Enterprise Miner (SAS EM)
  – 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 (Week)</th>
<th>日期 (Date)</th>
<th>內容 (Subject/Topics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>103/02/19</td>
<td>資料探勘導論 (Introduction to Data Mining)</td>
</tr>
<tr>
<td>2</td>
<td>103/02/26</td>
<td>關連分析 (Association Analysis)</td>
</tr>
<tr>
<td>3</td>
<td>103/03/05</td>
<td>分類與預測 (Classification and Prediction)</td>
</tr>
<tr>
<td>4</td>
<td>103/03/12</td>
<td>分群分析 (Cluster Analysis)</td>
</tr>
<tr>
<td>5</td>
<td>103/03/19</td>
<td>個案分析與實作一 (SAS EM 分群分析)：Case Study 1 (Cluster Analysis – K-Means using SAS EM)</td>
</tr>
<tr>
<td>6</td>
<td>103/03/26</td>
<td>個案分析與實作二 (SAS EM 關連分析)：Case Study 2 (Association Analysis using SAS EM)</td>
</tr>
<tr>
<td>7</td>
<td>103/04/02</td>
<td>教學行政觀摩日 (Off-campus study)</td>
</tr>
<tr>
<td>8</td>
<td>103/04/09</td>
<td>個案分析與實作三 (SAS EM 決策樹、模型評估)：Case Study 3 (Decision Tree, Model Evaluation using SAS EM)</td>
</tr>
<tr>
<td>週次 (Week)</td>
<td>日期 (Date)</td>
<td>內容 (Subject/Topics)</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>9</td>
<td>103/04/16</td>
<td>期中報告 (Midterm Project Presentation)</td>
</tr>
<tr>
<td>10</td>
<td>103/04/23</td>
<td>期中考試週 (Midterm Exam)</td>
</tr>
<tr>
<td>11</td>
<td>103/04/30</td>
<td>個案分析與實作四 (SAS EM 迴歸分析、類神經網路)：Case Study 4 (Regression Analysis, Artificial Neural Network using SAS EM)</td>
</tr>
<tr>
<td>12</td>
<td>103/05/07</td>
<td>文字探勘與網頁探勘 (Text and Web Mining)</td>
</tr>
<tr>
<td>13</td>
<td>103/05/14</td>
<td>海量資料分析 (Big Data Analytics)</td>
</tr>
<tr>
<td>14</td>
<td>103/05/21</td>
<td>期末報告 (Final Project Presentation)</td>
</tr>
<tr>
<td>15</td>
<td>103/05/28</td>
<td>畢業考試週 (Final Exam)</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

• 作業篇數
  – 3篇

• 學期成績計算方式
  – 期中評量：30 ％
  – 期末評量：30 ％
  – 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

• 3-4 人為一組
  – 分組名單於 2014/03/05 (三) 課程下課時繳交
  – 由班代統一收集協調分組名單
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
Databases

Data Warehouse

Task-relevant Data
Selection

Data Mining

Pattern Evaluation

Source: Han & Kamber (2006)
Data Warehouse

Data Mining and Business Intelligence

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
Paper, Files, Web documents, Scientific experiments, Database Systems

End User

Business Analyst

Data Analyst

DBA

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

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
A Taxonomy for Data Mining Tasks

<table>
<thead>
<tr>
<th>Learning Method</th>
<th>Popular Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>Classification and Regression Trees, ANN, SVM, Genetic Algorithms</td>
</tr>
<tr>
<td>Supervised</td>
<td>Decision trees, ANN/MLP, SVM, Rough sets, Genetic Algorithms</td>
</tr>
<tr>
<td>Supervised</td>
<td>Linear/Nonlinear Regression, Regression trees, ANN/MLP, SVM</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>Apriory, OneR, ZeroR, Eclat</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>Expectation Maximization, Apriory Algorithm, Graph-based Matching</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>Apriory Algorithm, FP-Growth technique</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>K-means, ANN/SOM</td>
</tr>
<tr>
<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
Business Intelligence Trends

1. Agile Information Management (IM)
2. Cloud Business Intelligence (BI)
3. Mobile Business Intelligence (BI)
4. Analytics
5. Big Data

Business Intelligence Trends: Computing and Service

- Cloud Computing and Service
- Mobile Computing and Service
- Social Computing and Service
Business Intelligence and Analytics

• Business Intelligence 2.0 (BI 2.0)
  – Web Intelligence
  – Web Analytics
  – Web 2.0
  – Social Networking and Microblogging sites

• Data Trends
  – Big Data

• Platform Technology Trends
  – Cloud computing platform

Business Intelligence and Analytics: Research Directions

1. Big Data Analytics
   - Data analytics using Hadoop / MapReduce framework

2. Text Analytics
   - From Information Extraction to Question Answering
   - From Sentiment Analysis to Opinion Mining

3. Network Analysis
   - Link mining
   - Community Detection
   - Social Recommendation

Data Scientist: 
*The Sexiest Job of the 21st Century*

Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren’t seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don’t know anyone. So you just stand in the corner sipping your drink—and you probably leave early.”
Big Data 2014

http://www.ettoday.net/events/bigdata2014/
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
  – Text and Web Mining
  – Big Data Analytics
  – Data Mining Using SAS Enterprise Miner (SAS EM)
  – Case Study and Implementation of Data Mining
Contact Information

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.

Min-Yuh Day, Ph.D.