

22:010:622
*Internet Technology and
E-Business*

Dr. Peter R. Gillett

Associate Professor

**Department of Accounting & Information Systems
Rutgers Business School – Newark & New Brunswick**

Overview

- Questions regarding Group Projects?
- Data Mining
- The Myth of Internet Time (by A. Odlyzko)
- SpeedTracer
- Data Attributes
- Electronics and the Future of Education?
- Mbone and Multicasting

Data Mining

- What is data mining?
- How is it different from sales analysis?
- What are the necessary components for data mining?
- Secondary data
- How can we supplement this secondary data with primary data?

Why Data Mining?

- Sell more!
- Build customer loyalty
 - * *What is customer loyalty?*
 - * *How can we measure it?*
- Sell more at higher margin!
- Sell to gain market share!
- Don't leave money on the table!
- Focus on your profit centers!

Why Data Mining?

- Just as with Time Series Analysis
 - * Descriptive: what happened
 - * Explanation: why this lead to a sale
 - * Predictive: anticipate what will happen next

Data Mining

- From “*The Information Gold Mine*”
Business Week: 26-July-1999
- Almost instant credit-card approval
 - * Does the usual checking via computer, the Internet and phone
 - * Perhaps even more consistent than humans
 - * If it is programmed wrong, then more damage may be done than with individual humans

Data Mining

- This almost-instant credit card is an example of the power and speed of complex information processing on the Internet
- Is this the tip of the Web Data Gold Rush?
- Wanamaker: “Half my advertising is wasted, I just don’t know which half”
- Does the Internet offer a good answer to Wanamaker?

Data Mining

- Targeting highest yielding customers: with one-to-one marketing
- We already know that banner ads don't do as much for us as we had hoped.
- How about other approaches?
- What if the Internet keeps converging with other media?
- Yahoo collects 400 billion bytes/day of such information! This is the data in 800,000 books!

Data Mining

- Focus on targeted personalization
- Is this the beginning of custom commoditization?
- That is, is customization becoming a commodity?
- What does this mean about manufacturing?
 - * Modularization
 - * Standards
 - * Just-in-time

Data Mining

- Privacy issues abound
- DoubleClick and Abacus Direct
 - * Web click tracking combined with (household) direct order purchase information
- Whose data is this anyway?
- E-commerce needs lots of development: only 25% of manufacturers track their web customer's data at all

Data Mining

- Tracking your window shoppers
- E.piphany, Responsys.com, Blue Martini
- Track, organize and analyze web visitor data
- Forrester Research says:
 - * 16% of large firms expected more efficient use of customer data to give cost savings in 1999
 - * 20% expected use of data to boost revenues
 - * 34% of large firms expect cost savings by now
 - * 74% expected boosts in revenues by now

Data Mining

- It's not just about *any* customer - it is about finding the good customers!
 - * High yielders
 - * Let the rest (more costly customers), go to your competition!
- What industries will likely be (or have been) the first to see gains from data mining over the Internet?

Data Mining

- Government Computer Sales, Inc.
- Created profiles of 3,600 Government Departments in 6 states from on-line & traditional sales
- 60% of GCSI sales are on-line
- This data make it possible for them to get their clients to more than double their purchases
- Raised the margins from 7% (1997) to 11% (1998)
- Where did the extra purchases come from?

Data Mining

- Consolidated Freightways
- Small customers may not know the ins-and-outs of their system
- They placed a pop-up help-agent for people to use their system better
- Small business brings 10% profit margin where large business gives only 3%
- They increased their small business customers

Data Mining

- A classic bond house: PIMCO Funds
- Store and segment site visitors' profiles
- Generate offers and suggest funds that suit the profile
- 700,000 investor records supplement their data
- About 30 proposals that average \$250,000 are generated every day!
- Sales proposals or bids computer generated for the customer's preferences!

Data Mining

- Click-through successes
- SmarterKids.com used AdKnowledge, Inc.
- Found that MSN, Yahoo and Family Education Network got the most clicks for the ad dollar
- Half of SmarterKids.com's customers bought after viewing on-line ads!

Data Mining

- Where is the value added?
 - * Click-throughs?
 - * Data?
- BabyCenter's experience:
 - * Shot-gun email links got 6% click through
 - * Target baby age groups: 25% click through with between 1% & 2% making a purchase

Data Mining

- Classical techniques
 - * Timely and relevant “subject” header
 - * Make it very easy to buy from you!
 - * The usual WIIFM!
- On-line modification of web sites to suit different visitors
- KeyCorp: use data on 7 million customers to generate different web offers for different customers

Data Mining

- Forrester Research surveyed 54 on-line retail firms
 - * 39% said combining data from different systems is difficult
 - * Firms are “drowning” in data
- Must address explicit and implicit privacy concerns!

Data Mining

- Try to understand:
 - * Needs
 - * Preferences
 - * Behaviors
- Data Mining
 - * Pricing
 - * Promotion
 - * Product/offer development

Data Mining

- Chase: lower minimum balance for checking account requirement
 - * Customers go if they have trouble maintaining the minimum balance
 - * Are such customers profitable anyway?
 - * Yes: these customers are more profitable than other segments
 - * Keep them!

Data Mining

- This last Chase example shows that data mining is useful for both finding interesting segments and analyzing these segments
- Other uses?

Data Mining

- CRM: Customer Relations Management
 - * Customer managed relationships!
 - * Selectively market to different customer segments:
Fleet Bank
 - ◆ Keep costs low
 - ◆ Keep chance of getting more business high!
- Credit Risk Management
 - * Default
 - * Bankruptcy

Data Mining

■ Bank America

- * Customer Service Call Center - Marketing
- * Not Product of the Week
- * Rather, data-segmented promotions on a one-to-one Basis

■ Also, manage credit portfolios better

- * Credit Risk Analysis
- * Credit Scoring

■ Manage loans better

Data Mining

- How does all this work?
 - * Statistical algorithms
 - * Neural Networks
 - * Genetic Algorithms
 - * Fuzzy Logic
 - * etc.
- How does this differ from OLAP etc?

Data Mining

- Collecting data
- Cleansing data
- Enforcing data consistency
- Operational databases v. data warehousing
- Levels of servers (drill-down, functional . . .)
- Parallel processing
 - * *Symmetric multiprocessing (SMP)*
 - * *Massively parallel processing (MPP)*

Data Mining

- Query-and-reporting tools
 - * Require databases
 - * Human direction
- Multidimensional analysis
 - * Less human guidance
 - * Data in special forms
- Intelligent agents

The Myth of Internet Time

- *A. Odlyzko: Technology Review*
- Is the Internet to the year 2000 as CB radio is to the 1980s?
- Over-optimism to over-pessimism
- Over 8 weeks in 1999, Venture Capitalists funded 7 pet portal firms
- Hindsight shows none of them could have really survived!

The Myth of Internet Time

■ Key myths:

- * Internet Time: product development and consumer acceptance happen much quicker than usual

■ Let's define the following:

- * First-mover advantage
- * Network effects
- * Ever increasing returns
- * Control of open standards
- * Standards lock-in

The Myth of Internet Time

- All this together overwhelmed due diligence
- All well-believed myths have *some* truth
- Examples of some truth behind:
 - * *First mover advantage*
 - * *Internet time*
 - * *Network effects*
- How about counter-examples?

The Myth of Internet Time

- Is it the product development cycle that has sped up and not the consumer demand?
- J.C.R. Licklider (1965): “A modern maxim says: ‘People tend to overestimate what can be done in a year and underestimate what can be done in five or ten years’ ”
- It is argued that it takes a decade for new products/services to be readily accepted.
- Adaptability of businesses is necessary for survival, in this light
- *People do not operate on Internet time!*

SpeedTracer: Web Usage Miner

- K.-L. Wu, P.S. Yu, and A. Ballman, IBM
- SpeedTracer is a WWW usage and analysis tool for understanding web surfers
- Web logs are often incomplete and only form a partial picture
 - * First reconstruct user traversal path
 - * No cookies required, no user privacy violation
 - * Data mining and algorithms to find pages that are used together and the most common paths to these pages

SpeedTracer

■ Typical Server logs contain

1. Domains
2. Countries
3. Companies
4. Individuals (machine's IP addresses)
5. Click-throughs
6. Purchases

SpeedTracer

- Also timing logs often contain:
 1. How long on site
 2. How long on each page
 3. When? Minute, hour, day, month, year
 4. Repeat visits

SpeedTracer

- The data in the typical logs does not tell enough about the surfer behavior!
- Also, because of proxy servers, firewalls, etc., much data is muted or not tied together!
 - * For example, different IP addresses for the same person (or vice versa!)
- There are existing tools for analyzing the web server logs, but they fall short
 - * Why?

SpeedTracer

- To really see what surfers are doing we need to identify them and associate their data in different sessions!
- First: separate user sessions from agent's sessions
 - * How can we distinguish these two?
 - * The order the branches are traveled through
 - * The thoroughness the branches are traveled

SpeedTracer

- Methods Agents or Bots use to traverse web pages (like searching through a maze)
 - * **Depth First Search**
 - ◆ Deep as possible, then backtrack
 - * **Breadth First Search**
 - ◆ One level at a time
- How should you design your web site relative to depth-first and breadth-first?
- Such analysis is useful for the structure of your page and the details too

SpeedTracer

■ Take user sessions

- * Top N referring web pages to your site
- * Top N pages most frequently visited
- * Top N pages most frequently exited from
- * Top N browsers most often used
- * Top N IP hosts from where most users visit
- * The usual web log statistics

SpeedTracer

- With this data, standard Data Mining algorithms can be applied!

Attributes (Based on Fishbein)

1. Price (acquisition cost and maintenance cost)
2. Quality
3. Reputation
4. Longevity
5. Location
6. Relationship
7. Offering
8. Terms/Conditions

Attributes for a Computer

1. Price
2. Quality
3. Disk Memory
4. RAM
5. Internet
6. Screen
7. Type (Desktop/Laptop/Notebook)
8. Service
9. Software

What can we do with Attributes?

- Primary data analysis
- Score the relative importance of each attribute
- Score how a handful of the most significant brands do on these attributes
- Take the dot-product: see which brands work best with the attributes!
- This is a simulation
- What about the Internet?

Attribute Based Loyalty

	X is the primary brand	X is NOT the primary brand
X Chosen in simulation	Loyal customers	Low cost acquisitions
X NOT chosen in simulation	Vulnerable customers	High cost acquisitions

Electronics and Education

- By A. Odlyzko
- Will the E-Revolution make education cheaper and the infrastructure smaller?
- Can a giant jump in bandwidth eliminate some large need for lecturing?
 - * Consider 3D imaging of famous lecturers blasted over the Internet
 - * Interactive discussions across continents

Electronics and Education

- This gives more opportunity for one-to-one teacher-student relationships
- It is noted that despite all the communication equipment (fax, electronic conference calls, email, etc), corporate travel budgets are up 20% over the last 6 years
- Certainly, some business trips have been avoided, but the new level of competition has pushed the envelope!

Electronics and Education

- As countries become wealthier, they spend more on education (something like medical care)
- The more complex society, the more education becomes necessary
- Education as value-added
- It will probably continue to grow in expense and size, despite the Internet

Multicasting and the Mbone

■ Unicast

- * Computer-to-computer

■ Broadcast

- * Computer-to-everyone

■ Multicast

- * Computer-to-select-group

Multicasting and the Mbone

■ Multicast Backbone (Mbone)

- * Virtual network layered on top of the Internet
- * An extension to the Internet to support *IP Multicasting* -- two-way transmission of data between multiple sites
- * Handles live multimedia messages
 - ◆ Video Conferencing
 - ◆ Real-time data distribution (e.g., Stock Prices)