Customer Analytics

Data Mining for CRM: an overview

MONASH
cssLAB
DECISION SUPPORT SYSTEMS LABORATORY

Architecture for Analytical CRM

Customer Data Warehouse

Operational systems and external info

Prediction and Discovery
- Data Mining

Operational systems and external info

Prediction and Discovery
- Query
- Reporting

Customer contact points

Customer analytics tools
- OLAP
- Query
- Reporting

Architecture for Analytical CRM

Customer Data Warehouse

Operational systems and external info

Prediction and Discovery
- Data Mining

Operational systems and external info

Prediction and Discovery
- Query
- Reporting
Data mining

“When it comes to data, many multichannel retail businesses face an embarrassment of riches. What’s missing is timely, actionable insight. Data mining is the centerpiece of an analytics strategy that can deliver business value”

(Usama Fayyad)

Data Mining and knowledge discovery, definitions

- Data mining, or knowledge discovery, is the process of discovering valid, novel and useful patterns in large data sets
- “Data mining is the process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules”. (Berry & Linoff 2000)
- The patterns can take many different forms:
  - If then rules, mathematical equations, clusters, decision trees, visual form

Knowledge Discovery and Data Mining (definitions)

- There is no consensus as to the scope and definition of the terms data mining and knowledge discovery
- Often the terms are used interchangeably
- In [Fayyad et al 1996] data mining refers to a particular step within the overall KDD process.
Predictive vs descriptive data mining

- There are two important goals of data mining: Prediction and description
- Predictive data mining
  - discovering patterns for the purpose of predicting future trends and behaviours
- Descriptive data mining (Discovery)
  - identifying understandable patterns for the purpose of obtaining an insight into the business domain

Types of data mining activity

- Classification
  - e.g. classifying cardholders into good and bad accounts
- Prediction
  - e.g. predicting which customer segment is likely to respond to a particular offer
  - Sales prediction
- Description and visualisation
- Segmentation
  - e.g. finding subgroups of customers with similar behaviours
- Affinity grouping, association analysis
  - Market basket analysis

Data mining can be used at each stage of Customer Life Cycle

- Prospect
- Responder
- Established Customer
- Former Customer
- High Value
- Voluntary Churn
- High Potential
- Low Value
- Former Churn

- Company History
- Product Usage
- Payment History
- Campaign Response

- Target Market
- Responsible
- High Value
- Voluntary Churn
- High Potential
- Low Value
- Former Churn

- Company History
- Product Usage
- Payment History
- Campaign Response

- Target Market
- Responsible
- High Value
- Voluntary Churn
- High Potential
- Low Value
- Former Churn

- Company History
- Product Usage
- Payment History
- Campaign Response

- Target Market
- Responsible
- High Value
- Voluntary Churn
- High Potential
- Low Value
- Former Churn

- Company History
- Product Usage
- Payment History
- Campaign Response
Typical data mining applications for CRM

• targeted marketing
  – customer acquisition
  – campaign optimisation
• market basket analysis
• trend analysis (e.g., purchases, sales)
• market segmentation
• Increasing customer lifetime value (cross-selling, up-selling)
• risk analysis
• customer retention (churn modelling)
• and more

Data mining tools and techniques

• Statistics
• Decision trees
• Neural networks
• Clustering
• Rule induction
• Fuzzy logic
• Rough sets
• Genetic algorithms
• Text mining
• Visualisation (aids understanding of data and results)
• Suites: SAS, IBM, SPSS, many see: knuuggets.com

Decision Trees

• Tree-shaped structures
• Can be converted to rules
• Can be used for prediction and description
• Problems: tree can become very large – difficult to understand
Clustering

- Clustering algorithms group records with similar characteristics
- k-Means Method
- Kohonen feature Maps
- Clusters have to be analysed (visualisation can help)
- Can be used for discovering segments

Application: Customer segmentation

- DM algorithms, such as clustering, decision trees, can help find customer segments
- These segments are different from the segments known to the business
- They are derived from data

Decision trees can be used for segmentation:

```
All Customers
  /     \ 
Young    Old
  |       |
income high     income low
  |       |
Segment 1     Segment 2
  |       |
Segment 3
```

(Zikmund et al 2003)
### Some Commercially Available Cluster Tags (from thearling.com)

<table>
<thead>
<tr>
<th>Name</th>
<th>Income</th>
<th>Age</th>
<th>Education</th>
<th>Vendor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Blood Estates</td>
<td>Wealthy</td>
<td>35-54</td>
<td>College</td>
<td>Claritas Prizm™</td>
</tr>
<tr>
<td>Shotguns and Pickups</td>
<td>Middle</td>
<td>35-44</td>
<td>High School</td>
<td>Claritas Prizm™</td>
</tr>
<tr>
<td>Somedock City</td>
<td>Poor</td>
<td>Mix</td>
<td>Grade School</td>
<td>Claritas Prizm™</td>
</tr>
<tr>
<td>Living Off the Land</td>
<td>Middle-Poor</td>
<td>Mix</td>
<td>School Age Families</td>
<td>Equifax MicroVision™</td>
</tr>
<tr>
<td>University USA</td>
<td>Very low</td>
<td>Young -</td>
<td>Medium to High</td>
<td>Equifax MicroVision™</td>
</tr>
<tr>
<td>Samsrt Years</td>
<td>Medium</td>
<td>Seniors</td>
<td>Medium</td>
<td>Equifax MicroVision™</td>
</tr>
</tbody>
</table>

### Neural Networks

- Neural networks are computing systems inspired by the structure and functioning of the brain.
- Neural networks “learn” patterns in data through the process of training on examples
- Neural networks are used for prediction

Neural Networks (more in week 9)
Visualisation

- Powerful tool for
  - Exploration
  - Analysis of input data for data mining
  - Analysis of data mining results
  - Aids understanding of data and results

Example of Histogram

What is the age of majority of customers?

Visualisation example (decision tree)

Visualisation can help understand data mining results

(Source: www.thearling.com/text/dmviz.htm)
Rule induction
(more next week)

- Many different algorithms
- Type of rules
  - Crisp rules
    - If Income is > 100,000 then risk = 0
  - Fuzzy rules (fuzzy logic)
    - If income is High then risk is Low
    - High and low are linguistic variables
  - Association rules, affinity grouping
    - If home loan then insurance policy

Market basket analysis

- Automated extraction of association rules from point-of-sales transaction data
- Example rules
  - IF bread then milk and bananas
  - IF soy milk then soy cheese
  - IF nappies then beer

Can assist in business decisions
- arranging items on shelves (put milk next to bananas)
- which items should be promoted together

Data mining process - general view

Many iterations may be required!
(adapted from Fayyad 1996)
Data mining process

- Define business problem
- Identify relevant data
- Explore data
- Select input/output variables
- Select data (may include sampling)
- Prepare data for modelling
- Develop model
- Evaluate model and interpret results
- Deploy model

Define business problem

- Business problem definition
  - Examples of CRM business problems
    - Who is likely to churn (churn modelling)
    - Credit risk
    - Market basket analysis
    - Acquisition campaign

Different models will be developed for different problems

Identify relevant data

- Identify relevant data and their sources
- Talk to relevant business people
  - For acquisition campaigns use campaign histories and demographic data
  - Example: for delinquency risk analysis the following data were selected from customer database to predict delinquencies:
    - Borrower financial situation, including current income, portion of income other than salary, amount of obligations other than the principal property
    - Mortgage data, including loan to value ratio, type of mortgage, ratio of income to mortgage payments, loan amount
    - Borrower status, including borrower’s credit rating, number of dependents, number of years employed, whether borrower is self employed, intention to occupy the property
    - Property status, including the property age, property location, number of units, property value
Explore the data

Understand your data
- statistical analysis (std deviation, averages, data distribution, etc.)
- clustering combined with visualisation
- combination of statistics and data visualisation
  - histograms, scatterplots
  - (see Visualisation examples from informationbuilders.com)
- AI methods: eg. rough sets’ attribute strength analysis

Example of Histogram

What is the age of majority of customers?

Predictive modelling

- Many DM applications for CRM involve predictive modelling
Select input/output variables

- Output: what do we want to predict?
  - E.g., bad customers, sales volume, customer (or segment) who is likely to respond to our offer
- Input: Many attributes may be available (or not enough)
- Which are relevant?
- Which input variables are the best predictors?
- How to represent the inputs?
  - It may be better to use financial ratios rather than raw variables, e.g., liability / assets instead of liability and assets

Select data

- Using all available data or a sample of data
- High volumes of data - it may take too long to mine all data
- Sampling may be required
- Must be properly selected, statistically significant, random sample

Prepare data for modelling

- Data cleansing, updating (data quality)
  - Missing data, outliers
    - Mostly correct if data sourced from DW
- Transforming – different data mining technique may require different representation of data
  - Must be transformed according to the requirements of the chosen technique
  - E.g., neural nets require numerical values so categorical information must be transformed, for rough sets inputs must be discreet
Developing models

- We have to divide the data used for modelling into training and testing set.
- Models are developed using training data and then tested using testing data.
- E.g. we train a neural network on training data and then test it on data it has not seen before.

Building models

Model building and evaluating

- Iterative process
- Evaluate results
- Typical criteria: prediction/classification rate, error rate, mean squared error, lift (performance improvement)
- Build several models using different techniques and compare performance.
- Different techniques may produce different results on the same data.
- Choose the best model.
Marketing campaigns:
- predictive models (e.g., neural networks or decision trees) can be built from historical data to predict which customer segments are likely to respond to a particular campaign

The prospects predicted are then selected for a targeted marketing campaign

Benefits: increased response rate, savings on marketing campaigns

Churn modelling

Objective is to reduce attrition rate of valuable customers

In telecommunications industry annual churn rates are around 25 to 30 percent, costing up to $10 billion worldwide (SAS Institute)

Predictive models using various data mining techniques can be built to predict which customers are likely to leave

Some churn statistics
- Mobile 25-25%
- Long distance phone 25%
- Computer on-line service 33%
- Retail banking 40%
The value of customer retention, example

Typical wireless carrier
- 500,000 users
- 25% churn rate
- $300 cust. acquisition cost
- $500/year avg. cust. Value

5% reduction in churn:
- savings acquisition
  - 125000 churners * 5% * $300/customer = $1.9 million
- savings in retained customer
  - 125000*5% * $500/customer = $3.1 million
- Total savings $5.0 million

References
- Groth R., Data Mining, Prentice Hall, 2000
  - www.theading.com
  - www.knuggets.com