Building Behavioural Profiles of Debtors

Alexandros Ladas

Supervisors:
Uwe Aickelin
Jon Garibaldi
Eamonn Ferguson

21/10/2013
IMA Seminars 2013/14
My PhD Plan

Data
- Personality Traits, Behaviours, Attitudes
- Demographics, Economic/Financial

Unsupervised Learning

Behavioural Profiles
- Economic/Financial Behaviours, Psychological Characteristics, Attitudes, Social groups

Supervised Learning

Classifications in Economic Applications
- Credit Risk Assessment
- Consumer Debt

Traditional Approach
CCCS Dataset

- Consumer Credit Counselling Service [1]
  - 75000 approx
  - 58 variables

Clustering CCCS

• Problems
  – Outliers
  – Asymmetry in Data
  – Noise
  – Mixed Data
  – High Dimensionality
Data Transformations

• Goal:
  – Tackle the inconsistencies in the data
  – Create a form of behavioural data

• Transformations
  – **Homogeneity analysis** [2] to transform categorical attributes to numerical
  – **Factor analysis** on Financial attributes to explore relations between income-assets-debt
  – **Clustering** on the correlations among spending items to create Spending Behavioural Clusters

Summary of Transformations

CCCS data

Demographics

Financial

Expenditure

Debt Details

2d Spatial Coords

3 Financial Factors

4 Spending Behavioural Clusters
Contribution of Transformations

- How do we assess the quality of Clustering?
  - Clustering validation
  - Experimental Validation Framework
Clustering Validation

• External Validation
  – Comparing to the ground truth
    • Rand Index, Jaccard Coefficient

• Internal Validation
  – Compactness
  – Separation
  – A way to determine the optimal number of clusters


Issues of Internal Clustering Validation

- Each of them suffers from their own assumptions and limitations

- Unsure whether they can be used to compare clusterings on different datasets
A more holistic approach to evaluation

• Compactness/Separation does not usually hold usually in real world data.[5]

• Clusters have to be meaningful for the application

• So our evaluation needs:
  – To be quantified by a validation measure
  – Verified by the patterns returned

Cohen’s Kappa

• Quantifies the level of agreement between 2 classifications

• Intuition
  – True patterns will be revealed by different clustering algorithms
  – Powered by the Consensus Clustering philosophy [6]

Experimental Framework

• 2 clustering algorithms
  – Kmeans
  – PAM

• Evaluation by Kappa’s index
  – Level of agreement

• Definition of core classes
  – Alignment of clusters
  – Common Elements

• Inspection of core classes
  – Statistical tests: ANOVA, z-tests, t-tests
### Experiments

<table>
<thead>
<tr>
<th>Experiment Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
</tbody>
</table>
Results
Non Scaled Data

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Kmeans Clusters</th>
<th>PAM Clusters</th>
<th>S_K</th>
<th>S_P</th>
<th>Kappa</th>
<th>Core Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>16</td>
<td>0.7</td>
<td>0.42</td>
<td>0.189</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>16</td>
<td>0.7</td>
<td>0.42</td>
<td>0.189</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>4</td>
<td>0.17</td>
<td>0.15</td>
<td>0.663</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>4</td>
<td>0.17</td>
<td>0.15</td>
<td>0.663</td>
<td>4</td>
</tr>
</tbody>
</table>

Data need to be scaled
### Results after scaling

Kappa Index increases constantly

Low silhouette values
Profiles of Debtors

- **Low Income**
  - Low levels of debt, cheap assets, low spending
  - Young, single, unemployed, renter
  - Older, retired

- **Average Income**
  - Average Spending, low levels of debt, cheap assets
  - Clothing+, food+, Cohabiting
  - p/t employment
  - Average levels of debt

- **High Income**
  - High levels of debt, High spending
  - Expensive houses
  - Mortgagees
  - Self employed
  - Cheap houses
  - Expensive houses
  - Mortgagees
  - Self employed

**Differences in Demographics**

**Differences in Spending**

**Differences in Financial Behaviours**
Importance of Findings

- Deviation from the rational model of Economics
- Self employed \sim Level of debt
Conclusions

• Transformations improved the quality of clustering
  – Quantified by Cohen’s kappa
  – Verified by significance of patterns

• Importance of scaling

• Novel way to transform categorical attributes to a meaningful numerical representation

• Cohen’s kappa as an evaluation measure
Low Silhouettes

• Questions about:
  – Separation/Compactness in real world data
  – Role of validation indices in determining the optimal number of clusters
Future Steps

• Supervised approach
  – Level of debt prediction

• A new consensus clustering approach

• New transformations in Spending

• Focus research on specific type of debt
  – Mortgage, credit, unsecured
References

1. R. Disney, and J. Gathergood, "Understanding consumer over-indebtedness using counselling sector data: Scoping Study", Report to the Department for Business, Innovation and Skills (BIS), University of Nottingham, 2009.


