Data driven approach for channel usage segmentation

Vera Helman

SEUGI 18
June 20-23, 2000
Dublin

Database Marketing
• Business objective: Channel Changing initiative
  - Transaction Costs
  - Traditional marketing segmentation scheme
• Cluster analysis - data mining approach to segmentation
  - Cluster analysis, overview
    - Iterative clustering
    - Hierarchical clustering
  - Combination of iterative and hierarchical clustering
    - 3 stage clustering
    - Bootstrap aggregating
    - Bagged clustering framework
  - Data reduction and transformation
• Application of bagged clustering to transaction migration
  - Resulting channel usage segmentation scheme
  - Examples of implementation
• Summary
Business objective

Our goal was to use a channel usage based segmentation scheme to help in developing marketing strategies for channel or transaction migration from more expensive to less expensive channels

- Use objective, data-driven approach to identify the distinguished groups or segments of Fleet’s retail bank customers characterized by specific patterns of transaction behavior

- Use developed segmentation scheme to help business lines to adopt specific strategies and to develop different marketing programs, which will lead towards improving bank’s profitability
Database Marketing

Business objective

Transaction Costs

230 MM Teller Transactions annually  
( based on 1998 data )

170 MM Teller Transactions “Get Cash” and “Make Deposit”  
( based on 1998 data )

Make Deposit

- Platform $3.38
- Teller $2.15
- ATM $0.78

Get Cash

- Teller $2.18
- ATM $0.35

all costs are from FMCG
Transaction Costs

Business objective

Balance Inquiry

- Platform: $2.04
- Teller: $1.15
- ATM: $0.35
- VRU: $0.28

Transfer Credit or Debit

- Teller: $2.91
- ATM: $0.50
- VRU: $0.28
Business objective

Transaction Costs: transactions by channel and type

For each household, we have the following channel by transaction type volume data (only 21 cells are feasible):

<table>
<thead>
<tr>
<th>Transaction Type</th>
<th>ATM</th>
<th>Teller</th>
<th>Platform</th>
<th>VRU</th>
<th>Live Ag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance inquiry: BAL</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Deposit: DEP</td>
<td>Y</td>
<td>Y</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Account information: AIF</td>
<td>N/A</td>
<td>N/A</td>
<td>Y</td>
<td>N/A</td>
<td>Y</td>
</tr>
<tr>
<td>Withdrawal: WD</td>
<td>Y</td>
<td>Y</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Statement: STM</td>
<td>N/A</td>
<td>N/A</td>
<td>Y</td>
<td>N/A</td>
<td>Y</td>
</tr>
<tr>
<td>Cash check - Fleet: CK</td>
<td>N/A</td>
<td>Y</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Transfer credit/debit: TR</td>
<td>Y</td>
<td>Y</td>
<td>N/A</td>
<td>Y</td>
<td>N/A</td>
</tr>
<tr>
<td>Other (e.g. maintenance, night deposit): OTH</td>
<td>N/A</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

How to find and group the customers whose transaction behavior by the channel and the type of transactions is similar?
**Traditional marketing segmentation scheme**

Segmentation is the foundation of marketing strategy. A market segment is defined to be a group of customers having common characteristics:

- Possess a common set of attributes which serve to identify the group – in our case, transaction behavior
- Require a separate marketing approach to maximize their potential

**Traditional segmentation analysis to develop a marketing strategy:**

- Judgmental, subjective criteria based on the knowledge of business and experience or prior information: frequency distribution, crosstabs

**Objective, data driven approach to customer segmentation**

- Cluster analysis or unsupervised learning - data mining tool for grouping customers into homogenous segments
Cluster analysis, overview

Clustering procedure is a way to find natural groupings in a set of data, where samples in one cluster are more like one another than samples in other clusters.
Question #1:

How should one measure the similarity between samples?

Answer:

- Distance is the most obvious measure of similarity (or dissimilarity).
- The way to begin cluster analysis is to define a suitable distance function and compute the matrix of distances between all pairs of samples.
- The distance between samples in the same cluster should be significantly less than the distance between samples in different clusters.
- Examples of distances: Euclidean, Mahalanobis, city-block metric, etc. Euclidean distance between objects I and J:

\[
d_{ij} = \left\{ \sum_{k=1}^{p} (X_{ik} - X_{jk})^2 \right\}^{1/2}
\]
Question #2: How should one evaluate a partitioning of the set of samples into clusters?

Answer:

• Define a criterion function that measures the clustering quality of any partition of the data.

• Once a criterion function has been selected, clustering becomes a well-defined problem in discrete optimization: find those partitions of the set of samples that extremize the criterion function.

• Examples of criterion functions: Sum-of-Squared-Error, minimum variance, the trace criterion, etc.

The Sum-of-Squared-Error Criterion:

\[ m_i = \frac{1}{n_i} \sum_{x \in X_i} x \]  
\[ J_e = \sum_{i=1}^{c} \sum_{x \in X_i} \| x - m_i \|^2 \]
**Iterative clustering**

**K Means Training Flow Chart**

1. Randomly select cluster centers $C$ from the data set.

2. For each sample $X_i$ in the data set, find the nearest cluster center $C$ and classify $X_i$ as a member of $C$.

3. For each cluster, recompute its center by finding the mean of the cluster:
   \[ M_k = \frac{1}{N_k} \sum_{j=1}^{N_k} X_{jk} \]

   where $M_k$ is the new mean, $N_k$ is the number of samples in cluster $k$ and $X_{jk}$ is $j$-th sample belonging to cluster $k$. 
Iterative clustering / SAS software implementation

**Advantages:**

- Can be used to cluster large data sets - CPU time proportional to the number of observations.
- More robust with respect to the presents of outliers and a choice of the distance metric.

**Disadvantages:**

- Number of clusters should be specified before cluster analysis. How to find the optimal number of clusters?
- Very sensitive to the choice of starting points - initial seeds.

**SAS software solution for Iterative clustering**

SAS product SAS/STAT® contains FASTCLUS procedure, which is based on K-means algorithm. Proc FASTCLUS combines an effective method for finding initial clusters with a standard iterative algorithm for minimizing the sum-of-squared distances from the cluster means.
Hierarchical clustering

Stepwise-Optimal hierarchical clustering - Find the pair of distinct clusters $X_K$ and $X_N$, whose merger would increase (or decrease) the criterion function as little as possible.
<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>--Clusters Joined--</th>
<th>Frequency of New Cluster</th>
<th>Semipartial R-Squared</th>
<th>Approximate Expected R-Squared</th>
<th>Cubic Clustering Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>OB15, OB18</td>
<td>199</td>
<td>0.002154</td>
<td>0.802880</td>
<td>0.681921</td>
</tr>
<tr>
<td>29</td>
<td>OB6, OB13</td>
<td>136</td>
<td>0.003796</td>
<td>0.799083</td>
<td>0.678294</td>
</tr>
<tr>
<td>28</td>
<td>OB14, OB28</td>
<td>226</td>
<td>0.003905</td>
<td>0.795178</td>
<td>0.674497</td>
</tr>
<tr>
<td>27</td>
<td>OB2, CL28</td>
<td>267</td>
<td>0.004336</td>
<td>0.790842</td>
<td>0.670515</td>
</tr>
<tr>
<td>26</td>
<td>OB12, OB19</td>
<td>685</td>
<td>0.004366</td>
<td>0.786476</td>
<td>0.666332</td>
</tr>
<tr>
<td>25</td>
<td>CL26, OB20</td>
<td>812</td>
<td>0.005145</td>
<td>0.781331</td>
<td>0.661928</td>
</tr>
<tr>
<td>24</td>
<td>CL30, OB27</td>
<td>242</td>
<td>0.005612</td>
<td>0.775720</td>
<td>0.657284</td>
</tr>
<tr>
<td>23</td>
<td>OB7, OB23</td>
<td>2172</td>
<td>0.006068</td>
<td>0.769651</td>
<td>0.652375</td>
</tr>
<tr>
<td>22</td>
<td>OB21, OB25</td>
<td>2150</td>
<td>0.006971</td>
<td>0.762680</td>
<td>0.647172</td>
</tr>
<tr>
<td>21</td>
<td>OB8, CL25</td>
<td>1796</td>
<td>0.007216</td>
<td>0.755464</td>
<td>0.641645</td>
</tr>
<tr>
<td>20</td>
<td>OB16, OB29</td>
<td>1315</td>
<td>0.007572</td>
<td>0.747892</td>
<td>0.635756</td>
</tr>
<tr>
<td>19</td>
<td>OB4, OB31</td>
<td>870</td>
<td>0.008261</td>
<td>0.739632</td>
<td>0.629461</td>
</tr>
<tr>
<td>18</td>
<td>OB17, OB26</td>
<td>4164</td>
<td>0.010076</td>
<td>0.729556</td>
<td>0.622709</td>
</tr>
<tr>
<td>17</td>
<td>OB1, OB3</td>
<td>9318</td>
<td>0.011462</td>
<td>0.718093</td>
<td>0.615437</td>
</tr>
<tr>
<td>16</td>
<td>OB10, OB30</td>
<td>1184</td>
<td>0.011501</td>
<td>0.706593</td>
<td>0.607572</td>
</tr>
<tr>
<td>15</td>
<td>CL18, OB24</td>
<td>13086</td>
<td>0.014474</td>
<td>0.692119</td>
<td>0.599024</td>
</tr>
<tr>
<td>14</td>
<td>CL19, OB11</td>
<td>917</td>
<td>0.014730</td>
<td>0.677389</td>
<td>0.589680</td>
</tr>
<tr>
<td>13</td>
<td>CL29, OB22</td>
<td>504</td>
<td>0.015237</td>
<td>0.662152</td>
<td>0.579403</td>
</tr>
<tr>
<td>12</td>
<td>OB5, OB9</td>
<td>21689</td>
<td>0.016437</td>
<td>0.645715</td>
<td>0.568014</td>
</tr>
<tr>
<td>11</td>
<td>CL17, CL22</td>
<td>11468</td>
<td>0.025566</td>
<td>0.620149</td>
<td>0.555285</td>
</tr>
<tr>
<td>10</td>
<td>CL27, CL23</td>
<td>2439</td>
<td>0.026111</td>
<td>0.594038</td>
<td>0.540912</td>
</tr>
<tr>
<td>9</td>
<td>CL21, CL24</td>
<td>2038</td>
<td>0.028824</td>
<td>0.565214</td>
<td>0.524485</td>
</tr>
<tr>
<td><strong>8</strong></td>
<td>CL14, CL15</td>
<td><strong>14003</strong></td>
<td><strong>0.042910</strong></td>
<td><strong>0.522304</strong></td>
<td><strong>0.505426</strong></td>
</tr>
<tr>
<td>7</td>
<td>CL11, CL20</td>
<td>12783</td>
<td>0.049580</td>
<td>0.472723</td>
<td>0.482896</td>
</tr>
<tr>
<td>6</td>
<td>CL10, CL12</td>
<td>24128</td>
<td>0.050281</td>
<td>0.422442</td>
<td>0.450533</td>
</tr>
<tr>
<td>5</td>
<td>CL6, CL8</td>
<td>38131</td>
<td>0.064763</td>
<td>0.357679</td>
<td>0.398816</td>
</tr>
<tr>
<td>4</td>
<td>CL7, CL9</td>
<td>14821</td>
<td>0.079822</td>
<td>0.277857</td>
<td>0.334202</td>
</tr>
<tr>
<td>3</td>
<td>CL5, CL16</td>
<td>39315</td>
<td>0.081908</td>
<td>0.195949</td>
<td>0.250676</td>
</tr>
<tr>
<td>2</td>
<td>CL4, CL13</td>
<td>15325</td>
<td>0.090281</td>
<td>0.105669</td>
<td>0.144500</td>
</tr>
<tr>
<td>1</td>
<td>CL2, CL3</td>
<td>54640</td>
<td>0.105669</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

The SAS System
Hierarchical clustering / SAS software implementation

**Advantages:**
- The optimal number of clusters can be found during the clustering by checking some statistical criteria - semipartial R-Squared.

**Disadvantages:**
- Not practical for the large data sets - CPU time varies as a square or cube to the number of observations.
- Sensitive to outliers.

SAS software solution for hierarchical clustering

SAS product SAS/STAT® contains CLUSTER procedure, which finds hierarchical clusters based on usual agglomerative clustering. Examples of 11 available methods include average linkage, the centroid method, single linkage, Ward’s minimum-variance method, etc.

The various clustering methods differ in how the distance between clusters is computed.
3 Stage Clustering

1st K Means
arbitrary large number of clusters (50)
random initial seeds

Eliminate outliers
Preliminary cluster solution

2nd K Means
use means of previous cluster solution as initial seeds

Hierarchical (Ward) clustering

Select optimal number of clusters based on dendrogram
Bagging ("Bootstrap Aggregating")

Ensemble methods can be apply to enhance the performance and produced robust clustering: bagging

Steps to incorporate bootstrap sampling in the cluster analysis framework:

1. Obtain a collection of training sets by sampling from the empirical distribution of the original data, i.e., by bootstrapping
2. Run any partitioning cluster algorithm (K-means) on each of this training sets
3. Combine the results of all partitioning methods into a new data set
4. Run a hierarchical cluster algorithm resulting in the usual dendrogram
Combination of iterative and hierarchical clustering

Bagged Clustering Framework

Original sample

1st bootstrap sample
1st K-means

2ndt bootstrap sample
2nd K-means

Nth bootstrap sample
Nth K-means

Hierarchical (Ward) clustering

Select optimal number of clusters based on dendrogram
Invariance to transformations:
Clusters should be invariant to transformations natural to the problem

Standardization:
to obtain invariance to displacement and scales changes - scale the axes so that all of the variables have zero mean and unit variance

Principal components:
to obtain invariance to rotation - rotate the axes so that they coincide with the eigenvectors of the sample covariance matrix
Transformation to principal components can be preceded and/or followed standardization for scale

Routine normalization is not necessarily desirable
Alternative to normalizing the data and using Euclidean distance: use normalized distance, such as the Mahalanobis distance
Application of bagged clustering to transaction migration

Resulting channel usage segmentation scheme

<table>
<thead>
<tr>
<th>CLUSTER SIZE</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit Cost</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATM - BAL</td>
<td>$0.35</td>
<td>0.67</td>
<td>0.23</td>
<td>5.73</td>
<td>0.64</td>
<td>0.43</td>
<td>1.02</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>ATM - DEP</td>
<td>$0.78</td>
<td>0.85</td>
<td>0.07</td>
<td>2.06</td>
<td>0.50</td>
<td>0.24</td>
<td>1.04</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>ATM - TR</td>
<td>$0.50</td>
<td>0.32</td>
<td>0.02</td>
<td>2.11</td>
<td>0.10</td>
<td>0.09</td>
<td>0.21</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>ATM - W D</td>
<td>$0.35</td>
<td>6.90</td>
<td>1.58</td>
<td>13.92</td>
<td>5.39</td>
<td>2.96</td>
<td>7.62</td>
<td>1.52</td>
<td>0.63</td>
</tr>
<tr>
<td>PLA - AF</td>
<td>$3.53</td>
<td>0.09</td>
<td>0.04</td>
<td>0.07</td>
<td>0.25</td>
<td>0.61</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>PLA - BAL</td>
<td>$2.04</td>
<td>0.18</td>
<td>0.10</td>
<td>0.15</td>
<td>0.27</td>
<td>0.99</td>
<td>0.06</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>PLA - OT</td>
<td>$4.10</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>PLA - STM</td>
<td>$7.25</td>
<td>0.25</td>
<td>0.17</td>
<td>0.22</td>
<td>0.38</td>
<td>0.87</td>
<td>0.12</td>
<td>0.17</td>
<td>0.05</td>
</tr>
<tr>
<td>TEL - BAL</td>
<td>$1.15</td>
<td>0.21</td>
<td>0.55</td>
<td>0.25</td>
<td>0.26</td>
<td>0.27</td>
<td>0.14</td>
<td>0.40</td>
<td>0.10</td>
</tr>
<tr>
<td>TEL - CK</td>
<td>$2.07</td>
<td>0.33</td>
<td>0.12</td>
<td>0.25</td>
<td>0.37</td>
<td>0.32</td>
<td>0.17</td>
<td>0.86</td>
<td>0.05</td>
</tr>
<tr>
<td>TEL - DEP</td>
<td>$2.15</td>
<td>2.78</td>
<td>2.50</td>
<td>2.10</td>
<td>2.63</td>
<td>2.08</td>
<td>1.64</td>
<td>3.61</td>
<td>0.74</td>
</tr>
<tr>
<td>TEL - OT</td>
<td>$2.15</td>
<td>0.04</td>
<td>0.08</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>TEL - TR</td>
<td>$2.90</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>TEL - W D</td>
<td>$2.18</td>
<td>0.22</td>
<td>3.48</td>
<td>0.16</td>
<td>0.10</td>
<td>0.10</td>
<td>0.05</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>TSV - AF</td>
<td>$2.08</td>
<td>0.07</td>
<td>0.02</td>
<td>0.06</td>
<td>1.14</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>TSV - BAL</td>
<td>$1.70</td>
<td>0.12</td>
<td>0.04</td>
<td>0.08</td>
<td>1.27</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>TSV - OT</td>
<td>$1.50</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.16</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TSV - STM</td>
<td>$6.08</td>
<td>0.19</td>
<td>0.08</td>
<td>0.11</td>
<td>0.71</td>
<td>0.12</td>
<td>0.08</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>VRU - BAL</td>
<td>$0.28</td>
<td>11.42</td>
<td>0.79</td>
<td>2.59</td>
<td>5.43</td>
<td>1.44</td>
<td>3.42</td>
<td>1.17</td>
<td>0.18</td>
</tr>
<tr>
<td>VRU - OT</td>
<td>$0.28</td>
<td>1.27</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>VRU - TR</td>
<td>$0.28</td>
<td>4.45</td>
<td>0.03</td>
<td>0.08</td>
<td>0.12</td>
<td>0.04</td>
<td>0.15</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Total</td>
<td>30.41</td>
<td>9.93</td>
<td>30.00</td>
<td>19.80</td>
<td>11.10</td>
<td>15.82</td>
<td>8.28</td>
<td>2.11</td>
<td>8.74</td>
</tr>
</tbody>
</table>

1Based on 10/97 - 4/98 data.
Examples of implementation

Average number of products per Household by Cluster

Regular saving
Pathbook saving
Teller Depositors

Segment profile:

- Profitable
- Highest average number of CD products
- Longest length of relationships with Fleet (13 years) and length of residence (5 years), oldest age of the head of the household (45 years old)
- Highest average balance of Interest Checking, low average number of POS product

Transaction cost issues:

- 4.59 teller deposit transactions per household - more than twice higher than overall level of teller deposits
- Only 0.1 deposit transactions through ATM per household - 4 times less than average
- Cost of teller deposit per transaction is $2.15, cost of ATM deposit - $0.78
- Saving per one deposit transaction is $1.37/month

Next steps for segment:

- Develop specific programs and initiatives which can help to migrate teller deposit type of transactions in this segment to ATM deposit
- Maximum potential cost saving for the segment is $(1.37*4.59*size)/month - excellent opportunity for transaction migration!
Cluster analysis framework based on the procedures from SAS Institute software package (SAS/STAT® product) was used to develop channel usage segmentation scheme for retail bank.

Suggested clustering approach is the data mining tool, which allows the combination of hierarchical and partitioning algorithms.

Bagging ("bootstrap aggregating") module embedded in the clustering framework was enable to provide robust solution which lead to develop very stable channel usage segmentation based on customers transaction behavior.
The author wants to thank Victor Hoffman and Victor Lo for initiation of the project and helpful discussions

The project was implemented based on SAS/STAT® software and approaches discussed in the following papers:


R. Duda, P. Hart “Pattern Classification and Scene Analysis”, John Wiley & Sons, 1973

