CHURN PREDICTION MODEL IN RETAIL BANKING USING FUZZY C-MEANS CLUSTERING

Džulijana Popović
Consumer Finance, Zagrebačka banka d.d.

Bojana Dalbelo Bašić
Faculty of Electrical Engineering and Computing
University of Zagreb
Overview

Theoretical basis

- Churn problem in retail banking
- Current methods in churn prediction models
- Fuzzy c-means clustering algorithm vs. classical k-means clustering algorithm
Study and results

- Canonical discriminant analysis in outliers detection and variables’ selection
- Poor results of hierarchical clustering and crisp k-means algorithm
- Very good results of the fuzzy c-means algorithm
  - Introduction of fuzzy transitional conditions of the 1st and of the 2nd degree and the sums of membership functions from distance of k instances (*abb. DOKI sums*)
- Final models’ results
- Conclusions
Churn problem in retail banking

• No unique definition - generally, term churn refers to all types of customer attrition whether voluntary or involuntary
• Precise definitions of the churn event and the churner are crucial

• In this study:
  ✓ moment of churn is the moment when client cancels ("closes") his last product or service in the bank
  ✓ churner is client having at least one product at time $t_n$ and having no product at time $t_{n+1}$
  ✓ If client still holds at least one product at time $t_{n+1}$ - non-churner
Current methods in churn prediction models

- Logistic regression
- Survival analysis
- Decision trees
- Neural networks
- Random forests

To the best of our knowledge – no fuzzy logic based clustering for churn prediction in banking industry!
Fuzzy c-means clustering algorithm vs. classical k-means clustering algorithm

Possible advantages of fuzzy c-means:

• More robust against outliers presence
• High true positives rate and acceptable accuracy after just a few iterations
• Additional information hidden in the values of the membership functions
• Fuzzy nature of the problem requires fuzzy methods
Canonical discriminant analysis (CDA) in outliers detection and variables’ selection

Final data set: 5000 individual clients of the retail bank
Classes: 2500 churners vs. 2500 non-churners

CDA helped a lot in:
• variable selection process
• outlier detection and their further analysis
• graphical exploration of different data samples
Results of CDA applied on the data set with churners (black), non-churners (red) and “returners” (green)
Results of CDA applied on the data set with only churners (black) and non-churners (red) and variables in $t_0$ and $t_2$
Results of hierarchical clustering and crisp k-means algorithm

• were very poor, especially for crisp k-means
• k-means algorithm broke on even modest outliers
• only Ward’s method and Flexible Beta method performed better

NOTE:
• removing outliers from the database will not always be possible and desirable in the real banking situations
• churn prediction becomes extremely important in periods of financial crises – models need to be robust, stable and fast
## Results of the classical clustering in terms of true positives, false negatives, accuracy and specificity

<table>
<thead>
<tr>
<th>CLUSTERING ALGORITHM</th>
<th>STANDARDIZATION METHOD</th>
<th>tp rate (recall)</th>
<th>fp rate</th>
<th>accuracy</th>
<th>specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Linkage</td>
<td>standard deviation</td>
<td>99.96%</td>
<td>100.00%</td>
<td>50.44%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Average Linkage</td>
<td>range</td>
<td>100.00%</td>
<td>99.47%</td>
<td>50.73%</td>
<td>0.53%</td>
</tr>
<tr>
<td>Centroid Linkage</td>
<td>standard deviation</td>
<td>99.96%</td>
<td>100.00%</td>
<td>50.44%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Centroid Linkage</td>
<td>range</td>
<td>100.00%</td>
<td>99.96%</td>
<td>50.48%</td>
<td>0.04%</td>
</tr>
<tr>
<td>Ward’s Minimum Variance</td>
<td>standard deviation</td>
<td>84.67%</td>
<td>66.92%</td>
<td>59.11%</td>
<td>33.08%</td>
</tr>
<tr>
<td>Ward’s Minimum Variance</td>
<td>range</td>
<td>73.58%</td>
<td>60.81%</td>
<td>56.55%</td>
<td>39.19%</td>
</tr>
<tr>
<td>Complete Linkage</td>
<td>standard deviation</td>
<td>99.96%</td>
<td>99.92%</td>
<td>50.48%</td>
<td>0.08%</td>
</tr>
<tr>
<td>Complete Linkage</td>
<td>range</td>
<td>87.39%</td>
<td>70.07%</td>
<td>58.93%</td>
<td>29.93%</td>
</tr>
<tr>
<td>Flexible Beta</td>
<td>standard deviation</td>
<td>81.55%</td>
<td>64.89%</td>
<td>58.55%</td>
<td>35.11%</td>
</tr>
<tr>
<td>Flexible Beta</td>
<td>range</td>
<td>72.18%</td>
<td>59.01%</td>
<td>56.73%</td>
<td>40.99%</td>
</tr>
<tr>
<td>McQuitty’s Similarity Analysis</td>
<td>standard deviation</td>
<td>99.96%</td>
<td>100.00%</td>
<td>50.44%</td>
<td>0.00%</td>
</tr>
<tr>
<td>McQuitty’s Similarity Analysis</td>
<td>range</td>
<td>98.03%</td>
<td>89.27%</td>
<td>54.81%</td>
<td>10.73%</td>
</tr>
<tr>
<td>Median Linkage</td>
<td>standard deviation</td>
<td>99.96%</td>
<td>100.00%</td>
<td>50.44%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Median Linkage</td>
<td>range</td>
<td>100.00%</td>
<td>99.96%</td>
<td>50.48%</td>
<td>0.04%</td>
</tr>
<tr>
<td>Single Linkage</td>
<td>standard deviation</td>
<td>99.96%</td>
<td>100.00%</td>
<td>50.44%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Single Linkage</td>
<td>range</td>
<td>100.00%</td>
<td>99.96%</td>
<td>50.48%</td>
<td>0.04%</td>
</tr>
<tr>
<td>Crisp k-means*</td>
<td>standard deviation</td>
<td>100.00%</td>
<td>99.96%</td>
<td>50.02%</td>
<td>0.04%</td>
</tr>
<tr>
<td>Crisp k-means</td>
<td>standard deviation</td>
<td>99.88%</td>
<td>80.67%</td>
<td>59.61%</td>
<td>19.33%</td>
</tr>
</tbody>
</table>

*performed on complete data set, without outlier removal*
Dendrogram of the Average Linkage method and standardization with range shows typical problem of hierarchical clustering: chaining
Dendrogram of Ward’s Minimum Variance method and standardization with range
Results of the fuzzy c-means

• were significantly better than the results of classical clustering, regarding true positives, false positives and accuracy (z-test)
• 10 different values of the fuzzification parameter $m$ were applied
• different number of iterations were tested – fast reaction is very important in banking industry!
• in order to improve the prediction results three definitions were introduced:
  ✓ fuzzy transitional condition of the 1st and of the 2nd degree
  ✓ distance of $k$ instances fuzzy sum (DOKI sum)
Results of the fuzzy c-means with different values of the fuzzification parameter $m$

<table>
<thead>
<tr>
<th>PARAMETER m VALUE</th>
<th>TRUE POSITIVES RATE</th>
<th>FALSE POSITIVES RATE</th>
<th>ACCURACY</th>
<th>SPECIFICITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m=1.25$</td>
<td>85.56%</td>
<td>58.23%</td>
<td>63.64%</td>
<td>41.77%</td>
</tr>
<tr>
<td>$m=1.30$</td>
<td>84.80%</td>
<td>56.76%</td>
<td>64.00%</td>
<td>43.24%</td>
</tr>
<tr>
<td>$m=1.40$</td>
<td>83.82%</td>
<td>55.34%</td>
<td>64.22%</td>
<td>44.66%</td>
</tr>
<tr>
<td>$m=1.50$</td>
<td>82.67%</td>
<td>54.59%</td>
<td>64.02%</td>
<td>45.41%</td>
</tr>
<tr>
<td>$m=2.00$</td>
<td>80.84%</td>
<td>53.04%</td>
<td>63.88%</td>
<td>46.96%</td>
</tr>
<tr>
<td>$m=3.00$</td>
<td>80.53%</td>
<td>52.55%</td>
<td>63.97%</td>
<td>47.45%</td>
</tr>
<tr>
<td>$m=5.00$</td>
<td>80.58%</td>
<td>52.37%</td>
<td>64.08%</td>
<td>47.63%</td>
</tr>
</tbody>
</table>

- Value $m=1.25$ chosen for application on training data set, due to the highest true positives rate (significance in difference tested)
Final models’ results

PE = Prediction Engine

PE-1: apply fuzzy c-means algorithm to the training dataset; find the best parameter m; add new clients from the validation set and reapply fuzzy c-means

PE-2: apply fuzzy c-means algorithm to the training dataset; extract only correctly classified clients; add new clients from the validation set and reapply fuzzy c-means;

PE-3: apply fuzzy c-means algorithm on the training dataset; new client from the validation set belongs to the cluster of his 1st nearest neighbor
PE-4: apply fuzzy c-means to the training dataset; for every new client from the validation set find k nearest neighbors and calculate DOKI sums; client belongs to the cluster with highest value of DOKI sum.

**Prediction Models in Comparison**

<table>
<thead>
<tr>
<th>Prediction Models in Comparison</th>
<th>Empirical z value for tp rate</th>
<th>Empirical z value for fp rate</th>
<th>Theoretical z (α=0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE-1 (non-balanced set) vs. PE-4 (non-balanced set)</td>
<td>2.42</td>
<td>1.10</td>
<td>1.96</td>
</tr>
<tr>
<td>PE-3 (non-balanced set) vs. PE-4 (non-balanced set)</td>
<td>6.51</td>
<td>1.91</td>
<td>1.96</td>
</tr>
<tr>
<td>PE-2 (non-balanced set) vs. PE-4 (non-balanced set)</td>
<td>1.74</td>
<td>0.37</td>
<td>1.96</td>
</tr>
<tr>
<td>PE-4 (balanced set) vs. PE-4 (non-balanced set)</td>
<td>0.19</td>
<td>0.37</td>
<td>1.96</td>
</tr>
</tbody>
</table>

PE-4 model applying DOKI sums performed best, no matter if tested on balanced or non-balanced test sets.

PE-2 had insignificantly lower tp rate, but is at least twice slower than PE-4 and every delay in the reaction increases the losses!
Conclusions

• classical clustering methods totally failed on the real banking data due to the modest outliers
• fuzzy c-means algorithm showed great robustness in outlier presence
• introduction of DOKI sums significantly improved churn prediction in comparison to other fuzzy models
• introduction of fuzzy transitional conditions revealed hidden information about product characteristics of these clients
• fuzzy methods can be successfully applied on banking data
Questions?

dzulijana.popovic@unicreditgroup.zaba.hr
bojana.dalbelo@fer.hr