How to Design Text Mining Services for Marketing Applications

Prof. Dr. Thorsten Teichert • Universität Hamburg • Arbeitsbereich Marketing und Innovation

Von-Melle-Park 5 • Raum 3078 (Sekretariat) • 20146 Hamburg
Tel: (040) 42838 4643 • Fax: (040) 42838 5250
Email: ami@econ.uni-hamburg.de
Agenda

- Text Mining Applications in Marketing
- Conceptual Background
- Text Mining Software
- The Study
- Results of Analysis and Conclusions
Agenda

- Text Mining Applications in Marketing
- Conceptual Background
- Text Mining Software
- The Study
- Results of Analysis and Conclusions
Application of text mining in in-depth interviews to elicit obvious and low-lying associations of test persons

- **Idea:** Eliciting of obvious and low-lying associations in in-depth interviews
- **Problem:** Transcription and analysis of in-depth interviews generates confusing amount of text
- **Solution:** Application of text mining to cluster automatically relevant associations and results of in-depth interviews
Example - A case Study of brand image dimensions for Nivea

Brand Image Analysis:
- Elicitation of brand image dimensions
- Connection of qualitative and quantitative interview techniques

Text Mining Integration:
- Using „Stop-Word-Lists“ to reduce text amount of interviews
- Elicitation of synonyms: Elicitation of core-associations
- Clustering of synonyms: Automatic processed assignment of associations into clustered synonym vectors
Agenda

- Text Mining Applications in Marketing
- Conceptual Background
- Text Mining Software
- The Study
- Results of Analysis and Conclusions
Assumptions made to use text mining in brand image studies

- Flow of speech follows the flow of thoughts during the interview
- Background (Human Associative Memory-Model): Spread of activation leads to a constant flow of thoughts
- Any two nodes in the network are inter-connected
- Frequency of word mentions is an indicator for concept salience
- Frequency of co-occurrence represents the strength of connection in the respondent's mind
Feature extraction procedure

start

read word

contained in stop list?

yes

save on clipboard

yes

contained in valence word list?

no

list as feature

no

next word

no

save on clipboard

delete clipboard content

assign valence word to feature

valence word on clipboard?

yes
General advantages of text mining

- Reduced time spending in analysis of in-depth interviews

- Maximum level of objectivity and replicability

- Objectivity: Especially in the context of the analysis for emotional terms and sentiments proven
Agenda

- Text Mining Applications in Marketing
- Conceptual Background
- Text Mining Software
- The Study
- Results of Analysis and Conclusions
Steps of feature clustering

- General preparation: Text formatting for automatic process

- Step 1: Text analysis
  - Elimination of stop words and valence words
  - Remaining features are assigned to their synonym vectors based on lemmatized features

- Step 2: Graph creation and clustering
  - Similarity of any two vectors are calculated (by dice coefficient)
  - Graph is clustered using Chinese Whisperers Algorithm

- Step 3: Output
  - Set of Output Files contains list of all extracted features and assigned clusters
A close look on our text mining tool
Agenda

- Text Mining Applications in Marketing
- Conceptual Background
- Text Mining Software
- The Study
- Results of Analysis and Conclusions
Background of the study for our analysis

- **Brand image study:**
  - Use of in-depth interviews: 120 probands interviewed by free association, sensory and projective techniques
  - Text mining tool was used to facilitate the coding and categorization of respondent associations with the brand

- **Additional step of manual recoding necessary:**
  - After Co-Word-Analysis:
    - Large amount of one-word cluster (2039 out of 2263 (90.1%))
  - Re-assignment of automatic cluster into new manual-clusters
  - Manually coding results in 195 Clusters
Descriptive statistics

Features: 8785

Unique Features: 3140

Lemmatized Features (Principal Forms): 2669

Eliminated Principal Forms: 872

Multiple-Word-Clusters (MWC): 224

One-Word-Clusters (OWC): 2039

Further Processed Automatic Clusters: 1443
(MWC: 213, OWC: 1230; corresponds to 1797 principal forms)

Manually Re-coded Clusters (MC): 195

MWC: 191

OWC: 4

Automatic clustering

Manual clustering
Agenda

- Text Mining Applications in Marketing
- Conceptual Background
- Text Mining Software
- The Study
- Results of Analysis and Conclusions
Match as an indicator for clustering quality

Match of automatic and manual clusters based on frequency

Lemmatized Features: Frequency of occurrence in text

Match of AC and MC overall
Match per frequency interval

Match of automatic and manual clusters based on frequency

Lemmatized Features: Frequency of occurrence in text

- Match of AC and MC overall
- Match of AC and MC per frequency
### Extract from comparison of top 30 automatic vs. manual clusters

<table>
<thead>
<tr>
<th>Automatic cluster (AC)</th>
<th>Rank AC</th>
<th>% (of max.)</th>
<th>Total mentioning frequency</th>
<th>Manual cluster (MC)</th>
<th>Rank MC</th>
<th>% (of max.)</th>
<th>Total mentioning frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golf</td>
<td>1</td>
<td>1.000</td>
<td>222</td>
<td>Golf</td>
<td>4</td>
<td>0.653</td>
<td>222</td>
</tr>
<tr>
<td>classical/characteristic</td>
<td>2</td>
<td>0.797</td>
<td>177</td>
<td>Classic</td>
<td>3</td>
<td>0.679</td>
<td>251</td>
</tr>
<tr>
<td>work</td>
<td>3</td>
<td>0.667</td>
<td>148</td>
<td>labour</td>
<td>6</td>
<td>0.693</td>
<td>205</td>
</tr>
<tr>
<td>family/family circle</td>
<td>4</td>
<td>0.622</td>
<td>138</td>
<td>family</td>
<td>1</td>
<td>1.000</td>
<td>840</td>
</tr>
<tr>
<td>manufactory/corporation</td>
<td>5</td>
<td>0.581</td>
<td>129</td>
<td>corporation</td>
<td>8</td>
<td>0.491</td>
<td>167</td>
</tr>
<tr>
<td>quality/stable value</td>
<td>6</td>
<td>0.586</td>
<td>130</td>
<td>quality</td>
<td>2</td>
<td>0.844</td>
<td>287</td>
</tr>
<tr>
<td>partner/husband</td>
<td>7</td>
<td>0.536</td>
<td>119</td>
<td>male</td>
<td>7</td>
<td>0.526</td>
<td>179</td>
</tr>
<tr>
<td>Wolfsburg</td>
<td>8</td>
<td>0.464</td>
<td>103</td>
<td>Niedersachsen</td>
<td>12</td>
<td>0.353</td>
<td>120</td>
</tr>
<tr>
<td>design/model</td>
<td>9</td>
<td>0.446</td>
<td>99</td>
<td>class</td>
<td>23</td>
<td>0.215</td>
<td>73</td>
</tr>
<tr>
<td>job/position</td>
<td>10</td>
<td>0.423</td>
<td>94</td>
<td>job</td>
<td>13</td>
<td>0.332</td>
<td>113</td>
</tr>
<tr>
<td>employer/boss</td>
<td>11</td>
<td>0.378</td>
<td>84</td>
<td>employer</td>
<td>20</td>
<td>0.235</td>
<td>80</td>
</tr>
<tr>
<td>emotional</td>
<td>12</td>
<td>0.306</td>
<td>68</td>
<td>emotion</td>
<td>17</td>
<td>0.265</td>
<td>90</td>
</tr>
<tr>
<td>expensive/costly</td>
<td>13</td>
<td>0.302</td>
<td>67</td>
<td>expensive</td>
<td>5</td>
<td>0.644</td>
<td>219</td>
</tr>
<tr>
<td>emerging/young</td>
<td>14</td>
<td>0.293</td>
<td>65</td>
<td>young</td>
<td>14</td>
<td>0.274</td>
<td>93</td>
</tr>
</tbody>
</table>
Top 30 automatic association network
Top 30 manual association network

Factor 1

Factor 2

Factor 3
Conclusions

- **Overall Match of automatic and manual clustering**
  - Match overall on acceptable level
  - Automatic clustering process able to assign the most relevant features to existing synonym vectors
  - Very good fit for clusters with a high occurrence in text

- **Need for manual recoding**
  - Remaining features are mostly clustered as one-word cluster
  - Inability of the tool to assign unknown synonyms to existing clusters
  - Especially relevant for colloquial expressions, compounded words and metaphorically used words

- **Implications**
  - Setting a threshold to reduce complexity
  - Reduction of manual clustering efforts to assure objectivity and replicability
  - Using synonyms databases in topic-specific applications
Thank you for your kind attention!!!

- Questions welcome!
Universität Hamburg
Lehrstuhl für Marketing und Innovation
Prof. Dr. Thorsten Teichert
Von-Melle-Park 5

20146 Hamburg

Tel. 040/42838-4643
ami@econ.uni-hamburg.de