Structure of this Talk

- Motivation: Why Automated Coding?
- Latent Semantic Analysis
- Algorithm I: headcount
- Algorithm II: termcount
- Evaluation
- Conclusion & Future Agenda

Motivation
Motivation

- Increased deployment of qualitative methods in marketing
- But: decrease of in-depth interviews due to high costs
- But: qualitative research has advantages: not feeding analysts expectations so much, open ended, spontaneous associations
- Problem: High Human Resource Costs
- Problem: inherent subjectivity in manual coding:
  - More interviews = more errors
  - More coders = more errors

Latent Semantic Analysis

Input (e.g., documents)

- Term = feature
- Vocabulary = ordered set of features
- $\{ M \}$ =...


Only the red terms appear in more than one document, so strip the rest.
Latent Semantic Analysis

- “Humans learn word meanings and how to combine them into passage meaning through experience with paragraph unitized verbal environments.”
- “They don’t remember all the separate words of a passage; they remember its overall gist or meaning.”
- “LSA learns by ‘reading’ paragraph unitized texts that represent the environment.”
- “It doesn’t remember all the separate words of a text it; it remembers its overall gist or meaning.”

(Landauer, 2007)

Singular Value Decomposition

\[ M = T S D^T \]

Latent Semantics

- Assumption: language utterances have a semantic structure
- However, this structure is obscured by word usage (noise, synonymy, polysemy, …)
- Proposed LSA Solution: map doc-term matrix using conceptual indices derived statistically (truncated SVD) and make similarity comparisons using angles
Similarity in a Latent-Semantic Space

\[
\sum_{i=1}^{m} \theta_{ii} = \sum_{i=1}^{m} \sum_{i=1}^{m} \cos \theta_{i}\nabla \theta_{1} \theta_{2}
\]

(Landauer, 2007)

Ex Post Updating: Folding-In

- SVD factor stability
  - SVD calculates factors over a given text base
  - Different texts – different factors
  - Challenge: avoid unwanted factor changes
    (e.g., bad essays)
  - Solution: folding-in of essays instead of recalculating

- SVD is computationally expensive
  - 14 seconds (300 docs textbase, this machine)
  - 10 minutes (3500 docs textbase, this machine)
  - … and rising!

Algorithm I: Headcount
Algorithm I: Headcount

- Calculate latent-semantic space from answers per brand per person
  - e.g. "bad advertisement focused on young target group first net in the market expensive"
- Fold-in concept of interest + synonyms & distinct paraphrases = 'seed terms' defining the concept
  - e.g. "big market share, established, known"
- (Several organised concepts = 'coding scheme')
- Headcount = $100 \times \frac{\text{number of answers correlating high with the concept}}{\text{number of answers}}$

\[ hc = \frac{|\{d \mid (\text{cor}(c,d) > t) \land (d \in M)\}|}{|M|} \times 100 \]

Algorithm II: Termcount

- Calculate latent-semantic space from answers
- Fold-in brand name
  - e.g. 'Mercedes'
- Fold-in 'seed-terms' for coding construct
  - e.g. 'secure safe stability'
- Measure distance between the two vectors = association strength (Pearson’s product moment correlation coefficient)
Evaluation

Methodology

- Pseudo Experiment to evaluate validity
- External validation: machine findings against human analysis results
- Two real-life data sets:
  - Set 1: Austrian Mobile Phone Market (Marketmind, Soja Ehrenberger, Wolfgang Rejzlik)
  - Set 2: German & US Automobile Sector (for Mercedes, Andreas Strebinger)

Data-Set 1: Mobile Phone Market

- 969 Interviews conducted by MarketMind
- Open questions to activate brand associations:
  - “Which image do you perceive if you consider brand X?”
  - “Please imagine brand Z. What do you associate?”
  - “What are your impressions and feelings you relate to brand Y?”
- Up to 10 short answers per interview
- Questions and answers in German
- Short answers (Ø: 103 chars, std. dev.: 61 chars, Ø: 14 words)
Data-Set 2: Automobile Sector

- 24 German interviews about brand 'Mercedes' in USA and Germany
  - Each interview had ~ 64 questions
    - "If I buy a Mercedes, I have a good feeling because . . ."
    - "Please characterise a typical Mercedes driver!"
    - "Please tell me three things you directly associate with Mercedes!"
  - length: long answers (each interview 3500 to 11,500 words, Ø: ~ 3500 words)
  - 1624 answers (for 1624 questions)

Results for Algorithm I

<table>
<thead>
<tr>
<th>Brand</th>
<th>D</th>
<th>T</th>
<th>ρ</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>0.5</td>
<td>0.6</td>
<td>0.52</td>
<td>0.000</td>
</tr>
<tr>
<td>One</td>
<td>0.5</td>
<td>0.6</td>
<td>0.51</td>
<td>0.000</td>
</tr>
<tr>
<td>T-Mobile</td>
<td>0.5</td>
<td>0.6</td>
<td>0.50</td>
<td>0.000</td>
</tr>
<tr>
<td>Dee</td>
<td>0.5</td>
<td>0.6</td>
<td>0.47</td>
<td>0.000</td>
</tr>
<tr>
<td>ThiRings</td>
<td>0.5</td>
<td>0.6</td>
<td>0.40</td>
<td>0.011</td>
</tr>
<tr>
<td>AI</td>
<td>0.5</td>
<td>0.5</td>
<td>0.41</td>
<td>0.000</td>
</tr>
<tr>
<td>One</td>
<td>0.5</td>
<td>0.5</td>
<td>0.40</td>
<td>0.000</td>
</tr>
<tr>
<td>T-Mobile</td>
<td>0.5</td>
<td>0.5</td>
<td>0.35</td>
<td>0.011</td>
</tr>
<tr>
<td>Dee</td>
<td>0.5</td>
<td>0.5</td>
<td>0.40</td>
<td>0.000</td>
</tr>
<tr>
<td>ThiRings</td>
<td>0.5</td>
<td>0.5</td>
<td>0.40</td>
<td>0.050</td>
</tr>
<tr>
<td>AI</td>
<td>0.5</td>
<td>0.5</td>
<td>0.55</td>
<td>0.000</td>
</tr>
<tr>
<td>One</td>
<td>0.5</td>
<td>0.5</td>
<td>0.52</td>
<td>0.000</td>
</tr>
<tr>
<td>T-Mobile</td>
<td>0.5</td>
<td>0.5</td>
<td>0.51</td>
<td>0.000</td>
</tr>
<tr>
<td>Dee</td>
<td>0.5</td>
<td>0.5</td>
<td>0.44</td>
<td>0.000</td>
</tr>
<tr>
<td>ThiRings</td>
<td>0.5</td>
<td>0.5</td>
<td>0.37</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- D: share of cumulative singular values
- T: Threshold
- ρ: Spearman's rho
- p-value: highly significant
- correlation with human judgement in a range slightly less than human-human interrater correlation
- Expl: TeleRing was very small data-set!

Results for Algorithm II

- Spearman's rho = .51
- p-value = .07
- Pearson can have negative values: outlier at 'security': seed terms very different from human coders interpretation
Conclusion

Conclusion & Future Work

- Acceptable Validity: near human results
- Eliminates coding subjectivity: High Reliability
- Proposal: headcount for large corpora, termcount for smaller and more lengthy ones
- Future work:
  - Fine tuning
  - Test with more data-sets
  - Ease applicability through provision of a software package
  - Ease Coding Construct Exploration: interpretable similarity value! (association strength?)

#eof.
Word Order Neglection?

- Educated adult understands ~100,000 word forms
- An average sentence contains 20 tokens.
- Thus 100,000^{20} possible combinations of words in a sentence
- \therefore \text{maximum of log}_2 100,000^{20}
  \quad = 332 \text{ bits in word choice alone.}
- \text{20!} = 2.4 \times 10^{19} \text{ possible orders of 20 words}
  \quad = \text{maximum of 61 bits from order of the words.}
- \frac{332}{61+332} = 84\% \text{ word choice}

(Landauer, 2007)

LSA Process & Driving Parameters

Parameter Settings

- Stopwords filtered
- Minimum word length = 2
- Share of .5/.4/.3 of the cumulative singular values
- No background corpus
- Pearson Correlation as similarity measure
- No weighting