Mining Banks Reviews in Google Maps with R

Gabriela A. Werb
Chair of Electronic Commerce
Faculty of Business and Administration
Graduate School of Finance, Economics and Management

efinance lab
at the HOUSE OF FINANCE
Agenda

- Background
- Important Concepts in Text Mining
- Preparatory Steps
- Sentiment Analysis
- Data
- Sentiment Analysis in R
Background

• User Generated Content (UGC): Word of mouth leaves a “digital trace”
  – Large amounts of unstructured data available from social channels, blogs, review sites
  – This data can contain relevant information for the business.

• Main questions
  – **What** are consumers talking about? – Discover underlying **topics** (topic modelling)
  – **How** are consumers talking? – Discover the emotional **tone** (sentiment analysis)
• Bag-of-words model (unigrams)
  – Focus on term frequencies (order does not matter)

“My bank is the best”
### Important Concepts in Text Mining

**Document**

- blog entry, review, tweet

**Corpus**

- Collection of documents

#### Document Term Matrix

<table>
<thead>
<tr>
<th>Terms</th>
<th>bank</th>
<th>bad</th>
<th>fees</th>
<th>good</th>
<th>service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document 1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Document 2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Document 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Document 4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Ben H, 4 reviews

★★★★★ a year ago
(Translated by Google) Very friendly and helpful. Top Consulting. Always have a Council in stock. My bank of trust!

(Original)

Jette Sievers, 6 reviews

★★★★★ 2 weeks ago
(Translated by Google) Service! Here is an ir not refer to any hotli other stores. Keep it

(Original)
Sehr freundlicher Kontakt, keine Verteufelung. Ich werde weiterhin bei anderen Filialen passen.

La Familia, 27 reviews

★★★★★ 4 months ago
(Translated by Google) TOP, perfect service, top Internet banking, top prices, think currently the best bank, thank you, keep it up!

(Original)
TOP, perfekter Service, Top Internet Banking, Top Konditionen, denke zur Zeit die beste Bank, herzlichen Dank, weiter so!
## Preparatory Steps

1. **Remove unwanted characters**
   - Special characters
   - Punctuation
   - Numbers
   - Words < 2 characters
   - "Bin schon 10 Jahre da und sehr zufrieden."
   - "Bin schon Jahre und sehr zufrieden"

2. **Lower case + Tokenization**
   - Break down sentences into words
   - "bin" "schon" "jahre" "und" "sehr" "zufrieden"

3. **Remove “stop words”**
   - Non informative common words
   - "jahre" "zufrieden"
Sentiment Analysis

• Tone = positive – negative

• Term weighting
  – Polarity based: “good” and “excellent” have the same (positive) weight
  – Valence based: “excellent” receives more weight
  – Tf-idf weighting: terms are weighted by their importance in the document and in the whole corpus

• Need a dictionary of positive and negative words
  – We will use SentiWS (University of Leipzig)
Data

- Subsample of Google Maps Reviews
- 41 branches in Frankfurt and surroundings (20 banks)
- 282 reviews in German
- Timespan: 2011 to 2018
- Source: Google Places API
SENTIMENT ANALYSIS IN R
Useful Links

• **Google Places API**
  – [https://developers.google.com/places/web-service/get-api-key](https://developers.google.com/places/web-service/get-api-key)
  – [https://developers.google.com/places/web-service/details](https://developers.google.com/places/web-service/details)

• **SentiWS Dictionaries (University of Leipzig)**
  – [http://wortschatz.uni-leipzig.de/de/download](http://wortschatz.uni-leipzig.de/de/download)

• **Dictionary of Emoticons**

• **Regex Cheatsheet for R**
BACKUP
Caveats

• Control for negation
  – “The service is awesome” vs “The service is far from awesome”
  – “The prices are the best” vs “The prices are not the best”

• Context
  – “I love this bank” vs “I love to wait forever in line in this bank”

• Natural Language Processing (NLP) is a rich and expanding field