Classification and Detection of Micro-Level Impact of Issue-Focused Films based on Reviews

Rezvaneh Rezapour, Jana Diesner
School of Information Sciences, University of Illinois at Urbana-Champaign

**Problem Statement and Contribution**
- Open questions in the field of impact assessment of information products:
  - What types of impact can an information product have on individuals?
  - Can this impact be reliably measured and predicted from user-generated text data?
- Our contributions—we developed:
  - A theoretically grounded classification schema for micro-level impact.
  - A codebook and annotation schema for labeling.
  - Identifying text level features that are indicative of these types of impact.
  - A probabilistic model for predicting the identified types of impact based on user-generated reviews of documentaries.

**Method**

**Overall workflow** (see also Figure 1):
1. Obtain permission from data provider.
2. Collect corpus of user-generated reviews of issue-focused documentaries.
3. Develop categorization schema for micro-level impact based on a) systematic review of literature from different fields (psychology, media studies, journalism, communication) and b) close reading of samples from our data.
4. Develop codebook for annotating reviews (Table 1).
5. Train two human coders to apply codebook to the data. Iterative annotation process until inter-rater reliability was sufficiently high.
6. Analyze the annotated data. Select features. Train a classifier that predicts the outlined impact categories. Assess accuracy. Conduct errors analysis.

**Data**

We collected (with permission from data provider) 2,290 user-generated reviews from eight issue-focused documentaries: “Fed Up,” “This Changes Everything,” “Pray the Devil Back to Hell,” “Darkly,” “Pandora’s Promise,” “SolarMamas,” “The House I Live In,” and “Pay to Play.”

**Results**

Table 3: Classification Result for SVM, RF, and NB

<table>
<thead>
<tr>
<th>Features</th>
<th>SVM</th>
<th>RF</th>
<th>NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unigram+Bigram</td>
<td>57.3</td>
<td>51.2</td>
<td>52.5</td>
</tr>
<tr>
<td>Unigram+Bigram+Trigram</td>
<td>57.3</td>
<td>51.5</td>
<td>52.7</td>
</tr>
</tbody>
</table>

**Conclusions**

- Information products, namely issue-focused documentaries, can impact peoples’ perception of an issue. This impact can be associated with changes in the understanding and attitudes toward societal problems.
- Micro-level impact on behavior, cognition, emotions and opinions can be predicted from user-generated text data.
- The developed codebook advances research in revision mining, e.g., by enabling the extraction of different types of impact from reviews.
- This work advances the field of social impact assessment by providing a data driven, computational and probabilistic solution to identifying and classifying impact.

**Acknowledgments**

This work is supported by the FORD Foundation, grant 0155-0370, and a faculty fellowship from the National Center of Supercomputing Applications (NCSA) at UIUC. We are grateful to Amazon for giving us permission to collect review data from their website. We thank Professor Corina Roxana Girju from the Linguistics department at UIUC for her helpful insight and direction for developing the codebook. We also thank Sanda Franco, Harathi Korrapati and Julian Chin, from the School at Illinois, for their help with this work.