Quantifying in-domain Distributed Word Representations from Privacy Policies

Vinayshekhar BK, Abhilasha Ravichander, Peter Story and Norman Sadeh

**Motivation**

Privacy policies are documents which describe what companies do with users’ data. These are very long and often written in a way that is difficult for an average user to understand. In the recent past there has been interest in the Natural Language Processing (NLP) community to analyze these policies using machine learning algorithms. We want to improve research in this domain by building word embeddings. We use 150000 privacy policies to build word vectors in an unsupervised manner.

**System Architecture**

- Play store corpus
- Fast Text
  - Word Embeddings
  - Embeddings Store
  - Text to word Vector
    - OPP 115 corpus
    - Statistics generator
- Various ML models
- Classification of segments
- Embeddings in vector space

**Deep CBOW Model**

- Softmax 2 units
  - Hidden Layer 2 (Relu - 64 units) (DropoutRate-0.2)
  - Hidden Layer 1 (Relu - 64 units) (DropoutRate-0.3)
- Average Embeddings - 300 dimensions

**Corpus Stats and Results**

<table>
<thead>
<tr>
<th>Category</th>
<th>Avg F1</th>
<th>Std Dev</th>
<th>Test-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Party Collection/Use</td>
<td>0.814</td>
<td>0.01</td>
<td>0.801</td>
</tr>
<tr>
<td>Third Party Sharing/Collection</td>
<td>0.791</td>
<td>0.02</td>
<td>0.79</td>
</tr>
<tr>
<td>User Choice Control</td>
<td>0.692</td>
<td>0.07</td>
<td>0.712</td>
</tr>
<tr>
<td>Data Security</td>
<td>0.838</td>
<td>0.01</td>
<td>0.837</td>
</tr>
<tr>
<td>Intl and Specific Audiences</td>
<td>0.898</td>
<td>0.03</td>
<td>0.871</td>
</tr>
<tr>
<td>Access, Edit and Delete</td>
<td>0.757</td>
<td>0.04</td>
<td>0.823</td>
</tr>
<tr>
<td>Policy Change</td>
<td>0.917</td>
<td>0.08</td>
<td>0.875</td>
</tr>
<tr>
<td>Data Retention</td>
<td>0.55</td>
<td>0.05</td>
<td>0.58</td>
</tr>
<tr>
<td>Do Not Track</td>
<td>0.949</td>
<td>0.07</td>
<td>0.941</td>
</tr>
</tbody>
</table>

By using in-domain word embeddings we improve the performance of our classifiers.

- Macro F1: Glove Embeddings (Previous best): 0.76
- Fast Text Embeddings: 0.8

Table 2: Feed-forward Network’s result on the dev and test set across different data-practices in the OPP-115 corpus.

**Fast Text Word Embeddings**

- ed, ti, d

**Glove Word Embeddings**

- data, collection, track, cookies, browser

**Amount of data needed**

More training data is needed for the Data Retention class which has a lower number of examples.