Multilingual detection of Fake News Spreaders via Sparse Matrix Factorization

Boshko Koloski
Senja Pollak
Blaž Škrlj

University of Ljubljana
Faculty of Computer and Information Science

Jožef Stefan Institute, Ljubljana, Slovenia
Task

Given Twitter feed of an author determine if the user is:

- Fake-news spreader
- Non-spreade

- Languages: English & Spanish
- 30 tweets per author, 150 negative & 150 positive cases for both languages
- Evaluation on classification accuracy
Motivation

- Fake news make a significant impact on society
- Analysis of representations' expressiveness learned via multilingual LSA
Preprocessing

- Concatenate author’s tweets
- Remove Punctuation
- Remove URL
- Remove hashtags
- Remove stopwords
- Cleaned data
Feature generation

Example tweet:

1) Character n-grams (1,2):
   - 1-gram: d, o, n; 2-gram: do, on, nt;
2) Word n-grams (2,3):
   - 2-grams: dont know; 3-gram: dont know where;
3) TF-IDF on generated features
Latent Semantic Analysis

TF-IDF on n-grams → Sparse Matrix Factorization → LSA space
Visualization of training data
Models

- Stochastic Gradient Descent based:
  - linear-SVM
  - logistic regression
- Monolingual vs Multilingual model
- 10-fold GridSearchCV on 90% on the data; evaluate on 10%
Optimization

- Grid search on:
  - Number of generated features, $n$: $[2500, 5000, 10000, 20000, 30000]$
  - Number of dimensions in the SVD, $d$: $[128, 256, 512, 640, 768, 1024]$

- Model fine-tuning (regularization):
  - ElasticNet regularization
    - Lasso
    - Ridge
Learning pipeline

Generate word + char n-grams → Perform SVD → LSA representation → Apply classifiers → Support Vector Machine → Evaluate @ Classification Accuracy

Logistic Regression

Evaluate on different SVD dimensions and number of features.
Learning
Alternative approaches

- Separate model for each language
- Doc2Vec & BERT representations
- Different Tokenizer: TweetTokenizer
- Tested AutoML methods, scored similarly to the proposed model
## Results on DEV

<table>
<thead>
<tr>
<th>name</th>
<th>type</th>
<th>#features</th>
<th>#dimensions</th>
<th>model</th>
<th>EN ACC</th>
<th>ES ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>tfidf_large</td>
<td>multi</td>
<td>5000</td>
<td>768</td>
<td>LR</td>
<td>0.9633</td>
<td>0.9867</td>
</tr>
<tr>
<td>tfidf_tweet_tokenizer</td>
<td>multi</td>
<td>5000</td>
<td>768</td>
<td>LR</td>
<td>0.9633</td>
<td>0.9533</td>
</tr>
<tr>
<td>tfidf_small</td>
<td>mono</td>
<td>5000</td>
<td>512</td>
<td>SVM,SVM</td>
<td>0.9700</td>
<td>0.4900</td>
</tr>
<tr>
<td>tfidf_cv</td>
<td>mono</td>
<td>10000</td>
<td>768</td>
<td>SVM,SVM</td>
<td>0.9100</td>
<td>0.9367</td>
</tr>
<tr>
<td>tfidf_no_hash</td>
<td>multi</td>
<td>10000</td>
<td>768</td>
<td>LR</td>
<td>0.9300</td>
<td>0.9067</td>
</tr>
<tr>
<td>doc2vec_baseline</td>
<td>mono</td>
<td>100</td>
<td>#</td>
<td>RF,SVM</td>
<td>0.6428</td>
<td>0.6971</td>
</tr>
<tr>
<td>tfidf_tpot_baseline</td>
<td>mono</td>
<td>30000</td>
<td>#</td>
<td>LR,SVM</td>
<td>0.7500</td>
<td>0.7400</td>
</tr>
<tr>
<td>tfidf_baseline</td>
<td>mono</td>
<td>10000</td>
<td>#</td>
<td>LR,LR</td>
<td>0.5567</td>
<td>0.7033</td>
</tr>
</tbody>
</table>

*Table 2. Final training data on TIRA.*
## Final evaluation results

<table>
<thead>
<tr>
<th>name</th>
<th>type</th>
<th>#features</th>
<th>#dimensions</th>
<th>model</th>
<th>EN ACC</th>
<th>ES ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>tfidf_large</td>
<td>multi</td>
<td>5000</td>
<td>768</td>
<td>LR</td>
<td>0.7150</td>
<td>0.7950</td>
</tr>
<tr>
<td>tfidf_cv</td>
<td>mono</td>
<td>10000</td>
<td>768</td>
<td>SVM, SVM</td>
<td>0.7000</td>
<td>0.7950</td>
</tr>
</tbody>
</table>

Table 3. Un-official evaluation on test data on TIRA

<table>
<thead>
<tr>
<th>POS</th>
<th>TEAM</th>
<th>EN</th>
<th>ES</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>bolonyai20</td>
<td>0.7500</td>
<td>0.8050</td>
<td>0.7775</td>
</tr>
<tr>
<td>1</td>
<td>pizarro20</td>
<td>0.7350</td>
<td>0.8200</td>
<td>0.7775</td>
</tr>
<tr>
<td>-</td>
<td>SYMANTO (LDSE) [1]</td>
<td>0.7450</td>
<td>0.7900</td>
<td>0.7675</td>
</tr>
<tr>
<td>3</td>
<td>koloski20</td>
<td>0.7150</td>
<td>0.7950</td>
<td>0.7550</td>
</tr>
</tbody>
</table>
Conclusion

- Space obtained by word and character n-grams is a good representation of the problem space.
- Semantic features don’t introduce significant improvements.
- Multilingual space maintains space structure and word patterns.
- Multilingual approach tackles the problem better compared to the monolingual approach.
Further work

- Explore and exploit the multilingual approach on more languages.
- Try to enrich the space with a background knowledge about entities appearing in the text.