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-spreader

- 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

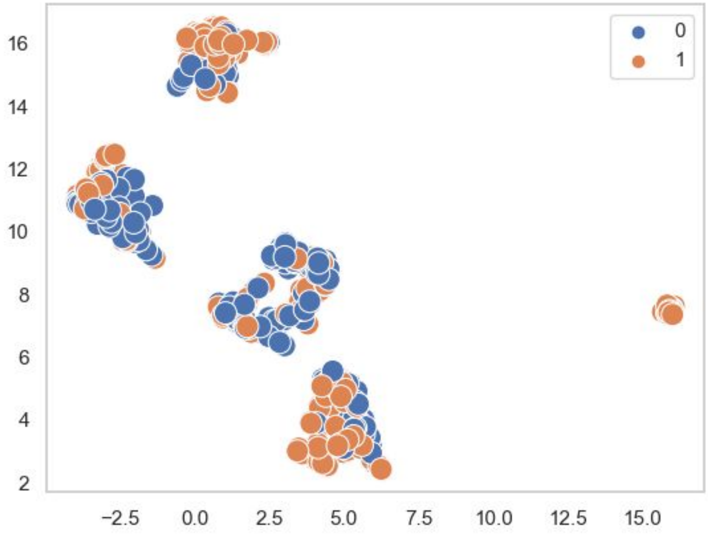
Feature generation

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

Visualization of training data



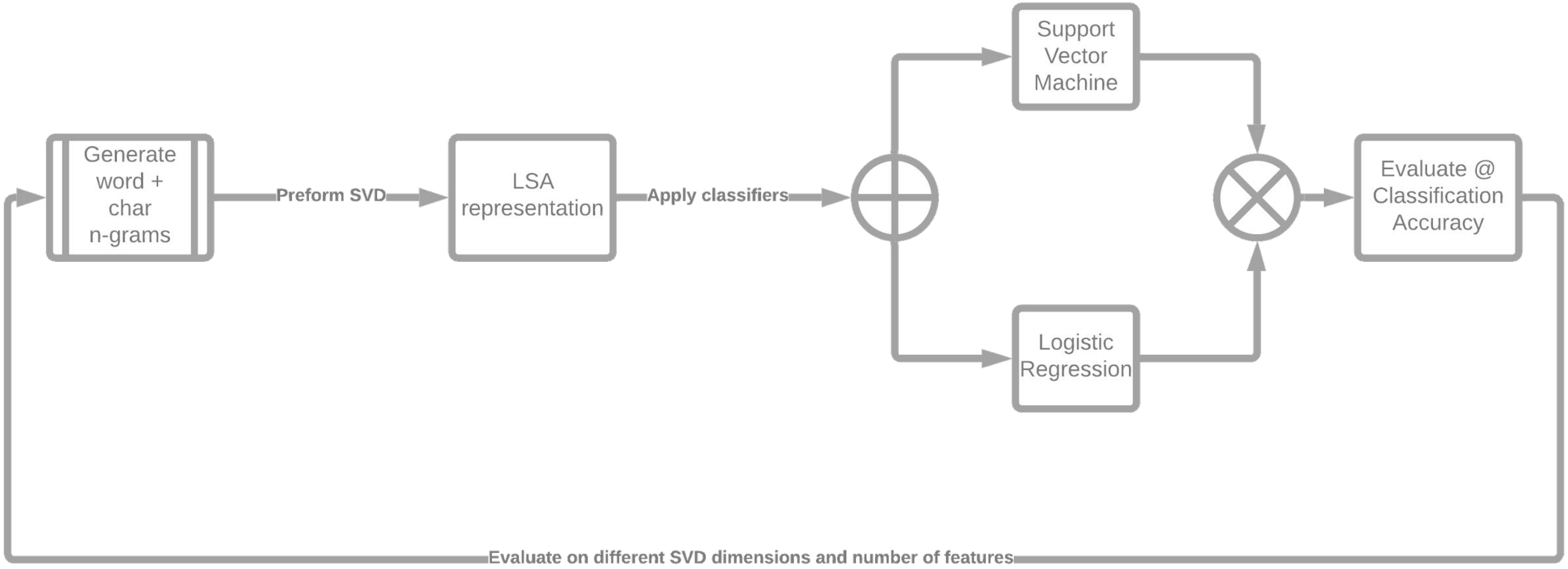
Models

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

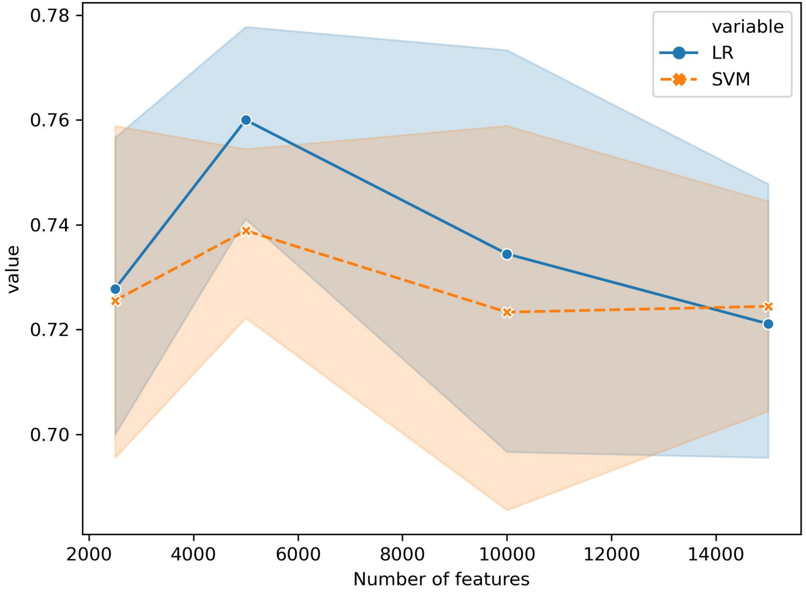
Optimization

- 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



Learning



variable
LR
SVM
0.78
0.76
0.74
0.72
0.70
value
2000
4000
6000
8000
10000
12000
14000
Number of features

Alternative approaches

- Separate model for each language
- Doc2Vec & BERT representations
- Different Tokenizer: TweetTokinzer
- Tested AutoML methods, scored similarly to the proposed model

Results on DEV

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

Table 2. Final training data on TIRA.

Final evaluation results

name	type	#features	#dimensions	model	EN ACC	ES ACC
tfidf_large	multi	5000	768	LR	**0.7150**	**0.7950**
tfidf_cv	mono	10000	768	SVM,SVM	0.7000	**0.7950**

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

POS	TEAM	EN	ES	AVG
1	bolonyai20	0.7500	0.8050	0.7775
1	pizarro20	0.7350	0.8200	0.7775
-	SYMANTO (LDSE) [1]	0.7450	0.7900	0.7675
3	koloski20	0.7150	0.7950	0.7550

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.