Preparing Stance Detection

Kristóf Vass
K.K.Vass@student.tudelft.nl

How do popular feature extraction methods compare to each other when using them on feature-based machine learning algorithms for stance detection?

1 - Background

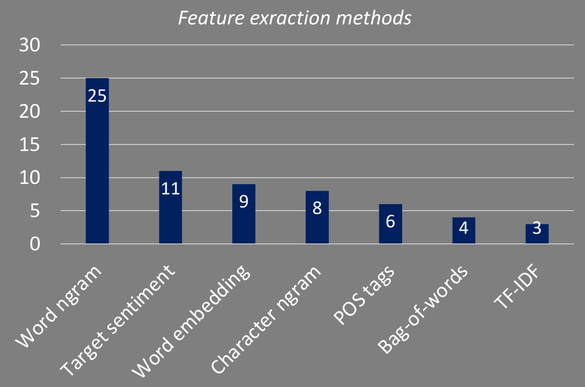
- **The research:** combining feature extraction methods and ML models to find the best combination evaluated on the SemEval-2016 Stance dataset
- **Stance Detection:** compare text-target pairs to determine their relation
- **The motivation:** Spread of fake news and disinformation is an urgent issue

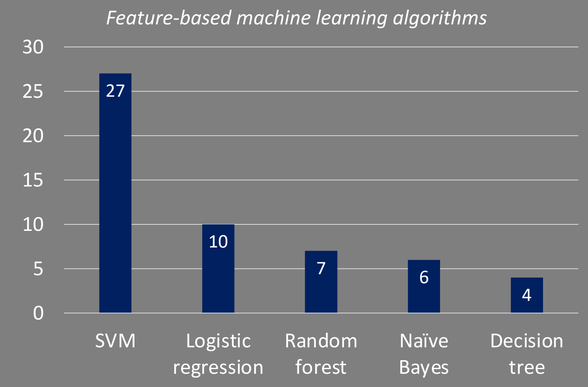
Examples of texts and their stance towards a target [Mohammad et. al, 2016]

Target: Climate change is a real concern	
Against	ONE Volcano emits more pollution than man has in our HISTORY!
Neutral	Climate change is currently a hot topic to talk about.

2 - Methodology

1. Literature study on most commonly used methods



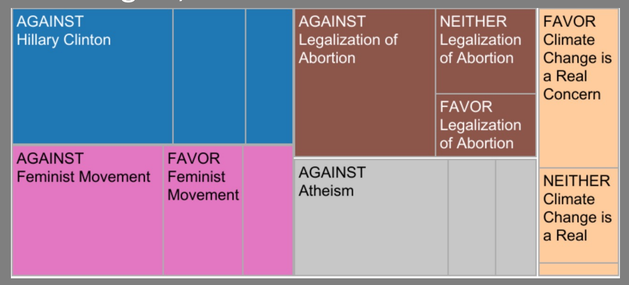


2. Combine and implement elements

"Be kind to the earth beneath your feet" → → FAVOR

3 - Dataset

- SemEval-2016 Stance dataset
- For training and evaluation
- Dataset created from tweets
- 5 targets, 3 labels



Divison of targets and stances [Mohammad et. al, 2016]

The dataset can be found through the QR code

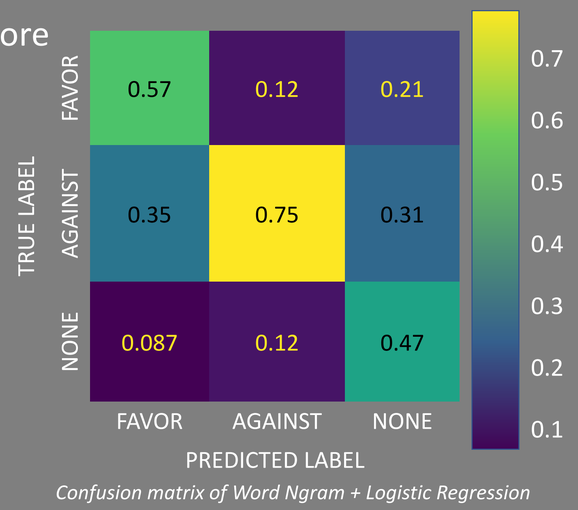
4 - Results

- Performance measured with accuracy and F-score
- F-score is the most commonly used

F-scores of the pairs

	SVM	Logistic Regression	Random Forest	Naïve Bayes
TF-IDF	0.5663	0.5684	0.5500	0.4379
WE	0.5978	0.5396	0.5359	0.4726
BOW	0.5062	0.5686	0.5672	0.5418
Ngram	0.5467	**0.5991**	0.5741	0.5536
POS	0.5482	0.5649	0.4793	0.4674

Best result: Word Ngram + Logistic regression
- F-score: 0.5991
- Accuracy: 0.6557



Confusion matrix of Word Ngram + Logistic Regression

- 3 highest values are on the diagonal
- "AGAINST" has the highest accuracy

References

[Mohammad et. al, 2016] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. Semeval-2016 task 6: Detecting stance in tweets. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pages 31–41, 2016.

GitLab repository of the source code - access can be requested through email.