SPRÅKBANKENTEXT

A First Attempt at Unreliable News Detection in Swedish

Ricardo Muñoz Sánchez, Eric Johansson,
Shakila Tayefeh, Shreyash Kad

Unreliable News Detection

Let's take news article A

Does it come from a source that usually presents COVID-19 information in an unreliable way?

Can we actually detect this?

Our Data

The original dataset comes from Kokkinakis (2021)

Contains news articles in Swedish related to the COVID-19 pandemic

The data comes from from medical and guvernmental outlets to conspiracy theory blogs

Our Data

- The original dataset has eight different domain labels

- We grouped them into two classes, according to their source:
  - Reliable
  - Unreliable

Our Data - Caveats

- There are no fact-checking agencies in Sweden

- The dataset is very small (~2K articles)

- The dataset is very unbalanced (one unreliable article for every ten reliable ones)

Our Solution – Data Augmentation

Subsampling — Generates a more balanced dataset by leaving out reliable articles but makes the dataset size problem worse
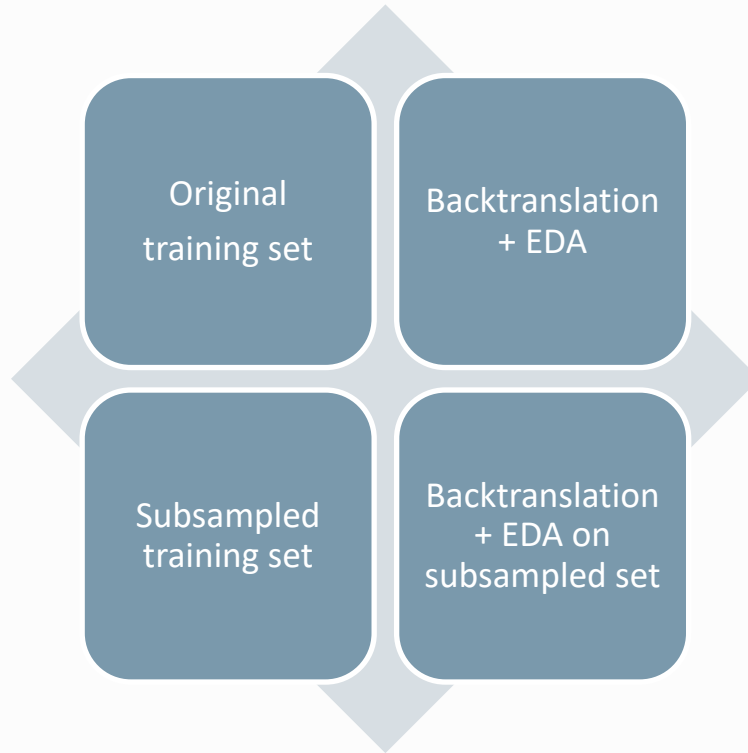
Backtranslation — Synthetic data is generated by translating the original article to English and then back into Swedish

Easy Data Augmentation (EDA) — Uses four simple operations to alter the text – synonim replacement and random insertions, swaps, and deletions

Training Datasets

Our Models

Classical Machine Learning

- Logistic Regression
- Support Vector Machine (SVM)

Deep Learning

- LSTM + word2vec
- BERT

Results

Model	Balanced	EDA	Accuracy	F1-Score
LogReg (Baseline)	No	No	0.922	0.631
LogReg	**Yes**	**Yes**	0.953	**0.866**
SVM	No	Yes	**0.956**	0.837
LSTM	No	No	0.943	0.824
BERT	Yes	Yes	0.952	0.837

Takeaways

- EDA greatly improved the performance of the classical machine learning approaches

- The F1 Score for the unreliable news was very low compared to the for reliable news ones

- Subsampling did not have an easily interpretable pattern

Future Directions

- Create a larger dataset of mis/disinformation in Swedish
  - More sistematic data retrieval
  - Article and source-level annotations
  - Different kinds of mis- and disinformation

- Use liguistic and psycholinguistic features
  - Have proven to be useful both with smaller and larger datasets
  - Allow for the use of feature importance

- Alter the EDA algorithm to better deal with longer form text and with sequential models

GÖTEBORGS UNIVERSITET

CHALMERS
UNIVERSITY OF TECHNOLOGY

SPRÅKBANKENTEXT

Ricardo Muñoz Sánchez

ricardo.munoz.sanchez@svenska.gu.se

rimusa.github.io