M2 Software Project

Oaks
Fake News Detector

Anna MOSOLOVA
Elisa LUBRINI

UNIVERSITÉ DE LORRAINE

IDMC Institut des sciences du Digital Management & Cognition

September 28, 2021

Contents

1. Introduction
2. Literature Review
3. Our Solution
4. The Product

Introduction

News plays an important role in shaping people's beliefs and opinions. The spread of fake news is being increasingly problematic. [**Zhou_2019**]

Example

The past election cycle for the 45th President of the United States. [**Zhou_2019**]

NLP can help through **fake news detection**:

- **reminding** newsreaders to be alert
- providing information on the **accuracy level** of each case

Literature Review

Knowledge base approach

- DEAP-FAKED [**mayank2021deap**]
  - Dataset: Kaggle Fake News[1]
  - biLSTM for article representation
  - ComplEx algorithm for knowledge base entity representation (WikiData)
  - Classification with Multilayer perceptron
  - F1: 88.66%

[1] https://www.kaggle.com/c/fake-news/overview

Literature Review

Neural network approach

- Exploring Text-transformers in AAAI 2021 Shared Task: COVID-19 Fake News Detection in English [**li2021exploring**]
  - Dataset: COVID19 Fake News Detection in English[2]
  - Bert, Ernie, Roberta, XL-net, and Electra ensemble with pseudo-labeling
  - F1: 98.59%

[2] https://competitions.codalab.org/competitions/26655

Literature Review

Linguistic features approach

- Linguistic feature based learning model for fake news detection and classification [**choudhary2021linguistic**]
  - Dataset: Buzzfeed Political News Data, Random Political News Data
  - Features:
    - Syntax-based
    - Sentiment-based
    - Grammatical-based
    - Readability-based
  - Ensemble of Gaussian Naïve bayes, Kernel Naïve Bayes, Linear SVM, Gaussian SVM
  - Accuracy: 72%

Our Solution

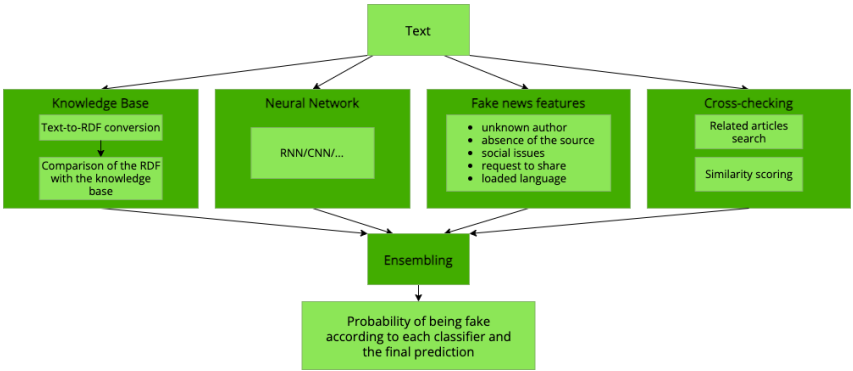

Figure: Ideal structure of the system's algorithm

Knowledge Base approach

- Text-to-RDF graph conversion
- Comparison with the WikiData knowledge base

Neural Network approach

Modern neural network architectures that show SoTA results in the various NLP tasks

- GRU
- LSTM
- CNN
- Transfer learning

Feature-based approach

As stated in [**fakenews_features**], fake news may be distinguished by the following features:

- unknown author
- absence of the source
- social issues
- requests to share the news
- loaded language usage

Cross-checking approach

- search of the similar articles on GoogleNews
- verification if the information is not contradictory

The Product

Various options are possible:

- **Web** App
- Browser **Extension**
- **Smartphone** App

The Product

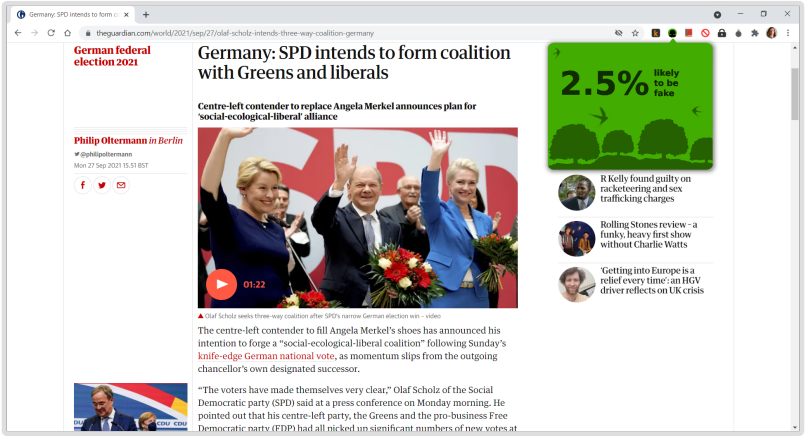

Figure: Draft mockup of the browser extension

The Product

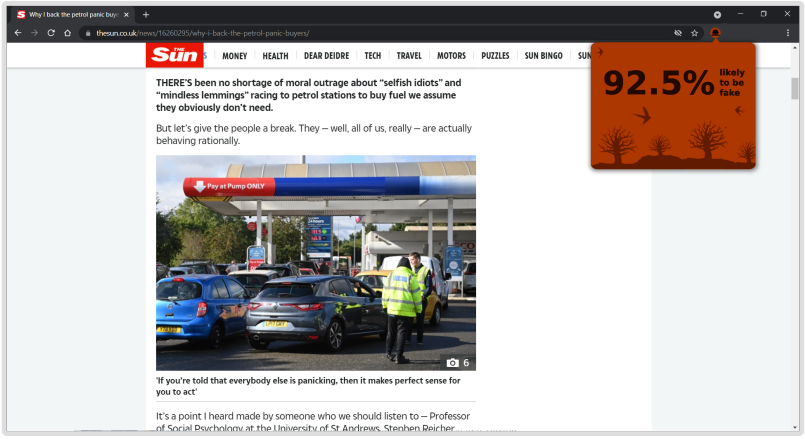

Figure: Draft mockup of the browser extension when detecting fake news

Bibliography I