COMP3900 Model proposal

Abstract: Integration with multiple approaches. Text analysis using LIWC with classification model + Google fact check tools API. Use LLAMA2 language model to integrate two different results and put them together in English.

Motivation: Combine these methods to complement each other's drawbacks.

1) LIWC-based analysis provides analysis of text in terms of contextualised psycho-social and cognitive factors, which aims to understand the nature of fake news. However, this cannot always guarntee correct detection of fake news and may generate false positives(negatives) as:

a) Complexity and subtlety of language itself.

b) Might not be able to distinguish between satire sarcasm, and actual misinformation.

c) Might contain bias inherited from training data.

> LIWC-only based existing systems show about 70% correctness, slight difference depends on the model (according to the articles below).

2) Database-based analysis may outperform in general cases and solve above problems as this involves human to detect fake news, but they also have several limitations:

a) Information could be outdated.

b) Imcomplete coverage of topics.

c) Dependence on Reliable sources.

d) Dataset has a risk of manipulation.

e) Human bias.

d) Scalability issue in terms of cost, effort and resources.

> Google fact check tools API mostly provides correct data, but cannot even answer very simple questions depends on the topic.

Solution: Generate our own LIWC-based model result, source related information from Google fact check tools API, which is one of the largest existing fake news dataset, and use this together to determine falisty. Large language model LLAMA2 can provide explanation in English.

Advantage:

1. Could solve issues that existing systems/works have.
2. Could provide user-interpretable explanations.
3. Could provide detailed information about the falsity based on various data, instead of true/false binary classification.
4. Could minimise bias by using multiple sources.

Limitation: Cannot highlight specific sentence in the article which is likely to be false, instead the language model can provide explanation in English as a whole article.

Plan:

1) Create data pipeline containing data sourcing (FakeNewsNet & LIAR), data preprocessing.

2) Connect LIWC CLI to the system.

3) Connect Google fact check tools API to the system via Google cloud.

4) Connect LLAMA2 via Huggingface transformer.

5) Train and evaluate which types of model performs the best for LIWC-based analysis (Random Forest, Logistic Regression, Support Vector Machine, Feedfoward Neural Network …).

6) Connect all backend components.

7) Connect to the frontend website.

Helpful articles:

https://ceur-ws.org/Vol-2696/paper_218.pdf

https://arxiv.org/pdf/1712.07709.pdf

https://sbp-brims.org/2017/proceedings/papers/challenge_papers/AutomatedFakeNewsDetection.pdf

https://cdn.aaai.org/ojs/5389/5389-13-8614-1-10-20200511.pdf

https://uu.diva-portal.org/smash/get/diva2:1782408/FULLTEXT01.pdf

https://web.eecs.umich.edu/~mihalcea/papers/perezrosas.coling18.pdf

https://arxiv.org/pdf/1708.07104.pdf