Phenom Detection

Scott Pramuk, Daniel Ching, Luke Fanguna

Our Approach

- Detect Organizations (Named Entity Recognition)
- Detect Sentiment
- Detect Toxicity
- User Interaction through IRC

Demo

Organizations + IRC contributions Approach

Organizations

- From [2401.10825] A survey on recent advances in Named Entity Recognition Hugging Face Bert-based models tended to perform the best on NER task across multiple datasets
- Some popular datasets for NER training: CoNLL 2003, WNUT 2017 (noisy text)

- Validate organizations found by comparing them to orgs in DB via Levenshtein edit distance (also include orgs not in DB with keywords)
- Challenge of evaluation: manually creating labeled data

IRC

- Show valid discussion ids to user from Digital Democracy DB

- Given a valid discussion id from the user, output all phenoms detected

Organizations + IRC contributions Implementation

Organizations

- Used pre-trained bert-based model that was fine tuned on the CoNLL-2003 dataset (dslim/bert-base-NER · Hugging Face)

- Created pipeline to process data for input to model and process output as needed for IRC (merge B-ORGS, I-ORGS), validate the output via edit distance

- Used this model to estimate proportion of utterances with orgs mentioned
- Created two samples of 100 utterances that matched calculated proportion, manually labeled all orgs I found

IRC

- Used existing chatbot as a foundation
- Incorporates various tables from DB:
  - BillDiscussion, BillAnalysis, Utterance, Person

- Expanded commands to include:
  - [[botname]: list [integer]] → outputs all discussion ids starting with integer
  - [[botname]: show [did]] → outputs all phenoms found from utterances with this did (if did is valid)

Organizations Results

Tested 3 models:

1. NLTK Max Entropy
2. bert-base model fine tuned on CoNLL-2003 dataset
3. bert-base model fine tuned on CoNLL-2003 then trained on WNUT 2017 dataset

- Model #2 tended to perform better than the others, though still relatively poorly
- Manual inspection of results also pointed to model #2

	Micro Precision	Macro Precision	Micro Recall	Macro Recall	Micro F1	Macro F1
NLTK	0.48	0.41	0.52	0.44	0.50	0.42
Bert CoNLL-2003	0.51	0.63	0.49	0.61	0.50	0.61
Bert WNUT 2017	0.50	0.24	0.50	0.24	0.50	0.24

Sentiment Analysis Overview

- Leverage sentiment analysis to provide insights into speakers' opinions
  - Identify most positive and negative speakers throughout discussion
  - Summarize these speakers to better understand their opinion
- Sentiment analysis for political speech
  - Issues with common sentiment analyzers
    - Most speech classified neutral, rarely positive or negative
  - XLM-T-Sent-Politics model trained on politician's tweets
  - Metric to represent sentiment for a speaker over course of discussion
- Summarization
  - Accurate and concise summary of speaker's opinion
  - led-large-book-summary
    - Required large input to realize context
  - facebook/bart-large-cnn
    - Output inconsistent with events that occurred during hearing
  - Generative Mistral-Nemo-Instruct-2407 LLM
  - Prompt engineering for desired output

Identifying Discussed Topics

- Approach
  - Identify similar groups of words
  - Generate political heading for word groups
- Latent Dirichlet allocation
  - Generative statistical model that leverages probabilistic methods
  - Requires tokenized text and stop words to be removed
  - Extract inputted k topics from text
- Generative Mistral-Nemo-Instruct-2407 LLM
  - Create headings from related words
  - Prompt engineering to output a political category
  - Remove duplicates

Toxic Approach

Problem: Want to identify “Toxicity” in a Bill Discussion.

Toxic Implementation

Toxic-Bert [1]:

- Biased: WIkipedia Comments
- Unbiased: Civil Comments
- Multilingual: Both
- Toxic, severe_toxic, obscene, threat, insult, identity_hate, sexual_explicit

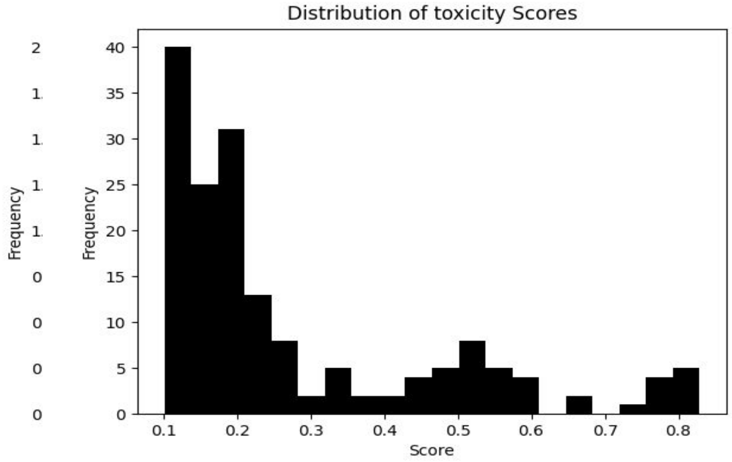
Analysis:

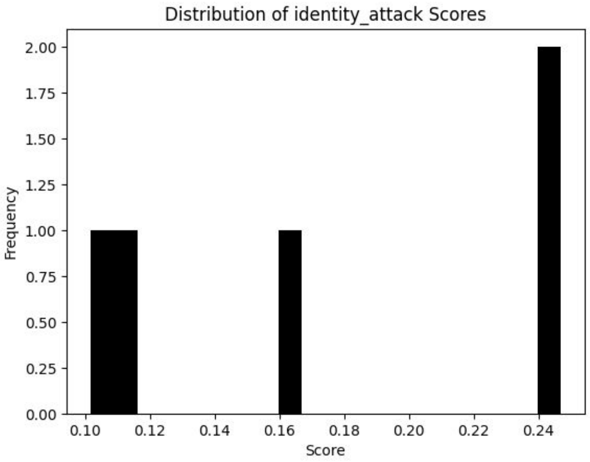
- Used Toxic-Bert to label data
- Interpret data to add a threshold
- Determine whether a given label and score is enough to trigger phenom

Toxic Results

Metrics:

- Attempted to use both models
    - Decided to use biased model because we want to find toxicity
    - Unbiased model was tagging too much





References

[1] Detoxify, Laura Hanu and Unitary team. Detoxify. Github. Available at:

https://github.com/unitaryai/detoxify, 2020.

[2] Keraghel, Imed, Stanislas Morbieu, and Mohamed Nadif. "A survey on recent advances in Named Entity Recognition." *Centre Borelli UMR9010 Université Paris Cité, Kernix Software*, Year.

[3] Barbieri, Francesco, et al. "XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond." Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, 2022, pp. 258–266. https://aclanthology.org/2022.lrec-1.27.

[4] Hugging Face. "Mistral-Nemo-Instruct-2407." Hugging Face, 2024, https://huggingface.co/mistralai/Mistral-Nemo-Instruct-2407. Accessed 11 Dec. 2024.

Questions