Sentiment Analysis Towards Named Entities – PoC

Patryk Rakus, Michał Tomczyk,
Filip Szympliński, Julia Kaznowska

Scientific goal and justification

- Enhancing sentiment analysis by putting an emphasis on Named Entity Recognition (NER) - by focusing on sentiments directed towards Named Entities (NEs) in textual data
- Contributing both to NLP and explainable AI (XAI), by working with interpretable machine learning methods, e.g. LIME, SHAP or attention based mechanisms
- The project will address whether integrating NEs focused sentiment classification with explainable models will increase the interpretability and usability
- The topic selection was motivated by the growing demand for explainable and ethical AI and existing gaps in the field of sentiment analysis

State of the Art - sentiment analysis

- Sentiment analysis has always been an important field within NLP
- First attempts were based on pattern-matching systems
- Over the time, classifiers and statistical models have started to be applied
- The development of deep learning methods and the increase of the computational power drastically changed this field of study

SOTA – large language models

- In recent years, transformer-based networks are commonly applied for sentiment analysis and NER
- A foundation for most state of the art solutions is BERT - Bidirectional Encoder Representations from Transformers, introducing bidirectional contextual sentiments
- Fine-tuned variations of it (e.g. SciBERT) show exceptional performance in domain-specific tasks, incorporating NER into a single transformer
- Another architecture with significant impact is GPT - Generative Pre-trained Transformer, combining unsupervised pre-training and supervised fine-tuning

SOTA - XAI Techniques

LIME, SHAP

Checking drops on prediction quality when removing one word

Rarely applied specifically for NE-focused sentiment analysis

Conterfactual explanations

Altering the input text to detect changes in output predictions

Attention mechanisms

Monitoring the attention weights can help determining which parts of sentences the model focuses on

Sentiment analysis – challenges

Adding context
One of the problems is defining, where the context starts and ends

Example:
"I absolutely loved the main character, Buzz, in Toy Story, but the ending of the movie was terribly disappointing"

Sarcasm and irony
It requires a deeper understanding of nuance and tone.

Example:
"Great, I love waiting in the queue for hours"

Domain-specific language
Sentiment can vary across domains.

Example:
“The plot was predictable"

Datasets

MultiCoNER 2022 dataset

- Took only the english part:
  - Train sentences count: 15300
  - Dev sentences count: 800
  - Test sentences count: 217818
- Up to 49 words per sentence
- Entities: CORP (Corporation), CW (Creative Work), GRP (Group) LOC (Location), PER (Person), PROD (Product), O (Other)
- All labels except O have 2 forms: starting with “B-” (e.g. B-PROD) and “I-” (e.g. I-PROD), signifying whether the word is the beginning of the entity, or a continuation of it, respectively

PoC - Named Entity Recognition

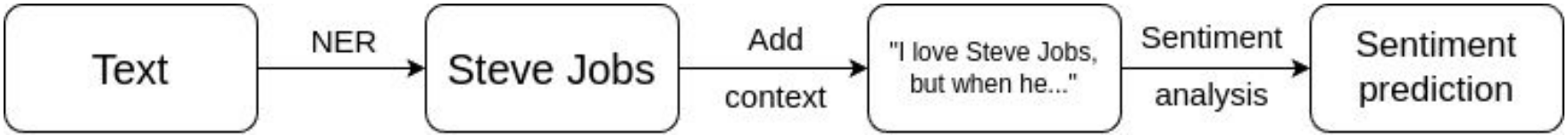
- The NER model used is a fine-tuned version of BERT
- The input words are tokenized with the tokenizer provided along the model, which can split words into sub-words. Because of that we specifically align the labels only to the first sub-token. The other sub-tokens, and padding are specially annotated by a [PAD] token.
- Within just 5 epochs:
  - Train accuracy: **0.9802**
  - Validation accuracy: **0.9655**

```
(I = O)
(love = O)
(steve = B-PER)
(jobs, = I-PER)
(but = O)
(when = O)
(he = O)
(created = O)
(the = O)
(iphone = B-PROD)
(15 = I-PROD)
(was = O)
(the = O)
(worst = O)
(phone = O)
(ever = O)
```

PoC - Sentiment Analysis

The sentiment analysis model uses DistilBERT fine-tuned on SST-2 (Stanford Sentiment Treebank) dataset. It is a distilled version of BERT, 40% smaller and 60% faster with minimal decrease in language understanding capabilities.

Proposed sentiment analysis pipeline



Example of model's capabilities

Input: "I love *steve jobs*, but when he created the *iphone 15* was the worst phone ever."

Output:

```
{'entity': 'steve jobs,', 'sentiment': {'label': 'POSITIVE', 'score': 0.9965392351150513}}
{'entity': 'iphone 15', 'sentiment': {'label': 'NEGATIVE', 'score': 0.9995965361595154}}
```

Thank you for your attention

References

- Voita, E., Talbot, D., Moiseev, F., Sennrich, R., & Titov, I. (2019). Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 5797–5808.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 4171–4186.