A comparative study of different approaches for Named Entity Recognition

GROUP 6

CONCEPT RECAP

NER, also known as entity chunking, extraction or identification, involves identifying and classifying important details (entities) in a given text.

By using NER, organizations can gain a deeper understanding of the data they collect, enabling them to make more informed decisions and identify valuable insights.

NER has numerous applications, including classifying content for news providers, implementing efficient search algorithms, powering content recommendations, and many other use cases.

OBJECTIVES

- To identify the strengths and weaknesses of different NER approaches.
- To assess and compare the efficacy of traditional NER methods and a transformer-based approach on a single dataset, which is not widely adopted as a standard benchmark for NER evaluations.
- To investigate the impact of different factors on NER performance, such as dataset size and complexity.
- To propose guidelines for selecting the most suitable NER approach for a given task.

WHAT WE DID

We initiated our comparison study by selecting the Hidden Markov Models (HMM), Conditional Random Fields (CRF), and BERT as our primary source of interest.

Afterwards, we chose our metrics for comparison, specifically F1, precision, and recall. While these metrics were taken over the dataset, F1 scores were also taken over each label/tag

Speaking of dataset, WikiNER was chosen which has a total of 7,200 labelled Wikipedia articles across nine different languages, including English, consisting of 4,853 articles.
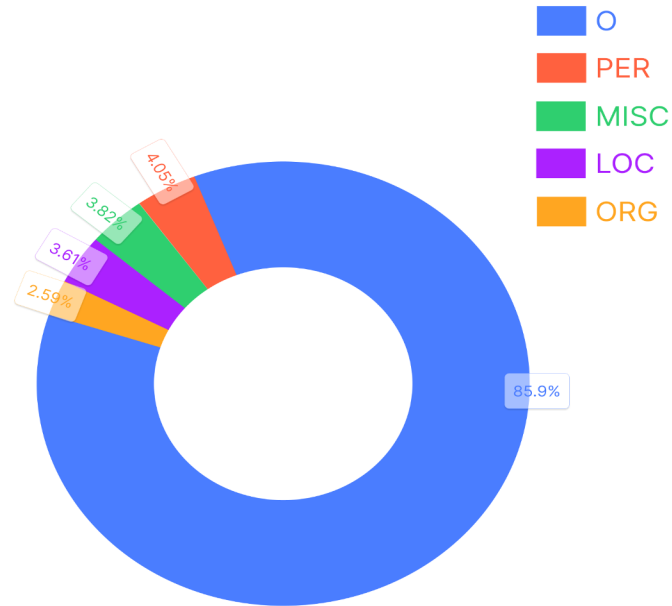
We then trained and tested the models, and subsequently stored, analyzed, and studied the results.

METHODOLOGY

The process we followed was simple and can be broken down into the following steps:

- Data collection
- Preprocessing (was not necessary as data was relatively clean)
- Model training
- Evaluation
- Interpretation

RESULTS

RESULTS

Take for example the following sentence,

"Some researchers , such as Michael Persinger of Laurentian University , Canada , have speculated that changes in geomagnetic fields could stimulate the brain 's temporal lobes and produce many of the experiences associated with hauntings ."

The correct output to the statement would be:

Michael: **I-PER**, Persinger: **I-PER**, Laurentian: **I-ORG**, University: **I-ORG**, Canada: **I-LOC**.

RESULTS

The models output as follows,

The HMM outputs: University: **I-MISC,** Canada**: I-PER;**

The CRF outputs: Michael: **I-ORG**, Persinger: **I-ORG**, of: **I-ORG**, Laurentian: **I-ORG**, University: **I-ORG**, Canada: **I-LOC;**

BERT outputs: Michael: **I-PER**, Persinger: **I-PER**, Laurentian: **I-ORG**, University: **I-ORG**, Canada: **I-ORG**;

RESULTS

Overall metrics

	Precision	Recall	F1	Accuracy
HMM	1.33%	1.08%	1.19%	78.51%
CRF	37.02%	31.27%	33.90%	90.49%
BERT	86.96%	88.23%	87.59%	98.09%

RESULTS

Class based metrics

	PER	ORG	LOC	MISC
HMM	2.15%	0.34%	0.70%	0.95%
CRF	39.72%	34.35%	35.05%	22.57%
BERT	95.01%	81.38%	89.36%	79.36%

CONCLUSION

To conclude, HMMs are useful for efficient text processing in low-resource settings but struggle with long-term dependencies and out-of-vocabulary words.

CRF's can handle long-term dependencies and out-of-vocabulary words, but may require more labeled data for training and be computationally expensive.

BERT is good for accuracy with complex language, but training and fine-tuning can be computationally expensive and require a lot of labeled data.

The approach to NER tasks depends on application requirements, such as computational resources, labeled data amount, and the desired balance between efficiency and accuracy.