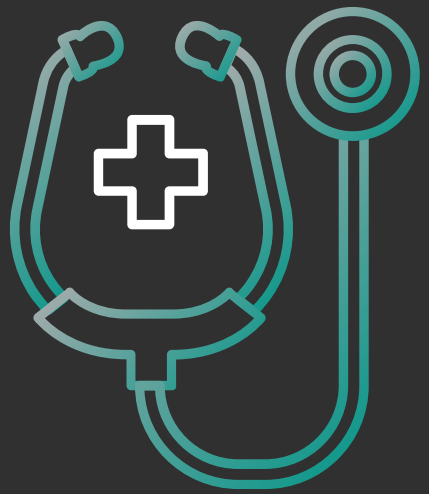
Medication Entity Extraction using Transformer Models and Mistral-7B: A Comparison

Frances Cue and Ashok Sundararaman



Motivation

- **Problem**: Electronic health records are unstructured, vary in semantics, overloaded with information causing medication errors that can lead to adverse drug events

- **Solution**: Utilize transformers and large language models to extract important medication information

NER Architectures

Rule Based

Bi-LSTM CRFs

BERT

Data

BIO Tagged for Transformers

Patient was taking
0 0 0

Penicillin 250 mg
B-Drug B-Dosage I-Dosage

every six hours for
B-Frequency I-Frequency I-Frequency 0

Infection. Now complaining
B-Reason 0 0

of upset stomach.

0 B-ADE I-ADE

JSON for Mistral-7B

```
{
  "text": "The patient is
     prescribed 20 mg of
     Prednisone B.I.D PO x 1 week
      for asthma.",
  "entities": {
    "Drug": ["Prednisone"],
    "Duration": ["1 week"],
    "Dosage": ["20 mg"],
    "Frequency": ["B.I.D"],
    "Strength": [],
    "Form": [],
    "Route": ["PO"],
    "Reason": ["asthma"],
    "ADE": []
  }
}
```

Transformer Models

BERT-base-case

Microsoft-BioMed-BioClinical-BERT(BBB)

Clinical Longformer

These models are fine-tuned for token-level classification to enhance the extraction of medication-related entities from clinical records.

Transformer Results

BERT-base-cased

High baseline at .86

BBB

Highest micro average F1 score of .88

Clinical Longformer

Highest F1 score for adverse drug event extraction at 0.54, overall F1 score of 0.87

Mistral-7B Configurations

Few-Shot

(Baseline)

Zero-Shot

(Fine-Tuned)

Few-Shot

(Fine-Tuned)

Extract the entities for the specified labels from the given medical text and provide the results in JSON format
- Entities must be extracted precisely as they appear in the text.
- Return each entity under its label without creating new labels.
- Provide a list of entities for each label. If no entities are found for a label, return an empty list.
- Prioritized accuracy and relevance in the identification of entities.

Here are the entity labels and their descriptions:
1. Drug: Extract any mentioned medications or drugs.
2. Duration: Extract the duration of treatment or medication usage.
3. Dosage: Extract dosages related to medications, including units.
4. Frequency: Extract how often the medication or treatment is to be taken or administered.
5. Strength: Extract the concentration or potency of the medication.
6. Form: Extract the form in which the medication is to be used.
7. Route: Extract the method of administration for a medication.
8. Reason: Extract the reason or condition the medication is prescribed for.
9. ADE: Extract adverse drug events or side effects mentioned.

Make sure to go through the text carefully and extract all entities mentioned above if they are present. Do not create fictitious data.

Mistral-7B Results

	Full Test Set			Removed Empty Predictions		
Model	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Few-Shot	0.17	0.18	0.15	0.44	0.47	0.41
Zero-Shot (Fine-Tuned)	**0.44**	**0.46**	**0.44**	**0.66**	**0.68**	**0.66**
Few-Shot (Fine-Tuned)	0.36	0.37	0.34	0.45	0.45	0.42

Conclusion

Limitations

- Data was annotated by humans
- LLMs can “hallucinate”
- Investigated only 1 LLM due to resource constraints

Future Work

- Exploration of more LLMs
- Further experimentation with prompt engineering
- Experimentation with languages other than English

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution