

Project Title: NLP Named Entity Recognition (NER) Using BERT

Week 3: Fine-Tuning BERT for NER

- By Hareetima Sonkar

Project Overview

This week, I focused on laying the foundation for the project titled "NLP Named Entity Recognition (NER) Using BERT (Bidirectional Encoder Representation from Transformers)". The goal was to familiarize myself with the basics of NER and BERT, set up the required environment, and explore sample datasets.

Objectives for Week 3

Fine-tune a pre-trained BERT model for the NER task.

Tasks Completed:

1. Load the Pre-trained BERT Model
We used the BertForTokenClassification model from Hugging Face's Transformers library.
Set the number of labels to match the NER classes (O, B-PER, I-PER, etc.).
Mapped the label IDs to labels and vice versa.
2. Modify the Model for Sequence Labeling
Adjusted the classifier head of BERT to fit the NER classification task.
3. Define a Loss Function
Used CrossEntropyLoss, suitable for multi-class classification problems like NER.
4. Train the Model on the Processed Dataset
Utilized the Trainer API from Hugging Face for model training.
Defined training arguments such as learning rate, batch size, evaluation strategy, number of epochs, and logging directory.
Fine-tuned the BERT model for 3 epochs using the CoNLL-2003 dataset.
5. Save the Fine-Tuned Model
Saved the model and tokenizer after training for future inference.

Training Results:

- The model was trained for **3 epochs**.
- Loss values improved across epochs:
 - **Epoch 1:** Training Loss = 0.1634, Validation Loss = 0.0404
 - **Epoch 2:** Training Loss = 0.0281, Validation Loss = 0.0381
 - **Epoch 3:** Training Loss = 0.0146, Validation Loss = 0.0374

Classification Report:

	precision	recall	f1-score	support
LOC	0.93	0.93	0.93	1666
MISC	0.78	0.84	0.81	702
ORG	0.88	0.91	0.90	1661
PER	0.97	0.96	0.96	1615
micro avg	0.91	0.92	0.91	5644
macro avg	0.89	0.91	0.90	5644
weighted avg	0.91	0.92	0.92	5644

Explanation of Classification Report:

- **Precision:** Measures how many predicted entities were correct. High precision means fewer false positives.
- **Recall:** Measures how many actual entities were correctly predicted. High recall means fewer false negatives.
- **F1-score:** Harmonic mean of precision and recall, representing the overall accuracy of the model.
- **Support:** Number of true instances for each class.
- The model performed best on **PER (Person)** entities with **97% precision and 96% recall**.
- The overall **weighted average F1-score is 92%**, showing strong performance on the NER task.

Deliverables:

- Fine-tuned BERT model for NER.
- Training script and saved model.
- Evaluation results and classification report.

The next step will be evaluating and optimizing the trained model's performance in **Week 4**.