Classifier 3: Fine-Tuned RoBERTa with BM25-Based Context Retrieval

Introduction

This component of the fake news detection system uses a transformer-based architecture, **RoBERTa**, for classification, paired with the **BM25 ranking algorithm** for retrieving contextually similar real news articles. The goal of this pairing is to maintain high classification accuracy while retrieving interpretable supporting evidence for the chatbot's explanations. RoBERTa provides deep semantic understanding of input texts, and BM25 ensures efficient context retrieval from a corpus of known real news.

RoBERTa: Robustly Optimized BERT Pretraining

RoBERTa (Liu et al., 2019) is a modification of BERT with optimized pretraining procedures including dynamic masking, longer training duration, and training on larger mini-batches with more data. It omits the Next Sentence Prediction (NSP) objective, which was shown to be less beneficial for downstream performance.

In our project, RoBERTa would be fine-tuned on a labeled dataset of real and fake news articles using the roberta-base model from Hugging Face Transformers. The model was trained using PyTorch with early stopping and the AdamW optimizer.

Advantages for This Task:

- Strong language modeling capability across news-style data
- Better generalization compared to base BERT under limited fine-tuning
- Particularly effective at capturing subtle cues and context that distinguish misleading from factual reporting

---

Context Retrieval Using BM25

To enhance the explanation component of the chatbot, we implemented **BM25** (Best Matching 25), a ranking function commonly used in information retrieval systems. BM25 scores documents in a corpus based on term frequency, inverse document frequency, and document length normalization.

How It Works in Our System:

1. Preprocess a set of verified real news articles into a BM25 index using rank_bm25.

2. At prediction time, take the user's input and compute BM25 similarity scores against the real article corpus.
3. Retrieve the top-k relevant documents as **supporting context** for interpretability.
4. Present these alongside the classification result to justify the model's decision.

Why BM25?

- Lightweight and fast to compute
- Does not require neural embeddings (transformer-free)
- Works well in retrieval scenarios where relevance is defined by keyword overlap
- Provides human-readable supporting documents for each prediction

---

Implementation Tools

- **RoBERTa Fine-Tuning**: transformers, torch, sklearn
- **BM25 Retrieval**: rank_bm25 from the rank-bm25 Python package
- **Data Processing**: pandas, nltk, re
- **Evaluation**: sklearn.metrics for accuracy, F1-score, precision, and recall

---

Sample Use Case

Input:

"Congress has passed a secret bill to give every citizen free cryptocurrency in 2025."

RoBERTa Output:

Prediction: Fake

BM25 Retrieved Context:

"Congress denies passing any bills related to public cryptocurrency distribution in recent sessions."

Explanation:
The claim references a significant legal decision with no real corroborating coverage. The retrieved document suggests no such law or policy exists, reinforcing the prediction.

Summary

The RoBERTa + BM25 pipeline offers a robust and modular approach to fake news detection. RoBERTa ensures high classification performance on linguistically rich news data, while BM25 facilitates lightweight, real-time context retrieval for explanation. This setup provides both model confidence and interpretability, aligning with the project's goal of transparency and user trust.