

MedCompare: Using Bio+ClinicalBERT for comparative analysis of medications

Group 06

Celine Florian — Documentation & Presentation

Lukas Brandner — Data Preprocessing & Evaluation

Moritz Legat — Embeddings & Evaluation

Nina Tschikof — Documentation & Presentation

<https://github.com/syrrex/MedCompare>

Introduction

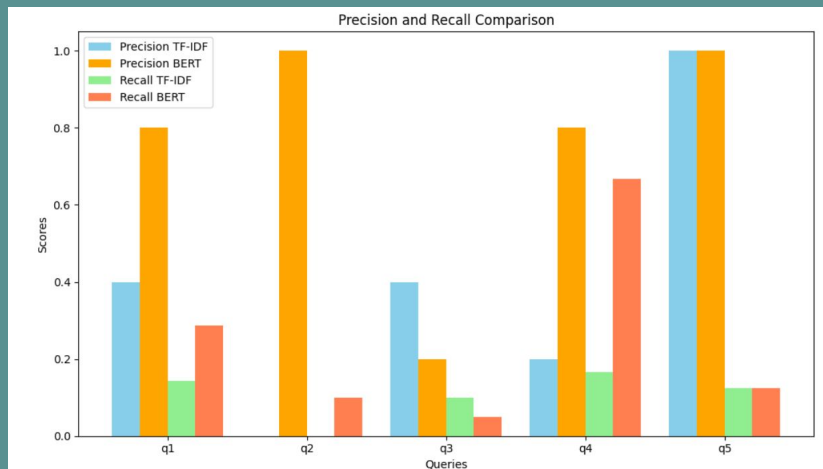
How can we use large language models (LLMs) and retrieval-based techniques to create reliable and user-friendly summaries in real time that compare medications with similar use cases, focusing on their uses and side effects?

- Clinical decision support systems (CDSS)
- Comparing medications based on
 - ingredients
 - uses
 - side effects
- Empower users in making informed treatment decisions
- Simplifying process of identifying suitable alternatives
- Enhancing overall decision-making process for treatment options

Data + Methods

- Data on medical information from HuggingFace (MattBastar/Medicine_Details)
- Bio+ClinicalBERT pre-trained model understanding medical entities and concepts
 - provides NER pipeline
- Embeddings created using Bio+ClinicalBERT and TF-IDF
 - Foundation for similarity ranking
- Cosine Similarity
 - using Sklearn

Results



Precision TF-IDF vs. BERT: BERT better at returning relevant results with fewer irrelevant ones

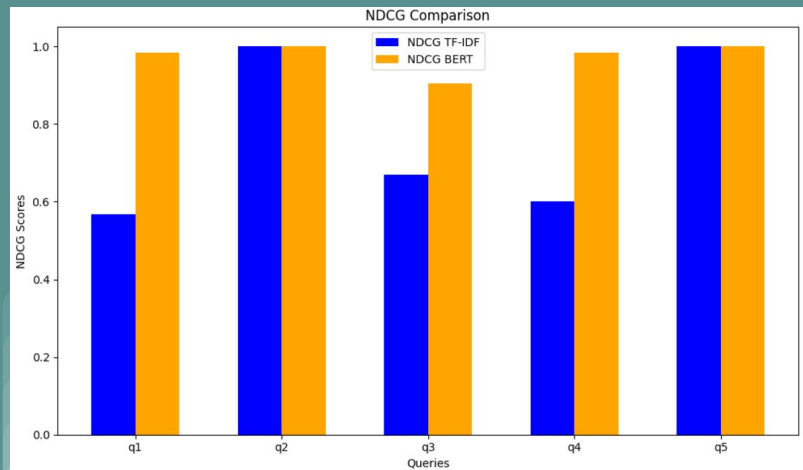
Recall TF-IDF vs. BERT: BERT retrieves more comprehensive set of relevant documents

→ BERT is the more reliable retrieval method

NDCG TF-IDF vs. BERT:

BERT achieves equal or better scores

TF-IDF is still competitive in ranking relevant documents for simpler queries



Conclusion

- No single source of truth
- Incorporate real-time updates from medical databases
- Task-specific fine-tuning of Bio+ClinicalBER
- Enhance evaluation and validation strategies
- Improve usability
 - Create intuitive and interactive interface
 - Integrate user-specific factors (age, medical history etc.)
 - Cross-Language Support