

250210 논문 피드백

250210 진행사항 보고

요약

Retrieval 및 RAGAS 평가 시스템의 주요 기능을 개선
새로운 분석과 실험 설계를 도입하여 시스템 성능과 논문의 방향성을 구체화

1. 데이터셋 및 초기 설정

- 방대한 데이터셋에서 특정 범위로 제한:
- 기존의 방대한 법률 데이터셋 대신, 형사 분야에 집중하여 질 높은 데이터셋을 확보.
(형사A,형사B분야의 데이터로 사용)
- 연구 범위를 좁힘으로써 도메인 특화 성능을 극대화할 가능성을 확인.
- **ground truth**는 법률 자문 전문가의 검증을 통해 보완 필요(현재는 자료 기반하여 임의로 구조생성)
- **Sparse Encoder** 활용 개선:
- 기존에 학습된 Sparse Encoder가 있을 경우, 이를 로드하여 불필요한 학습 과정 생략.
- 새로운 쿼리나 데이터셋일 경우 Sparse Encoder를 동적으로 학습 후 저장.
- 결과 로그에 [LOAD] 및 [FIT] 태그를 포함하여 실험 투명성 강화.

2. Context Relevance Check (검색 평가)

- **Hybrid Score** 계산:
- Sparse와 Dense 점수를 가중치(alpha)로 선형 결합하여 Hybrid Score 생성.
- 가중치는 0.5로 설정하여 일단 테스트중 가장 적합한 값 탐색. -> 추후 0.3~0.7 중 적합 값 탐색 필요
- **RAGAS** 평가 프로세스 개선:
- 정규화 전처리: 문서 및 쿼리의 대소문자, 공백 제거 등 전처리를 추가하여 평가 데이터의 신뢰성을 확보.
- 정규화 문제 해결: 일부 문서에서 평가 점수가 0으로 나타나는 문제를 해결.
- 상관관계 분석:
- RAGAS 평가 지표(Context Recall, Precision)와 Sparse/Dense/Hybrid 점수 간의 상관관계를 계산.
- 히트맵 및 2D 그래프를 통해 Sparse-Dense 관계와 Hybrid Score의 상호보완 효과 시각화.

결과

- Hybrid Score가 단일 검색 방식(Sparse, Dense) 대비 검색 품질(Recall, Precision)이 높음을 입증.
- Sparse와 Dense 점수의 분포 및 관계를 KDE(커널 밀도 추정)를 통해 시각적으로 분석.
- Sparse와 Dense 점수의 영향력을 균형 있게 결합한 Hybrid 방식의 필요성을 강조.

3. Top-k (n_docs) 파라미터 실험

- 목적: 검색 결과의 상위 문서 개수(n_docs)가 성능에 미치는 영향을 분석.
- 실험 설계:
 - 로 설정하여 Hybrid Score 및 RAGAS 지표를 평가.
 - K-Means 및 K-selection 기법을 참고하여, 최적의 값을 탐색.
- 현재 결과:
 - 일 때 가장 우수한 Hybrid Score와 사용자 평가 점수를 기록.
 - 가 지나치게 작거나 클 경우 성능 저하 경향 확인.

4. Boosting 방법 설계

- Sparse와 Dense 결합 방식 개선:
 - 현재는 선형 결합 방식($Hybrid_Score = \alpha \cdot Sparse + (1 - \alpha) \cdot Dense$) 사용.
 - 향후, **Linear-SVM**, **Gradient Boosting** 등 다양한 결합 방식을 도입하여 최적의 가중치 조합 탐색.
- 목표: RAGAS와 사용자 평가 점수의 안정성을 개선하여, 검색과 생성 단계를 전반적으로 Boosting.
- 추가 실험 계획:
 - Sparse/Dense/Hybrid 각각의 상관관계를 상세히 분석하고, 다른 데이터셋에서도 일관성 있는 결과를 검증.

종합 및 향후 계획

- 현재 성과:
 - Sparse와 Dense 점수의 조화를 통해 Hybrid 검색 평가 시스템 구현.
 - 데이터 전처리 및 평가 프로세스 개선으로 RAGAS 기반의 성능을 정량적으로 입증.
- 남은 과제:
 - 긴 시간동안 걸리는 실험 마무리, 최적의 가중치 탐색
 - Linear-SVM 및 Boosting 방식 도입을 통한 Hybrid Score 결합 방식 최적화.
 - K-Means 및 K-selection 기반의 최적 Top-k 파라미터 설정.
- 논문 방향:
 - Retrieval 중심 논문과 Generation 중심 논문을 별도로 작성.
 - 실제 법률 자문을 반영한 데이터셋 확장으로 Scopus 수준의 학술적 기여 달성.

다음 미팅까지

1. Sparse / Dense / Sparse+Dense 실험 완료 및 KDE 분석 마무리.
2. n_docs 와 Hybrid Score의 최적 조합 구체화.
3. Re-RAG 실험 코드를 완성하고, 재작성 전후의 성능 비교.
4. 논문 작성에 필요한 주요 데이터와 그래프 구성 완료.