

RAG 기술을 활용한 리뷰 및 실시간 데이터 기반 맛집 추천 챗봇

이화여자대학교 데이터사이언스대학원 김재은 | 김현지 | 한정화

I. 연구 배경 및 목표

[연구 배경]

- 직장인의 점심 메뉴 선택이나 특별한 날의 장소 탐색 등 맛집을 찾기 위해 인터넷 검색을 활용하는 경우가 점점 증가하고 있음
- 하지만 온라인 상단 노출의 대부분이 광고성 게시글로, 내 상황에 꼭 맞는 정보를 선별하기 어려우며 LLM 챗봇 모델은 할루시네이션으로 부정확한 추천을 제시

[연구 목표]

- 실제 방문자들의 리뷰 기반으로 신뢰도 높은 식당과 메뉴를 추천하는 모델 구축
- 사용자의 개별적 상황과 조건을 반영하여 메뉴 단위까지 정교하게 추천
- 대화형 챗봇 형태로 제공하여 사용자가 필요한 정보를 직관적이고 편리하게 획득 할 수 있도록 함

II. 데이터 수집

[음식점 기본정보 및 리뷰데이터]

- 출처 : 카카오톡, 구글맵
- 대상 : 서울숲역, 성수역, 건대입구역 인근 음식점 및 카페 2,465 점포
- 항목 : 상호명, 메뉴, 영업시간 등 51개 항목의 기본 정보와 리뷰 데이터 49,903건, 메뉴 데이터 29,051건

[실시간 데이터 API]

- 출처 : 서울 실시간 도시데이터
- 대상 : 서울숲, 성수카페거리, 건대 지역 실시간 날씨, 혼잡도, 공영 주차장 정보

III. 데이터 전처리

[리뷰단위 감성분석 (ABSA)]

- 5만 건의 리뷰 데이터 대상 맛·가격·분위기·서비스 항목에 대해 긍정/부정/중립 감성값 라벨링 (klue/bert+MLP 결과 F1- SCORE: 0.886, 정확도 91.9%)
- 감성 점수를 식당별 정량 지표로 변환하여 RAG 시스템 1차 필터링 및 메뉴 기반 추천 순위에 사용

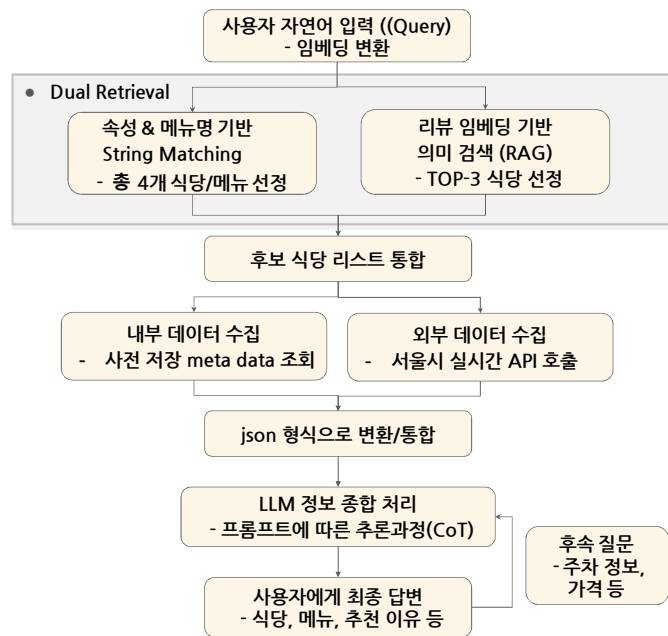
[리뷰 임베딩]

- Chunking : 각 리뷰 단위로 Split 함
- Embedding 모델 : BM-K/KoSimCSE-roberta-multitask 적용
- Vector DB : Chroma (metadata는 별도 DB에 저장하여 RAG 응답 속도 향상)

[메뉴 속성 태깅]

- 메뉴 이름을 토큰화하여 string matching을 통해 15개의 속성 태그 생성
- 예) 매운해물찜볶음 → 매움/해산물/뜨거움/면 특성 보유하도록 함

IV. 연구 방법



V. 연구 내용 및 결과

[속성 가중치 & 메뉴명 기반 문자열 매칭 (String Matching)]

- 메뉴속성 가중치 기반 탐색
 - 지역 필터링 → 사용자 의도(속성) 분석 → 속성 및 시간대 가중치 적용 80% 스코어링 + 베이지안 평균을 활용한 식당별 평점 20% 스코어링
- 메뉴명 기반 문자열 매칭 탐색
 - 지역 필터링 → 사용자 쿼리 토큰 추출 및 매칭 → 유사도 점수 산정(완전 일치 1점, 부분 일치 0.7점 차등 부여) 80% + 베이지안 평균 활용 식당별 평점 20%
 - 최종 랭킹 산정 기준 : 1차 - 각 탐색 모델의 최종 합산 점수 / 2차 - 리뷰 감성 지표 중 '맛' 점수 순

[리뷰 임베딩 기반 의미 검색 (RAG)]

- 1차 필터링 - Sparse Retrieval
 - meta data로 저장해둔 식당별 역/영업시간/업종/속성 항목을 단계적 pre filtering
 - 사용자 쿼리와 키워드 매칭을 통해 조건을 충족하는 식당만 1차 선별하여 전체 검색 시간 단축 및 정확도 향상
- 2차 필터링 - Dense Retrieval
 - 1차 필터링 선별된 리뷰 대상 cosine similarity 계산하여 쿼리와의 유사도를 측정
 - 3가지 모델 중 Time Weighted가 현실적이고 트렌드를 신뢰성있게 반영

기본 유사도 방식	Time Weighted (최신리뷰 가중치 부여)	MMR
0.87	0.89	0.81

[프롬프팅]

- System / Human / AI message : 입력 쿼리의 역할, 단계적 추론 과정 상세 설계
- ChatPromptTemplate, MPlaceholder: 챗봇이 대화 문맥 유지하는 구조 설계
- few-shot : API로 주차장 정보를 조회할 때 특정 KEY 값을 사용하는 방식에 적용
- Chat Message History : 과거 대화 이력을 저장하여 후속 질문 답변에 활용
- LCEL : 프롬프트 생성부터 모델 추론 후 응답까지 전체 파이프라인으로 연결

[연구 결과]

- 메뉴명 문자 매칭과 RAG 방식으로 추출한 식당 정보를 GPT-4o-mini LLM 적용
- 전통적 String Matching 방식 대비 RAG 기술의 성능이 높지만 동시 적용하였을 때 상호 보완적인 성격으로 최종 정확도 0.95의 성능을 보임

메뉴명 String Matching		RAG	최종 정확도
메뉴 속성 기반	문자열 매칭	Time Weighted	
0.33	0.43	0.89	0.95

VI. 시연 결과

- 챗봇 모델 : Gradio
- 최초 질문에 2개 이상의 식당을 추천하며, 후속 질문을 통해 추가 정보도 확인 가능



VII. 주요 성과 및 향후 연구

[주요 성과]

- 맛집 검색 패러다임을 기존 단순 키워드 매칭에서 RAG의 맥락기반 추천으로 전환
- '비오는날 국물'과 같은 추상적 니즈의 속성을 구체적인 검색 필터로 변환
- LLM의 할루시네이션을 실제 리뷰 데이터에 근거한 RAG 시스템으로 보정하여 사실 기반의 추천 구현

[향후 연구]

- 사용자 ID 단위 데이터 축적을 통해 초개인화된 추천엔진으로 고도화 가능
- 사용자 피드백을 기반으로 강화학습(RLHR)을 적용하여 추천 성능을 지속적으로 향상시키는 모델 구축