

텍스트마이닝 팀 프로젝트: 도로명 주소 자동완성 모델 개발

구성원: 242DBG01임천혜(Lin, Chien Hui) , 242DBG02 김예린, 242DBG04 정희수

1. 분석 목적 및 연구 문제

본 연구는 도로명 주소의 사용 불편함을 해소하고 사용자 편의성을 제공하고자 합니다. 저희가 분석한 도로명 주소 사용의 불편함은 크게 세 가지입니다. 첫째, 도로명 주소의 길이가 지번주소보다 길어 사용 시 시간이 오래 걸립니다. 둘째, 상호 소통 시 도로명 주소보다 지번주소가 용이하여 선호됩니다. 셋째, 지번주소 사용에 있어 실용적 문제를 크게 느끼지 않습니다. 따라서, 본 연구의 목적은 도로명 주소 인식의 정확도를 극대화하여, 사용자가 원하는 목적지를 신속하게 찾을 수 있는 자동완성 시스템을 개발하는 것입니다. 이 시스템은 다양한 긴급 상황에서도 응용될 수 있는 다목적 솔루션으로, 경찰이 신고 접수 시 신속히 위치를 파악하고, 현장에 적시에 도착하여 효과적으로 사건을 해결하는데 기여할 것입니다. 또한, 콜센터에서는 주소 변경 업무와 모바일 영업 고객 주소 입력 시간을 단축시켜 서비스의 전반적인 품질을 향상시킬 수 있습니다.

2. 연구 의의

본 연구는 도로명 주소 검색 및 입력의 불편함을 해결함으로써 사용자의 경험을 근본적으로 개선하고자 합니다. 자동완성 기능을 통해 주소 입력의 편리함을 제공하고, 이는 일상 생활뿐만 아니라 다양한 서비스 산업에 큰 영향을 미칠 것입니다. 택시나 우체국 같은 주소 정보가 핵심적인 서비스의 효율성을 높이고, 본 연구를 통해 개발된 자동완성 인터페이스는 데이터 입력이 요구되는 모든 분야에서 업무 처리 시간을 단축하여 소비자 만족도와 서비스 품질을 개선할 것입니다. 또한, 도로명 주소 자동완성 기능을 넘어, 음성 인식 기반의 주소 인식 기술로 확장 가능한 기반을 마련하여, 응급 의료 서비스, 배송 서비스, 택시 호출 등에서 신속하고 정확한 위치 정보 제공이 필수적인 서비스들의 위치 관련 문제 해결에 중요한 역할을 할 것으로 예상됩니다.

3. 데이터 선정 및 획득

AI-Hub에서 제공하는 "주소 음성 데이터"를 사용할 것입니다.

4. 탐색적 분석 및 전처리

불필요한 메타데이터 제거 및 필요한 feature의 추출

dialogID: 1의 ex_add (예시 주소)와 이에 대응하는 original address (원본 주소)를 매핑
오타 및 주소 형식의 불일치를 수정하며 데이터의 품질을 보장하기 위한 정제 과정 수행

5. 학습방법 및 변수 선정

본 연구는 지도학습 방법으로 진행

X 변수: 예시 주소(dialogID: 1의 ex_add)

Y 변수: 원본 주소(dialogID: 1의 original address)

6. 태스크 도출

본 연구의 주요 목표는 사용자가 입력을 시작할 때 정확한 전체 주소를 자동완성하는 것입니다. 모델의 성능을 측정하기 위하여 혼동행렬을 사용할 것이며, 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 점수(F1 Score)의 지표를 사용할 것입니다.

7. 분석 모델

본 연구에서 선정한 모델은 주소 예측의 정확도, 학습 시간 및 추론 속도, 그리고 모델의 복잡성과 유지 관리 용이성을 기준으로 선택하였습니다. 따라서 저희가 선택한 3가지 모델은 다음과 같습니다.

1. BERT (Ko-BERT): 단어의 문맥을 이해하는 능력이 뛰어나고, 주소 예측에 효과적일 수 있음.
2. LSTM: 시퀀스 데이터 처리에 강점을 가지며, 주소의 연속성 인식에 유용할 것.
3. T5: 텍스트를 이해하고 생성하는 능력이 뛰어나며, 주소 자동완성에 필수적인 특성을 지님.

모델 학습 절차로는 교차 검증을 통한 모델의 일반화 능력 평가하고, 과적합 방지를 위한 조기 종료(Early Stopping) 기법 적용할 것이며, 마지막으로 최적화된 하이퍼파라미터를 찾기 위한 그리드 서치 또는 랜덤 서치 수행할 것입니다.

8. 예상 결과 및 기대 효과

본 연구의 예상 결과는 사용자의 주소 입력 과정에서 정확하고 빠른 자동완성 모델을 만드는 것이며, 이에 따른 기대 효과에는 사용자 경험의 향상, 긴급 서비스의 응답 시간 단축, 다양한 서비스 분야에서의 활용성 증대 등과 같은 효과를 예상합니다.

9. 결론 및 추후 연구 방향

본 프로젝트의 성공적인 완성은 도로명 주소 입력의 효율성을 크게 개선할 것이며, 이는 사회적으로 매우 중요한 서비스의 품질 향상에 기여할 것입니다. 추후 연구에서는 음성 인식과 통합하여 더욱 편리한 사용자 인터페이스를 개발할 수 있는 기회를 모색할 것입니다.