2024 한양대학교

SW/ICT/AI 종합학술대회

[분야 : 산학연구개발/연구인턴]

제목: 대화 관리자를 활용한 사용자 친화적 LLM 기반 의료상담 시스템

정창엽(2020008686)
김우신(2024154867)

2024. 11. 15

한양대학교 ERICA 소프트웨어융합대학 / 한양대학교 ERICA SW중심대학사업단

서 론

• 연구의 필요성

▪ LLM을 활용한 의료 챗봇은 의료 서비스의 접근성을 향상시키며 환자와 의료 전문인 간 소통을 원활히 함

▪ 의료 챗봇이 진단에 필요한 단서 가 충분하지 않다고 판단하면 사용자에게 계속해서 질문을 하여 추가 정보를 요구

▪ 대부분의 LLM의 경우 **사용자가 명시적으로 진단을 내리도록 명령할 때 까지 계속해서 질문하는 문제**를 가지고 있음

▪ 명령을 통해 강제로 질문 생성을 중단하는 경우, 충분한 정보를 모으지 않았음에도 진단을 생성해야 하는 새로운 문제를 야기

• 연구 소개

▪ 본 연구에서는 **대화 관리자(Dialogue Manager)**를 활용하여 의료 챗봇이 정확한 진단을 적절한 시점에 생성할 수 있도록 하는 자동 의료 상담 시스템을 제안

▪ 대화 관리자는 의료 데이터를 사전 학습 하였으며, 이 방법은 횟수 기반 접근법이나 LLM의 자율 판단 에 의존한 대화 관리 방식보다 더 효율적으로 대화를 관리할 수 있고, 실제 의사와의 상담과 같은 챗봇 서비스를 제공하는 데 더 적합한 구조임

연구 방법

• 대화관리자

✓ 본 연구에서는 중국어로 학습된 BERT 모델을 대화 관리자로 선택하였으며 이를 의료 대화 데이터 셋으로 미세 조정

① 실제 의사와 환자 간의 대화와 유사한 대화 횟수와 정확한 진단을 목표로 학습

② 대화 관리자는 사용자의 입력 정보를 바탕으로 1 또는 0을 예측

③ 만약 0이 도출되면, 챗봇이 생성한 답변을 그대로 사용자에게 전달하고, 1이 도출되면 LLM의 생성을 중지 하고, 지금까지 얻은 정보를 바탕으로 진단을 내리게 함

✓ 학습 데이터 구성

▪ 학습 데이터는 중국어 의사-환자 대화 데이터 셋을 기반으로 하며, 총 14,864개의 대화가 포함

▪ 정답 라벨(1)은 의사 가 진단을 내리기 직전까지의 환자 발화로 구성하였고, 오답 라벨(0)은 의사가 진단을 내리기 전 두 발화의 내용을 누적하여 구성

▪ 전체 학습 데이터는 총 22,823개이며, 그중 정답 데이터는 7,179개, 오답 데이터는 15,649개로 구성

연구 방법 (cont'd)

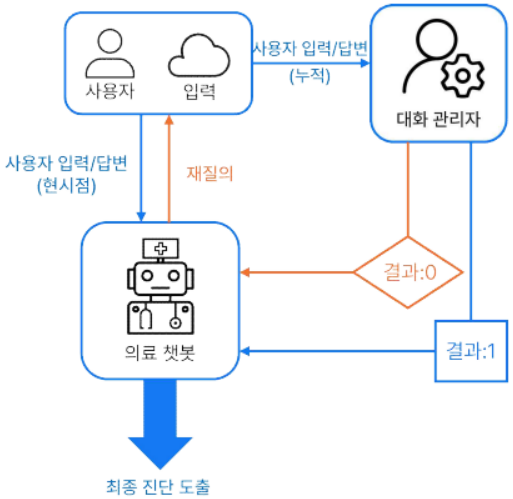
- **대화 관리자를 활용한 자동의료상담 시스템**

✓ 본 연구에서는 LLM이 생성한 질문 및 진단을 대화 관리자가 평가하여 자동 의료 상담 시스템을 설계하였으며, 다음과 같이 크게 3단계로 진행

① 사용자가 증상을 입력하면, 해 당 정보는 대화 관리자와 의료 챗봇에 전달

② 대화 관리자는 진단에 충분한 정보가 있는지(라벨 1), 또는 불충분한지(라벨 0)를 판단, 대화가 길어질수록 사용자 의 입력 정보는 누적 됨

③ 대화 관리자의 판단에 따라, 의료 챗봇은 추가 정보를 요청하거나, 현재까지의 정보를 바탕으로 진단을 생성



◀ 대화 관리자를 활용한 자동의료상담 시스템

실험 및 결과

• 실험 방법

▪ 실험에서는 512개의 테스트 데이터 셋으로 평가하였으며, 사용자의 첫 번째 입력에 따른 진단 정확도 및 모델의 응답속도와 효율성을 평가

▪ 자동의료상담 시스템의 정확도 평가를 위해 대화 관리자를 사용하지 않은 모델과(llama3-w/o-dm, gemma2-w/o-dm) 대화 관리자를 사용한 모델(llama3-dm, gemma2-dm)을 비교

• 평가지표

▪ 정확도

①BERT 스코어를 기반으로 한 코사인 유사도(cos_sim)를 사용해 진단 정확도 평가

②정밀도(precision), 재현율(recall), F1 스코어(F1) 지표를 사용해 문장 생성 평가

▪ 효율성

①모델이 진단을 생성하는 데 걸린 시간

② 모델이 진단을 생성하는 데 필요한 평균 대화 횟수

실험 및 결과

• 정확도 평가 결과

▪ 대화 관리자를 사용한 모델(llama3-dm, gemma2-dm)이 대화 관리자를 사용하지 않은 모델에 비해 **모든 지표에서 유사하거나 더 나은 성능을 보였음**

▪ 이는 대화 관리자가 **모델의 정확도를 향상 시키는 데 기여했음을** 시사

모델	cos_sim	precision	recall	F1
llama3-dm	**0.86**	**0.63**	**0.52**	**0.57**
llama3-w/o-dm	**0.86**	**0.63**	0.51	0.56
gemma2-dm	**0.87**	**0.64**	**0.50**	**0.56**
gemma2-w/o-dm	0.86	0.61	0.50	0.55

▲ 정확도 평가

실험 및 결과

• 효율성 평가 결과

▪ 대화 관리자를 사용한 모델이 대화 관리자를 사용하지 않은 모델에 비해

대화 생성 시간이 대략 10시간에서 12시간 정도 더 짧았음

▪ 대화 관리자를 사용한 모델이 대화 관리자를 사용하지 않은 모델에 비해

평균 대화 횟수가 현저히 적었음

▪ 이는 대화 관리자(dm)를 사용하는 모델들이 **더 짧은 시간과 더 적은 대화 횟수로도 정확한 진단을 도출하는 데 효과적이라는** 점을 시사

모델	데이터 수	대화 생성 시간	평균 대화 횟수
llama3-dm	513	**2시간 20분 35초**	**1.38**
llama3-w/o-dm	513	14시간 47분 47초	7.45
gemma2-dm	513	**4시간 6분 58초**	**2.15**
gemma2-w/o-dm	513	14시간 52분 43초	5.55

▲ 생성 시간과 대화 횟수를 바탕으로 한 효율성 평가

결 론

- **최종 결론**

  ▪본 연구에서는 사용자의 입력 정보를 바탕으로 추가적인 질문이 필요한지, 또는 진단을 내릴 준비가 되었는지를 효과적으로 판단하는 대화관리자에 대해 다룸

  ▪대화 관리자 적용 모델은 기존 모델에 비해 **더 높은 진단 정확도와 대화 효율성을 보여줌**

- **연구 의의 및 향후 연구**

  ▪대화 관리자를 활용한 자동 의료 상담 시스템은 사용자 경험을 개선하고, **의료 챗봇의 신뢰성과 진단 정확도를 높이는 데** 중요한 역할을 할 수 있음

  ▪ 이는 의료 서비스의 접근성을 높이고, 사용자에게 **더 나은 의료 상담 경험을 제공**할 수 있는 가능성을 제시함

  ▪ 향후 연구에서는 더 다양한 의료 데이터와 정교한 대화 관리 알고리즘을 적용하여, 더욱 향상된 자동 의료 상담 시스템을 개발할 수 있을 것으로 기대함