

사용자 맞춤형 강의 추천을 위한 대화 시스템 연구

최예린^{1*} · 연희연¹ · 김동근²

¹서강대학교 · ²서울대학교

Dialogue System for User Customized Lecture Recommendation

Yerin Choi^{1*} · Yeen-heui Yeen¹ · Dong-Geun Kim²

¹Sogang University · ²Seoul National University

E-mail : lakahaga@sogang.ac.kr / yeen214@sogang.ac.kr / kdg5188@snu.ac.kr

요 약

인공지능 대화 기술이 발달하며 다양한 분야에서 사용자의 의도에 따른 적절한 대응이 가능한 챗봇이 적용되고 있다. 특히 사용자 문의 및 상담이 많은 교육 도메인에서의 챗봇 도입 중요성은 증가하고 있다. 하지만 현재까지의 챗봇은 단순 대응이나 사용자가 자주 사용하는 기능 위주로 대응이 이루어지고 있다. 또한 교육콘텐츠가 다양화되고 증가함에 따라 인공지능을 통해 맞춤형 추천이 가능한 추천 시스템 연구가 활발해지고 있다. 이러한 대화 시스템과 추천 시스템에 대한 각 연구 분야는 같은 도메인에서 핵심적인 요소임에도 불구하고 별도로 진행되고 있는 한계점을 가지고 있다. 따라서 본 연구에서는 사용자 맞춤형 강의 콘텐츠를 추천할 수 있는 추천 시스템과 부가 기능에 대한 대응이 가능한 대화 시스템이 결합된 사용자 맞춤형 강의 추천 대화 시스템을 제안한다. 이를 통해 다양화 및 개인화되어가는 교육도메인에서의 챗봇 적용으로 업무 효율성 및 사용자 만족도가 향상되기를 기대한다.

ABSTRACT

Task-oriented chatbots prevail in various filed with the artificial intelligent dialogue system. The need for chatbots in customer services is growing, especially in education businesses given that there are many user inquiries and consultation requests. However, current dialogue systems only function as simple reactions or predetermined and frequently used actions. Meanwhile, the research about customized recommendation systems through artificial intelligence is very active with a wide variety of educational content. Although a dialogue system and a recommendation system is a core element in this domain, it has a limitation in that it is being conducted separately. Therefore, we present a study on a recommendation system that can recommend user-customized lectures combined with a dialogue system. With this combination, our system can respond to additional functions beyond these limitations. Through our research, we expect that work efficiency and user satisfaction will be improved by applying chatbots in education domains that are becoming more diversified and personalized.

키워드

Artificial intelligence, Dialogue System, Recommended system, Lecture recommendation

1. 서 론

많은 산업 분야에서 인공지능 기술이 적용됨에 따라 업무의 효율성과 편리함이 증가하고 있다. 특히 콜센터 등 상담 분야에서의 코로나 19로 인한 비대면 전환이 더욱 가속화 되어 가고 있으며, 기존 상담원 업무를 자동화 시스템으로 변환하는 추세다.

특히 교육 분야에서는 사용자가 다양하고, 그렇기 때문에 다양한 관심사에 맞게 서비스를 제공해야 한다. 이에 따라 교육 도메인에서의 사용자 맞춤형 강의 및 서비스에 대한 추천 시스템과 사용자 의도에 맞는 대응 시스템 대한 필요성이 대두되었다. 따라서 본 연구에서는 교육 분야에서 활용 가능한 사용자 맞춤형 강의 추천이 가능한 추천 알고리즘과 사용자 의도에 맞춘 적절한 대응이 가능한 대화시스템에 대한 방법론을 제안한다.

* corresponding author

II. 관련 연구

2.1 챗봇의 온라인 서비스 적용

여러 온라인 쇼핑몰이 사용자의 록인(lock-in) 효과를 위해 기존의 메신저 플랫폼을 활용한 챗봇을 사용하고 있다. 대량의 상품 데이터를 챗봇 빌더 서버(Server)가 주기적으로 수집하여 챗봇의 데이터베이스(Database)에 저장하고, 메신저에서 챗봇을 통해 소비자가 상품을 검색할 수 있도록 관리하는 시스템을 설계하는 연구가 진행된 바 있다. 하지만 이는 기존의 챗봇 빌더를 사용한 것이며, 일반화가 어렵다는 한계점이 있다.[1]

2.2 추천 알고리즘

추천 시스템 분야에서는 지식그래프를 사용하여 사용자에게 맞춤화된 추천 알고리즘에 관한 연구가 활발히 진행되고 있다. 지식그래프는 노드가 엔티티로 엣지가 엔티티 간의 관계를 나타내는 그래프이다. 이 특징을 추천 시스템에 이용하면 아이템과 아이템의 특징들을 그래프의 노드로 지식그래프에 매핑할 수 있고 노드의 상호관계 또한 표현할 수 있다. 이는 사용자와 아이템의 부가적인 정보뿐만 아니라 사용자-사용자, 사용자-아이템의 관계를 추천 모델에 포함시킴으로써 성능 향상에 도움이 된다고 이전 연구에서 밝혀진 바 있다. 따라서 본 연구에서는 지식그래프를 사용해서 사용자의 선호도를 기반으로 강의 추천을 할 수 있도록 연구를 진행하였다.[2]

III. 연구 방법

3.1 데이터셋

본 연구에서 사용된 데이터셋은 AI-hub에서 용도별 목적대화 코퍼스에서 학원과 온라인 교육 서비스 상담센터에서 수집한 교육 분야 목적 대화를 활용하였다. 이 데이터는 프로그램 문의, 등록 문의, 비용 환불 문의, 일정문의 데이터 총 4가지의 발화 의도로 구성되어 있다. 본 연구에서는 실제 교육 도메인에 적용을 위해 총 7개의 발화 의도 기준을 세워 재라벨링을 진행하였다.

3.2 강의 추천 대화 시스템 파이프라인 구축

본 논문에서 제안하는 강의 추천 대화 시스템의 구조는 그림 1과 같다. 이 구조를 통해 강의를 추천하고, 기타 사용자 의도에 적절한 대응을 할 수 있다.

언어 이해 단계에서는 사용자 발화 의도 구분과 이해를 위한 적절한 개체명 인식 모델로 개체명을 추출하는 부분으로 구성되어 있다. 그리고 다중 턴과 대화의 흐름을 기억하기 위한 대화 상태 추적(DST), 이에 따른 시스템의 적절한 대응을 결정하는 대화 정책(DP)로 이어진다. 마지막으로 이러한 대응을 자연어로 생성하는 대화 생성(NLG) 파트로 구성되어 있다. 사용자 의도 분류와 개체명 인식을

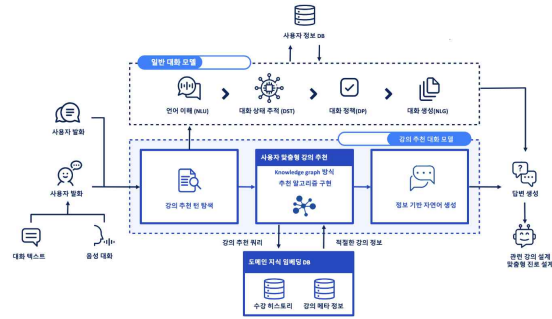


그림 1. 사용자 맞춤형 강의 추천 대화 시스템

위해서는 ELECTRA 기반의 한국어 사전학습 모델 KoELECTRA를 사용했다.[3] 추가적으로 일반 대화를 위한 KoGPT2모델 학습으로 강의 추천 대화 외에도 일반 대화에도 대응이 가능한 대화 시스템 연구를 진행하였다.[4]

3.2 지식 그래프 기반 추천 시스템

추천 시스템 구축을 위해서는 사용자, 강의, 강의의 카테고리, 강의 키워드를 노드로 지식그래프를 생성하였고 추천 모델의 학습을 위해 사용자가 실제 강의를 수강한 positive sample의 0.1의 비율로 사용자가 수강하지 않은 데이터에 대해서 negative sample을 생성했다.

모델의 훈련과정은 다음과 같다. 하나의 수강 데이터에 대해서 사용자 노드에서 수강한 강의의 노드로 가는 5개의 경로를 생성했다. 경로 생성은 파이썬의 networkx 라이브러리를 사용해 graph 내에서 길이가 3인 경로를 찾고 찾은 경로가 5개 이상이면 랜덤 추출하는 방식으로 진행하였다.[5] 얻은 경로들은 Bi-GRU와 Fully Connected Layer를 통과하고 최종적으로 소프트맥스 함수를 통해 얻은 확률값에 대해 positive sample 이면 1, negative sample 이면 0인 이진분류로 학습된다. 최종적으로 사용자 임베딩 값, 강의 임베딩 값의 내적으로 사용자의 모든 아이템들에 대한 추천 스코어를 계산할 수 있다.

IV. 연구 결과

4.1 사용자 의도 모델 및 개체명 인식 모델 결과

사용자 의도 모델은 3.1절에서 기술한 데이터셋에서 5,292개를 사용하였다. 개체명 인식 모델은 학습의 정확도 향상을 위해 기존에 공개된 네이버 개체명 인식 데이터를 추가적으로 활용하여 109,829개를 활용하였다. 각 모델의 Train, Dev, Test 데이터 구성과 성능은 표 1과 같다.

4.2 추천 알고리즘 성능

추천 모델의 결과는 표 2와 같다. 추천 리스트의 TOP 1, 3, 10 아이템에 대해서 precision과 TOP10

	사용자 의도 구분 모델	개체명 인식 모델
데이터셋 구분	Train : 4,233 Dev : 553 Test : 530	Train : 87,861 Test : 21,966
성능	80 (Accuracy)	89 (F1-score)

표 1. 사용자 의도 구분 및 개체명 인식 모델 성능

MRR로 측정하였다. Precision은 추천한 아이템 K 개 중에 실제 사용자가 관심있는 아이템의 비율을 나타낸다. MRR은 사용자가 선호하는 아이템이 추천 결과 리스트에서의 위치를 뜻하는 평가 지표다.

Datasets	Metric	Performance
강의 수강 데이터셋	precision@1	0.13
	precision@3	0.067
	precision@10	0.03
	MRR@10	0.179

표 2. 추천 모델 성능

4.3 대화 시스템 성능 종합분석

실제 업무 분야에의 적용 가능성을 점검하기 위해 사용자 의도별 대화 시스템의 대응 시간을 분석 하였다. 결과는 표 3과 같다. 최초로 강의 추천 결과를 로딩하는 경우(강의 추천의 Test1)를 제외하고 모든 프로세스 시간이 1초 이내임을 확인할 수 있다.

	Test1	Test2	Test3	Test4	Test5	평균 시간
강의 추천	1.01	0.26	0.24	0.24	0.23	0.396
거절	0.24	0.24	0.25	0.24	0.48	0.29
강의 정보조회	0.31	0.32	0.24	0.30	0.31	0.296
마이 페이지	0.22	0.24	0.23	0.23	0.23	0.23
쌤 기능 문의	0.23	0.23	0.36	0.23	0.23	0.256
일반대화	0.51	0.33	0.35	0.41	0.40	0.4
ODD	0.24	0.47	0.26	0.27	0.24	0.296

표 3. 발화 의도별 대화 시스템 대응 시간 분석

V. 결 론

본 연구를 통해 기존의 복잡한 교육 분야 대응 과정에 챗봇 도입하여 보다 효율적인 시스템을 구축하였다. 또한 사용자의 수강 이력을 기반으로 강의 추천뿐만 아니라, 기타 업무 및 일반 대화가 가능한 대화 시스템을 구현하였다. 기존의 공개 데이터셋을 실제 도메인에서 사용하는 사용자 발화 의도에 맞게 재라벨링하고, 개체명 인식 데이터 증강을 통해 실제 업무에 적용될 수 있도록 하였다. 추후 연구에서는 추천에 대한 성능 향상을 위한 난이도 기반 추천 및 사용자 관심사 설정을 통해 사용자 맞춤형 추천이 가능한 사용자 맞춤형 강의 추천 대화 시스템 연구를 진행하고자 한다.

Acknowledgement

이 논문은 2022년도 SK텔레콤의 지원으로 진행된 SKT AI Fellowship 과정에서 수행된 공동연구 결과임 (과제번호 11. AI큐레이션 강의 추천을 위한 챗봇 엔진 개발)

References

- [1] Seokjong Kang, Byeong Eon Hyun, Gil Seop Kim. (2020). A Study on the Integration Process between Chatbot Builder and Online Shopping Mall for Big Data Search - Focused on Kakao AI Platform -. Journal of East and Central Asian Studies, 31(3), 31-46.
- [2] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent knowledge graph embedding for effective recommendation. In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18). Association for Computing Machinery, New York, NY, USA, 297-305.
- [3] Clark, Kevin, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. "Electra: Pre-training text encoders as discriminators rather than generators," *arXiv preprint arXiv:2003.10555* (2020).
- [4] Radford, Alec, et al. "Language models are unsupervised multitask learners," *OpenAI blog* 1.8 (2019): 9
- [5] Mnetwork [Internet]. Available : <https://networkx.org/>