

Chat-Trip: 친구추천을 위한 대화 기반 사용자 임베딩

최윤정^{1‡}, 강민지^{2‡}, 이채연³, 강다빈⁴, 구고은⁵, 이규영⁶

¹ 서울대학교 연합전공 인공지능

² 성균관대학교 시스템경영공학과

³ 인천대학교 컴퓨터공학부

⁴ 서울여자대학교 정보보호학과

⁵ 숭실대학교 전자정보공학부

⁶ 한국과학기술원 전산학부 정보보호대학원

‡공동주저자

racheal0@snu.ac.kr, minji0801@g.skku.edu, sarah031024@inu.ac.kr, kchabin@swu.ac.kr,
gugoeun1234@soongsil.ac.kr, leeahn1223@kaist.ac.kr

Chat-Trip: Chat-Aware User Embedding for Friend Recommendation

Yun-Jeong Choi^{1‡}, Min-Ji Kang^{2‡}, Chae-Yeon Lee³, Da-Bin Kang⁴, Go-eun Gu⁵, Gyu-Young Lee⁶

¹ Interdisciplinary Major in Artificial Intelligence, Seoul National University

² Dept. of Systems Management Engineering, Sungkyunkwan University

³ Dept. of Computer Science, Incheon National University

⁴ Dept. of Information Security, Seoul Women's University

⁵ Dept. of Electric Engineering, Soongsil University

⁶ Graduate School of Information Security, KAIST

‡ These authors equally contributed to this work

요 약

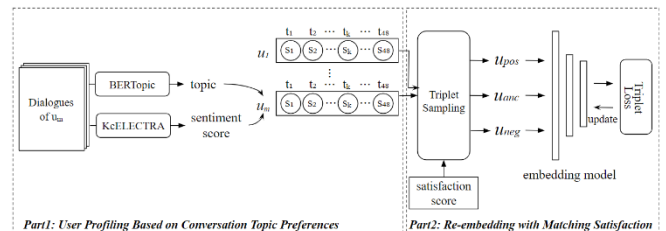
빅데이터 시대에 추천시스템은 끊임없이 진화하고 있으며 그 중요성도 나날이 높아지고 있다. 그럼에도, 친구추천시스템은 여전히 사용자의 맥락 정보에만 의존하는 제한적인 모습을 보이고 있다. 이에 본 연구는 사용자의 대화에 드러난 관심사와 대화 만족도 등을 직접 사용자 임베딩에 활용한 Chat-Trip 모델을 제안하고, 실험을 통해 그 효용성을 입증하였다.

1. 서론

빅데이터 시대에 추천 시스템은 다양한 분야에서 그 영역을 확장하고 있으며, 대화가 잘 통하는 친구를 추천해주는 서비스에 대한 수요가 소셜 미디어의 발전과 함께 증가하고 있다. 그러나 대부분의 추천 시스템 연구는 상업적인 서비스에 치중되어 있으며, 친구 추천에 대한 연구는 상대적으로 부족해 주로 전통적인 알고리즘에 한정되어 있다.[1] 현재의 친구추천 연구는 대부분 장소나 친구 목록과 같은 맥락 기반(Context-aware) 추천을 중심으로 진행되고 있으므로, 사용자의 대화 내용에 대한 고려 없이 사용자의 맥락 정보에만 의존하는 경향이 있다.[2][3]

본 연구는 이러한 문제를 극복하기 위해 사용자의 대화 내용에서 도출한 관심사를 직접 사용자 임베딩에 적용하는 방안을 제안한다. 주제 및 감성 분석을 통해 대화 내용에서 사용자 데이터를 형성하고, 만족

도를 기반으로 Triplet Loss를 적용하여, 대화 내용과 매칭 만족도를 모두 반영하여 임베딩한 Chat-Trip 모델을 아래 그림 1과 같이 제시한다.



(그림 1) Chat-Trip 모델 구조.

2. Chat-Trip 모델 제안

2.1. 대화에 드러난 관심사 기반 사용자 프로파일링

대화에 드러난 사용자의 관심사를 파악하기 위해, 대화의 주제를 분석한 후 해당 주제에 대한 사용자의 호/불호를 감성 점수를 통해 나타낸다. 즉, 사용자마

다 주제에 선호도를 대응시켜 성향 벡터를 만든다.

2.1.1. 주제 분석

사용자 간 대화를 BERTopic 의 입력으로 주어 해당 대화의 주제를 추출한다. BERTopic 은 BERT 기반의 사전학습 모델로, 임베딩-차원축소-군집화-군집표현의 순서로 주제를 추출한다. 본 모델에서는 3.1. 데이터셋의 대화-주제 데이터를 차원축소 단계의 semi-supervised learning 에 활용하여, 한국인의 대화에서 드러나는 가장 특징적인 대화 주제 카테고리 48 개를 추출하여 사용하였다.

2.1.2. 감성 분석

사용자의 발화를 KcELECTRA 의 입력으로 주어 해당 대화의 주제를 추출한다. KcELECTRA 모델은 마스크 복원 방식의 Masked Language Model 인 ELECTRA 의 한국어 사전학습 모델이다. 주어진 발화의 감성 점수를 산출할 수 있도록 fine-tuning 하여 사용하였다.

2.2. 대화 만족도를 반영한 사용자 임베딩 재조정

2.2.1. Triplet 샘플링

사용자 매칭 후 대화 만족도를 반영하여 임베딩을 재조정하기 위하여, 대화 만족도 기반 triplet 을 샘플링하였다. 각 사용자를 anchor 로 하고, 사용자가 만족한 상대는 positive, 만족하지 않은 상대는 negative 로 할당하여, 모든 조합에 대해 triplet 을 생성하였다. 이때 각 사용자의 평가 감도를 보정하기 위하여, 만족도 값을 그대로 사용하지 않고 anchor 만족도 평균을 빼서 양수면 만족, 음수면 불만족으로 할당하였다.

2.2.2. 사용자 임베딩 재조정

앞서 생성한 triplet 을 Triplet Loss 를 이용하여 임베딩하였다. Triplet Loss 는 anchor 와 positive 간의 거리는 최소화하고, anchor 와 negative 간의 거리는 최대화 하는 loss 이다. 대화 후 만족도가 높은 사용자들은 가까운 공간에, 만족도가 낮은 사용자들은 먼 공간에 배치하기 위하여, 완전연결 3 계층을 구축한 후 48 차원 공간에 새롭게 임베딩하였다.

3. 실험 결과

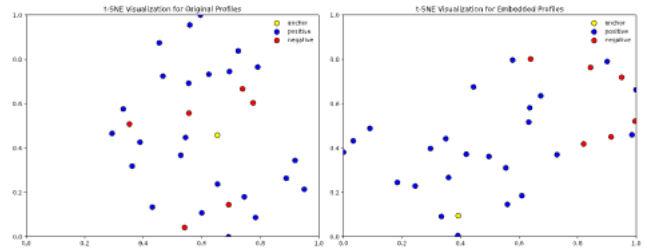
3.1. 데이터셋

충분한 양의 사용자 대화, 매칭, 만족도 정보를 직접 얻을 수 없기 때문에, AI hub 에서 제공하는 <페르소나 대화> 데이터셋을 재구성하여 사용하였다. 이는 20 개 주제에 대한 1366 개의 페르소나 간 대화 데이터로, 대화를 나누는 모두가 매칭되었다고 가정하고

대화의 앞 절반은 매칭 전 대화, 뒤 절반은 매칭 후 대화로 사용하였다. 즉, 앞 절반은 사용자의 성향 분석을 위한 데이터, 뒤 절반은 매칭 후 만족도를 파악하기 위한 데이터로 사용하였다. 만족 여부는 대화에 드러난 사용자의 감성 점수로 판단하였으며, 사용자 마다의 평균을 뺀 정규화 점수를 사용하였다.

3.2. 실험 결과

타겟 사용자와 매칭된 타 사용자들을 t-SNE 를 활용한 2 차원 그래프로 시각화한 결과, Triplet Loss 임베딩 이전의 분포는 그림 2 좌측과 같이 positive/negative 구분이 없는 랜덤한 모양을 띠는 반면, 임베딩 후에는 그림 2 우측과 같이 타겟 사용자를 기준으로 positive/negative 사용자가 명확히 구분되어 분포하는 것을 확인할 수 있다. 따라서, 대화 기반 성향벡터를 만족도에 기반하여 임베딩하면, 관심사 및 만족도를 반영하여 사용자를 표현할 수 있음을 확인하였다.



(그림 2) id.157 의 임베딩 전/후 t-SNE 분포 비교.

4. 결론 및 향후 연구

본 연구에서는 사용자 간 대화 데이터에 Triplet Loss 를 적용하여 최적화된 임베딩 모델을 대화 기반 친구 추천시스템에 적용하는 방안을 제시하였고, 상기 실험을 통해 효용성을 입증하였다. 따라서 본 연구 결과를 도입할 경우, 친구추천시스템의 성능이 크게 향상될 것으로 예상된다. 향후 보다 광범위한 실 사례 데이터셋을 확보하고 추가적인 성능평가를 진행하여 연구성과를 지속적으로 발전시켜 나갈 예정이다.

※ 본 논문은 과학기술정보통신부 정보통신창의인재 양성사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

참고문헌

- [1] M. SONI, et al., Friend recommendation system using machine learning method. *Journal of Scientific Research & Engineering Trends*, 6(5), 2020.
- [2] M. Xin, et al., Using multi-features to partition users for friends recommendation in location based social network, *Information Processing & Management*, 57(1), 2020.
- [3] R. Parveen, et al., Friend's recommendation on social media using different algorithms of machine learning, *Global Transitions Proceedings*, 2(2), 2021, 273-281.