

의약품 추천 방법에서의 의약품들 간의 다양한 관계를 효과적으로 융합하기 위한 방법 성능 평가

허지호¹, 김태리², 김상욱^{3*}

¹ 한양대학교 컴퓨터소프트웨어학과 석사과정

² 한양대학교 컴퓨터소프트웨어학과 박사과정

³ 한양대학교 컴퓨터소프트웨어학과 교수

linda0123@hanyang.ac.kr, taerik@hanyang.ac.kr, wook@hanyang.ac.kr

Fusion of Relations among Medications in Medication Recommendation

Jiho Heo, Taeri Kim, Sang-Wook Kim*

Dept. of Computer Science, Hanyang University

요 약

의약품 추천 방법은 환자의 현재 건강 상태가 주어졌을 때, 해당 환자에게 효과적이면서도 안전한 의약품들을 추천해주기 위한 방법이다. 최근 의약품 추천 방법들은 환자에게 효과적이면서도 안전한 의약품들을 추천해주기 위해, 환자의 방문 정보(즉, 환자에게 진단된 질병, 수술 정보) 뿐만 아니라 의약품들 간의 (1) 동시 처방 여부 관계와 (2) 부작용 유발 관계를 추가적으로 활용하고 있다. 구체적으로, 이들은 의약품들 간의 두 관계를 각각 임베딩으로 나타낸 뒤 하나의 임베딩으로 융합하여 활용한다. 그러나 기존 의약품 추천 방법들은 이 두 임베딩을 융합하는데 있어서, 두 임베딩 간의 관계를 깊게 고려하지 않고 단순히 뺄셈 연산을 적용하고 있다. 따라서 본 논문에서는 실세계 데이터 집합을 활용하여, 의약품들 간의 두 관계에 대한 임베딩들을 다양한 방법을 통해 융합한 뒤 각 방법이 추천 정확도에 어떠한 영향을 미치는지 비교 및 분석하여 더욱 정확한 의약품 추천을 제공하고자 한다.

1. 서론

최근 전자 건강 기록(Electronic Health Records, EHR)이 많이 축적됨에 따라 질병 진단, 생리학적 상태 모니터링 등 다양한 의료 태스크(task)에서 임상적 의사 결정을 도와주기 위한 여러 연구들이 진행되어 왔다 [1-3]. 이러한 의료 태스크들 중 하나인 의약품 추천 태스크는 의사가 환자의 현재 건강 상태에 따라 효과적이면서도 안전한 의약품들을 처방할 수 있도록 의사를 도와주기 위해 중요한 태스크가 되어 왔다.

기존 의약품 추천 방법들은 크게 두 가지 유형으로 구분될 수 있다. 첫 번째 유형은 instance-based 방법으로, 환자의 현재 방문 정보(즉, 환자의 현재 방문에서 진단된 질병, 수술 정보)만을 사용하여 환자에게 의약품들을 추천해주는 방법이다 [4]. 그러나 이러한 방법은 어떠한 질병에 대해 만성질환을 가진 환자와 급성질환을 가진 환자를 구분하지 못하고 같은 의약품

들을 추천한다는 문제를 지닌다 [5,6]. 두 번째 유형은 longitudinal-based 방법으로, 환자의 현재 방문 정보뿐만 아니라 과거 방문 정보들도 함께 사용하여 환자에게 의약품들을 추천하는 방법이다 [5-10].

최근, 대부분의 의약품 추천 방법들은 instance-based 방법의 한계를 극복하기 위해 longitudinal-based 방법에 초점을 두고 진행되어 왔다. 이들은 대개 순환 신경망(Recurrent Neural Network, RNN)을 기반으로, 환자의 과거 방문 정보들과 현재 방문 정보를 순서를 고려한 융합을 통해 하나의 임베딩으로 나타내고자 했다. 뿐만 아니라, 이들은 의약품들 간의 (1) 동시 처방 여부 관계와 (2) 부작용 유발 관계를 추가로 활용하여 두 관계 각각에 대한 의약품 특징 임베딩들을 생성한 뒤, 두 의약품 특징 임베딩들을 단순한 융합(구체적으로, 뺄셈 연산)을 통해 하나의 의약품 특징 임베딩으로 나타내고자 했다. 마지막으로, 이들은 앞

* 교신 저자

서 생성된 두 임베딩을 닷-프로덕트(dot-product) 연산하여 환자에게 효과적이면서도 안전한 의약품들을 추천했다.

비록 최신 의약품 추천 방법들이 의약품들 간의 두 관계(즉, 두 의약품 특징 임베딩)를 추가로 활용함으로써 환자들에게 더욱 효과적이고 안전한 의약품들을 추천했음에도 불구하고, 우리는 이들의 다음과 같은 한계를 지적한다: 기존 의약품 추천 방법들의 의약품 특징 임베딩 융합 방법은 두 의약품 특징 임베딩 사이의 복잡한 관계를 고려하지 못한다.

따라서 본 논문에서 우리는 실세계 데이터 집합과 최신 의약품 추천 방법들의 기반이 되는 의약품 추천 방법(구체적으로, GAMENet [5])을 활용하여, 두 의약품 특징 임베딩 사이의 복잡한 관계를 포착하는데 효과적인 의약품 특징 임베딩 융합 방법을 탐구하고자 한다.

2. 관련 연구

본 장에서는 최신 의약품 추천 방법들의 기반이 의약품 추천 방법인 GAMENet [5]에 대해 간단하게 리뷰하고 기존 방법들의 한계를 이야기하고자 한다.

GAMENet 은 다음 핵심 두 컴포넌트(component)로 이루어진다: (1) 인코더(encoder)와 (2) 예측기(predictor). 구체적으로, 인코더에서는 환자가 과거 방문들에서 진단받은 질병들, 수술들과 현재 방문에서 진단받은 질병들, 수술들을 각각 텐스한(dense) 임베딩들로 나타낸 뒤 이들을 RNN 의 입력 값으로 사용하여 순서를 고려한 하나의 환자 임베딩으로 융합한다. 다음으로, 예측기에서는 먼저 의약품들 간의 (1) 동시 처방 여부 관계와 (2) 부작용 유발 관계를 다음과 같이 각각 그래프로 구축한다: (1)에 대한 그래프에서 노드는 의약품들이고, 임의의 두 의약품(노드)이 임의의 환자의 한 방문에서 함께 처방된 적이 있다면 예시로 연결된다; (2)에 대한 그래프에서 노드는 의약품들이고, 임의의 두 의약품(노드)이 함께 복용했을 때 부작용을 일으킬 수 있다 [5,6,9] 고 알려져 있다면 예시로 연결된다. 이후, 구축된 두 그래프에 각각 그래프 합성곱 신경망(Graph Convolutional Network, GCN)을 적용하여 두 관계 각각에 대한 의약품 특징 임베딩을 획득한다. 그러고나서, 두 의약품 특징 임베딩에 뺄셈 연산을 적용하여 하나의 의약품 특징 임베딩을 생성한다. 마지막으로, 인코더에서 얻어진 환자 임베딩과 최종적으로 생성된 하나의 의약품 특징 임베딩을 닷-프로덕트 연산하여 환자에게 효과적이면서도 안전한 의약품들을 예측한다. 최종적으로, GAMENet 은 구축한 모델을 이진 교차 엔트로피 손실함수(Binary Cross Entropy loss), 다중 레이블 마진 손실함수(Multi-label Margin

loss), 의약품 부작용 손실 함수(Drug- Drug Interaction, DDI loss)를 통해 학습시킨다.

GAMENet 을 비롯한 기존 의약품 추천 방법들은 초기 의약품 추천 방법들에 비해 추천 정확도를 상당히 올렸음에도 불구하고, 서로 다른 특징을 나타내는 두 의약품 특징 임베딩들을 단순히 뺄셈 연산을 적용하여 융합할 뿐 다양한 융합 방법을 탐구하지 않았다.

3. 실험

3.1 실험 환경

본 장에서 우리는 두 의약품 특징 임베딩 사이의 복잡한 관계를 포착하는데 효과적인 의약품 특징 임베딩 융합 방법을 탐구하기 위해, 의약품 추천 태스크에서 널리 사용되는 실세계 공개 데이터 집합인 MIMIC-III 데이터 집합 [11]에 대해 실험을 진행한다. 구체적인 데이터 집합에 대한 통계치는 표 1 에 나타나 있다. 우리는 평가를 위해 데이터 집합의 환자들을 훈련 집합 (4/6), 검증 집합 (1/6), 시험 집합 (1/6)으로 나눈다. 또한, 우리는 각 환자에 대해 예측된 의약품들의 정확도를 측정하기 위해 의약품 추천 태스크에서 주로 사용되는 평가 지표인 Jaccard, PRAUC, F1 을 활용한다. 공정한 실험을 위해, 우리는 각 실험을 5 번 반복 수행한 뒤 평균 값을 나타낸다.

<표 1> MIMIC-III 데이터 집합 통계치

특성	값
환자들 수	5,442
방문들 수	14,124
질병들 수	1,958
수술들 수	1,430
의약품들 수	131
방문들 수의 평균 / 최대 값	2.59 / 29

3.2 비교 대상: 다양한 융합 방법들

우리는 두 의약품 특징 임베딩들을 가장 효과적으로 융합할 수 있는 방법을 확인하기 위해, 기존 의약품 추천 방법들에서 사용되는 융합 방법인 뺄셈 연산(이하 sub.) 과 다음과 같은 다양한 융합 방법들을 GAMENet 에 각각 적용하여 비교 실험을 진행한다: (1) avg.: 두 의약품 특징 임베딩들을 평균한다; (2) max-p.: 두 의약품 특징 임베딩들에 max pooling 을 적용한다; (3) concat-m.: 두 의약품 특징 임베딩들을 순차(concatenation)한 후 다층 퍼셉트론(Multi-Layer Perceptron, MLP)을 적용한다.

3.3 실험 결과

모든 실험 결과는 표 2 에 나타나 있다. 이 때, 평가 지표별로 가장 높은 정확도는 굵게, 두 번째로 높은 정확도는 밀줄로 표시하였으며, 가장 높은 정확도가 기존 융합 방법(즉, sub.) 대비 얼마나 개선되었는

지를 백분율로 표에 함께 나타내었다.

표 2 에서 보이는 것처럼 기존 융합 방법인 sub.은 모든 평가 지표에서 다른 모든 융합 방법들보다 낮은 추천 정확도를 보였다. 또한, 다른 모든 융합 방법들 중에는 모든 평가 지표를 고려해보았을 때, concat-m. 이 대게 가장 높은 정확도를 보였다. 이는 두 의약품 특징 임베딩 사이의 복잡한 관계를 고려하여 융합하는 것이 추천 정확도 개선에 효과적임을 의미한다.

<표 2> GAMENet [6]에서의 두 의약품 특징 임베딩들의 다양한 융합 방법에 따른 정확도

	Jaccard	PRAUC	F1
sub.	0.4350	0.6754	0.5902
avg.	<u>0.4600</u>	0.7412	<u>0.6270</u>
max-p.	0.4554	0.7465	0.6248
concat-m.	0.4671	<u>0.7419</u>	0.6279
개선	6.87%	9.52%	6.00%

4. 결론

본 논문에서 우리는 의약품들 간의 두 관계에 대한 임베딩들을 어떠한 방법으로 융합하는 것이 추천 정확도 개선에 가장 효과적인지 알아보기 위해 기존 융합 방법과 다양한 융합 방법들을 비교 실험하고 결과를 분석하였다. 그 결과, 두 의약품 특징 임베딩 사이의 복잡한 관계를 단순한 연산을 통해 융합하는 것보다 두 의약품 특징 임베딩들 각각의 의미를 고려하여 융합하는 것이 추천 정확도 개선에 가장 효과적임을 확인하였다.

사사

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 문화기술 연구개발 사업으로 수행되었음 (과제명 : 지능형 개인맞춤 재활운동 서비스 기술개발, 과제번호 : SR202104001, 기여율: 50%). 또한, 이 논문은 2023 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2022-00155586, 실세계의 다양한 다운스트림 태스크를 위한 고성능 빅 하이퍼그래프 마이닝 플랫폼 개발(SW 스타랩))

참고문헌

- [1] Weitong Chen, Sen Wang, Guodong Long, Lina Yao, Quan Z Sheng, and Xue Li. Dynamic illness severity prediction via multi-task rnns for intensive care unit. In International Conference on Data Mining. IEEE, IEEE Computer Society, Singapore, 2018, 917–922.
- [2] Yanda Wang, Weitong Chen, Bohan Li, and Robert Boots. 2019. Learning Fine-Grained Patient Similarity with Dynamic Bayesian Network Embedded RNNs. In DASFAA. Springer, Springer, Chiang Mai, Thailand, 2019, 587–603.
- [3] Lin Yue, Dongyuan Tian, Weitong Chen, Xuming Han, and Minghao Yin. 2020. Deep learning for heterogeneous medical data analysis. World Wide Web 23, 2020, 2715–2737
- [4] Yutao Zhang, Robert Chen, Jie Tang, Walter F. Stewart, and Jimeng Sun. LEAP: Learning to Prescribe Effective and Safe Treatment Combinations for Multimorbidity. In Proceedings of the ACM International Conference on Knowledge Discovery & Data Mining (ACM KDD), 2017, 1315–1324.
- [5] Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. GAMENet: Graph Augmented MEMORY Networks for Recommending Medication Combination. In proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2019, 1126–1133.
- [6] Chaoqi Yang, Cao Xiao, Fenglong Ma, Lucas Glass, and Jimeng Sun. Safe-Drug: Dual Molecular Graph Encoders for Recommending Effective and Safe Drug Combinations. In Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI), 2021, 3735–3741.
- [7] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter F. Stewart. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. In Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS), 2016, 3504–3512.
- [8] Junyuan Shang, Tengfei Ma, Cao Xiao, and Jimeng Sun. Pre-training of Graph Augmented Transformers for Medication Recommendation. In Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI), 2019, 5953–5959.
- [9] Yanda Wang, Weitong Chen, Dechang Pi, Lin Yue, Sen Wang, and Miao Xu. Self-Supervised Adversarial Distribution Regularization for Medication Recommendation. In Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI), 2021, 3134–3140.
- [10] Yanda Wang, Weitong Chen, Dechang Pi, Lin Yue, Miao Xu, and Xue Li. Multi-hop Reading on Memory Neural Network with Selective Coverage for Medication Recommendation. In Proceedings of the ACM International Conference on Information & Knowledge Management (ACM CIKM), 2021, 2020–2029.
- [11] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. Scientific data 3, 2016, 1–9.