

# 개체명 인식과 키워드 네트워크 분석을 활용한 약물 이상 반응 탐지 시스템 개발

이채연, 김현희  
 동덕여자대학교 정보통계학과  
 chland23@naver.com, heekim@dongduk.ac.kr

## Development of Detection of Adverse Drug Reactions based on Named Entity Recognition and Keyword Network Analysis

Chae-Yeon Lee, Hyon Hee Kim  
 Dept. of Statistics and Information Science, Dongduk Women's University

### 요 약

본 논문에서는 소셜 미디어 약물 리뷰 데이터로부터 약물 이상 반응을 탐지하는 모델인 FC-BERT 를 기반으로 소셜 네트워크 분석을 활용하여 웹 애플리케이션을 구현하였다. FC-BERT 모델을 거쳐 나온 개체명 인식 결과 중에 같은 의미를 가진 서로 다른 약물 이상 반응 표현들을 MedDRA 부작용 사전을 참고하여 하나의 MedDRA 용어로 표준화하여 매핑했다. 해당 결과에 소셜 네트워크 분석 기법을 적용하여 생성한 상위 15 개의 ADR 동시 출현 그래프를 상위 30 개의 워드 클라우드와 함께 시각화하여 보여주는 웹 애플리케이션을 개발했다. 동시 출현 그래프는 가장 많은 리뷰에서 동시에 나타나는 ADR 쌍을 보여준다. 본 논문에서 제안한 웹 애플리케이션은 사람마다 다르게 나타나는 다양한 약물 이상 반응을 사용자에게 좀 더 접근성이 좋게 제공할 수 있을 것으로 보인다.

### 1. 서론

본 논문에서는 일반 사용자들이 쉽게 약물 이상 반응을 검색하고 참고할 수 있도록 FC-BERT 모델[1]을 기반으로 ADR 표준화 매핑, 소셜 네트워크 분석을 적용하여 웹 애플리케이션으로 개발하였다. 본 연구에서 개발한 FC-BERT 모델 기반의 약물 이상반응 탐지 시각화 웹 시스템의 기능은 다음과 같다. 첫째, 사용자가 직접 작성한 약물 리뷰 데이터를 활용하여 사용자의 개인적인 특성을 포함한 약물 이상 반응을 보여준다. 둘째, 같은 의미의 다른 부작용 표현들을 MedDRA 부작용 사전을 통해 하나로 표준화 매핑한 후, 소셜 네트워크 분석을 적용한 결과를 제시한다. 셋째, 분석 결과는 상위 30 개의 워드 클라우드와 상위 15 개의 가장 동시 출현 그래프로 시각화하여 제공한다.

본 시스템을 활용하면 약물 부작용 탐지를 위해 장시간의 모니터링 시간 이전에 약물 사용자들의 실시간 리뷰 분석을 통해 약물의 부작용 시그널을 빠르게 파악하고, 이들간의 관계를 쉽게 알아볼 수 있다. 또한 사용자가 웹을 통해 원하는 약물의 ADR 탐지 시각화 결과를 간편하게 파악할 수 있다는 것이다.

2 장에서는 관련 연구를 약물 이상 반응과 소셜 네트워크 분석으로 구분하여 소개하고, 3 장에서는 FC-BERT 모델과 추출된 ADR 을 표준화하여 매핑하는 것에 대해 기술하였다. 4 장에서는 소셜 네트워크 분석을 적용한 결과를 사용자에게 접근성이 좋게 웹으로 구현하는 과정을 기술하고, 5 장에서는 결론에 대해 기술하였다.

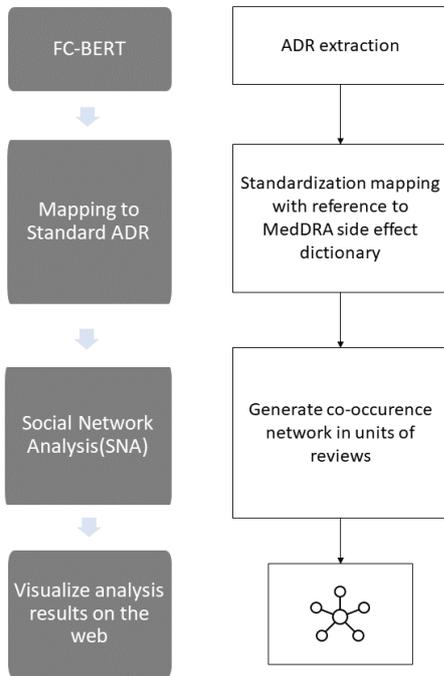
### 2. 관련 연구

약물 이상 반응 탐지 관련 연구로는 다중 측면 의미 정보 학습을 목표로 하는 Multihop Self-Attention Mechanism(MSAM)과 여러 예측 헤드를 결합하여 다중 작업 학습을 지원하고 멀티태스킹 학습을 사용하는 FARM 으로 미세 조정된 BERT, signal 감지를 위해 GNN 을 사용하여 약물-ADR 쌍을 예측하는 새로운 그래프 기반 접근 방식 등이 있다[2,3,4].

소셜 네트워크 분석 관련 연구로는 코로나 19 와 관련하여 현재의 협력 네트워크와 향후 구축되어야 할 네트워크를 비교 분석한 연구, 소셜 네트워크 분석을 통해 마약 관련 범죄 네트워크의 시사점을 발굴한 연구, 제 e-헬스의 주요 연구 주제와 동향에 대한 동시

발생 네트워크를 생성하고 분석한 연구 등이 있다[5].

### 3. 시스템 개요



(그림 1) Flowchart of the proposed web system.

그림 1 은 FC-BERT 기반의 약물 이상 반응 탐지 웹 애플리케이션의 흐름도이다. FC-BERT 를 거쳐 추출된 약물 이상 표현들을 MedDRA 부작용 사전에 기반으로 표준 ADR 로 매핑한 후, 리뷰를 단위로 소셜 네트워크 분석을 적용하여 결과를 웹으로 시각화한다.

#### 3.1 Filtering Clinical BERT(FC-BERT) 모델

FC-BERT 모델은 4 단계로 분류 가능하며, 입력은 약물 명을 포함한 텍스트와 포함하지 않은 텍스트로 분류할 수 있다. 첫 번째 단계는 입력으로 들어온 말뭉치를 문장 단위로 분류하는 단계로 약물 명 포함 여부와 상관없이 SentenceDetectorDL(SDDL) 모델을 사용한다. 두 번째 단계는 문장 리스트에서 약물 이상 반응 표현을 포함한 문장만을 추출하는 단계이며, 사용되는 모델은 ADE Classifier 이다. 약물명이 포함된 텍스트의 경우, 전체 문장에서 목표 약물명이 언급되기 전 모든 문장을 제거하는 필터링을 거친다. 이는 ‘약물을 복용하기 전 증상을 말하는 문장은 약물 명을 언급하기 전에 나온다.’를 전제로 하기 때문이다.

세 번째 단계는 목표 약물에 대한 약물 이상 반응 표현을 추출하는 단계이며, ADE Pipeline 모델을 사용한다. 약물명이 포함된 텍스트의 경우, ADE Relation Extraction 모델을 사용하여 추출된 약물 명-부작용 쌍에서 목표 약물에 대한 ADE 를 제거하는 필터링을 거친다. 이 과정을 거치면 목표 약물이 아닌 다른 약물에 대한 부작용 표현만을 얻게 된다. 두 번째 단계

의 약물명이 언급되기 전 모든 문장을 제거하는 필터링을 거친 ADE 에서 세 번째 단계에서 얻은 다른 약물에 대한 ADE 를 제거하는 필터링을 거친다. 마지막 단계에서는 모델을 통해 추출된 부작용 표현이 있으면 1 로 분류하고, 추출된 부작용 표현이 없으면 0 로 분류한다.

비교 실험한 결과, FC-BERT 모델의 정확도는 0.804 로 기존 모델보다 정확도가 약 19.3% 상승하였다. FC-BERT 에 대한 유사코드와 실험 결과는 [5]에서 확인할 수 있다.

#### 3.2 Adverse Drug Reaction(ADR) 표준화 매핑

약물 이상 반응 탐지 결과를 제공하기 위한 데이터로는 약물 리뷰 사이트인 drugs.com 를 크롤링 한 데이터와 everydayhealth.com, askapatient.com, iodine.com, webmd.com 에서 수집한 데이터, UCI ML Drug Review data 를 중복 제거하여 합한 약물 리뷰 데이터이다. celecoxib 는 676 개, ibuprofen 은 387 개, naproxen 은 1,175 개의 리뷰 데이터를 사용했다.

해당 ADR 은 약물 리뷰 데이터에 FC-BERT 를 적용한 결과로 리뷰에서 ADR 로 판명되는 부분을 그대로 추출한 것이라 같은 의미의 표현이라도 다양하게 표현되어 있어 이를 하나의 같은 표현으로 통일하는 작업이 필요하다. 따라서 해당 ADR 을 MedDRA 부작용 사전을 참고하여 MedDRA 용어로 표준화하여 매핑하는 작업을 실행하였다. 예를 들어, ‘abdominal pain’, ‘stomach upset’를 모두 ‘stomach pain’으로 매핑 한 것이다. 그 결과 ‘swelling of hands and feet’처럼 여러 개의 ADR 이 ‘and’로 묶여 하나의 ADR 로 합쳐지는 문제가 있었다. 따라서 ‘swelling of hands and feet’을 ‘swelling of hands’와 ‘swelling of feet’로 분리하는 작업을 진행했다. 최종적으로 celecoxib 는 263 개, ibuprofen 은 131 개, naproxen 은 481 개의 ADR 이 추출되었다.

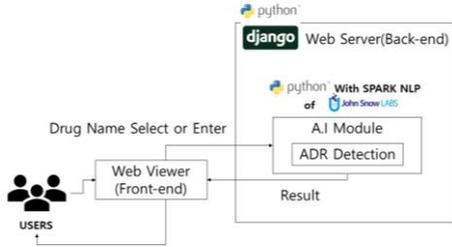
### 4. 시스템 구현

#### 4.1 소셜 네트워크 분석

최종적으로 얻어진 ADR Number, ADR Order, Review Number, ADR 등을 활용하여 각 리뷰에 동시에 나온 ADR 쌍의 빈도수를 구하여 동시 출현 그래프를 그렸다. 리뷰마다 얻은 ADR 들을 쌍으로 묶고 해당 쌍의 빈도를 계산한 후, 빈도를 가중치로 하여 동시 출현 그래프를 그리고 연결중심성, 매개 중심성, 근접 중심성을 계산하였다. celecoxib 에서는 ‘migraine’과 ‘muscle stiffness’, ‘tension headache’이 각각 3 개의 리뷰에서 같이 등장하는 것을 알 수 있다. ibuprofen 에서는 ‘heat’과 ‘cold’, ‘sweat’이 각각 6 개, 3 개의 리뷰에서 같이 등장하는 것을 알 수 있다. naproxen 에서는 ‘heartburn’이 ‘stomach pain’과 9 개의 리뷰에서 같이 등장하고, ‘dizziness’과 ‘stomach pain’, ‘nausea’이 각각 6 개, 4 개의 리뷰에서 같이 등장하는 것을 알 수 있다. 이는

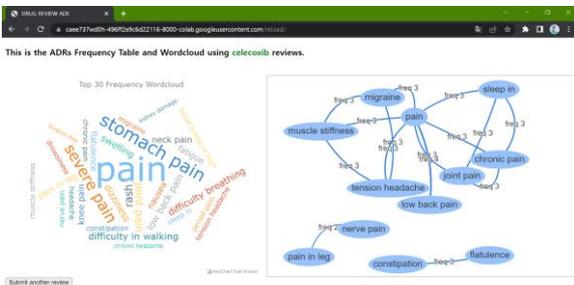
해당 ADR 들이 같은 리뷰에서 동시에 나타나는 것을 의미한다. 한편, ‘period pain’은 3 개 약물 모두에서 많이 등장하는데, 이는 증상이 필터링에 걸리지 않아 제거되지 않은 것으로 보인다.

4.2 시스템 구현



(그림 2) Architecture of the proposed web system.

FC-BERT 기반의 약물 이상 반응 탐지 웹 애플리케이션은 입력한 약물 명에 대한 사용자 약물 리뷰에 FC-BERT 모델과 소셜 네트워크 분석을 적용한 결과를 시각화하여 보여주며, 구조도는 그림 2 와 같다. 웹 서버는 Django 이며 웹 시스템은 MVT(Model-View-Template) 모델로 구현하였다. DB 는 SQLite 를 이용하였고, View 는 입력 데이터를 FC-BERT 에 적용한 후 적용 결과를 html 에 전달하며, Template 은 View 에서 받은 데이터를 html 로 화면에 구성한다. 결과는 상위 30 개의 ADR 워드 클라우드와 상위 15 개의 쌍으로 이루어진 가중 동시 출현 그래프로 보여준다. 워드 클라우드는 서버에서 전달받은 워드 리스트를 데이터로 하며, 워드 클라우드를 그려주는 자바스크립트 라이브러리인 AnyChart 라이브러리를 활용하였다. 동시 출현 그래프는 서버에서 전달받은 노드와 에지 리스트를 데이터로 하며, 동적 브라우저 기반 시각화 라이브러리인 vis.js 를 활용하였다. 본 연구에서 제안하는 웹 서비스는 구글 Colab 내에서 구동되며, 웹 서버는 파이썬으로 만들어진 오픈소스 웹 애플리케이션 프레임워크인 Django 를 사용하였다. Django 는 3.2.15 버전을, 파이썬(python)은 3.7.13 버전을 사용하였다.



(그림 3) Result Page of Celecoxib.

그림 3 은 ‘celecoxib 의 결과 페이지이다. ‘stomach pain’, ‘severe pain’이 각각 빈도수 12, 10 으로 가장 많이 언급된 약물 이상반응 표현이다. 또한 ‘constipation’, ‘flatulence’가 3 개의 리뷰에 같이 등장했다는 것을 알 수 있다.

5. 결론

본 논문에서는 필터링을 활용하여 약물 이상 반응에서 약물 복용 전 증상을 구분한 FC-BERT 모델을 기반으로 약물 리뷰에서 ADR 을 추출하고 표준화 매핑한 후, 소셜 네트워크 분석을 적용한 결과를 Django 웹 애플리케이션 프레임워크를 사용하여 웹 애플리케이션으로 구현하였다. 같은 의미라도 다른 단어들로 표현된 ADR 을 MedDRA 사전을 기반으로 같은 표현으로 표준화하여 매핑 한 결과를 상위 30 개의 워드 클라우드와 리뷰를 단위로 한 상위 15 개 동시 출현 그래프로 시각화하여 보여주었다. 실제 약물을 복용한 사용자들이 자유롭게 작성한 리뷰를 적용하였으므로 사용자의 개인적인 특성이 고려된 포괄적인 약물 이상반응을 탐지할 수 있을 것으로 보인다. 또한, 사용자가 접근하기 쉽게 웹 애플리케이션으로 구현하였으며 이를 실생활에서 활용할 수 있을 것으로 보인다. 향후 다양한 온라인 리뷰 사이트 리뷰를 활용하여 웹 애플리케이션을 구현한다면, 더욱 다양하고 새로운 약물 부작용 표현을 웹 사이트에서 쉽게 확인할 수 있을 것이다.

참고문헌

- [1] Chae-Yeon Lee, and Hyon Hee Kim, “Filtering Clinical BERT (FC-BERT): An ADR Detection Model for distinguishing symptoms from adverse drug reactions”, Proceedings of the Korea Information Processing Society Conference, pp. 549–552, 2022.
- [2] Tongxuan Zhang, Hongfei Lin, Yuqi Ren, Liang Yang, Bo Xu, Zhihao Yang, Jian Wang and Yijia Zhang, “Adverse drug reaction detection via a multihop self-attention mechanism.”, BMC Bioinformatics 20, 479 (2019).
- [3] Sajid Hussain, Hammad Afzal, Ramsha Saeed, Naima Iltaf, Mir Yasir Umair, "Pharmacovigilance with Transformers: A Framework to Detect Adverse Drug Reactions Using BERT Fine-Tuned with FARM", Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 5589829, 12 pages.
- [4] Kwak, H., Lee, M., Yoon, S., Chang, J., Park, S., Jung, K. Drug-Disease Graph: Predicting Adverse Drug Reaction Signals via Graph Neural Network with Clinical Data. In: Lauw, H., Wong, RW., Ntoulas, A., Lim, EP., Ng, SK., Pan, S. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2020.
- [5] Oh, Juyeon, Jinjae Kim, Taeho Lee, and Woojong Suh. “A Study on Collaborative Network for Coping with COVID-19 Using Social Network Analysis.” Journal of the Korea Industrial Information Systems Research 27, no. 3, 89–108. doi:10.9723/JKSIS.2022.27.3.089.