

빅데이터 기반 실시간 의약품 부작용 위험도 평가 모델

이채은, 배은지, 윤숙영, 임수연, 김용민, 김웅섭

동국대학교 정보통신공학과

e-mail : chaem516@naver.com , yhn01130@naver.com, dbstnrdud62@daum.net,
yeonm22@gmail.com, yongmin701@gmail.com, woongsup@dongguk.edu

Big data-based real-time drug side-effect risk assessment model

Chae-Eun Lee, Eun-Ji Bae, Sook-Young Yoon, Su-Yeon Lim, Yong-Min Kim

Woong-Sup Kim,

Dept of Information and Communication Engineering, Dongguk University,
Seoul, Republic of Korea

요 약

대한민국 인구의 고령화는 점점 더 심화되며 노인 환자의 수도 증가하는 추세이다.

노인환자들은 의약품 부작용 위험도가 높기 때문에 노인에게 부적절한 약물과 처방 이후의 부작용 발생 현황에 대한 통계적인 분석이 필요하다고 판단하였다. 의약품 부작용관련 문헌 정보와 실제

병원의 전자 의무 기록을 이용해 데이터 베이스를 구축하고 Python을 사용해 부작용 탐지 알고리즘을 설계했다. 노인 환자가 특정 약품을 투약한 이후에 그 약품에 부작용에 해당하는 진단을 받는다면 부작용이 발생한 것으로 판단한다. 알고리즘을 기반으로 사용자들이 쉽게 접근할 수 있도록 웹을 구축했다. d3.js를 통해 직관적으로 부작용 확률을 확인할 수 있도록 구현했고 새로운 진단 또는 처방을 입력하여 실시간으로 확률에 반영하였다. 실시간으로 데이터를 확인할 수 있기 때문에 부작용 이슈에 신속하게 예방 및 대응이 가능할 것으로 기대된다.

1. 서론

대한민국 인구의 고령화는 급속히 진행되어왔고, 앞으로 더욱 급격히 진행될 것으로 예측된다.

노인환자는 여러 가지 의약품을 동시에 복용하는 경우가 많고, 약물에 대한 순응도가 낮고 생리 기능이 저하되어 같은 약을 처방을 받더라도 그 약에 의한 부작용 발생 확률이 다소 높다. 따라서 노인에게 부적절한 약물과 처방 이후의 부작용 발생 현황에 대한 통계적인 분석이 필요하다고 판단하였다.

본 논문에서는 의약품 부작용에 대한 지식 베이스를 구축하여 전문가들의 의약품 부작용 정보 활용도를 높이고자 한다. 기존 의약품 부작용에 대한

자료들은 그 양이 매우 방대하고 표준이 없어 상호 호환성 문제로 의사나 약사들이 약의 부작용을 쉽게 인지하고 활용하기 어려웠다. 이를 개선하기 위해 본 논문의 연구에서는 의약품 부작용 실마리 탐지 알고리즘 설계를 통해 의약품의 부작용 정보를 실시간으로 제공하고, 전문가 뿐만 아니라 일반 사용자 모두가 쉽게 활용할 수 있도록 하여 의약품의 안전 사용 및 피해 최소화에 기여한다.

2. 요구 사항

- 특정 연령군을 대상으로 진단 데이터와 투약 데이터가 필요하다. 부작용 발생률을 구하기 위해서는 환자들의 진단 데이터와 투약 데이터가

필요하다. 노인을 대상으로 부작용 확률을 구하는 연구이기 때문에 수집한 진단 데이터와 투약 데이터 중 부작용 위험 약물을 처방받았던 노인들의 정보만 필터링 한다.

2) 수집한 자료를 바탕으로 사용자에게 부작용 발생 확률을 나타내주는 구체적인 방법이 필요하다. 진단 데이터와 처방 데이터를 이용하여 해당 약물을 투약한 노인환자만 분류하고, 해당 약품을 투약한 후의 진단 데이터를 이용하여 해당 약품의 부작용과 일치하는지 비교한다. 일치한다면 부작용이 발생한 것으로 확인하고 확률을 계산한다.

3) 약품에 대해 추가적인 부작용이 발생했을 경우 이를 실시간으로 데이터에 업데이트하는 기능이 필요하다.

약품에 대한 부작용 발생 정보가 많을수록 데이터의 신뢰도가 높아지기 때문에 부작용이 추가적으로 발생한 경우에 처방정보와 진단정보를 입력할 수 있도록 한다. 입력 받은 처방정보와 진단정보를 데이터에 추가하여 데이터베이스가 실시간으로 업데이트되고 정확한 결과가 출력될 수 있도록 한다.

4) 웹을 이용하여 사용자가 해당 약품의 부작용 빈도를 한눈에 파악하기 쉽고 원하는 정보를 빠르게 얻을 수 있도록 해야 한다.

부작용 발생 빈도를 통하여 투약하는 약의 위험도를 알고 싶은 노인 환자나 의사들이 약물 입력을 통해 쉽게 데이터에 접근하여 원하는 정보를 얻을 수 있도록 하기위해서 웹 페이지를 간단하고 이용하기 쉽게 설계하고 가장 중요한 정보인 부작용 발생률을 보기 쉽게 만드는 웹 설계가 필요하다.

3. 시스템 설계

1) 데이터베이스 설계

요구사항 1을 만족하기 위해서 Beer's Criteria 2015라는 미국에서 발표한 노인에게 부적절한 약물 등의 문헌정보를 바탕으로 하여 부작용 정보를 수집하고 데이터베이스를 생성하였다.

또한 부작용 발생유무를 확인하기 위해 1년간의 샘플 병원 데이터인 환자의 인적 기록부터 진료, 처방 등의 모든 의료 정보가 기록되어 있는 EHR(Electronic Health Record)기록을 데이터베이스화하였다. EHR 정보는 병원 내 시스템으로 다른 병원과 서로 호환되지 않지만 이후에 빅데이터 활용 및 의료기관 공유를 위해 만들고 있는 CDM(Common Data Model)이 일반화된다면

더 많은 병원 데이터를 함께 이용할 수 있을 것으로 예상된다.

데이터를 데이터베이스화하기 위하여 Python의 sqlite3 라이브러리를 이용하였고, sqlite 데이터베이스 파일을 생성하고 서버에 올려서 로컬뿐만 아니라 웹상에서 접근할 수 있도록 하였다.

2) 부작용 분석 알고리즘 설계

요구사항 2를 만족하기 위해 환자의 처방데이터와 진단데이터를 이용하여 부작용이 발생했다고 판단하기 위한 기준을 세웠다.

특정 약물을 투약한 환자에 이후 진단데이터를 확인하여 특정 약품을 처방받은 이후에 부작용에 해당하는 진단을 받는 경우, 부작용이 발생했다고 판단한다. 또한 약물에 대한 부작용 정보는 굉장히 방대할 뿐만 아니라 실질적으로 의약품에 의한 부작용이 발생할 확률은 생리 기능이 저하된 노인 계층에서 높게 나타나기 때문에, 노인의 데이터만을 사용한다. 이러한 데이터 필터링을 위해 환자의 생년월일 데이터를 이용하여 노인환자인지 확인하는 알고리즘을 설계하였다. 필터링 된 데이터를 바탕으로 의약품에 취약한 계층인 노인에게 부적절한 약물을 기준으로 삼아 알고리즘을 설계하였다.

부작용 분석 알고리즘 과정은 다음과 같다. 사용자에게 입력받은 약품명이 부작용 약물 데이터베이스에 존재하는지를 확인한다. 입력받은 약물이 부작용 약물에 해당한다면, 해당 약품을 투약한 환자가 있는지 확인한다. 약품을 투약한 환자가 있다면 해당 환자번호를 저장한 후 환자번호를 기준으로 노인인지 확인하고, 그 후의 진단에서 이전 약물로 인한 부작용에 해당하는 진단이 있었는지 확인한다. 이 과정은 진단명과 그에 상응하는 약물 부작용의 데이터를 매핑하여 구축해놓은 지식 데이터를 이용함으로써 이루어진다. 위와 같은 과정을 반복하면서 부작용이 발생할 확률을 직관적으로 보여주기 위한 확률계산을 한다.

3) 웹 서버 및 인터페이스 설계

요구사항에서 분석한 결과를 바탕으로, 효과적인 부작용 정보 제공을 위한 기능 및 목표를 설정하고, 이에 따른 웹 서버 및 인터페이스를 설계했다.



(그림 1) 시스템 구조도



(그림 2) 메인 화면

웹 인터페이스의 메인 페이지는 데이터 시각화 프레임워크 d3.js를 사용하여 약물을 부작용의 레벨을 색깔에 따라 구분할 수 있도록 구성하였다. 그림 1은 시스템 구조도로써 웹 UI로 부터 약물명을 입력받으면 웹 서버는 사전에 데이터베이스화 작업한 EHR(Electronic Health Record)과 미국의 Beers Criteria 문헌정보를 바탕으로 만 65 세 이상 노인들의 데이터만 필터링 한 후 약물의 부작용 확률을 계산하여 부작용 확률과 증상 및 기타 정보를 출력한다. 정보 제공 페이지는 긴요한 데이터인 약물명과 부작용 확률을 화면에 크게 띄워 사용자로 하여금 가시성을 증대시키는 방향으로 설계하였다. 또한 요구사항 4를 만족하기 위해 웹서버가 실시간으로 부작용 발생 정보를 추가하도록 하는 기능을 설계하였다.

메인페이지에서 진단 및 처방 추가 페이지로 이동하여 진단 정보 및 투약 정보 데이터를 입력하면 서버로부터 데이터베이스가 업데이트 된다. 사용자는 실시간으로 부작용 정보를 확인할 수 있다.

4) 웹 페이지 구축 상세

웹 언어인 HTML(Hypertext Markup Language)를 이용해 웹을 구축하고 스타일시트인 CSS(Cascading Style Sheets)를 사용하여 웹을 전반적인 디자인을 관리했다[3][5]. 또한 Javascript를 이용해 웹의 동작을 구현하여 풍부한 효과를 나타내었다. 노인 대상 약물 부작용 정보를 사용자에게 보여주는데 있어 데이터가 데이터로서만 존재하게 될 경우 가독성이 매우 떨어지게 된다. 이를 해결하기 위해 인터랙티브 데이터 시각화 라이브러리 D3.js를 이용하여 데이터의 전체적인 경향을 파악할 수 있도록 하였다. 사용자는 부작용 확률 레벨에 따라 나뉜 노드(원)의 색상을 통해서 직관적으로 위험도를 파악할 수 있다.

4. 구현 결과

4.1 메인 페이지

웹 메인 페이지에서 d3를 이용한 그래프를 통해 각 약품의 부작용 확률을 직관적으로 확인할 수 있도록 했다[1][6][7]. 부작용 발생 확률을 다섯 단계로 나누고 색깔로 구분했다. 적색에 가까울수록 부작용 확률이 높고 청색에 가까울수록 부작용 확률이 낮다. 페이지 오른쪽에서 확률에 따른 색상을 확인할 수 있다.

그리고 본 연구에서는 원하는 약물에 대한 정보를 찾기 위한 검색 기능을 구현했다. 검색을 위해 사용자로 하여금 직접 약품 이름을 입력 받을 수도 있고 그래프의 원을 클릭하면 해당하는 약품 이름이 검색창에 자동으로 입력되도록 설계하여 빠르고 간편하게 약물 검색이 가능하다.

4.2 검색 결과 페이지



(그림 3) 검색 결과

데이터베이스에서 쿼리문으로 검색한 약물의 데이터를 찾아 출력한다. 투약 정보 테이블에서 검색한 ‘Aspirin’을 투약한 환자의 환자 번호를 찾는다. 환자 번호를 이용하여 해당 환자를 찾고 65 세 이상인 경우, 투약한 노인 환자의 수에 포함시킨다. 투약한 노인 환자 중에서 ‘Aspirin’에 부작용에 해당하는 ‘Gastroenteritis and colitis 위장염과 대장염’을 진단받은 환자를 찾는다. 쿼리 결과를 이용해 부작용 발생 확률을 구한다. 사용자가 찾는 주요 정보인 약물명과 부작용 발생 확률을 크게 표시하여 사용자는 원하는 정보를 빠르게 확인할 수 있다.

4.3 정보 입력 페이지

(그림 4) 데이터 입력

환자 번호, 방문 유형, 진료일, 진단 코드, 진단명, 주상병 여부 등의 진단 정보를 입력 받는다.



(그림 5) 저장 완료 확인

입력한 정보의 저장을 확인하고 입력된 정보에 오탈자는 없는지 다시 확인할 수 있다.

의약품 부작용 이슈에 신속하게 대응하기 위해, 새로운 진단이 있는 경우 새로운 데이터를 입력 받기 위한 페이지다. 형식에 맞게 데이터를 입력하고 Add 버튼을 클릭하면 서버에 ‘POST’ 요청을 보내 입력 받은 데이터로 데이터베이스에 INSERT 한다. 투약 정보도 같은 방식으로 입력하여 데이터베이스에 저장한다. 새로운 정보는 실시간으로 확률 계산에 반영된다.

5. 결론

본 논문에서는 Beers Criteria와 같은 문헌 정보와 주로 병원에서 쓰이는 환자 관리 시스템 데이터 EHR를 바탕으로 데이터베이스를 구축하고 웹으로 구현하였다. 의약품 부작용 탐지 알고리즘 설계를 통해 의약품의 부작용 확률을 빠르게 예측하여 약품 안전 사용과 피해 최소화에

기여하도록 했으며, 웹 서비스를 제공하여 현장의 의사나 약사들이 약품에 대한 부작용 정보를 쉽게 접할 수 있고, 이를 통해 노인 부작용에 대한 정보 활용도를 높일 수 있었다. 의사나 약사들의 노인 부작용에 대한 정보 활용도가 높아짐에 따라서 노인 환자들은 안전하게 의약품을 사용하고 의약품 사용에 대한 불안도가 줄어들게 된다. 또한, 방대한 양의 문헌 정보를 이용하여 빠르게 원하는 정보를 얻을 수 있기 때문에 시간적인 면에서 이득이 있다.

의약품의 종류가 방대하여 본 논문에서는 가장 필요성이 높은 대상인 노인으로 국한하였으나 추후, 영유아 및 임산부로 대상을 확장할 수 있을 것으로 예상된다. 또한, 현재 거의 모든 병원이 전자 의무 기록(EMR)을 사용하고는 있으나 초기 가이드라인의 부재로 통일된 데이터 형식 없이 병원마다 각기 다른 EHR 소프트웨어를 사용해 본 연구에서는 호환성의 문제로 한 병원의 데이터만을 사용하지만 훗날 CDM(Common Data Model)이 사용된다면 모든 병원에서 공통적으로 사용 가능할 것으로 예상 된다.

6. 사사

"본 연구는 과학기술정보통신부 및
정보통신기획평가원의
SW 중심대학지원사업의 연구결과로
수행되었음"(2016-0-00017)

참고문헌

- [1] 스캇 머레이 저, 변치훈 역, 『D3.js』, 인사이트, 2014
- [2] 배프 저, 『배프의 오지랖 파이썬 웹프로그래밍』, 디지털북스, 2019
- [3] 최옥경 저, 『쉽게 배우는 HTML5 & CSS & JavaScript』, 북스홀릭퍼블리싱, 2019
- [4] 김석훈 저, 『파이썬 웹 프로그래밍:Django(장고)로 배우는 쉽고 빠른 웹 개발』, 한빛미디어, 2015
- [5] 장현희 저, 『(통)HTML&JavaScript&Css 기본+활용 지대로 배우기』, 웨북, 2007
- [6] 후루하타 카즈히로 저, 『D3.js 입문 : 웹을 매력적으로 만드는 그래프와 차트 만들기』, 프리렉, 2014
- [7] 일리야 미크스 저, 『D3.js 인 액션 : 기본 차트부터 빅데이터 시각화까지 데이터를 시각화하는 최고의 방법』, 한빛미디어, 2016