

약물들의 유사성을 이용한 숨겨진 약물 효과 추론

김신*, 김정우*, 박상현*†

연세대학교 컴퓨터과학과*

jcynico@gmail.com, jwkim2013@cs.yonsei.ac.kr, sanghyun@cs.yonsei.ac.kr

Inferring Hidden Drug Effects Using Similarity between Drugs

Shin Kim*, Jeongwoo Kim*, Sanghyun Park*†
Dept. of Computer Science, Yonsei University *

요 약

신약재창출(drug repositioning) 기술은 큰 비용과 많은 시간을 요구하는 기존 신약개발 과정보다 효율적인 신약개발 전략이다. 많은 약물 정보와 의약학 문헌자료들이 공개적인 데이터베이스로(database) 제공됨에 따라 생물학적 정보를 기반으로 하는 신약재창출 기술 개발의 길이 넓어졌다. 본 논문에서는 이러한 생물학적정보를 활용하여 약물 사이의 유사도(similarity)를 계산하고, 그 값을 기반으로 약물들의 알려지지 않은 효과를 추론하는 방법론을 제안한다. 특정 효과가 있는 약물들을 하나의 군집(cluster)으로 구성한다. 약물 데이터베이스와 생물학 문헌데이터를 이용하여, 효과(effect)/부작용(side effect) 측면에서 이 군집과 가장 유사한 약물을 추출한다. 위의 방법으로 추출된 약물과 약물 군집이 공유하는 특정 효과 사이의 중요한 생물학적 관련이 있을 것이라고 추론한다.

1. 서 론

신약개발은 연구 단계와 개발 단계라는 복잡한 과정을 거쳐야 하므로 많은 시간과 자본이 소요된다. 신약개발 관련 기술의 급격한 발전에도 불구하고 약물의 안정성과 예측 가능성에 대한 요구도 함께 증가하여 아직은 신약 개발에 드는 비용이 감소하고 있지 않았다. 이로 인해 이미 안정성을 검증받은 기존 약물들의 가치를 높이는 방법이 주목받고 있다. 이러한 방법을 신약재창출이라고 한다.

신약재창출 기술은 안정성이 검증된 약물들의 새로운 의학적 용도를 발굴하는 개발 전략이므로 시간과 자본을 절약할 수 있다. 신약재창출 기술은 기존 약물의 유용성을 증대시킬 뿐만 아니라 해당 약물에 대한 특허 독점 기간 연장에 도움을 주는 매우 유용한 전략이다.

하나의 약물은 다양한 질병에서 치료, 예방, 지연 효과를 가지기도 하고 다양한 부작용을 가지고 있기도 하다. 이처럼 약물이 가지고 있는 다양한 작용들을 활용하여 신약재창출 전략을 제시하고자 한다. 생물학 문헌데이터를 활용하여 작용 간 유사도를 계산하고, 이를 통해 약물 간 유사도를 계산하여 특정 작용을 가질 것이라 예상되는 약물을 추론하는 것이 연구의 목표이다.

특정 작용을 하는 약물들을 모아 하나의 약물 군집을 만들고, 이 약물 군집이 가지고 있는 작용들을 가장 많이 공유하는 새로운 약물은 높은 확률로 이 약물 군집에 포함될 수 있다는 가설을 기반으로 한다. 본 논문의 구성은 다음과 같다. 2장에서는 본 연구의

기반이 되는 관련 연구들에 관하여 기술하고, 3장에서는 본 논문이 제안하는 방법론을 설명한다. 마지막으로 4장에서 본 논문의 결론 및 발전 방향에 관하여 기술한다.

2. 관련 연구

유전자, 단백질, 질병 등의 생물학 개체들은 상호작용하므로, 생물학 분야에서 생물학적 개체들 사이의 관계를 찾는 작업은 매우 중요하다. 이러한 생물학 개체들 사이의 관계를 찾는 대표적인 방법으로 Swanson의 ABC model[1]이 있다.

Swanson은 의약학 문헌에서 A와 B의 연관성이 확인되고, B와 C의 연관성이 확인되었다면 A와 C의 연관성도 충분히 존재할 수 있음을 제안했다. 예를 들어, 작용 A를 하는 임의의 약물을 복용하면 작용 B가 나타나고 작용 B를 하는 또 다른 약물 C가 존재한다고 할 때, 작용 A와 약물 C의 연관 관계를 추론할 수 있다. 작용 A와 약물 C를 잇는 B와 같은 작용이 많을수록 더 큰 연관관계가 있다고 판단할 수 있다. 이러한 ABC model을 발전시켜 질병과 유전자의 관계를 추론하는 방법들에 대한 연구도 진행되고 있다.

Özgür[2]은 특정 질병과 연관이 있는 유전자와 같은 문헌의 같은 문장 안에 속해있는 유전자들을 추출하고, 두 유전자를 연결하여 유전자 네트워크를 구축하고, 구축된 네트워크에서 중심성(centrality)이 높은 유전자일수록 특정 질병과 관련이 깊을 것이라고 제안했다. 전립선암을 질병 데이터를 기반으로 실험하였고, 중심성을 기준으로 상위 20개의 유전자들 추출하였다. 추출된 20개의 유전자 중 최대 19개의 유전자가 질병과 관계가 있음을 검증하였다. 두 개의 개체가 문헌상에서 같은 문장에 동시에 출현하는 것을

† 교신저자(Corresponding Author)

* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015R1A2A1A05001845).