[BIOINFO 2021 정기학술대회]

A Supervised Markovian Random Walk Model for Investigating Hepatotoxicity Signatures of Chemical Drugs with Structural Alerts

Sangsoo Lim[1], Youngkuk Kim[2], Jeonghyeon Gu[3], and Sun Kim[1,2,3,4,5,*]

[1]*Bioinformatics Institute,* [2]*Department of Computer Science and Engineering,*
[3]*Interdisciplinary Program in Artificial Intelligence,* [4]*Institute of Engineering Research, Seoul National University*
[5]*Aigendrug, Co., Ltd*
**Corresponding author: sunkim.bioinfo@snu.ac.kr*

In early stages of drug discovery, Drug-induced Liver Injury (DILI) prediction only using structural information is desirable for cost-effective drug discovery because screening of candidate compounds can be done before expensive *in vitro* or *in vivo* experiments are performed. Our approach uses graph representation as is, rather than linear representation formats of drugs, i.e., SMILES. To effectively characterize graphs of variable sizes, we combine state-of-the-art graph mining technologies into a single computational framework for identifying toxic signatures of graphs. Our approach adopts a generate-test-refine strategy that explicitly enumerates subgraphs on small molecule drugs to predict DILI only using structural information. The first step of graph mining requires generation of subgraphs that are related to toxicity of drug. We use the supervised learning framework of Markovian random walk by optimizing subgraphs with respect to DILI labels. Then, the subgraphs are tested and refined by trimming subgraphs until they are over-represented either in toxic and nontoxic compounds. As a result, our method achieved superior performance, more than 15% improvements over previous DILI prediction methods and state-of-the-art GNN approaches. In addition to the performance improvement, DILI-related SAs are derived both at the single subgraph level and at the multi-subgraph level. To the best of our knowledge, the explicit use of graph structures for investigating liver toxicity of drugs is the first of its kind and explainability of prediction with SAs is specific and clear.