Pharmaceutical Regulatory Affairs 2012: 3D Equivariant Graph Neural Networks for Drug Discovery -University of Amsterdam and Qualcomm Technologies, Netherlands

Max Welling
University of Amsterdam and Qualcomm Technologies, Netherlands, E-mail: MaxWelling@gmail.com

Abstract

In this talk I will first review a very useful machine learning tool for pharmaceutical applications: the Graph neural Network (GNN). A GNN analyses a signal on graph structured data, such as molecules or knowledge graphs. We then define the concept of equivariance, an organizing principle that embues the neural network with the right transformation properties under symmetry transformations. For instance, a molecule has the same properties whether we rotate it in space or not. We will finally apply the new 3D equivariant graph neural network to molecular data and show that it can successfully predict properties that are relevant for drug screening. If time permits we will briefly describe further relevant generalizations of this technology. Graph machine learning (GML) is receiving growing interest within the pharmaceutical and biotechnology industries for its ability to model biomolecular structures, the functional relationships between them, and integrate multi-omic datasets — amongst other data types. Herein, we present a multidisciplinary academic-industrial review of the topic within the context of drug discovery and development. After introducing key terms and modelling approaches, we move chronologically through the drug development pipeline to identify and summarize work incorporating: target identification, design of small molecules and biologics, and drug repurposing. Whilst the field is still emerging, key milestones including repurposed drugs entering in vivo studies, suggest GML will become a modelling framework of choice within biomedical machine learning. Drug development is a time-consuming process which might take decades to approve the final version of the drug .It starts from the initial stage of drug discovery where it identifies certain groups of molecules that are likely to become a drug. Then, it goes through several steps to eliminate unsuitable molecules and finally tests them in real life. Important features that we look at during the drug discovery stage are ADME (Absorption, Distribution, Metabolism, and Excretion) properties. We can say that drug discovery is an optimization problem where we predict the ADME properties and choose those molecules that might increase the likelihood of developing a safe drug. Highly efficient computational methods that find molecules with desirable properties speed up the drug development process and give a competitive advantage over other R&D companies.

It was only a matter of time before machine learning was applied to the drug discovery. This allowed to process molecular datasets with a speed and precision that had not been seen before .However, to make the molecular structures applicable to machine learning, many complicated preprocessing steps have to be performed such as converting 3D molecular structures to 1D fingerprint vectors, or extracting numerical features from specific atoms in a molecule.