Abstract

This dissertation is a collection of deep learning methods developed with a particular focus on applications for pharmaceutical problems. My goal was to develop neural network-based tools at various stages of drug discovery to shorten this process and reduce its cost. I focused particularly on exploring the capabilities of graph models.

Medical, biological and chemical data have specific structure that should be considered when creating machine learning models. In particular, chemical compounds are often represented as graphs annotated with additional vertex and edge features that correspond to the atoms and bonds of the molecule. Graph neural networks, which have been significantly advanced in recent years, are increasingly being used to process this type of data. These models are now being used, for example, to predict the properties of chemical compounds and propose new drug candidates.

This dissertation includes six publications related to modeling molecules using graph networks. The first one introduces a new graph network architecture, taking into account the spatial positions of vertices. The second deals with the representation of compounds that are given as input to this type of graph networks. The third work is a review paper describing the generation of new chemical compounds that can be further developed as potential drugs. In the fourth paper, we show the application of graph networks in a case study related to the modulation of metabolism. The fifth paper is a proposition of a new interpretable graph network that explains its predictions based on the use of examples. Finally, the sixth paper shows how graph networks can be used to optimize chemical compounds.

Another type of data widely used in the field of medical and pharmaceutical sciences is image data. These can be natural images (photographs of objects taken with regular cameras) or, more commonly, images coming from a variety of medical instruments or specialized microscopes, such as X-rays or fluorescence microscopy. The second part of the series includes publications aimed at improving current image processing techniques that may be applied in the fields of chemistry, medicine and pharmacy. One of these papers describes the process of data acquisition and the use of neural networks on mobile devices for the rapid preliminary classification of lichen species. The other two papers present graph-based approaches to image processing, which refers technically to the first part of the series.