Drug Discovery Using Neural Networks

Problem Statement

Previously neural networks have become an efficient method for data analysis in the field of drug discovery because the volume of data is very large so to handle it efficiently and do the discovery of drugs deep learning is preferred. Results generated by deep Q-Learning are more accurate. The early problems encountered with neural networks such as overfitting and overtraining have been addressed resulting in a technique that surpasses traditional statistical methods. Deep Q-Learning have thus mainlysurvived up its promise, that is to overcome QSAR statistical problems. So, in this research areinforcementlearningtechnology,calledDeepQ-learning,togetherwith an RNN is to be proposed for drug discovery which will give more accurate results.

Background

There were many techniques used in drug discovery all aimed at shortening the time to identify development candidates.Previously the data is of Largervolumesso as the technology developed the use ofmachinelearning techniques are preferred. Besides establishedmethodslike support vectormachines(SVM), neuralnetworks(NN)and random forest which have been utilized to develop QSAR models for a long time, methods like matrix factorization and deep learning are also used in this field to do the drug discovery. CNN are also used to discover drugs.Self-Organizing Map (SOM) or Kohonen neural network is also used.

Methodology

Step 1: Data collection and dataset preparation

This will involve collection of data with chemical structures and activity data after this preprocessing is applied to extract features.

Step 2: Developinga Deep Q-learning technique for Drug Discovery.

A reinforcementlearningtechnology,calledDeepQ-learning,togetherwith an RNN is to be developed for drug discovery.

Step 3: Training and experimentation on datasets

The Deep Q-learning model is then trained on the training data set to do drug discovery accurately.

Step 4: Deployment and analysis on real life scenario

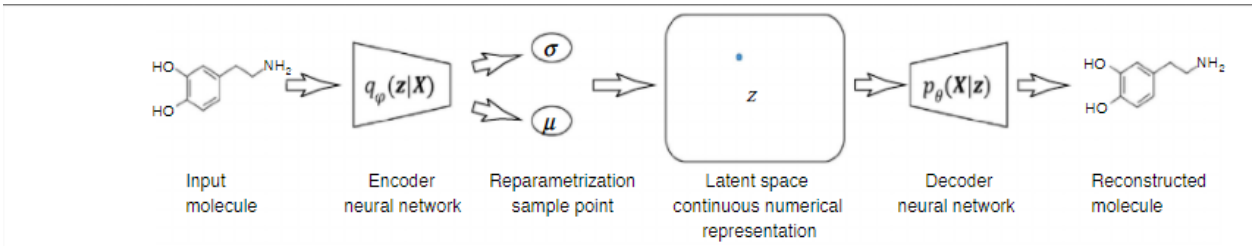

Figure.1 Work flow in discovering molecules of drugs [Chen, Hongming, et al. "The rise of deep learning in drug discovery." *Drug discovery today* (2018).]

The trained and tested discovering molecules of drugs model will be deployed in a real-life scenario for further analysis where both positive and failure cases will be leveraged for further improvement in the methodology and it will follow the above workflow.

Experimental Design

Dataset

Dataset for drug discovery using neural networkswith chemical structures and activity data are available for free i.e., ChEMBL or ChEMBLdb which can be used for experimentation and evaluation. URL to download this dataset is https://www.ebi.ac.uk/chembl/

Evaluation Measures

Measures such as Classification error, Computational cost, Accuracy, Loss, EOR (enrichment over random)can be used for calculating the accuracy of drug discovery using neural network.

Software and Hardware Requirements

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as, Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the Deep Q-learning technique for Drug Discovery.