

Gene Expression Prediction based on Deep Learning

Rui Xie

Dr. Jianlin Cheng, Thesis Supervisor

Abstract

Gene expression is a critical process in a biological system that is influenced and modulated by many factors including genetic variation. Thus, it is important to understand how genotypes affect the gene expression levels. Although several approaches have been implemented, we proposed a deep learning regression model to learn complex feature representation and to deal with over-fitting. In our experiment, the deep learning model produced results that are comparable to results generated by other methods by applying an independent test data set.

This thesis has several contributions. First, we propose an accurate predicting model based on deep learning to extract useful features with multilayer perceptron and stacked denoising auto-encoders after preprocessing the input data. Second, we ran a test on an independent dataset for several approaches to evaluate the performance of a multilayer perceptron with stacked denoising auto-encoders. Third, we further improved our model by adding a dropout technique to prevent overfitting. The result shows that dropout improved the model when we compared the result of our model with results of other existing approaches to evaluate its performance with a test data set. Finally, we present a software package that allows users to train the model with their own data and make predictions. An instruction on how to use this software package was also provided.