Final Project Proposal: Using Improved Deep Neural Network to Perform Dimension Reduction and Cell Classification by Analyzing scRNA-seq data

Jason Chen, Xuanyu Dong, Zhiran Xie

Department of Biological Science, University of California, San Diego

BIPN 162: Neural Data Science

Dr. Mikio Aoi

May 20, 2022

Background:

Single-cell RNA Sequencing is a useful technique to reveal cell heterogeneity. However, previous methods to analyze scRNA-seq data such as UMAP would produce results that are hard to interpret with biological features. Supervised methods including SVM are also with demerits like time-consuming. Though previous Deep Neural Network (DNN)'s outstanding performance for processing large amounts of data helps us classify cell types and states, it uses protein-protein interactions or gene regulation constraints and yields the result of low interpretability. In order to maintain the high performance of DNN and relatively higher interpretability, a refined pathway-driven DNN is proposed. The refined DNN applies the cell signaling pathway into its architecture to reduce the dimensionality of scRNA-seq data. Moreover, using signaling pathways can also improve the accuracy and interpretability of the DNN, since different cells have their own distinct signaling pathways, which means that it is easier to classify cell types through their signaling pathways. Specifically, how the refined DNN works is that the first layer serves as a filter (it will lock the weight to zero if no signaling pathway between nodes is found) and reduces the number of node interconnections required for a DNN. By analyzing the activation score of the first hidden layer, we can infer cell signaling pathways. Overall, the improved DNN will have the following benefits: smaller architecture (less computing time), outstanding performance, and better interpretability.

Proposed work

We will use human melanoma tumor cells from the original dataset to reproduce the result from the paper. Based on the paper, several cell types including B-cell and NK are tested. 6 types of cells are used with the selection of parameters, functions, and preprocessing steps. We expect the result to be the high-accuracy classification of B-cell, Macrophage, T-CD4, T-CD8, NK, and

some unknown cells in the samples with improved DNN trained with the input dataset. The original data set will be tested, as we mentioned above. In addition, we will use an additional dataset to train this DNN and analyze the performance of the model. For the additional dataset, we will use scRNA-Seq data from Allen Brain Map. Specifically, the dataset contains scRNA-Seq data from mouse ALM (anterior lateral motor cortex) with corresponding cell types. In order to show that we have successfully implemented our method, we will generate 2-D TSNE and Clustering analysis plots that represent our classification results for the cells in our additional dataset. We expect the improved DNN will also work on our additional dataset because deep neural networks are artificial neural networks that are designed to deal with data with similar forms. Since our additional data and the original data are all scRNA-seq data in the same form, the DNN will have no trouble handling our additional dataset. Also, we will explore the signal pathways revealed by the DNN. Specifically, we will extract the most weighted nodes learned by the improved DNN because they represent the most important cell signaling pathways and they may contain important information about how the cells regulate their activities. We propose that the improved pathway-oriented DNN can help us reveal the signaling pathways of different cells in mouse ALM, which is crucial for neuroscience study because it potentially helps us to understand cellular activity regulations in mouse ALM.

Delegation of duty

Group members should collaborate and discuss steps during the project process. Respect and integrity should be obeyed by group members. We decided to collect data and write code together, and each of the group members should accomplish their assigned tasks such as testing code or editing bugs. Zhiran Xie is responsible for introducing data and parameters. XUanyu Dong and Jason Chen will be responsible for editing and running the code.

References

Gundogdu, P., Loucera, C., Alamo-Alvarez, I., Dopazo, J., & Nepomuceno, I. (2022). Integrating pathway knowledge with deep neural networks to reduce the dimensionality in single-cell RNA-seq data. Biodata Mining, 15(1). doi: 10.1186/s13040-021-00285-4