Analysis of Omics Data using OmiEmbed

URVISH PUJARA - 2020101032

SYED IMAMI - 2020113012

Introduction

Multi-omics analysis is an emerging field that aims to integrate multiple types of biological data, such as genomics, transcriptomics, proteomics, and metabolomics, to gain a more comprehensive understanding of biological systems.

VAE learns to encode omics data into a lower-dimensional representation and then decodes this representation back into the original input data.VAE can be trained to generate new samples of gene expression data that are consistent with other types of biological data, such as protein expression data.

By learning joint representations of multiple types of biological data, VAEs can help to identify new biological insights and biomarkers that would be difficult to detect using single-omics analysis like cancer classification.

These techniques will help us understanding the complex interactions between different types of biological data.

Problem Statement

OmiEmbed supports multiple tasks for omics data including dimensionality reduction, tumour type classification, multi-omics integration, demographic and clinical feature reconstruction, and survival prediction. This can further be extended to predicting the growth rate of tumour and providing feature importance for each type of cancer. Using the age predictions, demographic analysis can be performed such as the analysis frequency of people at various stages of cancer in all types of cance. By incorporating clinical features, the model aims to provide a more comprehensive analysis that takes into account factors that can affect patient outcomes. Overall, the new model aims to improve patient care by providing accurate and personalized predictions of survival outcomes.

Goals

01

Relating and analyzing properties for multi-omics

02

Generate a lower dimensional representation of omics data

03

Performing different demographic task, providing feature importance, growth rate of tumor

Resources

- OmiEmbed: A Unified Multi-Task Deep Learning Framework for Multi-Omics Data by Xiaoyu Zhang, ORCID, Yuting Xing, Kai Sun and Yike Guo
  Published on:18 June 2021
  Link:
  Journal: Cancers - Volume 13, Issue 12

link

- Performance Comparison of Deep Learning Autoencoders for Cancer Subtype Detection Using Multi-Omics Data, by Edian F. Franco
  Published on: 22, April 2021
  Journal: Cancers (Basel)

link

- Multi-omics Data Integration, Interpretation, and Its Application Indhupriya Subramanian, Srikant Verma, Shiva Kumar, Abhay Jere2 and Krishanpal Anamika
  Published in: 2020
  Journal: Bioinform Biol Insights , v.14; 2020

link

Data set

Data set reference link : link

- GSE109381 BTM dataset for DNA Methylation
- TCGA Pan Cancer dataset
  - Tumour types: 33 +1(normal)
  - Omics data type: Gene expression, DNA methylation, miRNA expression
  - No of features 60483, 485577, 1881 respectively
  - No of samples 11538, 9736, 11020 respectively

Implementation

- Data preprocessing
  - first processed by the Bioconductor R package minfi to obtain the beta value of each CpG probe.
- Feature selection.
  - probes were filtered out in DNA methylation and gene expression data. (ex. the ones targeting y chromosome)
- Dimensionality reduction using autoencoders.
  - one-dimensional convolutional neural network (CNN) and the fully connected neural network (FC) for encoder and decoder in deep embedding module
- Downstream networks
  - diagnostic task
  - prognostic task
  - demographic task

Work Done

- Implementation of the baseline paper
- Started working on age predictions

```
TEST] [Epoch: 199   Iter:   32]   recon_A: 0.491   kl: 0.069   classifier: 0.012   accuracy: 1.000
TEST] [Epoch: 199   Iter:   64]   recon_A: 0.493   kl: 0.077   classifier: 0.338   accuracy: 0.953
TEST] [Epoch: 199   Iter:   96]   recon_A: 0.489   kl: 0.090   classifier: 0.014   accuracy: 0.969
TEST] [Epoch: 199   Iter:  128]   recon_A: 0.491   kl: 0.170   classifier: 0.008   accuracy: 0.977
TEST] [Epoch: 199   Iter:  160]   recon_A: 0.492   kl: 0.101   classifier: 0.060   accuracy: 0.975
TEST] [Epoch: 199   Iter:  192]   recon_A: 0.491   kl: 0.076   classifier: 0.127   accuracy: 0.974
TEST] [Epoch: 199   Iter:  200]   recon_A: 0.485   kl: 0.125   classifier: 0.094   accuracy: 0.975
TEST] [Epoch: 199]      recon_A: 0.490   kl: 0.101   classifier: 0.093   accuracy: 0.975   precisio
esting time used: 0.018s
TRAIN] [Epoch: 200   Iter:   32   Load_t: 0.001   Comp_t: 0.008]   recon_A: 0.466   kl: 0.114   cla
TRAIN] [Epoch: 200   Iter:   64   Load_t: 0.000   Comp_t: 0.008]   recon_A: 0.467   kl: 0.117   cla
TRAIN] [Epoch: 200   Iter:   96   Load_t: 0.000   Comp_t: 0.008]   recon_A: 0.467   kl: 0.124   cla
TRAIN] [Epoch: 200   Iter:  128   Load_t: 0.000   Comp_t: 0.008]   recon_A: 0.464   kl: 0.127   cla
TRAIN] [Epoch: 200   Iter:  160   Load_t: 0.000   Comp_t: 0.008]   recon_A: 0.465   kl: 0.113   cla
TRAIN] [Epoch: 200   Iter:  192   Load_t: 0.000   Comp_t: 0.007]   recon_A: 0.465   kl: 0.104   cla
TRAIN] [Epoch: 200   Iter:  224   Load_t: 0.000   Comp_t: 0.007]   recon_A: 0.467   kl: 0.123   cla
TRAIN] [Epoch: 200   Iter:  256   Load_t: 0.000   Comp_t: 0.008]   recon_A: 0.465   kl: 0.125   cla
```

THANK YOU