

CLASSIFICATION OF MICRORNA USING DEEP-LEARNING METHOD

Project proposal of Computational Genomics Spring 2017

Han Den (Andrew Id: hdeng1)

Luyi Ma (Andrew Id: luyim)

Project Idea

- A microRNA (abbreviated miRNA) is a small non-coding RNA molecule (containing about 22 nucleotides) found in plants, animals and some viruses, that functions in RNA silencing and post-transcriptional regulation of gene expression. While the majority of miRNAs are located within the cell, some miRNAs, commonly known as circulating miRNAs or extracellular miRNAs, have also been found in the extracellular environment, including various biological fluids and cell culture media [1]. So it's of great value to distinguish miRNAs from the whole genome and make accurate classifications.

The deep learning network will be worked as a classifier and we basically use CNN to make the classification. We will use some natural language processing methods to parse DNA sequence and make classifications based on these data. We choose both miRNA and non-miRNA data to train our model and test accuracy.

- We plan to use data of miRNA from an open-source database [2]. All the miRNA sequence in this database will be changed into DNA sequence based on the complementary rule and used as our positive input training and testing data. We will use permutation methods to generate our negative input training data and use real data in a sample genome as our negative input testing data. The output will be either 0 or 1, indicating it's miRNA or not.

Software Design

- We plan to use Python as our preferred programming language. We will implement algorithms involved in deep learning and natural language processing in Python. And we also plan to use some deep learning frameworks like Tensorflow to help us in feature extraction. To process data and train the model, we plan to use AWS platform. Finally, we may develop a software to wrap out core scripts implementing our models.

Papers to Read

- **List fo papers**
 - DNA Sequence Classification by Convolutional Neural Network [3].
 - Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations [4].

References

- [1] <https://en.wikipedia.org/wiki/MicroRNA>
- [2] <http://www.mirbase.org/>
- [3] Nguyen N G, Tran V A, Ngo D L, et al. *DNA Sequence Classification by Convolutional Neural Network*[J]. Journal of Biomedical Science and Engineering, 2016, 9(05): 280.
- [4] Lee H, Grosse R, Ranganath R, et al. *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations*[C]Proceedings of the 26th annual international conference on machine learning. ACM, 2009: 609-616.