Public Abstract

First Name:Renzhi

Middle Name:

Last Name:Cao

Adviser's First Name:Jianlin

Adviser's Last Name:Cheng

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:SS 2016

Department:Computer Science

Degree:PhD

Title:Genome data analysis, protein function and structure prediction by machine learning techniques

The raw information of a typical human genome has been generated at 2001 by Human Genome Project. However, since there are a huge amount of data, it is still a big challenge for people to understand them, and extract useful structure and function information, such as the function of genes, the structure of proteins encoded by gene, and the function of proteins. Understanding these information is crucial for us to improve longevity and quality of life, and has a lot of applications, such as genomic medicine, drug design, and etc. In the meantime, machine learning techniques are growing rapidly and are good at processing large datasets, but many of them are limited for the impact on larger real world problems.

In this thesis, three major contributions are described. First of all, we generate gene-gene interaction network from human genome conformation data by Hi-C technique, and the relationship of gene function and gene-gene interaction has been discovered. Second, we introduce a novel framework SMISS, which uses new source of information from gene-gene interaction network and uses a new way to integrate difference sources of information for protein function prediction. Finally, we introduce a tool called DeepQA which use machine learning technique to evaluate how well is the predicted protein structure, and a method MULTICOM for protein structure prediction. All of these protein structure and function prediction methods are available as software and web servers which are freely available to the scientific communities.