若手研究者インターナショナル・トレーニング・プログラム(ITP)

バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム 応募書類

Name: Peiying Ruan

Title: Prediction of heterodimeric protein complexes using protein-protein interaction networks

Institute: Bioinformatics Center (Akutsu Laboratory), Kyoto University Institute for Chemical Research; Kyoto, Uji, Gokasho 611-0011, Japan

Partner institute: Centre for Computational Biology, Mines ParisTech

Duration : September 4, 2013 – November 24, 2013

Plan :

Background

Since many proteins express their functional activity by interacting with other proteins and forming protein complexes, it is very useful to identify sets of proteins that form complexes. For that purpose, many prediction methods for protein complexes from protein-protein interactions have been developed such as MCL, MCODE, RNSC, PCP, RRW, and NWE. These methods have dealt with only complexes with size of more than three because the methods are often based on some density of subgraphs. However, heterodimeric protein complexes that consist of two distinct proteins occupy a large part according to several comprehensive databases of known complexes.

In our previous research, we propose a method with several feature space mappings and domain composition kernel to predict heterodimeric protein complexes. The results suggest that our proposed kernel considerably outperforms the naive Bayes-based method, which is the best existing method for predicting heterodimeric protein complexes. Although this work has been published in PLoS ONE, we need to further improve prediction accuracy in the future.

Research objectives

Dr. Jean-Philippe Vert, whose laboratory I'm going to study in, is an expert in machine learning. Researches in his laboratory mainly focus on the development of statistical and machine learning methods for computational biology. I believe this three-month's visiting would provide me a chance to learn a lot from him and propose a statistical method to improve the accuracy for predicting heterodimeric protein complexes.

At the Mines ParisTech, I will collaborate with Dr. Jean-Philippe Vert on the above objectives by the following steps:

(1) Try other kernels: Domain composition kernel that we proposed is a binary kernel related to domain composition. We wonder if some existing pairwise kernels may perform better by comparing domain composition of each two complexes, such as tensor product pairwise kernel (TPPK) and metric learning pairwise kernel (MLPK).

(2) Add other information: In the previous work, the features we designed mostly based on the interactions between protein pairs. However, much other information such as sequence, expression,

localization and phylogenetic profile, is not considered yet. Therefore, we plan to collect these data from databases and employ proper kernel function for each.

(3)Combine multiple kernels: Construct a kernel model to combine these kernels, with optimal weighting coefficients for each base kernel.

(4)Design a new kernel: We have proposed domain composition kernel, which improved the prediction accuracy a lot. However, domain composition kernel is a simple binary kernel. Hence, if possible, we will try to design a new kernel with more information.

In addition to researches, I am looking forward to enjoying the life in Paris in the coming three months. The literary culture of France has been attracting me a lot since I was a child. In my spare time, I plan to visit the famous architectures and museums. Also, I like beautiful French small towns very much. I really appreciate this rare opportunity to experience the local life in another country for months.