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

Name: Timothy Hancock

Title: Machine Learning Approaches for Computation Systems Biology

Institute: 京都大学化学研究所バイオインフォマティクスセンター

Partner Institute of your choice: Theoretical Biophysics Laboratory, Institute of Biology, Humboldt-Universität zu Berlin

Duration of your choice: 21st October 2011 to 8th January 2012

Report:

The Theoretical Biophysics Laboratory is located within Humboldt-Universität in the center of Berlin. The group of Professor Edda Klipp is quite large with approximately 30 regular members (6 staff, 10 post-doctoral fellows, ~ 16-20 students) and many visitors at all times of the year. Within the laboratory each person is attached to one or many projects with its own source of funding. The projects objectives range from the experimental “wet lab” biological approaches, through to combined experimental and computational methodologies and to finally pure theoretical and computational systems biology model creation and validation.

Group photo of everyone at the Theoretical Biophysics Laboratory in November 2011.

The laboratory has many events during the week beginning with the weekly progress meeting where everybody over all projects provides a quick five minute summary of their progress. Following, every week one member of laboratory is chosen a one hour seminar providing a more detailed progress of their's and the projects progress. In conjunction with this laboratory specific events, there are separate discipline specific educational interest groups, such as “Computational Systems Biology”, “Information Theory in Systems Biology” and “Molecular Systems Biology”, which hold regular weekly seminars where guest speakers present a tutorial style review of their current research. Through the attendance of these weekly events I was able to learn the research objectives currently being studied in Professor Klipp's laboratory and the processes which they follow to meet these objections.

During my stay I presented an overview of my research twice two different laboratories both related to the ITP project. Firstly to Professor Klipp's laboratory in the first week of my arrival. In this presentation I reviewed the topics related to my pathway mining through metabolic networks. This project constitutes the major part of my research, of which the most recent contribution was in part inspired by papers published with Professor Klipp's laboratory by Dr. Thomas Handorf (Basler et al. 2008, Christian 2009). From my presentation I received many useful comments from Dr. Handorf and others. The comments focused on the applicability of my proposed pathway mining approach to their specific networks and experimental purposes. These comments raised some interesting avenues for

future research in extending my current approach to more accurately capture the complexities inherent within biological networks, such as the action of protein complex, phosphorylation processes and inhibition reactions.

1. Basler G., Nikoloski, Z., Ebenhöh, O. & Handorf, T., "*Biosynthetic potentials from species-specific metabolic networks.*" (2008), Genome Informatics 20, 135-48.
2. Christian, N., May, P., Kempa, S., Handorf, T. & Ebenhöh, O. "An integrative approach towards completing genome-scale metabolic networks." (2009), Molecular BioSystems, 5, 1889-1903.

The second presentation I gave to the laboratory of Professor Ernst-Walter Knapp in the Freie Universität of Berlin. The focus of Professor Knapp's laboratory is more in the informatics direction, specifically chemoinformatics, than Professor Klipp's laboratory and as such I presented topics related to constructing network classification models. The goal of this work is to use a combination of a known network structure and experimental observations in order to identify network features which strongly relate to specific experimental conditions. From my presentation I was able have an informative discussion with Professor Knapp and his students on topics related constructing classifiers which directly encode the known structure of the biological or chemical problems. It was interesting to observe that many of the same problems we faced in our network classification project were also encountered within Professor Knapp's laboratory, although different solutions were employed to overcome these problems.

My research during my stay in Professor Klipps laboratory focused on two related application to Gaussian Process (GP) latent factor models for microarray data. The first model is for multi-task GP regression and the second was for latent protein profile estimation. I selected these two models after discussions with members of Professor Klipp's laboratory and realizing the need for structured predictive models for unobservable network features. Whilst in Berlin I derived and implemented these two approaches and applied them in a biological setting. During this process I discovered key limitations to these benchmark methods which have to be overcome before they are applicable to large scale systems biological modeling. These are two major problems which I quickly describe here:

1. **Convergence**: both algorithms employ a simple gradient decent optimization strategy and as such often gets caught in sub-optimal solutions.
2. **Multi co-linearity**: there is likely to large amounts of highly correlated variables in biological networks. In fact these are the targets of our latent factor approach. However when the correlation get too strong the methods fail.

There are well known methods within the machine learning literature to overcome these problems. However this requires customization of these algorithms which in turn offers an interesting avenue for future research. In addition to these more technical modifications, I discussed with members of Professor Klipp's Laboratory ways to better represent the biology within these models. These discussions centered on current methods of modeling known biophysical properties within differential equation models. During my stay I read several papers on these topics with an overall research objective to encoding enzyme kinetic equations into GP kernels which I can then include in GP regression and latent factor models.

On the non-research side Berlin is a famous old European city with a very interesting history. Additionally during my stay the annual Christmas markets were held which are a German tradition and allowed me to try traditional German food. In the photographs below there is a selection of some of my favorite Berlin tourist attractions.

The Brandenburg Gate

The road to Schloss Charlottenburg in fall.

Lions guarding the Babylonian Ishtar Gate in the Pergamon museum.

Ice skating rink at the Alexanderplatz Christmas markets during the night.

Alexanderplatz Christmas markets during the day.

The last remaining stretch of the Berlin Wall.