

Research plan for 2011 Fall semester

The research work for my first semester is divided into four parts.

(1) MS of Edin data (one month time 9/11-10/9)

This part is more about the project. Based on Chen's previous work, the program package is well developed. Two important things should be done. One is to subtract the sub-network from the results of MS. The other is using NetSVM to test it. This part may need one month, from learning to testing. To better understand the whole process, first, I need one week time to do some preparation, mainly about Chen's dissertation. After this, I could get to know more details and avoid making mistakes when I run the program or make experiments on the real data.

(2) Reading papers and books.(the whole semester)

I think I have collected enough books from Dr. Xuan, and Dr. Rebecca.

The Elements of Statistical Learning: Data Mining, Inference, and Prediction.

Pattern Classification (2000) - Duda & Hart.

W. J. Ewens and G. Grant, **Statistical Methods in Bioinformatics: An Introduction.**

N. C. Jones and P. A. Pevzner, **An Introduction to Bioinformatics Algorithms.**

Lehninger Principles of Biochemistry, 5th edition [David L. Nelson and Michael M. Cox]

Hallmarks of Cancer, Hanahan and Weinberg, Cell. 2011 Mar 4;144(5):646-74.

Review

These materials combined with the topics and papers related to our projects would help me build up a solid base for the long research career. Although it is impossible to finish all, I will choose the parts which have intense relationship with our projects, like significant analysis, bootstrap, cross validation, and SVM. And the materials from Rebecca play an important role in my start. Engineering analysis is just a tool, and I need to get more about the biology background information. About this part, I will keep in touch with Dr. Rebecca from Georgetown University. She could help me understand the biology background of our work much better. This reading may also include some papers.

(3) Prediction work on breast cancer research using NetSVM. (9/4-9/18)

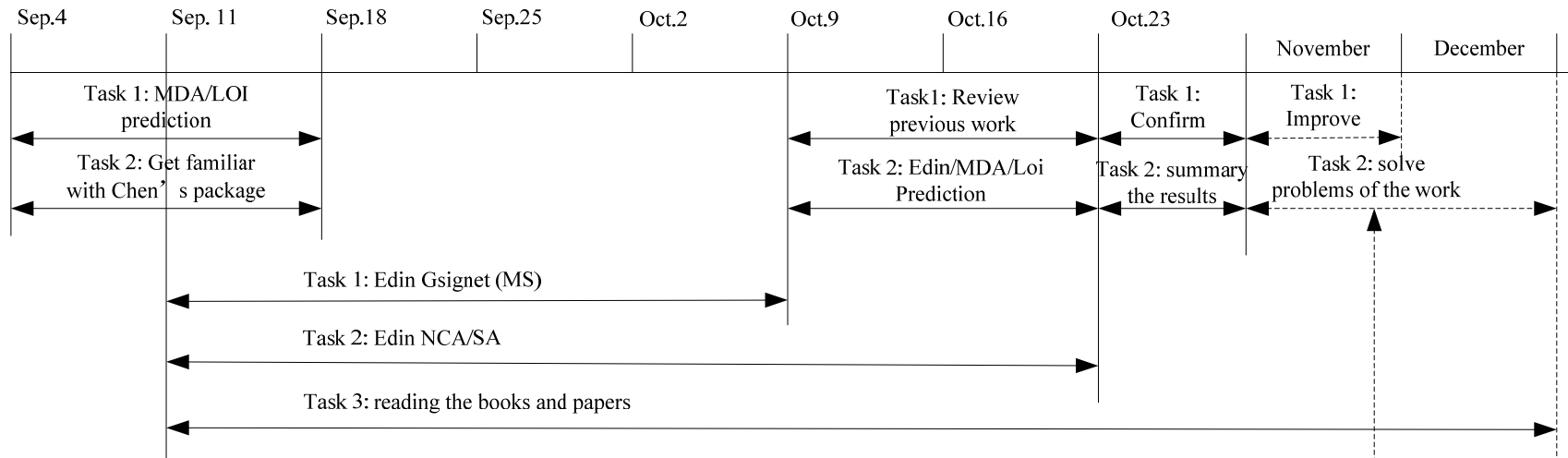
I have got a basic idea about NetSVM and the program package as well as the data works well now. So I plan to use the sub-networks from Chen to do more tests. Chen would prepare several networks for each data set (gene number ranges from 68 or 73 to around 200), so I have many choices to do independent test by selecting one sub-network from the train data set and then applying it to the test data set. This work is simple but would cost me much time.

Also, I need to learn more about the standards used to evaluate the work, like ROC, p-value, positive/negative prediction accuracy. The standards are very important, because I should know what I have done makes sense.

(4) Learning and simulation about NCA and SA (9/11-10/23)

This part serves to prepare me to pass QE, but not in hurry. I may need several weeks to learn about Chen's dissertation and get familiar with the program tools. Under the help of Chen, I will try to apply these tools to analyze our real breast cancer data, the in house data. Many problems would appear during this period, about one year. I wish to have good results and obtain some my own ideas about the research work. Now, I have read some papers in this field. Also, Chen has passed to me his package. So, in the following month, I will learn about this step by step. I wish to find something new and make some improvement on the methods. Then, I can publish a paper and make some preparation about my QE.

In the November and December, there are some holidays. It also comes to the end of the semester. All my work is about review and summary. Not only does our project need such a process, but also very important it is for my training in this new field. I can discuss about what I have done with our group members and collaborators. Only in this way, I can get to know what I need to improve in the following semester.



There would be many problems appearing in the projects and research, due to the limited time, some of which would not be solved fully. This period we have some holidays, so I could think about these problems and plan my next step