Project Requirements and Suggestions

Overview

Here are some suggestions for students who are considering doing a project.

As noted at the start of term it is expected that grad students in CS683 will do a project that involves the write up and presentation of a topic in structural bioinformatics. This may be based on one or more research papers, for example, a comparison of research results or a verification of some research result involving computations associated with protein structure.

General guidelines for content

As stated earlier, the topic must deal with 3D structure of biomolecules. For example:

- Proteins
- Proteins interacting with proteins
- Proteins interacting with a ligand (docking), or
- RNA.

It is expected that the topic will likely be related to some algorithm or procedure that serves some practical purpose related to the 3D structure or provides some insight or analysis about such a structure.

Note that such a study need not involve the entire protein. Scientists studying proteins will often concentrate most of their efforts on the binding pocket of the protein.

Project options

- Option A (literature survey):
  - Pick a problem that you consider to be interesting.
  - Search the literature for algorithms dealing with this problem.
  - Discuss the relative strengths of each approach.

- Option B (empirical evaluation):
  - Pick a problem that you consider to be interesting.
  - Implement and compare two or more algorithms to solve this problem.

Time lines

All projects are to be handed in on April 7, 2009 at the start of the last "extra" lecture (grads only).

Your project write up should be about 20 pages, double spaced, say 5000 words in length. Project presentations will be scheduled for an extra lecture to take place on April 7. These will be attended by graduate students but undergrads may attend if they wish to do so. All presentations should be 15 minutes in length followed by 5 minutes for questions.

By Feb. 26 at the latest, you should send me a one page description of the project so that I can offer suggestions and so that I can tell you whether or not the topic is acceptable. Students will be allowed to pick their presentation slot with a priority that depends on the submission of their proposal (so if you are first to submit a proposal then you get first pick of the presentation slots, second student get second pick, etc.).

Your project proposal should include the following:

- State the option that you have chosen.
- Clearly state the problem being studied with an emphasis on the structural aspects.
- What are the algorithms involved in this study?
- Be sure to include a set of references to papers that you will be using.

References and leads for project ideas

The following list is an assortment of web pages that may provide some starting points for you when you search for a topic:

1. Molecule of the month: http://www.pdb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/index.html. This is an excellent starting point for various studies in proteins. Goodsell usually relates function to structure and you can read the research papers associated with the protein to get more information. Recall that every protein in the PDB has at least one reference to a research paper dealing with that protein.

2. If you are interested in structure from a biophysics point of view here is a good starting point: http://blanco.biomol.uci.edu/WWWResources.html.

3. Look at: Journal of Computer-Aided Molecular Design Volume: 14, Issue: 3. **Computational methods for the structural alignment of molecules** Lemmen, Christian; Lengauer, Thomas pp. 215-232. This is a good survey of structural alignment.

4. Check out: Virtual Computational Chemistry Laboratory: http://146.107.217.178/ (There are various leads for software that works with structure, etc.)

5. Drug design & QSAR:
   Software (some of it free for academics) is listed at:
   http://www.moleculardescriptors.eu/softwares/softwares.htm.
   Drug discovery: http://www.netsci.org/Science/Combichem/feature09.html
   CoMFA: http://www.wiley.co.uk/ecc/samples/sample05.pdf
   Combinatorial Informatics: **D. K. Agrafiotis***, V. S. Lobanov, and F. R. Salemme, "Combinatorial informatics in the post-genomics era, *Nature Rev. Drug Discov.*, **2002**, *1*, 337-346.

6. Compare DOCK: http://mdi.ucsf.edu/DOCK_availability.html with some other docking software. Use ATP in 1kay, for example, as a test case.

7. RNA secondary structure:
   http://cmgm.stanford.edu/biochem218/Projects%20Spring%202003/Silverman.pdf

8. Anti-HIV drugs: http://www.mdpi.org/ijms/papers/i5020048.pdf
   (includes neural net and QSAR analysis).

9. If you want to explore techniques for the prediction of protein secondary structure, here are some starting points and various overviews:
   a. Prediction of Protein Secondary Structure. Arjunan, Deris, Illias. *Jurnal Teknologi* 35(C) Dec. 2001 pp 81-90.
   b. Secondary Structure Prediction with Support Vector Machines. Ward, McGuffin, Buxton, Jones. *Bioinformatics* Vol. 19, (13), 2003 pp. 1650-1653.
   c. Protein Secondary Structure Prediction Based on an Improved Support Vector Machine Approach. Kim, Park. Protein Engineering, Vol. 16 (8), 2003 pp. 553-560.
   d. A Novel Method for Protein Secondary Structure Prediction Using Dual-Layeer SVM and Profiles. Guo, Chen, Sun, Lin. *Proteins*; 54, 2004 pp. 738 – 743.
   e. A Novel Method of Protein Secondary Structure Prediction with High Segment Overlap Measure: Support Vector Machine Approach. Hua, Sun. *JMB*, 308, pp. 397-407.

10. Protein structural class determination:
    Protein Structural Class Determination Using Support Vector Machines.
    Isik, Yanikoglu, Sezerman.
    http://people.sabanciuniv.edu/~berrin/papers/fold-classification-iscis04.pdf.

The Contents of the Project Report

The following points are suggestions for the content of your report:

Option A: Literature survey

1. Introduction and motivation
    a. Clearly state the problem or topic.
    b. How is the problem or topic important?
2. Literature survey
   What are the techniques or algorithms used by various researchers? It will be important for you to organize this material in a way that makes it understandable and clear.
3. Critical evaluation
    a. How well do the techniques or algorithms achieve their objectives?
    b. What are the strengths and weaknesses of the various approaches?
    c. What problems or issues remain to be studied?
4. Conclusion
    a. What can we learn from these papers?
    b. Can you suggest any new avenues of investigation?
5. References

Option B: Empirical evaluation

1. Introduction and motivation
    a. Clearly state the problem or topic.
    b. How is the problem or topic important?
2. Algorithms to be studied
   Clearly describe the algorithms that you are comparing. Be sure to justify the choice of these algorithms. In other words: Why do they hold any promise of being successful?
3. Critical evaluation and comparison
    a. How do the algorithms compare?
    b. What are their strengths and weaknesses? The comparison should involve a performance evaluation, assessment of complexity, ease of use, and any other measures that you consider to be relevant.
4. Conclusion
    a. Considering your evaluation, what algorithm is the best?
    b. What can we learn from these algorithms?
    c. Can you suggest any new avenues of investigation?
5. References

Project Marks

Here is a project marking template. **Note**: I would highly recommend that students do their project using Microsoft Word. The grammar checker is very useful. Marks will be deducted for extensive bad grammar that is so bad that it could easily have been detected by Microsoft Word.

Technical Style of the report:

Grammar and Spelling:	_____ (/8)
Organization:	_____ (/6)
Clarity:	_____ (/6)

Content:

Introduction and motivation:	_____ (/10)
Background material:	_____ (/10)
Critical evaluation:	_____ (/25)
Discussion/Conclusion:	_____ (/10)
References	_____ (/5)

Presentation style:

Grammar and Spelling:	_____ (/4)
Organization:	_____ (/4)
Clarity (speaking & pacing):	_____ (/4)
Design of slides (font size, pictures, examples):	_____ (/4)
Interaction with audience: (enthusiasm, handling questions)	_____ (/4)

Total: _____ (/100)