CSCI2950-C: Topics in Computational Biology: Fall 2011

Meeting time: Mondays 3-5:20pm CIT 241
Instructor: Ben Raphael, CIT 505, braphael@brown.edu
Website: http://cs.brown.edu/courses/csci2950-c/

Description

The past decade has witnessed tremendous technological improvements in measuring the fundamental molecules of life: DNA, RNA, and protein. In particular, *next-generation DNA sequencing technologies* are producing exponentially increasing volumes of DNA sequence data. Interpreting this data presents numerous computational challenges requiring algorithmic, statistical, and machine learning techniques. This course will examine several problems arising in the sequencing of human and cancer genomes including: genome assembly and variant detection; identification of functional mutations/variants from multiple genome sequences; discovery of combinations of mutations/variants that influence an observed trait/phenotype.

The course will be organized in seminar style where students will read and present recent papers on the topics listed below. Each topic will be introduced with background lectures. Students will undertake a project to further study one of the topics. To the extent possible, projects will be adjusted to the background/interest of the student and could range from theoretical (e.g. designing a new algorithm and proving its correctness), to the practical (a software implementation). The project will include a written proposal, midterm report, and final presentation.

Prerequisites

- Undergraduate-level knowledge of probability: random variables, distributions, etc.
- Undergraduate-level knowledge of algorithms and/or statistics

No biology background is assumed. Necessary background will be introduced in lectures and reading.

Grading

Papers: 45% total
- Reviews 20%
- Presentations 15%
- Discussion Summaries 10%

Project: 45% total
- 10% Written proposal.
- 10% Written midterm report.
- 15% Final written report.
- 10% Final presentation.

Computer Assignment: 5%
Participation: 5%

Papers, Presentation, and Discussion

Presentations: Each student will present at least two papers from the reading list. The exact number will depend on class size. Presentations should be designed to facilitate a class discussion of the paper, and may use slides, whiteboard, or a combination of both. A long presentation is expected to be about 40 minutes, and a short presentation 20 minutes with the

remaining class time for questions and discussion. Students are *strongly* encouraged to discuss presentation with instructor beforehand.

Reviews: Each student (except presenter) will submit a written review for each discussion paper, with the exception of those that you present. Reviews will be submitted online (details to come) by **Monday 10am** of day that the paper is presented. Reviews will be graded on a 3-point scale: 2 = mastered the key issues introduced in the paper 1 = read the paper and understood the basics; 0 otherwise.

Discussion Summary: One student will be the appointed scribe for each presentation. Following the paper discussion, the presenter and scribe will prepare a written (document or slide) summary of the merits and deficiencies of the paper. This summary should incorporate material from the presentation (e.g. slides), the class paper reviews, and the discussion. The summary will be posted on the course webpage.

Project

The project is a semester-long effort to further study one of the class topics. The project could range from theoretical (e.g. designing a new algorithm and proving its correctness), to the practical (a software implementation) depending on the interest of the student. Projects can be undertaken in groups of 1-3 students. Students are encouraged to propose a project matching their research interests and discuss with instructor. A list of suggested projects will also be distributed.

Participation

All students are expected to contribute to paper discussions by asking questions, making observations, identifying strengths and weaknesses of the approaches under discussion. Critical discussion and analysis are key prerequisites for research.

Course Credits

PhD: Area B (Algorithms)
ScM: "Theory" or "Practice"* course.
Significant Programming*
*With appropriate class project.