Data analytics for personalized genomics and precision medicine: Course introduction

Yu LI (李煜)

iyu@cse.cuhk.edu.hk

Department of Computer Science and Engineering (CSE)

The Chinese University of Hong Kong (CUHK)

ESTR3605

ESTR3605 is a mirror course of BMEG3105, it will have an 45-minutes additional lecture per week.

Course Logistics

1. Lectures and tutorials

The lecture would be held on Wednesday morning from 9:30 a.m. to 11:15 a.m. at Science Center L4 and Friday morning from 9:30 a.m. to 01:15 a.m at Mong Man Wai Building room 703. The tutorial section would be held right after Friday lecture from 10:30 a.m. to 11:15 a.m.

2. Learning materials

Lecture slides would be posted in the link https://yu979.github.io/BMEG3105-Fall-2024/ the day before the lecture. Video recordings would not be available after the lectures unless reasonable reasons and evidences could be provided.

3. Grading

According to the majority of students' opinion, BMEG3105 are going to be graded by homework (20%), scribing (10%), in-class quiz (10%), midterm (20%), project (20%) and final (20%).

All quizzes, midterm and exam are going to be open book format.

A. Homework

There would be 3 grading homework and 1 non-grading homework, each of them would be 5%. The non-grading homework would be a programming assignment.

B. Scribing

Summarize one lecture and submit it one week after the lesson.

C. In-class quiz

There would be 2 quizzes, which would be held on the 18th October, 2024 and 27th October, 2024.

D. Midterm

There would be one bonus in the midterm exam, which would be an extra 2% for the grading.

E. Project

The project includes a mid-term report (5%), final report (7%) , presentation (3%) and the implementation (5%).)

What are data analytics?

In this modern generation, there are massive of different types data is being collected and warehoused every single day. There are web data, which are from Facebook, Google, Amazon or Twitter. From the study of biological field, there are biological data, such as DNA sequence and protein structure. In the economical field, there are bank or credit card transaction data, such as Alipay and Paypal. Last but not least, from our daily life, there mobile data, such as China Mobile and CSL. Those requires data analytics to keep on running their work. By time to time, computers become cheaper and more powerful that can finish more tasks and make things become more intelligent. It is useful for data analysis, especially for the aggregation of data, the generation of hypothesis, and the

supporting of the conclusion.

There are tons of data related to our DNA sequencing and health can be analyzed that allow us to learn more in personalized geonomics and precision medicine.

What data can we collect and analyze to measure a person?

- Gene and mutations
- Gene expression (Transcriptome)
- Proteome
- Metabolome
- Molecular network & Cellular network
- Microbiome (Oral and gut)
- Organ (Biomedical imaging)
- Hospital test (Blood test)
- Electrocardiography (ECG)
- Demographic information (Age, gender, location)
- Drug history and disease history
- Personal statement and doctor diagnosis
- Living habit (Exercise)
- Diet
- Family history
- Communications and social media data
- Environment (Pollution
- Travelling history (Global pandemic)

Conclusion

By this course, we are going to learn about ht fundamental concept of data analytics and the various data in genomics and medicine. Last but not least, at the end of this course, we would be able to apply the data analytics techniques to process the data and resolve the problems in biology.