

Algoritmos em Bioinformática

Projetos I e II

Semestre de verão 2023/2024

Objetivos de Aprendizagem

As aulas desta unidade curricular concentram-se no lado algorítmico da bioinformática: modelar problemas biológicos computacionalmente e, em seguida, formular algoritmos para resolver os problemas, às vezes sob certas suposições. O objetivo destes dois projetos é ter a oportunidade de explorar um desafio prático da vossa escolha, executando uma análise computacional em um conjunto de dados biológicos reais e, em seguida, interpretando os resultados.

Dentro deste objetivo, existem duas direções principais. Podem optar por implementar um algoritmo que resolva um problema biológico e, em seguida, aplicar este algoritmo para analisar um conjunto de dados, ou podem utilizar software existente para analisar o conjunto de dados. Mediante o caminho escolhido (mais/menos implementação) a análise dos resultados poderá ser (menor/maior).

Neste semestre, em ambos os projetos os grupos deverão ser de 2 estudantes.

Projeto I

Terão de elaborar e submeter via moodle dois entregáveis:

Entregável 1: Proposta de Projeto

Devem primeiro encontrar um problema biológico que gostariam de abordar. Há uma espécie em particular que vos interessa? Uma doença humana que gostariam de investigar? Algo da aula que já vos despertou o interesse?

Devem escrever uma proposta de projeto de no máximo 3 páginas. Esta proposta deve:

- clarificar o problema que propõem abordar;
- explicar por que o projeto proposto é interessante científica e computacionalmente;
- discutir brevemente a vossa abordagem para alcançar este objetivo e por que acham que isso é viável;
- identificar quaisquer dados ou recursos externos que usarão para o projeto;
- indicar as tarefas desta proposta e como cada membro do grupo irá contribuir individualmente.

Poderá depois existir alguma mudança durante o decorrer do segundo projeto em relação ao primeiro projeto, mas o primeiro projeto deve tentar planear o melhor possível.

Entregável 2: Apresentação do Projeto

Uma apresentação curta (de 15 a 20 minutos) na sala de aula (e simultaneamente submetida via moodle) explicando qual é a proposta do vosso projeto. Devem explicar cada proposta para uma audiência que não está familiarizada com o vosso projeto.

Projeto II

Entregável 1: Relatório + Documento individual

Um relatório (de 8 a 10 páginas) explicando o contexto do problema e os algoritmos necessários para abordar o problema. O texto deve ser claramente escrito, autocontido e conter citações para trabalhos relevantes. Todos os algoritmos que implementarem ou ferramentas que utilizarem devem ser explicadas. No entanto, o código desenvolvido deverá estar disponível, por exemplo num repositório, de modo a ser possível consultá-lo. O relatório também deve incluir uma análise da aplicação da técnica computacional a um conjunto de dados reais. Se alguma dificuldade surgir que faça com que o projeto precise ser adaptado, também deve ser mencionada. A parte da nota atribuída à componente escrita será parcialmente avaliada com base na qualidade da sua exposição e parcialmente com base na qualidade do seu trabalho. Como parte da entrega, cada membro do grupo deverá entregar individualmente um documento de uma página onde está descrito o que cada membro do grupo fez durante o projeto.

Entregável 2: Apresentação do Projeto

Uma apresentação curta (de 15 a 20 minutos) na sala de aula, explicando o que o grupo fez e demonstrando os resultados. Devem explicar o vosso trabalho para uma audiência que não está familiarizada com o vosso projeto e demonstrar os seus resultados.

Datas importantes

- **11/03/2024** Publicação deste documento.
- **24/03/2024** Entrega da proposta de projeto I (entregável 1 do Projeto I)
- **25/03/2024** Apresentação e entrega da apresentação do projeto I (entregável 2 do Projeto I)
- **30/05/2024:** Entrega do relatório do projeto II e documento individual (entregável 1 do Projeto II)
- **03/06/2024** Apresentação e entrega da apresentação do projeto II (entregável 2 do Projeto II)

Linhas Orientadoras

Algumas diretrizes que deverão ser tidas em conta:

- Existe uma introdução escrita que explica o problema ?
- Todas as figuras são claramente explicadas através de legendas e referenciadas adequadamente no texto principal?

- A proposta, relatório e apresentações estão estruturados logicamente, com uma fluidez clara?
- Estão articulados os objetivos e os resultados?
- Os problemas computacionais abordados no projeto são formulados de forma clara?
- Os principais algoritmos utilizados são explicados de modo compreensível?
- Existe uma secção de resultados?

Exemplos de projetos

O contexto dos projetos é deliberadamente aberto; os grupos podem escolher temas da bioinformática que considerem interessantes. Além disso, é perfeitamente razoável começar o projeto examinando pesquisas existentes. Na verdade, em alguns casos, a replicação dos resultados de um artigo existente pode ser significativamente desafiadora para constituir um projeto. No entanto, sugere-se alguns tópicos para os projetos a desenvolver:

- Compare a conjunto de programas de software para o *assembly* genomas em diferentes tipos de organismos e conjuntos de *reads*;
- Construir árvores evolucionárias para alinhamentos múltiplos ou dados de tipagem;
- Classificar um conjunto de imagens celulares ou médicas usando aprendizagem profunda com TensorFlow ou PyTorch;
- Construir uma rede de fatores de transcrição conectando fatores de transcrição aos genes que regulam e inferir quais conclusões biológicas podem ser obtidas das propriedades do grafo resultante (ex:<http://www.yeastract.com/>)

Ferramentas Bioinformáticas

- *De novo* assemblers de genoma: https://en.wikipedia.org/wiki/De_novo_sequence_assemblers
- Ferramentas de alinhamento de sequências (pairwise/multiple alignments, profile HMMs, read mapping, metagenomics aligners, etc.): https://en.wikipedia.org/wiki/List_of_sequence_alignment_software
- Ferramentas de predição de genes: https://en.wikipedia.org/wiki/List_of_gene_prediction_software
- Ferramentas de filogenia: https://en.wikipedia.org/wiki/List_of_phylogenetics_software
- Ferramentas de sequenciação de RNA: https://en.wikipedia.org/wiki/List_of_RNA-Seq_bioinformatics_tools