SORS: Genomic analysis pipeline: overview, challenges, and proposed solutions

Objectives

To download the prentation, please click here.

Abstract: In this talk we will give an overview of the genomic analysis pipeline, from data generation to its analysis. In doing so, we will identify the main challenges arising in the genomic setting. These include dealing with errors introduced during the sequencing process, designing state-of-the-art specialized compressors to deal with the ever growing amount of genomic data being generated, as well as improving the accuracy of the current tools used for the analysis.

We will emphasize on some of the effort being carried out by the international community to design a standard under the International Standardization Organization (ISO), denoted MPEG-G, for genomic information representation. We will also introduce a new filtering tool intended to improve the accuracy of variant calling, the last step of the genomic analysis pipeline whose output is generally the starting point for analysis in the personalized medicine paradigm. We will conclude the talk with some thoughts of where the community is going and the challenges that we will face in the near future.

Short bio:

Before joining UIUC, Prof. Ochoa obtained a Ph.D. from the Electrical Engineering Department at Stanford University, in 2016. She received her M.Sc. from the same department in 2012. Prior to Stanford, she graduated with B.Sc. and M.Sc. degrees in Telecommunication Engineering (Electrical Engineering) from the University of Navarra, Spain, in 2009. During her time at Stanford she conducted internships at Google and Genapsys, and served as a technical consultant for the HBO's TV show "Silicon Valley". Prof. Ochoa's main interests lie in the field of bioinformatics and computational genomics, and she uses a multidisciplinary approach that combines tools from information theory, signal processing, and machine learning, among others.
Her main contributions include the design of several lossless and lossy compression schemes tailored to raw and aligned genomic data, as well as denoising schemes to reduce the noise present in such data. She has also developed compression schemes for other types of omics data, as well as schemes to perform similarity queries on compressed databases without the need of decompression. Finally, she has developed new methods for the discovery of gene networks specific to different cancer types.
Prof. Ochoa is also part of the group of experts who is developing, under the International Standardization Organization (ISO), the new MPEG-G standard for genomic information representation. She is also part of the Center for Science of Information, an NSF Science and Technology Center, and she is the recipient of several US-based grants.



Speakers

Idoia Ochoa, Assistant Professor University of Illinois at Urbana-Champaign (UIUC) IL, USA

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

Source URL (retrieved on *29 Mar 2024 - 11:24*): https://www.bsc.es/es/research-and-development/research-seminars/sors-genomic-analysis-pipeline-overview-challenges-and-proposed-solutions