



Mathematics and CS Seminar

Data science of biology

Katarzyna Bozek

OIST

Host: Gasper Tkacik

Before large datasets became pervasive in all domains, life sciences and biology among them were generating copious and complex data. In my talk I will present through a range of example problems, the evolution and growth of data analysis in this field. HIV-host interaction, human evolution, or novel ways of quantifying natural behavior—the broad spectrum of applications that I will describe is tied by a common challenge of finding patterns and meaning in large and complex data. Currently, new machine learning methods to extract information from images have opened vast new opportunities to data analysis in biology. Traditionally, categories or states of objects in bioimages are determined based on human-defined relevant elements and the recent advances in deep learning have allowed to largely automate this task by providing adaptable and precise methods for image segmentation, object recognition, and classification. However, what if the human eye is not capable of perceiving potentially meaningful visual features, or when the observed objects are inherently complex? In my talk I will demonstrate through my most recent work on marker-less honeybee tracking and quantifying behavior of *C. elegans* how the new machine learning methods for image analysis are capable of quantifying patterns defying human vision. The methods I will present allow to distinguish apparently identical bees, quantify worm posture, predict temporal patterns in movement and behavior. The new ways of extracting signals from visual data, offer an unprecedented opportunity to transform image data into another type of large and complex data type that is intractable to human perception. Computational analysis can bring insights into biological questions, while challenging datasets in biology can give rise to innovative computational solutions. Biology is a source of rich and complex datasets which serve not only answering important scientific questions but also offer opportunities for creation of innovative solutions for data analysis. Similar to how large sequence datasets gave rise to efficient sequence alignment algorithms that in turn allowed for quantitative sequence analysis at an unprecedented scale, currently, biological images and new machine learning methods for their analysis offer an opportunity for another quantitative leap in biological data analysis.

Tuesday, February 12, 2019 10:00am - 11:00am

Mondi Seminar Room 2, Central Building



This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station.
Please find a schedule of the ISTA Shuttle on our webpage:
<https://ista.ac.at/en/campus/how-to-get-here/> The ISTA Shuttle bus is marked ISTA Shuttle
(#142) and has the Institute Logo printed on the side.