Seminars in Biotechnology BTEC 591 & BTEC 691

"Machine Learning for Molecules and Diseases"

Thursday, December 15, 2022
13.30
GTU Institute of Biotechnology, Lecture Hall

Dr. Öznur TAŞTAN

Faculty of Engineering and Natural Sciences, Sabancı University

Dr. Öznur Taştan holds a BSc in Biological Sciences and Bioengineering from Sabanci University and received her MSc in 2007 and her Ph.D. in 2011, from Carnegie Mellon University, School of Computer Science. Since 2018, she has been with the Sabanci University Computer Science and Engineering and Molecular Biology, Genetics and Bioengineering departments. Before joining Sabanci, she worked as a faculty member at Bilkent University, Department of Computer Engineering, and as a post-doctoral researcher at Microsoft Research New England Lab (Cambridge, MA, USA). She has worked on diverse problems in machine learning for computational biology. She is a recipient of the Young Scientist Research Award of the Science Academy (BAGEP), the UNESCO-L'OREAL National Fellowship for Young Women in Life Sciences, and METU Prof. Mustafa Parlar Foundation Research Incentive Award.

Machine learning has become an essential tool in computational biology, providing the ability to analyze and interpret large biological datasets. In this talk, I will give examples of methods developed in our group at the intersection of machine learning and computational biology. In the first part, I will focus on cancer subtype identification and introduce PAMOGK (Pathway based Multi Omic Graph Kernel clustering), a method that integrates multi-omics patient data with existing biological knowledge on pathways to classify patients into molecular subgroups. Accurate classification of patients into molecular subgroups is critical for the development of effective therapeutics and for deciphering what drives these different subgroups to cancer. The availability of multi-omics data catalogs for large cohorts of cancer patients provides multiple views into the molecular biology of the tumors with unprecedented resolution. We design a graph kernel that evaluates patient similarities based on a single molecular alteration type in the context of a pathway. To corroborate multiple views of patients evaluated by hundreds of pathways and molecular alteration combinations, we use multi-view kernel clustering. We will show that PAMOGK outperforms other state-of-the-art methods in terms of its ability to partition cancer patients into groups with different survival distributions. The discovered patient subgroups also differ with respect to other clinical parameters such as tumor stage and grade and primary tumor and metastasis tumor spreads. The pathways identified as important are highly relevant to KIRC. In the second part of the talk, I will present machine learning-based projects applied to different biological problems.