Mechanical Engineering Dissertation Defense

"Topological Machine Learning for High-Dimensional Data Analysis"

School for Engineering of Matter, Transport and Energy

Nan Xu

Advisor: Yongming Liu

Abstract

In an age where Big Data dominates across various sectors, the complexity and sheer volume of datasets present a formidable challenge, especially in domains employing advanced high-resolution imaging and extensive sensor networks. This proliferation of high-dimensional data significantly complicates the task of extracting meaningful insights, as the vast amounts of information often mask the valuable insights within, erecting substantial barriers to effective analysis. Traditional data analysis methodologies, which primarily depend on the manual selection of features, are increasingly seen as insufficient. This insufficiency stems from their reliance on the subjective expertise of practitioners, a reliance that often leads to substantial information loss and a reduced capability to adapt across diverse environmental conditions.

This dissertation presents a pioneering approach that synthesizes topological machine learning (TML) with topological data analysis (TDA), aimed at advancing feature extraction within high-dimensional datasets. Leveraging the capabilities of TDA, this strategy uncovers the intrinsic topological structures hidden within data, facilitating a smooth transition to a more manageable low-dimensional representation through the process of manifold learning. This crucial transformation permits the employment of conventional statistical methods, such as Gaussian Processes (GP), within this compressed space, thus effectively countering the curse of dimensionality that hampers many modern analytical endeavors.

The TML methodology distinguishes itself by its proficiency in accurately capturing and representing the complex intra-cluster and inter-cluster correlations found within these lower-dimensional embeddings. This ability not only aids in the classification and visualization of the data but also guarantees the retention of essential information imperative for the identification of varied operational conditions during the feature extraction phase. The method's application and subsequent validation across a wide array of datasets have highlighted its efficacy, consistently demonstrating unparalleled accuracy in classifying an extensive range of scenarios.

Furthermore, this work elucidates the transformative capacity of TML to address and surmount the enduring obstacles related to manual feature selection and data representation. By forging a connection between the convoluted, often disordered realm of high-dimensional data and the practical necessities of analysis and interpretation, the TML methodology signifies a notable progression in the field of data science. Through detailed investigation and empirical substantiation, this dissertation not only validates the theoretical value of the TML approach but also establishes its practical utility across a multitude of domains. This represents a groundbreaking solution to the intricate challenges posed by high-dimensional data analysis in today's digital age, promising a new frontier in the effective management and utilization of Big Data.

April 12, 2024; 11:00 AM; ECG G214

Zoom Link:

https://asu.zoom.us/j/87249560933?pwd=WHhHZ3pFVGgxZU9nU3VUZHBOd2s3UT09

Arizona State University

Copyright © 2021 Arizona Board of Regents