



Department of Computer Science

Local Intrinsic Dimensionality and Machine Learning:
The Dynamical Systems Perspective

Michael Houle
Computer Science, NJIT

Hosted by Vincent Oria

Date: Monday, November 20, 2023

Coffee: 2:15 PM – 2:30 PM

Time: 2:30 PM – 4:00 PM

Location: GITC 4402 (4th floor Seminar Lecture Hall)

WebEx Link: <https://njit.webex.com/njit/j.php?MTID=m5be6b710996f265690c9c3ab9750e52d>

<http://cs.njit.edu/seminars>

Abstract:

Researchers have long considered the analysis of similarity applications in terms of the intrinsic dimensionality (ID) of the data. Although traditionally ID has been viewed as a characterization of the complexity of discrete datasets, more recently a local model of intrinsic dimensionality (LID) has been extended to the case of smooth growth functions in general, and distance distributions in particular, from its first principles in terms of similarity, features, and probability. Since then, LID has found applications — practical as well as theoretical — in such areas as similarity search, data mining, and deep learning. LID has also been shown to be equivalent under transformation to the well-established statistical framework of extreme value theory (EVT). In this presentation, we will survey some of the wider connections between ID and other forms of complexity analysis, including EVT, power-law distributions, chaos theory, and control theory, and show how LID can serve as a unifying framework for the understanding of these theories. Finally, we will reinterpret recent empirical findings in the area of deep learning in light of these connections.

Bio:

Michael Houle obtained his PhD degree in 1989 from McGill University in Canada, in the area of computational geometry. Since then, he developed research interests in algorithmics, data structures, and relational visualization, first at Kyushu University and the University of Tokyo in Japan, and from 1992 at the University of Newcastle and the University of Sydney in Australia. From 2001 to 2004, while at IBM Japan's Tokyo Research Laboratory, he first began working on approximate similarity search and shared-neighbor clustering methods for data mining applications. From 2004, at the National Institute of Informatics, Tokyo, his research interests expanded to include dimensionality and scalability in the context of fundamental AI / machine learning / data mining tasks such as search, clustering, classification, and outlier detection. In 2021, he relocated to Vancouver, BC, Canada. Currently he is with the New Jersey Institute of Technology in Newark, NJ, USA, and divides his time between Newark and Vancouver.