Department of Computer Science

Local Intrinsic Dimensionality: An Extreme-Value-Theoretic Foundation

Michael Houle, Principal Fellow
The University of Melbourne, Australia

Hosted by: Vincent Oria

Date: Wednesday, November 16, 2022
Coffee: **2:15** PM – 2:30 PM
Time: 2:30 PM – 3:30 PM
Location: GITC 4402 (4th floor Seminar Lecture Hall)
WebEx Link: **https://njit.webex.com/njit/j.php?MTID=m60a4163a9c56e3f4a3b65d4fac46fd07**

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. This presentation is concerned with a generalization of a discrete measure of ID, the expansion dimension, to the case of smooth functions in general, and distance distributions in particular. A local model of the ID of smooth functions is first proposed and then explained within the well-established statistical framework of extreme value theory (EVT). Moreover, it is shown that under appropriate smoothness conditions, the cumulative distribution function of a distance distribution can be completely characterized by an equivalent notion of data discriminability. As the local ID model makes no assumptions on the nature of the function (or distribution) other than continuous differentiability, its generality makes it ideally suited for the learning tasks that often arise in data mining, machine learning, and other AI applications that depend on the interplay of similarity measures and feature representations. The talk will conclude with a discussion of recent applications of local ID to deep learning.

Bio: Michael Houle obtained his PhD degree in Computer Science 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 as a research associate 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, he was a Visiting Scientist at IBM Japan's Tokyo Research Laboratory, where he first began working on approximate similarity search and shared-neighbor clustering methods for data mining applications. From 2004 to 2021, he was a Visiting Professor at the National Institute of Informatics (NII), Japan. During this time, 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. Currently, he is a Principal Fellow at the University of Melbourne, Australia.