SIS Colloquium Series →

UNIVERSITY OF PITTSBURGH
School of Information Sciences
Information Sciences Building, 135 N. Bellefield Ave.

Wednesday, September 2, 2015
IS Building, room 522 | 1 pm, Refreshments | 1:30 pm, Lecture

Jaakko Peltonen
Associate Professor of Statistics, School of Information Sciences
University of Tampere, Finland

An Information Retrieval Approach to Visualization of High-Dimensional Data

Nonlinear dimensionality reduction methods are often used to visualize high-dimensional data, but many proposed methods have been designed for other related tasks such as manifold learning. It has been difficult to assess the quality of visualizations since the task has not been well-defined. We give a rigorous definition for a specific visualization task resulting in quantifiable goodness measures and new visualization methods. The task is information retrieval given the visualization: to find similar data based on the similarities shown on the display. The fundamental tradeoff between precision and recall of information retrieval can then be quantified in visualizations as well. The resulting family of visualization methods, called NeRV (Neighbor Retrieval Visualizer), performs very well in unsupervised visualization tasks and has been extended in many ways to supervised visualization, parametric visualization, fast visualization scalable to big data, and interactive visualization, and has been incorporated as part of exploratory information seeking systems.

Jaakko Peltonen is an associate professor of statistics (data analysis) at the School of Information Sciences, University of Tampere, Finland. He is also currently an academy research fellow at Aalto University, Finland, where he is a PI of the Probabilistic Machine Learning research group. He is an associate editor of Neural Processing Letters and an editorial board member of Heliyon. He has served in organizing committees of seven international conferences, one international summer school, in program committees of 28 international conferences/workshops, and has performed referee duties for numerous international journals and conferences. He has 74 publications and has 730 citations so far (h-index 14). He is an expert in statistical machine learning methods for exploratory data analysis, visualization of data, and learning from multiple sources.