

Information Science and Technology Seminar Speaker Series



Joachim M. Buhmann
Institute for Machine Learning ETH Zurich

Information theory of algorithms: How precisely should we compute in Big Data settings?

Wednesday, January 13, 2016

3:00 - 4:00 PM

TA-3, Bldg. 1690, Room 102 (CNLS Conference Room)

Abstract: Algorithms map input spaces to output spaces where inputs are possibly affected by fluctuations. Beside run time and memory consumption, an algorithm might be characterized by its sensitivity to the signal in the input and its robustness to input fluctuations. The achievable precision of an algorithm, i.e., the attainable resolution in output space, is determined by its capability to extract predictive information in the input relative to its output. I will present an information theoretic framework for algorithm analysis where an algorithm is characterized as computational evolution of a (possibly contracting) posterior distribution over the output space. The tradeoff between precision and stability is controlled by an input sensitive generalization capacity (GC). GC measures how much the posteriors of two different problem instances generated by the same data source agree despite the noise in the input. Thereby, GC objectively ranks different algorithms for the same data processing task based on the bit rate of their respective capacities. Information theoretic algorithm selection is demonstrated for minimum spanning tree algorithms and for greedy MaxCut algorithms. The method can rank centroid based and spectral clustering methods, e.g. k-means, pairwise clustering, normalized cut, adaptive ratio cut and dominant set clustering. Applications are in cortex parcellations based on diffusion tensor imaging and for the analysis of gene expression data.

Biography: Joachim M. Buhmann heads the Institute for Machine Learning in the Computer Science Department at ETH Zurich. He has been a full professor for Information Science and Engineering since October 2003. Since 2014, he serves as vice rector for study Programs at ETH Zurich. He studied physics at the Technical University Munich and obtained his PhD in Theoretical Physics. As postdoc and research assistant professor, he spent 1988-92 at the University of Southern California, Los Angeles, and the Lawrence Livermore National Laboratory. He held a professorship for applied computer science at the University of Bonn, Germany from 1992 to 2003. His research interests spans the areas of pattern recognition and data analysis, including machine learning, statistical learning theory and information theory. Application areas of his research include image analysis, medical imaging, acoustic processing and bioinformatics. Since 2009, he serves as president of the German Pattern Recognition Society DAGM e.V..

For more information contact the technical host Stephan Eidenbenz, eydenben@lanl.gov, 667-3742.

Hosted by the Information Science and Technology Institute (ISTI)