VirginiaTech

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Computer Science Seminar Series

National Capital Region

Data Compression via Quantized Random Projections: A Statistical Perspective

Speaker: Dr. Martin Slawski
George Mason University
Friday, October 6, 2017
1:00PM- 2:00PM, NVC T3

Abstract

The methods of Random Projections (RPs) has found numerous applications in modern data processing and analysis ranging from compressed sensing over numerical linear algebra, machine learning, and similarity search to data privacy. In recent years, the aspect of data compression going along with RPs has been developed further in terms of subsequent scalar quantization. This gives rise to a trade-off between the bit depth of the quantizer and the dimension of the reduced data representation when given a total budget of bits to be used for data transmission or storage. In this talk, we present a statistical analysis of this trade-off in one-scan compressed sensing and in similarity search with respect to cosine similarity. It is shown that for a wide range of scenarios, coarse quantization (five bits or less) causes only a small loss in statistical accuracy. Depending on the situation, it is even well possible that the aforementioned trade-off is optimized in the extreme case single-bit quantization.

Biography

The speaker is an Assistant Professor in the Department of Statistics at George Mason University. Prior to that, he was a postdoctoral associate at Rutgers University, affiliated with both the Department of Statistics and the Department Computer Science. He received his PhD from Saarland University, Germany, with a thesis in the field of statistical machine learning. His research interests span areas at the interface of computer science, mathematics, signal processing and statistics such as structured and compressed representations of high-dimensional data, cost-accuracy trade-offs, and nonlinear programming. Application-wise, he has developed computational approaches to address problems in the life sciences.