A Constrained Maximum Entropy Principle for Data Reduction via Pattern Selection

Advanced Seminar on Optimization

Renato Budinich

24.11.2015

Abstract

Following previous works from M. del Viva, G. Punzi and D. Benedetti, we propose a model that claims to operate a type of lossy compression in a most efficient manner, when we wish to preserve the most relevant parts of the data while taking into account limited storage capacity and output bandwidth of the system. We'll introduce the two motivational examples of the model, which are Associative Memories (electronic architectures used in certain High Energy Physics particle detectors) and the visual cortex in the human brain. Both these systems receive huge quantities of data (of the order of Terabyte/s and Gigabyte/s respectively) and need to operate an aggressive filtering of the input before passing it on to the successive processing phases. Our model proposes a purely statistical criteria for deciding which parts of the data are to preserve and which to discard, based on a principle of maximizing the output Shannon entropy, taking into account the limits of the system. Mathematically, finding this criteria means solving an Integer Programming 0-1 problem, which is a particular case of Knapsack with cardinality constraint. We show how the Dynamical Programming approach to solving the Knapsack problem can be adapted to our case, including the FPTAS scheme. We discuss the shortcomings of this approach (which performs much worse than general IP solvers), and how for certain values of the model parameters a greedy heuristic gives the exact solution. Finally we'll mention what is missing from the model in order for its effectiveness to be tested on other applications.