APPROXIMATE BAYESIAN INFERENCE BY ADAPTIVE QUANTIZATION OF THE HYPOTHESIS SPACE

Mathias Johansson[1,2]
(1) Signals and Systems Group, Uppsala University, Sweden,
e-mail: mj@signal.uu.se
(2) Dirac Research AB, Sweden, e-mail: mj@dirac.se

Abstract

We introduce a simple and practical method for making approximate Bayesian inference based on quantizing the hypothesis space and repartitioning it as observations become available. The method relies on approximating an optimal inference by using a probability distribution for quantized intervals of the unknown quantity, and by adjusting the intervals so as to obtain higher resolution in regions of higher probability, and vice versa.

We repartition the hypothesis space adaptively with the aim of maximizing the mutual information between the approximate distribution and the exact distribution. It is shown that this approach is equivalent to maximizing the entropy of the approximate distribution, and we provide low-complexity algorithms for approximating multi-dimensional posterior distributions with tunable complexity/performance.

The resulting quantized distribution for a one-dimensional case can be visualized as a histogram where each bar has equal area, but in general unequal width. The method can be used to provide adaptive quantization of arbitrary data sequences, or to approximate the posterior expectation of for instance some loss function by summing over a pre-specified number of terms. A useful feature is that the method adapts optimally to incoming data using Bayes' theorem and is able to infer interdependencies between different variables.

We give explicit formulas for computing approximate expectations and posterior distributions for cases where only the number of outcomes in different classes, and not which order they occur in, is relevant, leading to a formulation based on Laplace's rule of succession. The method's usefulness is illustrated with an application in resource allocation with uncertain demands.

Finally, we discuss the problem of how to select the number of intervals to use in the general multi-dimensional case, and suggest a method for increasing and decreasing the granularity of the quantization based on the relative entropy of different quantizations.