A New Measure for the Accuracy of a Bayesian Network

Alexandros Pappas, Duncan Gillies
Department of Computing, Imperial College of Science, Technology and Medicine,
180 Queen's Gate, London SW7 2BZ, United Kingdom
{ap297, dfg}@doc.ic.ac.uk

Abstract. A Bayesian Network is a construct that is used to model a given joint probability distribution. In order to assess the quality of an inference, or to choose between competing networks modelling the same data, we need methods to estimate the accuracy of a Bayesian network. Although the accuracy of a Bayesian network can be easily defined in theory, it is rarely possible to compute it in practice for real-world applications due to the size of the space representing the variables. Instead, alternative characteristics of a Bayesian network, which relate to and reflect the accuracy, are used. A popular formalism that adopts such methods is the Minimum Description Length (MDL). It models the accuracy of a Bayesian network as the probability of the Bayesian network given the data set that it models. However in the context of Bayesian Networks, the MDL formalism is flawed, exhibiting several shortcomings. In its place, we propose a new framework for Bayesian Networks. We specify a measure, which models the accuracy of a Bayesian network as the accuracy of the conditional independencies implied by its structure. Experiments have been conducted, using real-world data sets, to compare MDL and the new measure. The experimental results demonstrate that the new measure is much better correlated to the real accuracy than the MDL measure. These results support the theoretical claims, and confirm the significance of the proposed framework.

KeyWords: Bayesian networks, MDL, Model Accuracy