Explaining inference in Bayesian networks for clinical decision support

Evangelia Kyrimi and William Marsh

Queen Mary University of London, UK.

Abstract
Bayesian networks (BNs) are often developed as prognostic tools in medicine. However, few of them have been applied to improve decision-making in practice. One reason is that clinicians tend to reject a model's advice if they do not trust it. There are many aspects of this trust; some, such as the model's predictive accuracy, are well known but others have been neglected. We focus on how the prediction of a BN can be explained to a user who is not an expert in probabilistic inference.

Many different types of explanation in a BN have been proposed. One type explains how the unobserved variables justify the observations, for instance the disease that best explains the symptoms; another type of explanation focuses on the knowledge behind the network structure, without considering observations. Our focus is on the reasoning process from a set of observations to the prediction of a target variable.

The prediction is presented as a single probability without any further information on the other variables included in the network. For example, a predicted probability of death shows nothing about how the observation of being a woman more than 65 years old having a major heart attack affected this prediction. Clinicians find it hard to contextualise a probability and may prefer a simple argument linking the observations to the prognosis.

Our approach to explaining inference in a BN assumes the network is made up from causal relationships, rather than just associations, and then answers two key questions (i) what is the impact of each observation in the marginal probability of the target variable and (ii) through which causal pathways do the observations affect our target. The first question is answered by the difference in the posterior probability distribution when each observation is removed, measuring KL distance. The second question is more challenging as in a realistic network the causal pathways can overlap. For this reason our approach focuses on the path from the observations to the target variable through only its Markov Blanket (MB) and not through every variable in the chains of reasoning, as the MB contains all the information needed to predict the value of a variable. Focusing our explanation in only a small number of variables is more helpful to clinicians trying to understand the causal story explaining a probabilistic inference.

Keywords: Bayesian network, clinical decision support