Question from Marek J. Druzdzel

A diagnostic system based on a Bayesian network model can do extremely well in computing the posterior probability over modeled disorders given a set of observations. The key here is "modeled" -- if a disorder in a given case is not modeled, the system can essentially get confused and produce meaningless and confusing results. It is of high importance for the user to realize that the system is incapable of diagnosis of a given case, i.e., that it does not model the disorder at hand. Propose and discuss possible measures that the system could derive from its model and the evidence entered by the user in order to self-reflect whether it is capable of diagnosing a given case or not.

Answer

A diagnostic system based on a Bayesian network model may well produce a meaningless result when applied to a case containing a disorder it does not modeled. The unmodeled disorder may correcpond to a new variable, or to a new value in an existing variable that represents mutually explusive disorders which means that the node's probability distribution in the model is incorrect.. Either way the case is trawn from a distribution not represented by the network.

One way for the system to detect the error is related to the joint posterior probabilities of the alternative diagnoses and observed evidence values. In typical domains, the distributions conditional probabilties of individual variables are skewed. Since the joint probability distribution represented by the model is formed by multiplying the conditional distributions, the joint contains a combination of the individual skews [Druzdzel 1994]. Thus, we expect that a small number of all the joint probability states would occupy most of the total probability. In typical domains, there are orders of magnitude differences between joint probabilities of different states. For example, for the ALARM database, this difference spreads over 22 orders of magnitude, and the 11 most likely of the 525,312 states occupy 3/4 of the total probability.

However, small differences in joint probabilies do not automatically imply that the case is unmodeled, because there exist atypical domains, whose joint probabilities are quite symmetrically distributed [Druzdzel 1994]. The prior joint probabilities would then differ little.

Conversely, if the prior joint probabilities differ little, then the states consistent with the evidence would also have probabilities somewhat similar to each other.

Thus, if the distribution prior joint probabilities are very different but the posterior joints are similar, we should be suspicious about whether the observed evidence corresponds to a modeled case. Now, by Bayes' Theorem the joint posterior distribution is proportional to the posterior of the disorders given the evidence:

$$p(\,Disorders \mid evidence\,) \propto p(evidence \mid Disorders)\, p(Disorders) = p(evidence, Disorders).$$

If all the posterior joint of probabilities are equal, each equals $1/k$ where k is the total number of disorder states.

Therefore, the system could compare the greatest prior and posterior joint probabilities to $1/k$. If the prior is much greater than k, but the posterior is not, the system would decline to diagnose the given case. It is not clear how much greater these need to be, but judging by the ALARM numbers, one order of magnitude may be a good threshold.

References

Druzdzel (1994). Some Properties of Joint Probability Distributions. *Conference on Uncertainty in Artificial Intelligence*, 187-194.