Determining whether a Belief Net is Consistent with Auxiliary Information[1]

Russ Greiner, Chris Darken *University of Alberta, Siemens Corporate Research*

A belief net (BN),[2] whether produced by interviewing domain expert(s) or learned from a body of data, can be incorrect. This report presents two tools for evaluating a given BN, by determining whether that network is statistically consistent with a body of accurate auxiliary information, I_A. It also suggests how this information can be used to improve the BN.

Tool#1: Error bars: The first tool provides "error bars" around the answers produced by a BN — *e.g.*, allowing the BN-based system to state that $P(\,cancer \,|\, headache\,) = 0.72 \pm 0.02$ with 95% confidence. These error bars are based on the training sample that was used to instantiate the CPtables for a given BN-structure.

Here, we start with a structure for a belief net, and then use the observed frequencies over a collection S of typles to fill-in the CPtables [Hec95, CH92]. Then given the query $P(\,X \,|\, Y\,)$, the instantiated BN can compute an answer; call it $\hat{P}^{(S)}(\,X \,|\, Y\,) \in [0,1]$. We prove that,

Theorem 1 *If the belief net structure is correct, then for any query "$P(\,X \,|\, Y\,)$" (whose correct answer is p), with probability at least $1-\delta$,*

$$|\hat{P}^{(S)}(\,X \,|\, Y\,) - p| \quad \le \quad \sqrt{\frac{1}{2\,|S|}\ln\frac{2K}{\delta}} \times \frac{1}{P(\,Y\,)} \sum_{\langle\, q, \mathbf{r}\,\rangle} \sqrt{P(\,\mathbf{r}\,)}\,[P(\,X, Y \,|\, q, \mathbf{r}\,) - P(\,X \,|\, Y\,)P(\,Y \,|\, q, \mathbf{r}\,)]$$

when there are K CPtable entries (each of the form $\langle\, q, \mathbf{r}\,\rangle$) that participate in this $P(\,X \,|\, Y\,)$ computation, where each $\langle\, q, \mathbf{r}\,\rangle$ in the summation corresponds to the CPtable entry for $P(\,Q = q \,|\, \mathbf{R} = \mathbf{r}\,)$, for node Q with parents $\mathbf{R}$. ∎

If the auxiliary information I_A states that the answer to this $P(\,cancer \,|\, headache\,)$ query should be, say $p = 0.65$, then we have reason to question the given structure of the network, as we can be confident that the particular training sample used cannot explain the 0.72 vs 0.65 discrepancy. One can often use a set of queries — some of whose answers were "within tolerances" and others not — to isolate which parts of the network appear problematic.

Tool#2: Does the Sample Support Existing Links? The second tool provides a more fine-tuned probe into the structure of the network. The I_A here corresponds to

[1] For more information, see `http://www.cs.ualberta.ca/~greiner/BN-results.html#consistent`

[2] We assume the reader knows that a belief net (aka Bayesian network, probability net, causal net) is a graphical model of a (factored) distribution, whose nodes correspond to random variables, which each include a "CPtable" that specifies the conditional probabilities of that node given each possible assignment to its parents; see [Pea88]. We also say a *BN*-structure is "correct" if it correctly models the (in)dependencies of the underlying distribution.

a body of sample tuples, and the probe determines whether these samples support a specified link between a pair of nodes.

Definition 2 *1. Let G_b be a belief net structure, and G_s be a substructure of G_b, produced by deleting an existing link. Given any sample S, let BN_s (resp., BN_b) be the maximum-likelihood belief net corresponding to G_s (resp., G_b) [produced by filling the CPtables using the observed frequencies], and let P_s (resp., P_b) be the associated probability measure. Let $\nu = \nu(G_s, G_b, S) \;=\; -2\,\ln(P_s(\,S\,)/P_b(\,S\,))$.*[3]
2. Let $\langle\,\phi, \theta\,\rangle$ be a parameterization of P_b with the property that $\langle\,\phi, \vec{0}\,\rangle$ is a parameterization of the constrained P_s.
3. Let H_s (resp., H_b) be the hypothesis that the data S was generated (iid) by a distribution representable by G_s (resp., by G_b, but not by G_s). ∎

We expect ν to be close to $-2\ln 1 = 0$ when H_s holds, and to blow up towards infinity otherwise. In fact,

Proposition 3 ([SO91]) *When H_b holds, ν asymptotically has a* non-central χ^2 *distribution with $n = |\theta|$ degrees of freedom and non-central parameter $\lambda \;=\; \theta^T\ M\ \theta$ where $M_{ij} = -E_{X_1,\ldots,X_m}\big(\frac{\partial^2 \ln P_b(S)}{\partial\theta_i \partial\theta_j}\big)$.* ∎

Given a body of data and a specified link, we can first compute associated λ and ν parameters. Next, for any specified confidence value $\alpha \in (0, 1)$, we can then determine the threshold $t_\alpha = t_\alpha[n, \lambda]$ such that $1 - \alpha \;=\; P(\,X\, <\, t_\alpha\,)$, where the r.v. X is drawn from a χ^2 distribution with n degrees of freedom and non-centrality parameter λ. We then reject H_b (*i.e.*, decide to *exclude* the specified link) with probability $p \geq \alpha$ if $\nu \leq t_\alpha$.
The major challenge is computing the non-centrality parameter λ. Fortunately,

Theorem 4 *Let the belief net G_b include only nodes A and B, where A is the only parent of B, and the BN G_s differs from G_b only by deleting this one link. Then after observing the (complete tuple) sample S, the non-centrality parameter is*

$$\lambda \quad = \quad |S| \sum_{a,b} [P(\,B = b\,) - P(\,B = b \,|\, A = a\,)]^2 \ \frac{P(\,A = a\,)}{P(\,B = b \,|\, A = a\,)}$$

where each $P(\,\cdot\,)$ value is based on the empirical distribution. ∎

Obvious corollaries show that this result scales up, to handle the situation where (1) B has > 1 parents; (2) we are considering simultaneously removing the links from B to several parents; and (3) we are considering simultaneously removing the links from several nodes to several (respective) parents. In addition to deciding whether to delete an existing link, we can use a similar approach (based on *standard* χ^2 test) to decide whether to *add* new links.

[3] It is easy to compute this ν quantity, as most of the terms involved in the $P_s(\,S\,)/P_b(\,S\,)$ computation cancel out.

Open problems: These tools were originally motivated by the goal of helping a domain expert to evaluate and revise an existing belief net, based on auxiliary information. (As that information was not necessarily correct, we left the expert in the loop.) Of course, these ideas can also be used to help *learn* a belief net, by "revising" the 0-link belief net. Here, we need ways to automate many of the decisions — *e.g.*, decide which link to consider adding (or deleting), and also deciding how to use the "error-bar"-based analysis to decide which specific links (or CPtable entries) are problematic.

References

[CH92] G. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[Hec95] D. Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, 1995.

[Pea88] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, 1988.

[SO91] A. Stuart and J. Ord. *Kendall's Advanced Theory of Statistics*, volume 2. Oxford University Press, 5th edition, 1991.

[Russell Greiner; Department of Computing Science; University of Alberta; 615 General Service Bldg; Edmonton, Alberta T6G 2H1; Canada
greiner@cs.ualberta.ca 780 492-5461]

Contributed for [either oral/poster] presentation