Process of Simulating the Experts’ Opinions Regarding Bayesian Network Structures

In this short manuscript, we explain the process of simulating the experts’ opinions regarding Bayesian network structures based on the experts’ accuracy parameters.¹

Based on the original Bayesian network structure and the accuracy parameters of each expert, we can simulate his/her opinion about each pair \((X, Y)\). If there exists an edge \(X \to Y\) in the true graph, this expert selects \(X \to Y\) and \(X \leftarrow Y\) with probabilities \(\gamma_1\) and \(\gamma_2\), respectively. Therefore, the probability that this expert selects the absence of an edge between these variables is \(1 - \gamma_1 - \gamma_2\). On the other hand, if there is no edge between \(X\) and \(Y\) in the true graph, the expert correctly mentions the absence of an edge with probability \(\gamma_3\), and selects one of the edges \(X \to Y\) or \(X \leftarrow Y\) with equal probabilities \((1 - \gamma_3)/2\). These probabilities are equal because when we want to know the opinion of one expert about the relationship between \(X\) and \(Y\), one of the two ordered pairs \((X, Y)\) and \((Y, X)\) is randomly presented. Therefore, if there is no edge between \(X\) and \(Y\) in the true graph and this expert wrongly selects the existence of an edge, the probabilities of selecting each possible directions are the same.

The first step in the simulation process is to determine the number of opinions provided by each expert. The total number of opinions is controlled by a parameter \(\nu \in [0, 1]\). If there are \(N\) pairs of variables and \(R\) experts, the total number of opinions provided by all experts is \(M = \text{round}(\nu \times R \times N)\), where \(\text{round}(x)\) outputs the closest integer to \(x\).

First, to avoid the bias in the generated numbers, the experts are selected in a random order, and the number of opinions are generated according to this sequence. Assume that \(n_i\) is the number of opinions of expert \(i\), and \(S_i\) is the total number of opinions provided by experts 1 through \(i\), i.e. \(S_i = \sum_{j=1}^{i} n_j\). To generate the number of opinions for expert \(k\), a random integer is selected from this range:

\[
M - S_{k-1} - N \times (R - k) \leq n_k \leq M - S_{k-1}.
\]

¹ This is an auxiliary document to the following paper: Hossein Amirkhani, Mohammad Rahmati, Peter Lucas, and Arjen Hommersom, “Exploiting experts’ knowledge for structure learning of Bayesian networks,” Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.

Therefore, to understand the notations used here, please refer to the above paper.
This is because the total number of opinions that must be provided by experts $k$ through $R$ is $M - S_{k-1}$. Therefore, $n_k$ must be lower than or equal to $M - S_{k-1}$. On the other hand, because each expert can provide at most $N$ opinions, experts $k + 1$ through $R$ are able to provide at most $N \times (R - k)$ opinions. Therefore, $n_k$ must be at least equal to $M - S_{k-1} - N \times (R - k)$.

After determining the number of opinions for each expert, we must generate the opinions. To generate the opinions of expert $i$, $n_i$ pairs are randomly selected from $N$ possible pairs. Then, for each selected pair, the opinion is generated according to the status of the edge between those two variables in the true graph and the accuracy parameters of this expert. For example, if the selected pair is $(X, Y)$ and there exists an edge like $X \rightarrow Y$ in the true graph, a random number $r \in [0, 1]$ is generated and the opinion is produced according to the following conditions:

- If $r \leq \gamma_1$ then $X \rightarrow Y$ is selected as the opinion of this expert.
- If $\gamma_1 < r \leq \gamma_1 + \gamma_2$ then $X \leftarrow Y$ is selected.
- If $r > \gamma_1 + \gamma_2$ then the absence of any edge between these two variables is generated as the opinion of the expert.

To generate the opinion for the state where there is no edge between $X$ and $Y$ in the true graph, we follow a similar procedure and use the parameter $\gamma_3$. 

2