Building Bayesian Influence Ontologies
Annotated Bibliography

Tamlin Love
1438243

March 2, 2019

References

[1] R. Ajoodha and B. Rosman, "Tracking influence between naïve bayes models using score-based structure learning," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech).* IEEE, November 2017.

> **Aim:** To present a method that learns the high-level influence structure present between a set of independently learned naïve Bayes models.
>
> **Summary:** This paper presents an algorithm for learning the influence structure between naïve Bayes models (NBMs). The algorithm achieves this by first learning a set of independent NBMs (i). It then computes a score used to evalutate the fitness of the network (ii). This approach makes use of the Bayesian information criterion (BIC) for scoring, which provides an acceptable trade-off between model complexity and data fitting. The algorithm then refines the model given the new influence structure using expectation maximisation (iii). After this, the candidate network is subjected to a graph operation (edge addition, reversal or deletion) chosen to optimally improve the network's score (iv). This is achieved using a greedy hill-climbing heuristic, which guarantees monotonicaly improving score between iterations. Finally, steps (ii) to (iv) are repeated until an optimum is found. The result is a method which, in the authors' tests achieved 60-82% accuracy when compared to the ground truth structure. Additionally, the method outperformed the random structure and the structure with no conditional independece assertions, and tended towards the true structure as the number of samples increased.

[2] ——, "Learning the influence structure between partially observed stochastic processes using iot sensor data," in *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

> **Aim:** To present an algorithm for learning the influence structure between a set of stochastic processes represented as hidden Markov models (HMMs).
>
> **Summary:** This paper presents a method, referred to as the Greedy structure search (GESS), for recovering the delayed influence structure between a set of HMMs. It does so by first learning each HMM independently using partially observed Internet-of-Things (IoT) data. It then sets the independece assumptions between the models

and uses expectation maximisation to learn the associated influence network. The algorithm then evaluates the candidate network's score. The authors empirically test the algorithm using both the Akaike information criterion (AIC) and a modified Bayesian information criterion (BIC) for delayed dynamic influence networks (DDINs). Then the algorithm applies the graph operator (edge addition, deletion or reversal) to result in the best improvement of the network's score with respect to the data. This is done using greedy hill-climbing, which works by applying a change which increases the score, until no such changes can be made. The above steps are repeated until no improvement can be made to the score or the algorithm exceeds the maximum number of iterations.

In the authors' tests, the DDINs produced by the GESS algorithm with the aforementioned scoring criteria more closely recovered the ground truth structure than the tree structures and no structure for a large number of observations. However, for fewer (less than 200) observations, the tree structures and no structure performed better than the GESS-produced structures in this regard.

[3] D. Koller and N. Friedman, *Probabilistic Graphical Models - Principles and Techniques - Chapter 3: The Bayesian Network Representation.* The MIT Press, 2009.

Aim: To present the notion of a Bayesian network (BN), to prove some fundamental properties of BNs, to present the notion of I-equivalence and show that a partially directed acyclic graph (PDAG) can be used to represent all members of an I equivalence class, and to present an algorithm for constructing such a PDAG.

Summary: This chapter presents the concept of a Bayesian network and shows that it can be used to reduce the number of parameters needed to represent a joint distribution. The book provides two definitions of a BN, one as a data structure for compact representation of a joint distribution, and the other as a representation of the set of conditional independence assumptions that hold for such a distribution, and then shows that these definitions are in fact equivalent.

The authors then present the definition of I-equivalence: that two graphs belong to the same I-equivalence class if and only if they represent the same set of independence assumptions. The authors show that a PDAG can be used to represent an I-equivalence class, in which all the undirected edges of the graph can be oriented in any way to produce a graph that belongs to the class. The authors also provide a definition for an immorality between three variables.

Finally, the chapter provides a set of algorithms used to construct a PDAG for a given set of random variables and a distribution over said variables. It consists of an algorithm which constructs an undirected skeleton of the final graph, an algorithm which identifies the immoralities in the class and applies them to the skeleton, and finally an algorithm which applies the previous two algorithms and further directs any edges which could result in the creation of new immoralities or of cycles.