Notes on Lauritzen & Spiegelhalter 1988: Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems

Gray Stanton and Austin Ellingworth

November 4, 2020

Discussion Qs

- Discuss Figure 1. Check understanding
- The described MUNIN system and other medical expert systems via bayesian networks were implemented decades ago. How frequently do they influence the decision-making of practitioners. Are their still-existing barriers to adoption?
- Building off that, how do practitioners use them. In practice, are these networks used in both direction at the same time? As an example both to predict future symptoms and to categorize the disease. As a comfirmation tool? etc.
- In the analysis of the computational complexity of the algorithms given, they mention that maximum number of clique nodes, total state space, and largest state space of a clique have significant impact on run time (in addition to the obvious factors of total number of nodes/edges). Are these still constraints on evidence propagation through Bayesian networks? What is the biggest factor determining if an expert system of this type is feasible for real-world applications?
- The authors briefly mention extending their work to nodes with continuous measurements. Is this an easy jump, or are there important hurdles?
- Identification of influential items, if you were to explain these influence statistics to someone from a different field, how would you do that?
- Authors mention "imprecision" of probabilities, has there been any development on trying to incorporate error bounds on probabilities?

General Notes

- 4 Views on expert systems: Logical (symbolic reasoning only), linguistic (matching common language statements to exact propositions) Legal (Dempster-Shafer Theory), Statistical (probability)
- Problems posed by MUNIN expert system, which uses strict probabilistic reasoning (not ad-hoc). MUNIN is a ccausal network with 25 nodes.
- Need ability to efficiently handle 6 problems for causal network such as MUNIN: Initialization, Absorption of evidence, global propagation, hypothesizing and propagating single items of evidence, planning, infuluential findings. All require efficient calculation of conditional and marginal probabilities for nodes. Handled with local methods.
- Brute force evaluation of joint probabilities is very inefficient. More effective to exploit topology of graph.
- Moral graph: parents of nodes are married and directionality is removed. Next, graph needs to be triangulated/filled in, so that there are no cycles of length > 3.
- Joint distribution is then proportional to product of clique functions, all of at most 3 arguments.
- This produces local representation of joint distribution, ensuring that clique marginals make retrieval of post-evidence single node probabilities easy.
- Set chain using cliques of moral graph possesses running intersection property. This is useful as it ensures that the next separator marginal can be obtained from previously calculated clique marginal.
- This procedure is done in two passes through the graph: a forward pass to transform potential representation from conditional probabilities into new ones based on specific set chain, and second to calculate the clique marginals associated with that.These are basic operations for other problems of interest.
- Absorption of Evidence: More efficient to "save up" evidence and propagate all at once. Each observed node is removed from the graph and a new potential representation is computed by projecting potentials involving observed nodes onto new reduced clique/neighbor if clique entirely observed.
- For single items of evidence (i.e. setting single node to a particular observed value), there is an efficient/simple propagation procedure, which also covers the most frequent use case
- Done in a single pass, by computing update ratio through the clique intersections.

- Planning: For cases when particular node is of interest but unobservable, want to know which observed nodes would have greatest effect. Can be done by simulating away from node of interest.

- Influential observations: for interpretation purposes, we may want to understand how important different observed values are to the predicted value of a node of interest. This can be done by removing observations one by one, and computing "deconditioned" distribution in an efficient way.

- A clique is defined as a maximal subset of a graph where there are edges between each node in the subset (complete).

- A core concept is that of a perfect numbering of an undirected graph. A numbering is perfect if for all i, the boundary of the node i intersect all the preceding nodes in the numbering is a complete subgraph. In particular, a graph is triangulated if and only if it has a perfect numbering.

- Their procedures also depend on "maximum-cardinality search", which is a simple way to test is a graph is triangulated, and if so obtain a perfect numbering.

- Feasibility of procedures depend on small clique size, which as causal DAGs will tend to be sparse/irregular and so have small cliques, is not too much of a hurdle.

- Importance placed on local representation of the joint probability density, which is a system of functions where each depend only on a small number of variables. Then computing the entire joint distribution can be avoided.

- First local representation: Conditional probability tables. For each node, there is a table which expresses for each combination of states of parent nodes, the conditional probability of each state for that node. This is standard factorization of joint probability via conditional independence assumptions (causal Markov property). Also called transition kernels.

- Evidence potentials: Given a list of subsets of nodes, corresponding evidence potentials are functions which depend only on the states of variables in each subset. They are then assumed to have that the product of the potentials (when normalized correctly) is the joint probability of the network.

- There re many possibly potential representations for a given probability distribution, and this can be used to find efficient representation. They are also a superset of the conditional probability table representation.

- Set chain representation: An ordered chain of sets whose union is the whole network, and who has the running intersection property. Similar to conditional probability tables except involve sets of nodes instead of single nodes.

- Clique Marginal representation: A list of subsets of nodes whose union is the full network, and such that the hypergraph defined by the node list is acyclic. Interested in marginal probabilities on the subsets, who also satisfy consistency condition for intersecting subsets. If joint probability is assumed Markov, then there is a unique joint probability with given marginals on the subsets. This representation is leveraged to calculate current beliefs about states at a given node.

- Next, recursive procedures for updating marginal and potential representations. Conditioning on potentials is the easy task, simply requires using given values for particular nodes and computing reduced potential via multiplication by full potential with given values inserted. Normalizing constant is ignored.

- Conditioning in marginals is hard. Requires some restrictions on evidence set to be feasible, version in paper requires evidence nodes all to be part of same clique. Running intersection property is then leveraged to do this in single pass.

- Marginalizing in potentials refers to marginalizing over the set of evidence nodes for a particular subset of the graph. Also, there is marginalizing over marginals, which is difficult in general but refers to finding marginal distribution over a particular subset of nodes given a marginal representation over the whole graph.

- In the next section, methods to transfer between the four representations of the joint distribution are presented.

- Computational complexity of the previous methods are discussed. In particular, total state space, maximum number of nodes in clique, and largest clique state space have subtle and important implications for performance.

- Lastly, a few extensions/special cases are considered.