swissQuant
Mastering Complexity

Mathematical Challenge September 2017

Bayesian networks for decision making under uncertainty

References

- [1] Koller, Daphne, and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [2] Heckerman, David. "A Bayesian approach to learning causal networks." *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1995.

Description

Introduction

Bayesian networks are graphical models, which represent multivariate probability distributions [1]. Nodes correspond to variables and directed edges represent the probability factorization

$$P(X_1,..,X_n) = \prod P(X_i|Pa(X_i))$$

Where $Pa(X_i) = \{X_j | \exists\ edge\ X_j \to X_i\}$

The network structure essentially incorporates conditional independence assumptions. Among the many possible network structures which represent the same probability distribution, it is in general considered advantageous to choose one that has a causal interpretation (i.e. whose edges $A \to B$ have the semantic “A causes B”). However, while this usually provides sparse and interpretable structures, this is not strictly required for probability inference.

The situation is different when instead of being interested in the probability of an event $B = b$ conditional on another event $A = a$ being observed: $P(B = b|A = a)$, we are interested in the effect of the intervention of setting the variable A to the value a on B: $P(B = b|set\ A = a)$. These intervention queries are of particular interest, having a variety of applications: diagnosis and treatment choice in medicine, marketing decisions, policy decision of regulators,…

For intervention queries indeed the answer depends on the causal relationship between variables. For these queries the Bayesian network should represents a causal model and not just a probabilistic model. In the first case indeed the query $P_N(B = b|set\ A = a)$ can be solved as $P_{N-e_A}(B = b|\ A = a)$ where the network $N - e_A$ corresponds to the network N where all edges pointing to A are eliminated (mutilated network)

To understand the difference between probabilistic and causal relationship, consider the case of A and B being dependent variables. This can arises because A causes B, B causes A or A and B have a common (possibly latent) cause C. These three causality scenarios have different implications regarding the query: $P(B = b|set\ A = a)$. It is also clear that to determine the correct causality structure is paramount to explicitly consider all possible latent confounding factors. Often, however this cannot be done just relying on observational data, leading to the problem of identifiability of causal models.

- Q1: Consider the Simpson's Paradox: A medical study about the efficacy of a drug for treating a particular disease reports the following main result

  - 57.5% of all patients treated with the drug are cured, compared to just 50% for those not treated with the drug

  This seems to indicate that the usage of the drug should be recommended. However, following more refined statistics were also available:

  - 70% of all male patients treated with the drug are cured, compared to 80% for male patients not treated
  - 20% of all female patients treated are cured, compared to 40% for not treated females

  Explain this apparent paradox and conclude in a sound way if the drug should be recommended as a treatment for the disease

Influence diagrams (ID)

A generalization of Bayesian networks for general decision making under uncertainty purposes, is represented by influence diagrams [2]. Influence diagrams have

- Nodes: chance variables C, decision variables D, utility variables U (U are deterministic function of their parents and have no children)
- Edges:
  - Information edges : point to decision variable, specify the available information when decision is made
  - Relevance edges: point to chance variables with the same semantic as in standard Bayesian networks

In many cases it also assumed that the ID structure is compatible with a total ordering on the decision nodes, which represent the decision flow.

A decision rule δ_D for the ID consists in the specification of all probability distribution $P(D|Pa(D))$

Once the ID is fully specified: i.e. the structure of the network, the probability distributions and the utility functions $U(Pa(U))$ have been determined, every decision rule δ_D turns an ID into a Bayesian network, which can be used a.o. to determine the expected utility of δ_D and as consequence the maximum expected utility strategy.

The ID framework is particularly suited to handle causal networks. Indeed by defining two special ID nodes as possible parents of a chance variable C:

- Set decision variable $\hat{C}$: A decision variable whose domain is $\{\text{"set C} = \text{c"}\} \cup \text{"}do\ not\ set\ C\text{"}$
- Mapping variable $M^X\{Y\}$: A chance variable whose domain is the set of all possible mappings from $Domain(\{Y\})$ to $Domain(X)$, for $\{Y\} \subset Pa(X)$

interventions can be considered on the ID (also called augmented network) with nodes $\{X\} \cup \{\hat{X}\}$ and causal models can be described as functional causal models with nodes $\{X\} \cup M^X$.

Furthermore, the concept of *cause* can be defined as

$$\{CC\}\ causes\ C\ given\ the\ decision\ D \Leftrightarrow$$

$$C \notin \{CC\}\ and\ \{CC\}\ is\ a\ minimal\ set\ such\ that\ M^C(\{CC\})\ is\ unresponsive\ to\ D$$

A causal relationship given the decision D can then be casted in a ID diagram by making the chance variable C a deterministic function depending on CC and $M^C(\{CC\})$. This construct when applied iteratively on a set of variables gives an ID representing all set of causal relationships and satisfying the properties

- All chance nodes that are responsive to D are descendants of the decision nodes
- All nodes that are descendants of decision nodes are deterministic nodes

An ID satisfying these properties is said to be in *canonical form*

- Q2: The distribution of a mapping variable $M^X(Y)$ are related to conditional distribution of explicit (observed) variables X,Y by the condition

$$P(x_i|y_i) = \sum_{x_{-i} \in Val(X)^{m-1}} P\left(\mu^X_{y_i \to x_{i,} y_{-i} \to x_{-i,}}\right)$$

Where $\mu^X_{y_i \to x_{i,} y_{-i} \to x_{-i,}}$ is the map assigning $y_i \to x_{i,} y_{-i} \to x -_{i,}$

Consider a randomized clinical study for the treatment of a disease, described by the variables

- A: patient assignment (treatment or placebo)
- T: patient treatment compliance (always, almost-always, rarely, never)
- O: patient outcomes (cured, improved, no-change, worsened)

and by the network $\mathrm{A} \to \mathrm{T} \to \mathrm{O}$. Assume that the observed data allow the estimation of all conditional distribution. Since the disease severity S is a latent effect for both compliance (very ill patients tend to follow differently the protocol) and outcomes, the causal relationship between T and O is however not identifiable. Using the previous relation determine bounds for the treatment effect defined as $P(O|A = treatment) - p(O|T = placebo)$

- Q3: Provide a procedure to formulate general causal intervention networks as influence diagrams in canonical form.

We look forward to your opinions and insights.

Best Regards,

swissQuant Group Leadership Team