Soft Evidential Update for Probabilistic Multiagent Systems

Marco Valtorta
mgv@cs.sc.edu

Department of Computer Science and Engineering
University of South Carolina
November 9, 2000

This is joint work with Young-Gyun Kim and Jiří Vomlel.

Preview

- Motivation: Autonomous Agents Communicate by Passing Probabilities on Common Variables

- Problem: How to Update a Bayesian Network Upon Receipt of a Probability Distribution

- Issue: Conditional Independence Structure Must Be Maintained

- Solution Component: Observation Variables

- Solution Component: Big Clique Algorithm

- Implementation: BC-Hugin

Overview

- Agent Encapsulated Bayesian Networks
- Evidence
- Observation Variables
- The Big Clique Algorithm
- Summary and Conclusions

Characteristic of Agents

Agents are autonomous, possess a local view of the world, possess a local set of goals.

An agent system is a group of agents that communicate.

An agent system is therefore characterized by:

- A method for knowledge representation, so that each agent can maintain a view of the world and take rational decisions.

- A communication language and methods to facilitate the sharing of information.

Bayesian Networks

A Bayesian network is an efficient representation of the joint probability distribution on a set of variables.

A Bayesian network includes:

- a DAG $G = (V, E)$

- for each $v \in V$, a conditional probability table $P(v \mid \pi(v))$, where $\pi(v)$ is the set of parents of v.

$P(V) = \prod_{v \in V} P(v \mid \pi(v))$

Each variable in the DAG is independent of any subsets of its non-descendants given all of its parents.

Evidence

A *(hard) finding* on a variable is a function from the states (values) of a variable to $\{0, 1\}$, in which exactly one of the states has value 1. The meaning of a finding is that the value of a variable is observed (and therefore known).

Evidence on a set of variables is a set of findings, one per variable.

A *soft finding* is a distribution on the states of a variable. The meaning of a soft finding is that we are told authoritatively the distribution of the values of a variable.

Soft evidence on a set of variables is a set of soft findings, one per variable.

Jeffrey's Rule

Suppose we are interested in the updated distribution of variable A upon receipt of soft evidence on B.

Jeffrey's rule is the following update rule:

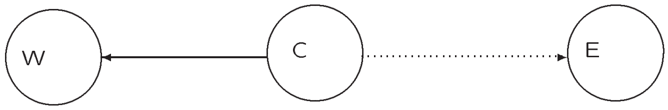
$Q(A) = \sum_i P(A \mid B_i)Q(B_i)$, where

- $Q(B)$ is the soft evidence, and
- $P(A \mid B_i)$ is the conditional probability of A given B *before* soft evidence.

A "Good" Example for Jeffrey's Rule

The following example is an elaboration of one by Jeffrey, and is due to Pearl [1988, 1990].

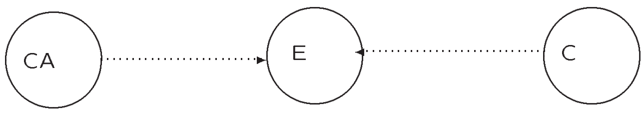
The worth of a cloth.



The worth of a cloth in the presence of soft evidence.

A “Bad” Example for Jeffrey’s Rule

Candle type and cloth color.



Candle type and cloth color in the presence of soft evidence.

Jeffrey's Rule and D-separation

The rule applies only if $P(A \mid B)$ is invariant with respect to evidence. "This condition is an internal or psychological condition that must be checked or accepted at each stage. Mathematics has nothing to offer here" [Diaconis and Zabellis, 1982].

There is, however, a principled way to check whether the condition holds, which is based on the theory of Bayesian networks.

Augment the Bayesian net with a node that represents the evidence (E) and the appropriate edges.

Check whether B *d-separates* E and A. If so, Jeffrey's rule can be applied. If not, it should not be applied.

Observation Variables

Upon addition of observation variables, the qualitative independence structure of the domain is encoded correctly.

Previous attempts to deal with soft evidential update have failed, because they have followed the "something for nothing" principle: in this case, the "mechanical solution" of Jeffrey's rule [Diaconis and Zabellis, 1982] is used in place of careful structural modeling of the domain.

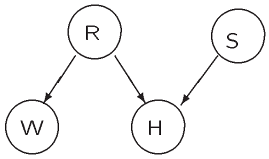
The Junction Tree Algorithm

The Junction Tree Algorithm is used to compute the marginal probability of every variable in a Bayesian network structure G with node set $V = \{V_1, \ldots, V_i, \ldots, V_n\}$ upon receiving evidence e.

$P(V_i, e) = \sum_{V_k \in V \setminus V_i} \prod_{V_j \in V} P(V_j \mid \pi(V_j))$,
where variables are instantiated according to the evidence e.

$P(V_i \mid e)$ can be obtained by normalizing $P(V_i, e)$.

Example of Update with the Junction Tree Algorithm

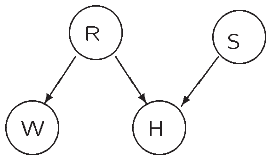


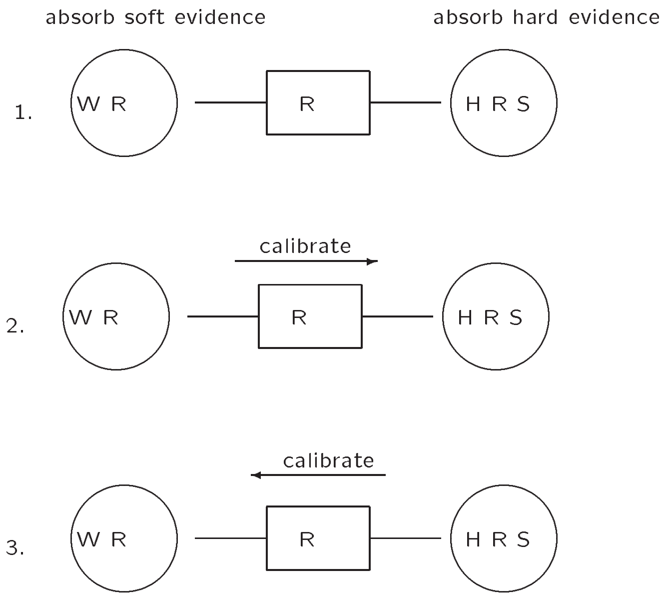
The wet grass Bayesian network structure.

Steps of the Junction Tree Algorithm

- Moralization
- Triangulation
- Identification of Cliques
- Construction of a Junction Tree
- Evidence Absorption
- Collect Evidence and Distribute Evidence from the Virtual Root

The Junction Tree Algorithm and Soft Evidence: Example





The junction tree algorithm and the wet grass network

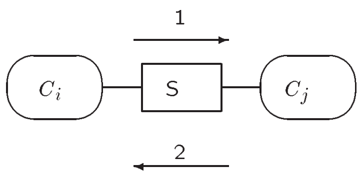
The Junction Tree Algorithm and Soft Evidence

In the example, HRS is the virtual root. There is hard evidence on H and soft evidence $Q(W)$ on W.

HRS calls CollectEvidence: this results in an update of P(H,R,S) via calibration.

Then, HRS calls DistributeEvidence: this results in an update of $P(W,R)$ to $P'(W,R)$ by calibration.

After this update, $\sum_R P'(W,R) \neq Q(W)$!



Generally, a soft finding $Q(V_i)$ ($V_i \in C_i$) is not treated as evidence.

Iterative Modification of the Junction Tree Algorithm

1. Enter hard evidence and propagate normally. Call the resulting distribution Q.

2. Iterate the following cycle (consisting of k applications of the pair of steps (2a) and (2b)) until convergence is achieved.

3. For each piece of soft evidence, $P(O_j), j = 1, \ldots, k$:

   (a) Absorb $P(O_j)$ in a clique C_{l_j} that contains O_j using the Iterative Proportional Fitting Procedure (IPFP) formula:

   $$Q(C_{l_j}) = \frac{P(C_{l_j}) \cdot P(O_j)}{Q(O_j)}$$

   (b) Propagate normally.

The Big Clique Algorithm

1. Build a junction tree that includes all variables for which soft evidence is given in one clique, the *big clique* C_1. (These variables may appear in other cliques as well.)

2. Update $P(V)$ to a distribution $P^*(V)$ by executing the junction tree algorithm using only hard evidence.

3. Absorb all soft evidence in C_1.

4. Call the routine DistributeEvidence from C_1.

Absorption of Soft Evidence

We define absorption in the special big clique C_1 as the process by which the joint probability $P(C_1)$, is updated to conform to soft evidence on variables $A \subseteq C_1$, where $A = \{A_1, A_2, \ldots, A_k\}$.

Absorption of soft evidence in clique C_1 is carried out by using the Iterative Proportional Fitting Procedure (IPFP) and consists of cycles of k steps, one per finding. Each step corresponds to one soft finding. The appropriate formulae are:

$$\begin{aligned} Q_{(0)}(C_1) &= P(C_1) \\ Q_{(i)}(C_1) &= \frac{Q_{i-1}(C_1) \cdot P(A_j)}{Q_{i-1}(A_j)} \end{aligned}$$

where $j = (i-1) \bmod k + 1$.

Example of Soft Evidence Absorption

Clique $\{A, B\}$ with joint probability:

	A	
B	y	n
y	.56	.03
n	.14	.27

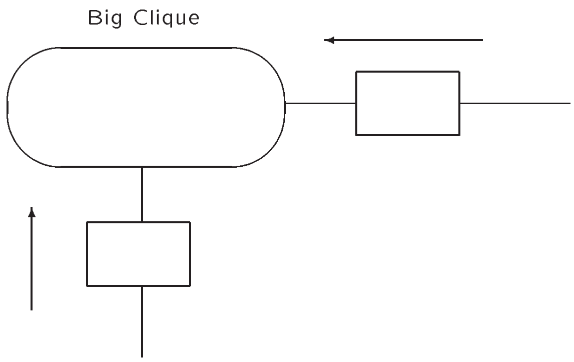
Soft evidence is $P(B) = (.7,\ .3)$. We compute the updated joint probability $Q(B)$ in two steps. The result of multiplication by $P(B)$ is given below:

	A	
B	y	n
y	.392	.021
n	.042	.081

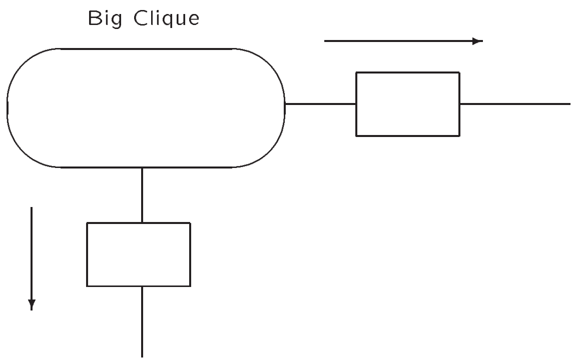
The result of division by $Q_{(0)} = (.59, .41)$ is:

	A	
B	y	n
y	.664	.036
n	.102	.198

Propagation of Soft Evidence

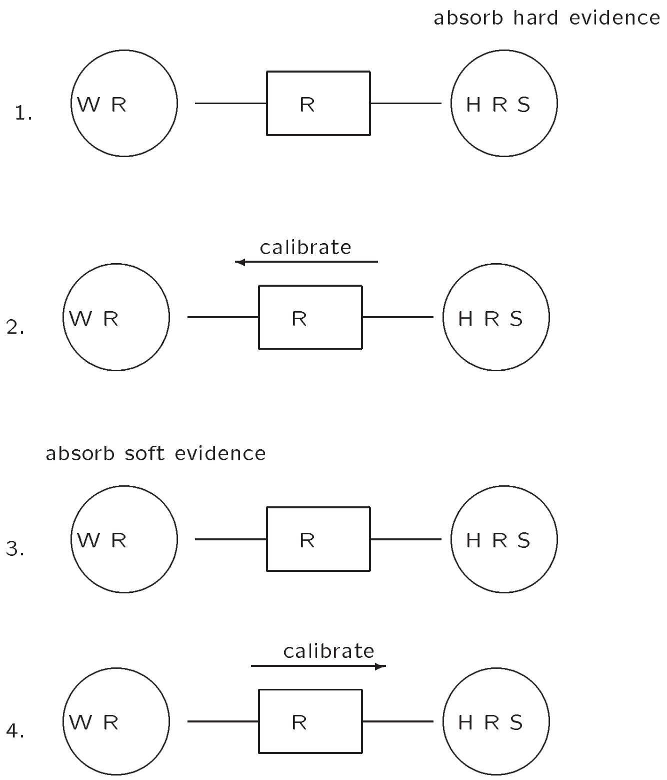


The big clique calls CollectEvidence.



The big clique calls DistributeEvidence.

Example: Wet Lawn



Operation of the big clique algorithm on the lawn example.

Why IPFP?

In the hard evidence case, independence of the observed variables is necessary, because when $Q(O_1), Q(O_2), \ldots, Q(O_n)$ are hard findings, the joint distribution $Q(O_1, \ldots, O_n)$ necessarily is

$$Q(O_1) \cdot Q(O_2) \cdot \ldots \cdot Q(O_n),$$

since no other joint distribution has the required marginals.

When the evidence consists of soft findings, however, the joint marginal distribution

$$Q(O_1, \ldots, O_n)$$

is not uniquely specified by the observations.

Deming and Stephen's Example

An agent models age (X_1) and education (X_2) of U.S. residents.

A small survey results in $P(X_1, X_2)$.

Communication with other agents gives authoritative information about age and education: $Q(X_1)$ and $Q(X_2)$.

Should we replace $P(X_1, X_2)$ with $Q(X_1, X_2) = Q(X_1) \times Q(X_2)$?

Deming and Stephen's Example, Ctd.

The agent should try and find a joint probability distribution $Q(X_1, X_2)$ such that:

- The marginals from the U.S. Census Bureau agent are respected.

- $Q_2(X_1, X_2)$ is the distribution that is closest to the distribution $P(X_1, X_2)$. As a measure of distance among distribution, Kullback-Leibler divergence (also known as cross-entropy) could be used:

$$I(P \parallel Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \qquad (1)$$

IPFP computes precisely this distribution, which is called the I-projection of P.

The Big Clique Algorithm and IPFP

The IPFP computation may be decomposed in such way that the Big Clique algorithm is obtained.

A companion paper (under review) contains the proof of this result.

The paper contains another algorithm for the IPFP computation suitable for soft evidential update in Bayesian network, which typically is slower than the Big Clique algorithm, but more space efficient.

Summary

- Application of BNs to multiagent systems uncovers the limitations of simple conditioning.

- Soft evidential update is possible in BNs without abandoning the commitment to the primacy of qualitative (causal or independence) structure.

- The big clique algorithm extends the junction tree algorithm to soft evidential update, thus opening new avenues for application.

Conclusions and Future Work

- BC-Hugin

- Some related issues are addressed in Young-Gyun Kim's dissertation

- The Rumor Problem in AEBNs

- Applications in Multiagent Systems

Thanks!

"I stand by all the misstatements that I've made."
. . . Governor George W. Bush, Jr. to Sam Donaldson, 8/17/93

Not really! Questions?