Related Work

[Breese, 1992] – Construction of belief and decision networks

[Wellman, Breese, Goldman, 1992] – From knowledge bases to decision models

[Goldman & Charniak, 1993] – A language for construction of belief networks

[Poole, 1993] – Probabilistic Horn abduction and Bayesian networks

[Bacchus, 1993] – Using first–order probability logic for the construction of Bayesian networks

[Haddawy, 1994] – Generating Bayesian Networks from Probability Logic Knowledge Bases

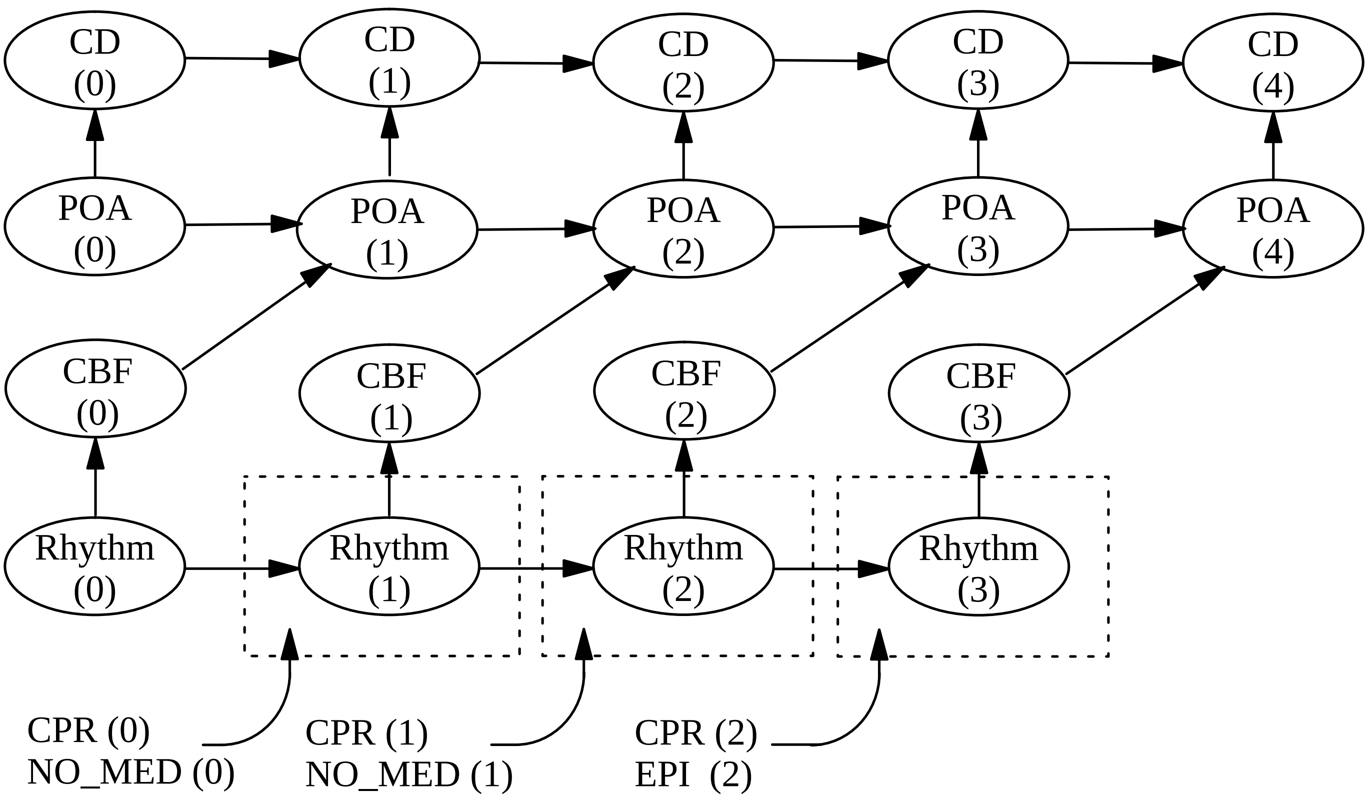
[Glesner & Koller, 1995] – Constructing Flexible Dynamic Belief Networks from First–Order Probabilistic Knowledge Bases

Drowning Victim

E = {rhythm(john,0,a), poa(john,0,5min), cd(john,0,none), cbf(john,0,absent)}

C = {CPR(john,0), CPR(john,1), CPR(john,2), EPI(john,2)}

Q = cd(john,4,V)



P(cd(john,4,none)) = 0.84
P(cd(john,4,mild)) = 0.16
P(cd(john,4,moderate)) = 0.00
P(cd(john,4,severe)) = 0.00

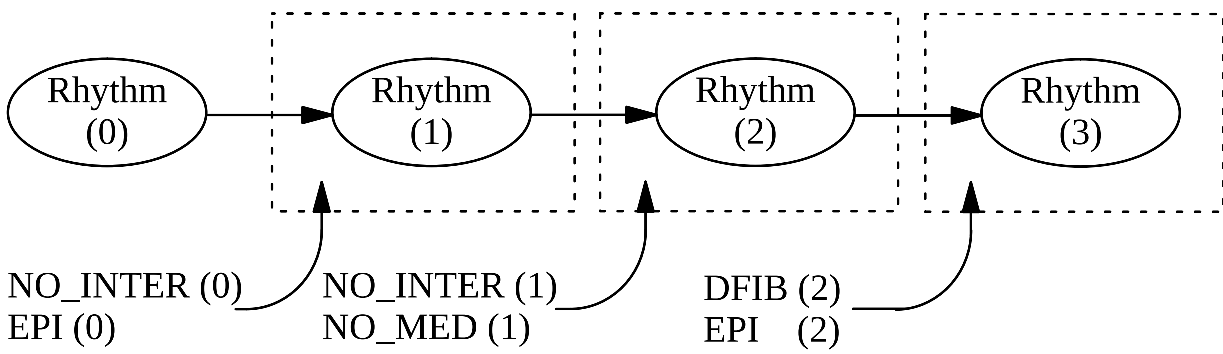
Medical Code Domain Examples

Heart Attack

E = {rhythm(john,0,vf), poa(john,0,none), cd(john,0,none), cbf(john,0, present)}

C = {EPI(john,0), EPI(john,2), DFIB(john,2)}

Q = rhythm(john,3,V)



P(rhythm(john,3,nsr)) = 0.41
P(rhythm(john,3,vf)) = 0.09
P(rhythm(john,3,vt)) = 0.04
P(rhythm(john,3,af)) = 0.00
P(rhythm(john,3,svt)) = 0.01
P(rhythm(john,3,b)) = 0.00
P(rhythm(john,3,a)) = 0.44

Soundness & Completeness

Theorem

Given a complete query $P(Q)=?$, where Q is at time t, $\iota \leq t \leq \tau$, a set of evidence E, a set of context information C, and a KB,

if

- the COMBINE function always generates finite sets of sentences,
- CRPB is (ι,τ)–bound completely quantified and (ι,τ)–bound consistent,
- the proof procedure for $C \cup CB$ is sound and complete wrt any query generated by Q–procedure, and
- Q–procedure stops after a finite amount of time,

then Q–procedure is sound and complete.

Theorem

In any framework $\langle P(Q)=?, E, C, KB \rangle$ where $P(Q)=?$ is a complete query and Q is at constant time t, $\iota \leq t \leq \tau$,

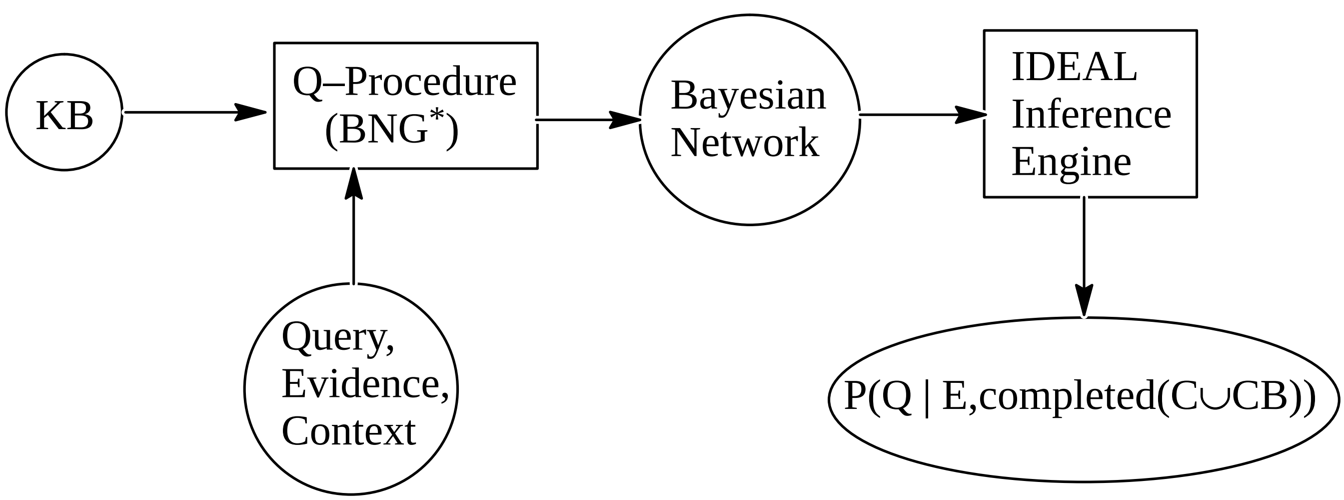
if

- PB and CB are both acyclic,
- $\langle E, C, KB \rangle$ is allowed,
- CRPB is (ι,τ)–bound completely quantified and (ι,τ)–bound consistent,

then Q–procedure is sound and complete.

Constructing the network by building supporting networks is justified by the fact that atoms that influence neither the evidence nor the query are irrrlevant.

The supporting networks are constructed via calls to a BUILD–NET function. BUILD–NET receives as input an atom, whose supporting network needs to be explored. It augments the Bayesian network and returns a set of substitutions such that for each substitution there exists a supporting network for the ground instance of the atom corresponding to the substitution. BUILD–NET frequently calls SLDNF to answer queries on $C \cup CB$.



*Available via www at http://www.cs.uwm.edu/faculty/haddawy

Query Answering Procedure

A **complete query** is a query of the form P(Q) = ?, where the last argument of Q is a variable and the other arguments may also contain variables.

Q–Procedure

1) Build the necessary portion of the Bayesian network, each node of which corresponds to an obj(A), where A is a ground p–atom in RAS.
2) Update the Bayesian network using the set of evidence E.
3) Output the updated beliefs of the query nodes.

Main idea: Build a supporting network for each random variable corresponding to a ground instance of an evidence atom or the query.

Let A be a ground p–atom and consider the set of all ground p–atoms B such that A is influenced by B in CRPB.
The **supporting network** for obj(A) is a Bayesian network consisting of obj(A) and the set of all obj(B), with the relevant influenced–by relations reprsented as links or sequences of links.

Proposition

The set of (ι,τ)–models of CRPB corresponds to a partition of:
1) the set of all (t_0,t_1)–models of CRPB if $t_0 \leq \iota, \tau \leq t_1$, and
2) the set of all possible models of CRPB.

Induced Probability Distribution

A probability distribution which is (ι,τ)–bound induced by the set of evidence E, the set of context information C, and KB is a probability distribution on possible (ι,τ)–models of CRPB satisfying (ι,τ)–CRPB and the independence assumption implied by (ι,τ)–CRPB.

Consistent Completely Quantified CRPB

A completely quantified CRPB is consistent if
1) it contains no cycles, and
2) for all $P(A_0 \mid A_1, \ldots, A_n) = \alpha$ in CRPB,

$$\sum \{\alpha_i \mid P(A_0' \mid A_1, \ldots, A_n) = \alpha_i \in \text{CRPB and } \text{obj}(A_0') = \text{obj}(A_0)\} = 1$$

Theorem

If the (ι,τ)–RAS is finite and the (ι,τ)–CRPB is completely quantified and consistent, then there exists one and only one (ι,τ)–bound induced probability distribution.

Completely Quantified CRPB

A CRPB is completely quantified if
1) for all ground atoms A in RAS, there exists at least one sentence in CRPB with A as the consequent, and
2) for all ground sentences S in CRPB we have a specifica–tion of the probabiltiy of conse(S) given all possible combinations of values of the atoms in ante(S).

Probabilistic Independence Assumption

If $P(A_0 \mid A_1, ..., A_n) = \alpha$ is in CRPB then for all ground p–atoms B which are not in Ext(A_0) and not influenced by A_0, A_0 and B are probabilistically independent given $A_1, ..., A_n$.

Model Theory

We define the semantics by using possible worlds on the Herbrand base. Characterizing the semantics by canonical Herbrand models is widely used in work on logic programming.

Possible (ι,τ)–Model

Given a set of evidence E, a set of context information C, and a KB, a possible (ι,τ)–model M of the corresponding CRPB is a set of atoms in (ι,τ)–RAS such that for all A in (ι,τ)–RAS, Ext(A)$\cap$M has exactly one element.

E = {rhythm(john,0,vf), poa(john,0,none), cd(john,0,none), cbf(john,0,present)}
C = {EPI(john,0), EPI(john,2), DFIB(john,2)}
ι = 0, τ = 3

RAS =
rhythm(john,0,nsr), rhythm(john,0,vf), rhythm(john,0,vt), ...
rhythm(john,1,nsr), rhythm(john,1,vf),rhythm(john,1,vt), ...
cbf(john,0,present), cbf(john,0,absent)
cbf(john,1,present), cbf(john,1,absent), ...
poa(john,0,none), poa(john,0,1min), poa(john,0,2min), ...
cd(john,0,none), cd(john,0,mild), cd(john,0,moderate), ...
...

RPB =
P(rhythm(john,0,nsr)=0.001, P(rhythm(john,0,vf)=0.74, ...
P(poa(john,0,none)=0.99, P(poa(john,0,1min)=0.005, ...
P(cd(john,0,none)) = 0.99, P(cd(john,0,mild) = 0.005, ...
P(cbf(john,0,present) = 0.99, P(cbf(john,0,absent)) = 0.01, ...
...
P(rhythm(john,1,nsr)|rhythm(john,0,nsr))=.05
P(rhythm(john,1,nsr)|rhythm(john,0,vf))=.01
P(rhythm(john,1,nsr)|rhythm(john,0,vt))=.01
P(rhythm(john,1,nsr)|rhythm(john,0,af))=.01
...
P(rhythm(john,1,vf)|rhythm(john,0,nsr))=.10
...
P(rhythm(john,2,nsr)|rhythm(john,1,nsr))=1.0
P(rhythm(john,2,nsr)|rhythm(john,1,vf))=.05
...
P(poa(john,1,3min)|cbf(john,0,present),poa(john,0,2min))=0.0
P(poa(john,1,3min)|cbf(john,0,absent),poa(john,0,2min))=1.0
...
P(cd(john,2,mild)|poa(john,2,3min),cd(john,1,mild)=0.0
P(cd(john,2,mild)|poa(john,2,sustained),cd(john,1,mild)=.98
P(cd(john,2,moderate)|poa(john,2,sustained),cd(john,1,mild)=.02
...

Relevant Probabilistic Sentences (RPB)

Given a set of evidence E, a set of context information C, and a KB, the set of relevant probabilistic sentences (RPB) is defined as the set of all prob(S), where S is a ground probabilistic sentence such that completed($C \cup CB$) $\models$ context(S), cons(S) $\in$ RAS, and ante(S) $\subseteq$ RAS.

Combined Relevant PB (CRPB)

The combined relevant PB (CRPB) is constructed by applying the corresponding combining rules to each maximally coherent set of sentences in RPB which have the same atom in the constequent.

(ι,τ)–Bounded RAS, RPB, CRPB

Given two integers ι,τ ($\iota \leq \tau$), a set of evidence E, a set of context information C, and a KB, the (ι,τ)–RAS is the set $\{A \mid A \in \text{RAS}$ and if A is timed then it is timed at t, $\iota \leq t \leq \tau\}$. The (ι,τ)–RPB and (ι,τ)–CRPB are confined versions of RPB and CRPB on (ι,τ)–RAS.

We can condition the sentences in PB on completed($C \cup CB$) by simply eliminating those where P(context)=0 and by eliminating the contexts from those where P(context) = 1.

P(DFIB(john,1))=1

P(ATRO(john,1))=1

P(NO_INTER(john,1))=0

P(rhythm(X,t,vf) | {rhythm(X,t–1,af)}, {DFIB(X,t–1), ATRO(X,t–1)})=.35

P(rhythm(X,t,nsr) | {rhythm(X,t–1,vt)}, {NO_INTER(X,t–1), EPI(X,t–1)})=.01

P(rhythm(john,2,vf) | {rhythm(john,1,af)}) = .35

Relevant P–Atoms (RAS)

Given a set of evidence E, a set of context information C, and a KB, the set of relevant p–atoms (RAS) is defined recursively:

1) ground(E) $\subseteq$ RAS
2) if S is a ground instance of a probability sentence such that completed ($C \cup CB$) $\models$ context(S) and ante(S) $\subseteq$ RAS, then conse(S) $\in$ RAS.
3) if a p–atom A is in RAS then Ext(A) $\subseteq$ RAS.
4) RAS is the smallest set satisfying the above conditions.

This is similar to the construction of Herbrand least models for Horn programs.

Inference problem:

- Query – Q
- Set of evidence (p–atoms) – E
- Set of context information (c–atoms) – C
- Knowledge base – KB

Compute P(Q|E) within context C.

Semantics

Let $S = P(A_0 \mid A_1, ..., A_n) = \alpha \leftarrow C_1, ..., C_n$

ante(S) = $A_1, ..., A_n$

conse(S) = A_0

context(S) = $C_1, ..., C_n$

Let $A = q(t_1,t_2,v)$

obj(A) = (q,t_1,t_2)

Ext(A) = $\{q(t_1,t_2,v_i), 1 \leq i \leq n\}$

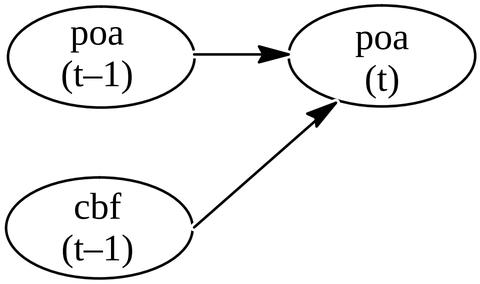
PB: $P(\text{conse} \mid \text{ante})=\alpha \leftarrow \text{context} \equiv$
$\forall x\ P(\text{conse} \mid \text{ante}, \text{context})=\alpha$

CB:

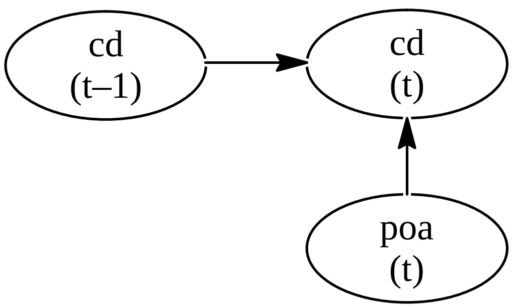
Conceptually, the context information C is a set of observations which we elaborate with the context base CB.

Apply completion semantics to the Herbrand universe of $C \cup CB$ and take completed($C \cup CB$) to hold with probability 1. So for every c–atom C_i either $P(C_i)=0$ or $P(C_i)=1$.

Now computing P(Q|E) within context C amount to computing P(Q | E, completed($C \cup CB$)).



P(poa(X,t,3min)|cbf(X,t–1,present),poa(X,t,2min))=.0
P(poa(X,t,3min)|cbf(X,t–1,absent),poa(X,t,1min))=.0
P(poa(X,t,3min)|cbf(X,t–1,absent),poa(X,t,2min))=1.0
...



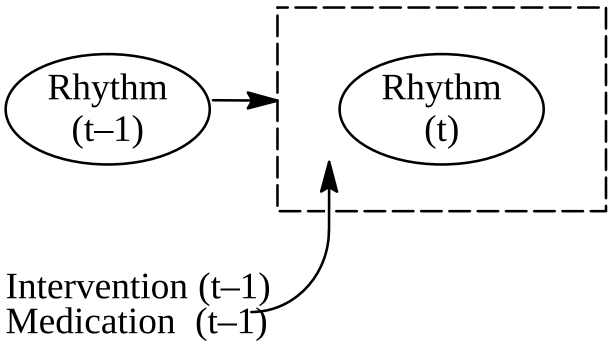
P(cd(X,t,mild)|poa(X,t,3min),cd(X,t–1,mild))=.0
P(cd(X,t,mild)|poa(X,t,sustained),cd(X,t–1,severe))=.0
P(cd(X,t,mild)|poa(X,t,sustained),cd(X,t–1,mild))=.98
P(cd(X,t,moderate)|poa(X,t,sustained),cd(X,t–1,mild))=.02
...

The rules in PB will typically not be a complete specification of a probabiltiy distribution over the ramdom variables represented by the p–atoms.
The specification of the probability of a variable given combinations of values of two or more variables that influence it may not be given. For real applications this information can be difficult to obtain.

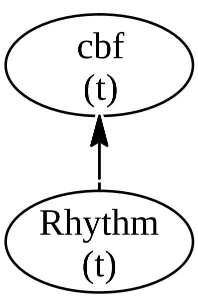
We associate a **combining rule** with each p–predicate.
e.g. generalized noisy–OR

The relations among the p–predicates are defined in the **probabilistic base (PB)**:

P(rhythm(X,0,nsr)) = 0.001, P(rhythm(X,0,vf)) = 0.74, ...
P(poa(X,0,none)) = 0.99, P(poa(X,0,1min)) = 0.005, ...
P(cd(X,0,none)) = 0.99, P(cd(X,0,mild)) = 0.005, ...
P(cbf(X,0,present)) = 0.99, P(cbf(X,0,absent)) = 0.01



P(rhythm(X,t,nsr)|rhythm(X,t–1,nsr))=.05←NO_INTER(X,t–1), EPI(X,t–1)
P(rhythm(X,t,nsr)|rhythm(X,t–1,vf))=.01 ← NO_INTER(X,t–1), EPI(X,t–1)
P(rhythm(X,t,nsr)|rhythm(X,t–1,vt))=.01 ← NO_INTER(X,t–1), EPI(X,t–1)
...
P(rhythm(X,t,vf)|rhythm(X,t–1,af))=.35 ← DFIB(X,t–1), ATRO(X,t–1)
...
P(rhythm(X,t,a)|rhythm(X,t–1,vf))=.15 ← NO_INTER(X,t–1), NO_MED(X,t–1)
...



P(cbf(X,t,present)|rhythm(X,t,nsr))=1.0
P(cbf(X,t,absent)|rhythm(X,t,nsr))=.0
P(cbf(X,t,present)|rhythm(X,t,vf))=.0
P(cbf(X,t,absent)|rhythm(X,t,vf))=1.0
...

The last attribute of a p–predicate represents the value of the corresponding random variable. The values must be mutually exclusive and exhaustive.
This is specified with **predicate declarations (PD):**

rhythm(X: Person, t: time,V)
VAL(rhythm)={nsr, vf, vt, af, svt, b, a}

cbf(X: Person, t: time, V)
VAL(cbf)={present, absent}

poa(X: Person ,t: time, V)
VAL(poa)={none, 1min, 2min, 3min, 4min, 5min, sustained}

cd(X: Person, t: time, V)
VAL(cd)={none, mild, moderate, severe}

Representation Language

Two types of predicates:

- Context predicates – deterministic
- Probabilistic predicates – represent random variables

The relations among c–predicates are defined in the context base, an acyclic normal logic program.

C–predicates:

DFIB(X,t)
CPR(X,t)
NO_INTER(X,t)

LIDO(X,t)
ATRO(X,t)
EPI(X,t)
NO_MED(X,t)

CB:

NO_INTER(X,t) $\leftarrow$ $\neg$DFIB(X,t), $\neg$CPR(X,t)
NO_MED(X,t) $\leftarrow$ $\neg$LIDO(X,t), $\neg$ATRO(X,t), $\neg$EPI(X,t)

Rhythms

- NSR – normal sinus rhythm
- VF – ventricular fibrillation
- VT – ventricular tachycardia
- AF – atrial fibrillation
- SVT – super ventricular tachycardia
- B – brady
- A – asytole

Interventions

- CPR – cardiopulmonary resuscitation
- DFIB – defibrillation

Medications

- LIDO – lidocane
  Slows down heart; restores rhythm
- ATRO – atrophine
  Speeds up heart
- EPI – epinephrine
  Speeds up heart; increases blood pressure

Medical Code Domain

Biomedical Process

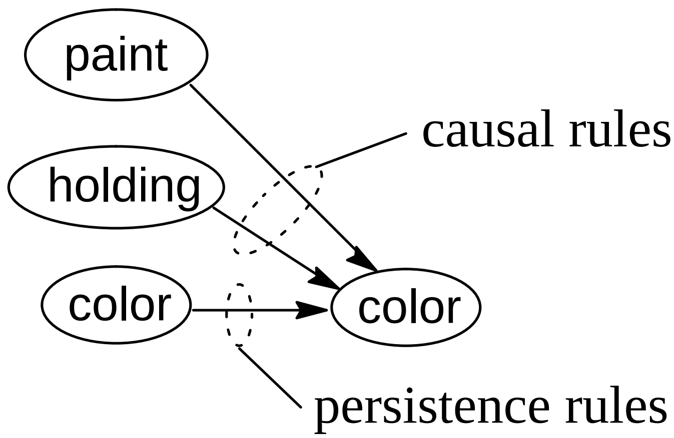
The cardiac arrest can present with several heart rhythms. The particular rhythm determines the amount of cerebral blood flow. The length of time without cerebral blood flow is known as the period of anoxia. Once the period of anoxia has exceed five minutes, cerebral damage may begin to occur. The level of probable cerebral damage depends on the existing level of damage and the length of the period of anoxia.

The Projection Problem

Given a current observed cardiac rhythm, a set of planned medical interventions, a set of planned medications, infer the future rhythm and extent of cerebral damage.

Example: Plan Projection

Actions are typically represented as nodes in the network. This often results in networks with large numbers of nodes and large link matrices. For each node we need both causal rules and persistence rules.



But since when evaluating a plan, the performance of one's own actions is deterministic knowledge, actions can be used as context information.

$$(P(painted(x,t) \mid holding(x,t-1)) = .99) \leftarrow paint(x,t-1)$$

$$(P(painted(x,t) \mid painted(x,t-1)) = .95) \leftarrow \neg paint(x,t-1)$$

The link matrix for the network representing both action effect and persistence is divided between the two types of rules and only one type of rule is applicable at any time.

General Approach: Represent a class of temporal Bayesian networks with a KB of probabilistic rules augmented with context constraints.

Problems

- Inference remains NP–hard
- Cannot effectively do temporal reasoning with a static representation.

Solution

- Treat networks as sets of ground instances of Horn–clauses.

Knowledge–Based Model Construction

- Represent a class of networks with a set of schematic rules
- Construct networks dynamically

Two extreme approaches:

- Emphasis on practical model construction algorithms
  [Breese, 1992], [Goldman & Charniak, 1993]

- Emphasis on formal semantics
  [Poole, 1993], [Bacchus, 1993]

Temporal Probability Model Construction

Common approach: Represent time discretely and create an instance of each time–varying random variable for each point in time.

We can greatly reduce the size of the network models if we can identify some deterministic knowledge and use it as context to index the probabilistic information.

Introduction

Temporal Probabilistic Reasoning

- Diagnosis
- Prediction
- Planning

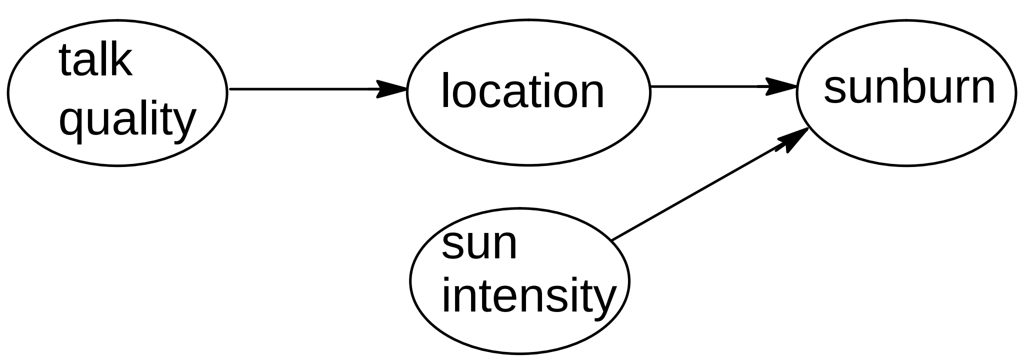
Bayesian Networks

Graphical representation of a probabiltiy distribution:
A DAG in which nodes represent random variables and edges represent probabilistic dependencies.

With each node is associated a conditional probability or link matrix, that encodes the probability of the values of that node given the values of its immediate parents.

Probabilistic independencies are encoded in the network topology.

- Reduces complexity of model specification
- Facilitates model building
- Allows for efficient inference algorithms
- Facilitates explanation generation



Temporal Reasoning with Context–Sensitive Probability Logic

Peter Haddawy

Liem Ngo
Robert Krieger
James Helwig

Dept. of Elect. Eng. & Computer Science
University of Wisconsin–Milwaukee