Bayesian Networks

Knowledge Representation

- Andreas Sauter
- Dec. 2023
- (Content adapted from Erman Acar)

Who am I?

- Ph.D. student at the *Learning and Reasoning* group.
- Working on *RL* and *Causality*
- Background: *Cognitive Science* (Tübingen) and *AI* (VU Amsterdam)
- Online contact: a.sauter@vu.nl or on Slack
- Offline contact: Find me in a random room on the 10th floor @NU

Module Content

1. Foundations

   Degrees of Belief, Belief Dynamics, Independence, Bayes Theorem, Marginalization

2. Bayesian Networks

   Graphs and their Independencies, Bayesian Networks, d-Separation

3. Tools for Inference

   Factors, Variable Elimination, Elimination Order, Interaction Graphs, Graph pruning

4. Exact Inference in Bayesian Networks

   Posterior Marginal, Maximum – A-posteriori, Most Probable Explanation

5. ~~Approximate Inference~~

6. ~~Learning Bayesian Networks~~

Lecture 1: Foundations

Lecture Overview

Introduction

Motivation, Degrees of Beliefs

Belief Dynamics

Properties of Belief, Belief Revision

Independence

Independence, Conditional Independence

Further Properties of Belief

Case Analysis, Chain Rule, Bayes Theorem, Marginalization

Introduction

Motivation

Uncertainty plays a major role in many AI applications like: Decision Making, Image Segmentation, Spam Filtering, Medical Diagnoses, NLP, …

Need for models that allow inference over probabilistic knowledge.

Examples:

- When a patient was treated with a drug, how likely is it they recover?
- Considering the car in front of me is breaking, is it more likely I will crash into it or not?

Bayesian Networks (BN) are such models.

Some Terminology

A world can be thought of as a set of statements

e.g.: There is an earthquake, a burglary is happening and the alarm is ringing

In classical knowledge bases, we have a binary classification of the world

Possible and not possible, true or false, …

A finer classification through a degree of belief or probability

For each world ω, the belief/probability is $\Pr(\omega) \in [0, 1]$

The probability/belief for a given event α:

$$\Pr(\alpha) \overset{\text{def}}{=} \sum_{\omega \models \alpha} \Pr(\omega)$$

Examples

event

$$\Pr(Earthquake)$$
$$= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4)$$
$$= 0.1$$

$$\Pr(Burglary) = 0.2$$

$$\Pr(\neg Burglary) = 0.8$$

$$\Pr(Alarm) = 0.2442$$

Joint probability table

world

world	Earthquake	Burglary	Alarm	Pr(.)
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

Properties of Belief

- $0 \leq \Pr(\alpha) \leq 1$ for any event α
- $\Pr(\alpha) = 0$ when α is inconsistent
- $\Pr(\alpha) = 1$ when α is valid
- $\Pr(\alpha) + \Pr(\neg\alpha) = 1$
- $\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta) - \Pr(\alpha \wedge \beta)$

Example – Partitioned World

Recall: $\Pr(\alpha) + \Pr(\neg\alpha) = 1$

The worlds that satisfy α and those that satisfy ¬α form a partition of the set of all worlds.

world	Earthquake	Burglary	Alarm	Pr(.)
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

$$\Pr(Earthquake)$$
$$= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4)$$
$$= 0.1$$

$$\Pr(\neg Earthquake)$$
$$= \Pr(\omega_5) + \Pr(\omega_6) + \Pr(\omega_7) + \Pr(\omega_8)$$
$$= 0.9$$

Belief Dynamics

Motivation

Now suppose we know that an alarm is triggered. I.e. Alarm = true.

This is called evidence e. We call worlds $\omega \vDash e$, consistent worlds.

While we are certain that Alarm=true, the belief state from the previous examples tells us $\Pr(Alarm = true) = 0.2442$, while we expect it to be 1.

Given evidence α, our goal is to update the state of belief $\Pr(.)$ to $\Pr(.|\alpha)$.

This is called the conditional probability (table).

How to Update Beliefs – Inconsistent Worlds

Assuming we have the evidence α

We expect $\Pr(.|\alpha)$ to assign a belief of 1 to α, hence: $\Pr(\alpha|\alpha) = 1$

This implies that $\Pr(\neg\alpha|\alpha) = 0$

In other words, every world ω that entails $\neg\alpha$ must be assigned the belief 0, hence:

$$\forall\omega \vDash \neg\alpha\colon\ \Pr(\omega|\alpha) = 0$$

How to Update Beliefs – Constraints for Consistent Worlds

But what about worlds that do entail α?

We already know that the sum of all beliefs of these worlds should sum to 1.

$$\Sigma_{\omega \vDash \alpha} \Pr(\omega|\alpha) = 1$$

Further, impossible worlds should stay impossible.

$$\forall \omega \; where \Pr(\omega) = 0 \colon \Pr(\omega|\alpha) = 0$$

And the relative probability of positive probability worlds should stay the same

$$\forall \omega, \omega' \vDash \alpha, \Pr(\omega) > 0, \Pr(\omega') > 0 \colon \frac{\Pr(\omega)}{\Pr(\omega')} = \frac{\Pr(\omega|\alpha)}{\Pr\left(\omega'\middle|\alpha\right)}$$

Updating Beliefs – Consistent Worlds

These constraints leave us with the following option:

$$\Pr(\omega|\alpha) = \frac{\Pr(\omega)}{\Pr(\alpha)} \; for\ all\ \omega \vDash \alpha$$

So we normalise the old belief w.r.t. α.

Together with the case of inconsistent worlds, we update the state of belief as:

$$\Pr(\omega|\alpha) \stackrel{\text{def}}{=} \begin{cases} 0, & if\ \omega \vDash \neg\alpha \\ \dfrac{\Pr(\omega)}{\Pr(\alpha)}, & if\ \omega \vDash \alpha \end{cases}$$

Example

world	Earthquake	Burglary	Alarm	Pr(.)	Pr(.|Alarm)
ω_1	true	true	true	.0190	.0190/.2442
ω_2	true	true	false	.0010	0
ω_3	true	false	true	.0560	.0560/.2442
ω_4	true	false	false	.0240	0
ω_5	false	true	true	.1620	.1620/.2442
ω_6	false	true	false	.0180	0
ω_7	false	false	true	.0072	.0072/.2442
ω_8	false	false	false	.7128	0

$\Pr(Burglary) = 0.2$

$\Pr(Burglary|Alarm) \approx 0.741$

$\Pr(Earthquake) = 0.1$

$\Pr(Earthquake|Alarm) \approx 0.307$

Both probabilities increase when we observe an alarm.

Bayes Conditioning

Sometimes we are only curious in how the belief in a certain event changes without having to consider all consistent worlds.

For two events $\alpha, \beta,$ we can use Bayes conditioning to express

$$P(\alpha \wedge \beta) = \Pr(\alpha|\beta)\Pr(\beta)$$

$$\Pr(\alpha|\beta) = \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)}$$

Note that this is only defined if $\Pr(\beta) \neq 0$

Further Examples

Scenario 1

Evidence: Earthquake = true

$$\begin{aligned}
\Pr(\text{Burglary}) &= .2 \\
\Pr(\text{Burglary}|\text{Earthquake}) &= .2
\end{aligned}$$

$$\begin{aligned}
\Pr(\text{Alarm}) &= .2442 \\
\Pr(\text{Alarm}|\text{Earthquake}) &\approx .75 \uparrow
\end{aligned}$$

Observation: Belief in Burglary does not change while Alarm increases.

Scenario 2

Evidence: Burglary = true

$$\begin{aligned}
\Pr(\text{Alarm}) &= .2442 \\
\Pr(\text{Alarm}|\text{Burglary}) &\approx .905 \uparrow
\end{aligned}$$

$$\begin{aligned}
\Pr(\text{Earthquake}) &= .1 \\
\Pr(\text{Earthquake}|\text{Burglary}) &= .1
\end{aligned}$$

Observation: Belief in Alarm increases while Earthquake stays the same.

Further Examples

Scenario 1

We know: Pr(Burglary) ↑ if Pr(Alarm) ↑and Alarm = true.

Question: How does our belief (on Burglary) change with the new evidence of an Earthquake?

Answer:

$\Pr(\text{Burglary}|\text{Alarm}) \approx .741$

$\Pr(\text{Burglary}|\text{Alarm} \wedge \text{Earthquake}) \approx .253 \downarrow$

Observation: Our belief in Burglary decreases, when we have an explanation for the Alarm.

Scenario 2

We know: Pr(Burglary) ↑ if Pr(Alarm) ↑and Alarm = true.

Question: How does our belief (on Burglary) change with the confirmation of no-Earthquake?

Answer:

$\Pr(\text{Burglary}|\text{Alarm}) \approx .741$

$\Pr(\text{Burglary}|\text{Alarm} \wedge \neg\text{Earthquake}) \approx .957 \uparrow$

Observation: The new evidence of no-Earthquake strengthens our belief on Burglary (as an explanation for the Alarm).

Independence

Independence

Recall that $\Pr(Burglary) = 0.2$ and $\Pr(Burglary|Earthquake) = 0.2$

We said Burglary is independent (denoted $\perp\!\!\!\perp$) of Earthquake.

We formalize this for two events α, β as:

$$\alpha \perp\!\!\!\perp \beta \; iff \; \Pr(\alpha|\beta) = \Pr(\alpha) \qquad (or \; \Pr(\beta) = 0)$$

Alternatively we can say

$$\alpha \perp\!\!\!\perp \beta \; iff \; \Pr(\alpha \wedge \beta) = \Pr(\alpha)\Pr(\beta)$$

Independence can Change

Independence is a dynamic notion: two events which are independent can become dependent after a new evidence.

Example:

Recall that $Burglary \perp\!\!\!\perp Earthquake$

Now consider $\Pr(Burglary|Alarm) \approx 0.741$

Adding evidence $Earthquake$: $\Pr(Burglary|Alarm \land Earthquake) \approx 0.253$

So, given Alarm, Burglary is not independent of Earthquake anymore.

Conditional Independence

The formalization of this dynamic notion is called conditional independence.

Given a state of belief Pr, an event α is conditionally independent from an event β given γ written $\alpha \perp\!\!\!\perp \beta \mid \gamma$ iff:

$$\Pr(\alpha|\beta \wedge \gamma) = \Pr(\alpha|\gamma) \qquad (or\ \Pr(\beta \wedge \gamma) = 0)$$

Alternatively:

$$\Pr(\alpha \wedge \beta|\gamma) = \Pr(\alpha|\gamma)\Pr(\beta|\gamma) \qquad (or\ \Pr(\gamma) = 0)$$

Furthermore: conditional independence is symmetric ($\alpha \perp\!\!\!\perp \beta \mid \gamma \Leftrightarrow \beta \perp\!\!\!\perp \alpha \mid \gamma$)

Further Properties of Belief

Chain Rule

Recall Bayes conditioning: $\Pr(\alpha|\beta) = \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)} \Leftrightarrow \Pr(\alpha \wedge \beta) = \Pr(\alpha|\beta) \Pr(\beta)$

With multiple events $\alpha_1, \alpha_2, \dots, \alpha_n$ we can generalize to the chain rule:

$$\Pr(\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_n) = \Pr(\alpha_1 | \alpha_2 \wedge \cdots \wedge \alpha_n) \Pr(\alpha_2 | \alpha_3 \wedge \cdots \wedge \alpha_n) \dots \Pr(\alpha_n)$$

This rule will be very important for Bayesian networks later on

Disjointness and Exhaustiveness

Important terms for a set of events $\beta_1, \ldots, \beta_n$:

The events are called mutually exclusive (or logically disjoint) iff:

$$\{\omega | \omega \vDash \beta_j\} \cap \{\omega | \omega \vDash \beta_k\} = \emptyset, \text{ for } j \neq k$$

where $\{\omega | \omega \vDash \beta_i\}$ is the set of worlds which entail β_i.

The events are called exhaustive iff:

$$\bigcup_{i=1}^{n} \{\omega | \omega \vDash \beta_i\} = \Omega, \text{ where } \Omega \text{ is the set of all worlds}$$

Marginalization/Law of Total Probability

Assume a set of events $\beta_1, \ldots, \beta_n$ that are mutually exclusive and exhaustive, then the law of total probability/marginalization states:

$$\Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha \wedge \beta_i) = \sum_{i=1}^{n} \Pr(\alpha|\beta_i)\Pr(\beta_i)$$

Examples:

$$\Pr(\alpha) = \Pr(\alpha \wedge \beta) + \Pr(\alpha \wedge \neg\beta)$$
$$\Pr(\alpha) = \Pr(\alpha|\beta)\Pr(\beta) + \Pr(\alpha|\neg\beta)\Pr(\neg\beta)$$

The examples hold, because β and $\neg\beta$ are mutually exclusive and exhaustive.

This is useful since in many cases it is easier to compute the belief for a specific case than for the whole α.

Bayes‘ Theorem

Sometimes we want to know the belief in a cause, given its effect, $\Pr(\alpha|\beta)$

e.g. The belief in a medical condition given a symptom.

Typically the effect given its cause, $\Pr(\beta|\alpha)$ is more readily available

e.g. Given a medical condition, we know the probability of a symptom.

With Bayes’ theorem, we can compute $\Pr(\alpha|\beta)$ from $\Pr(\beta|\alpha)$ by

$$\Pr(\alpha|\beta) = \frac{\Pr(\beta|\alpha)\Pr(\alpha)}{\Pr(\beta)}$$

Lecture 1: Summary

- We defined beliefs in a statement.
- We considered, how beliefs change when we have evidence.
- We formalized how two events can be (conditionally) independent.
- We considered further properties of beliefs which will later help us to deal with Bayesian networks.

Proof of Bayes Conditioning

Recall

For two events $\alpha, \beta,$ we can use Bayes conditioning to express

$$\Pr(\alpha|\beta) = \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)}$$

Note that this is only defined if $\Pr(\beta) \neq 0$

Bayes Conditioning – Why does it work?

$$
\begin{aligned}
\Pr(\alpha|\beta) &= \textstyle\sum_{\omega \models \alpha} \Pr(\omega|\beta) && \text{, computing the probability of an event} \\
&= \textstyle\sum_{\omega \models \alpha \wedge \beta} \Pr(\omega|\beta) + \sum_{\omega \models \alpha \wedge \neg\beta} \Pr(\omega|\beta), && \text{through partitioned worlds} \\
&= \textstyle\sum_{\omega \models \alpha \wedge \beta} \Pr(\omega|\beta) + 0 && \text{, because of impossible worlds} \\
&= \textstyle\sum_{\omega \models \alpha \wedge \beta} \Pr(\omega)/\Pr(\beta) && \text{, through belief update rule} \\
&= \frac{1}{\Pr(\beta)} \textstyle\sum_{\omega \models \alpha \wedge \beta} \Pr(\omega) && \text{, since } \Pr(\beta) \text{ is a constant} \\
&= \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)} && \text{, definition of the probability events}
\end{aligned}
$$

Bayes’ Theorem Example

Bayes' Theorem - Example

Suppose that we have a patient who was just tested for a particular disease and the test came out positive. We know that one in every thousand people has this disease. We also know that the test is not completely reliable: it has a false positive rate of 2% and a false negative rate of 5%.

What should be our belief in the patient having the disease given that the test came out positive?

Example – Solution 1

Our prior belief in the patient having the disease before we run any tests: $\Pr(D) = \frac{1}{1{,}000}$

The false positive rate of the test is $\Pr(T|\neg D) = \frac{2}{100}$ hence, $\Pr(\neg T|\neg D) = \frac{98}{100}$

Similarly, the false negative rate is $\Pr(\neg T|D) = \frac{5}{100}$ hence, $\Pr(T|D) = \frac{95}{100}$

Example – Solution 1

Using Bayes rule, we get $$\Pr(D|T) = \frac{\frac{95}{100} \times \frac{1}{1{,}000}}{\Pr(T)}$$

$\Pr(T)$ is not readily available, but can be obtained by case analysis:

$$\begin{aligned}\Pr(T) &= \Pr(T|D)\Pr(D) + \Pr(T|\neg D)\Pr(\neg D)\\ &= \frac{95}{100} \times \frac{1}{1{,}000} + \frac{2}{100} \times \frac{999}{1{,}000} = \frac{2{,}093}{100{,}000}\end{aligned}$$

which yields,

$$\boxed{\Pr(D|T) = \frac{95}{2{,}093} \approx 4.5\%}$$

Example – Solution 2

Because we have only two events of interest, T and D, leading to only four worlds, this solution is feasible.

world	D	T	
ω_1	true	true	has disease, test positive
ω_2	true	false	has disease, test negative
ω_3	false	true	has no disease, test positive
ω_4	false	false	has no disease, test negative

which gives rise to …

$$\Pr(\omega_1) = \Pr(T \wedge D) = \Pr(T|D)\Pr(D)$$

$$\Pr(\omega_2) = \Pr(\neg T \wedge D) = \Pr(\neg T|D)\Pr(D)$$

$$\Pr(\omega_3) = \Pr(T \wedge \neg D) = \Pr(T|\neg D)\Pr(\neg D)$$

$$\Pr(\omega_4) = \Pr(\neg T \wedge \neg D) = \Pr(\neg T|\neg D)\Pr(\neg D).$$

Example – Solution 2

$$\Pr(\omega_1) = \Pr(T \wedge D) = \Pr(T|D)\Pr(D)$$
$$\Pr(\omega_2) = \Pr(\neg T \wedge D) = \Pr(\neg T|D)\Pr(D)$$
$$\Pr(\omega_3) = \Pr(T \wedge \neg D) = \Pr(T|\neg D)\Pr(\neg D)$$
$$\Pr(\omega_4) = \Pr(\neg T \wedge \neg D) = \Pr(\neg T|\neg D)\Pr(\neg D).$$

..whose values are readily available in the problem setting, which yields:

world	D	T	$\Pr(.)$			
ω_1	true	true	$95/100$	$\times$	$1/1{,}000$	$= .00095$
ω_2	true	false	$5/100$	$\times$	$1/1{,}000$	$= .00005$
ω_3	false	true	$2/100$	$\times$	$999/1{,}000$	$= .01998$
ω_4	false	false	$98/100$	$\times$	$999/1{,}000$	$= .97902$

$$\boxed{\frac{Pr(\omega_1)}{Pr(\omega_1) + Pr(\omega_3)} \approx 4.5\%}$$