Inferenz in Credalnetzen

Seminar: (Imprecise) Probabilistic Graphical Models

Tobias Steinherr

Betreuung: Paul Fink

Institut für Statistik, LMU München

16. Januar 2016

Überblick

Gegeben sei ein gerichteter azyklischer Graph (**DAG**) mit binären Zufallsvariablen $X_i \in X$ mit den Ausprägungen $X_i = x_i$ (X_i tritt ein) oder $X_i = x_i^c$ (X_i tritt nicht ein)

Überblick

Gegeben sei ein gerichteter azyklischer Graph (**DAG**) mit binären Zufallsvariablen $X_i \in X$ mit den Ausprägungen $X_i = x_i$ (X_i tritt ein) oder $X_i = x_i^c$ (X_i tritt nicht ein)

- Jede Variable X_i kann von einer oder mehreren weiteren Variablen $pa(X_i)$ ('Eltern') direkt beeinflusst werden

Überblick

Gegeben sei ein gerichteter azyklischer Graph (**DAG**) mit binären Zufallsvariablen $X_i \in X$ mit den Ausprägungen $X_i = x_i$ (X_i tritt ein) oder $X_i = x_i^c$ (X_i tritt nicht ein)

- Jede Variable X_i kann von einer oder mehreren weiteren Variablen $pa(X_i)$ ('Eltern') direkt beeinflusst werden

$\Rightarrow$ $P(X_i|pa(X_i))$ bezeichnet die bedingte Wahrscheinlichkeit für eine Ausprägung von X_i gegeben einer Ausprägung seiner Eltern (Hat X_i keine Eltern, ist $P(X_i|pa(X_i)) = P(X_i)$)

Überblick

Gegeben sei ein gerichteter azyklischer Graph (**DAG**) mit binären Zufallsvariablen $X_i \in X$ mit den Ausprägungen $X_i = x_i$ (X_i tritt ein) oder $X_i = x_i^c$ (X_i tritt nicht ein)

- Jede Variable X_i kann von einer oder mehreren weiteren Variablen $pa(X_i)$ ('Eltern') direkt beeinflusst werden

$\Rightarrow$ $P(X_i|pa(X_i))$ bezeichnet die bedingte Wahrscheinlichkeit für eine Ausprägung von X_i gegeben einer Ausprägung seiner Eltern (Hat X_i keine Eltern, ist $P(X_i|pa(X_i)) = P(X_i)$)

- NEU: Bedingte und unbedingte Wahrscheinlichkeiten können nun nicht nur einen präzisen Wert annehmen, sondern beispielsweise intervallwertigen (Unterschied zwischen herkömmlichen Bayes- und Credalnetzen)

Überblick

Gegeben sei ein gerichteter azyklischer Graph (**DAG**) mit binären Zufallsvariablen $X_i \in X$ mit den Ausprägungen $X_i = x_i$ (X_i tritt ein) oder $X_i = x_i^c$ (X_i tritt nicht ein)

- Jede Variable X_i kann von einer oder mehreren weiteren Variablen $pa(X_i)$ ('Eltern') direkt beeinflusst werden

$\Rightarrow$ $P(X_i|pa(X_i))$ bezeichnet die bedingte Wahrscheinlichkeit für eine Ausprägung von X_i gegeben einer Ausprägung seiner Eltern (Hat X_i keine Eltern, ist $P(X_i|pa(X_i)) = P(X_i)$)

- NEU: Bedingte und unbedingte Wahrscheinlichkeiten können nun nicht nur einen präzisen Wert annehmen, sondern beispielsweise intervallwertigen (Unterschied zwischen herkömmlichen Bayes- und Credalnetzen)

$\Rightarrow$ Welche Auswirkungen hat das vor allem auf die Inferenz?

Motivation für impräzise Wahrscheinlichkeiten (I)

- Wahrscheinlichkeit für das Auftreten einer Zufallsvariable R ist von Interesse
- Zu 10 Zeitpunkten wird untersucht, ob R eintritt

Motivation für impräzise Wahrscheinlichkeiten (I)

- Wahrscheinlichkeit für das Auftreten einer Zufallsvariable R ist von Interesse
- Zu 10 Zeitpunkten wird untersucht, ob R eintritt

Zeitpunkt i	R_i
1	r
2	r^c
3	r^c
4	r
5	r
6	r^c
7	r^c
8	r^c
9	r

Motivation für impräzise Wahrscheinlichkeiten (I)

- Wahrscheinlichkeit für das Auftreten einer Zufallsvariable R ist von Interesse
- Zu 10 Zeitpunkten wird untersucht, ob R eintritt

Zeitpunkt i	R_i
1	r
2	r^c
3	r^c
4	r
5	r
6	r^c
7	r^c
8	r^c
9	r
10	???

Motivation für impräzise Wahrscheinlichkeiten (II)

Zum letzten Zeitpunkt konnte kein Wert beobachtet werden. $P(R = r)$ soll über die relative Häufigkeit von r bestimmt werden. Welchen Wert angeben?

Motivation für impräzise Wahrscheinlichkeiten (II)

Zum letzten Zeitpunkt konnte kein Wert beobachtet werden. $P(R = r)$ soll über die relative Häufigkeit von r bestimmt werden. Welchen Wert angeben?

$\Rightarrow$ Verfügbare Daten verwenden: $P(R = r) = \frac{4}{9}$

Motivation für impräzise Wahrscheinlichkeiten (II)

Zum letzten Zeitpunkt konnte kein Wert beobachtet werden. $P(R = r)$ soll über die relative Häufigkeit von r bestimmt werden. Welchen Wert angeben?

$\Rightarrow$ Verfügbare Daten verwenden: $P(R = r) = \frac{4}{9}$

$\Rightarrow$ Beide Möglichkeiten für zehnten Wert in Betracht ziehen und z.B. ...

- $\Rightarrow$ $P(R = r) \in \{0.4, 0.5\}$ oder

Motivation für impräzise Wahrscheinlichkeiten (II)

Zum letzten Zeitpunkt konnte kein Wert beobachtet werden. $P(R = r)$ soll über die relative Häufigkeit von r bestimmt werden. Welchen Wert angeben?

$\Rightarrow$ Verfügbare Daten verwenden: $P(R = r) = \frac{4}{9}$

$\Rightarrow$ Beide Möglichkeiten für zehnten Wert in Betracht ziehen und z.B. ...

- $\Rightarrow$ $P(R = r) \in \{0.4, 0.5\}$ oder
- $\Rightarrow$ $P(R = r) \in [0.4, 0.5]$

angeben und damit weiterarbeiten (Impräzise Wahrscheinlichkeiten)

Motivation für impräzise Wahrscheinlichkeiten (II)

Zum letzten Zeitpunkt konnte kein Wert beobachtet werden. $P(R = r)$ soll über die relative Häufigkeit von r bestimmt werden. Welchen Wert angeben?

$\Rightarrow$ Verfügbare Daten verwenden: $P(R = r) = \frac{4}{9}$

$\Rightarrow$ Beide Möglichkeiten für zehnten Wert in Betracht ziehen und z.B. ...

- $\Rightarrow$ $P(R = r) \in \{0.4, 0.5\}$ oder
- $\Rightarrow$ $P(R = r) \in [0.4, 0.5]$

angeben und damit weiterarbeiten (Impräzise Wahrscheinlichkeiten)

- Mit impräzisen Wahrscheinlichkeiten soll Ausmaß der Unsicherheit (Ambiguität) über das Eintreten von Ereignissen ausgedrückt werden

Updating

- Sei nun in einem Bayesnetz die Ausprägung einer oder mehrerer Variablen X_E bekannt (Es liegt Evidenz vor)

Updating

- Sei nun in einem Bayesnetz die Ausprägung einer oder mehrerer Variablen X_E bekannt (Es liegt Evidenz vor)
- Von Interesse sei nun die Eintrittswahrscheinlichkeit einer oder mehrerer weiteren Variablen X_Q gegeben dieser Zusatzinformation, also $P(X_Q|X_E)$

Updating

- Sei nun in einem Bayesnetz die Ausprägung einer oder mehrerer Variablen X_E bekannt (Es liegt Evidenz vor)
- Von Interesse sei nun die Eintrittswahrscheinlichkeit einer oder mehrerer weiteren Variablen X_Q gegeben dieser Zusatzinformation, also $P(X_Q|X_E)$

$$P(Q|E) = \frac{P(X_Q, X_E)}{P(X_E)} = \frac{\sum\limits_{X_i \in X \setminus \{X_Q \cup X_E\}} \prod\limits_{X_i \in X} P(X_i|pa(X_i))}{\sum\limits_{X_i \in X \setminus \{X_Q \cup X_E\}} \prod\limits_{X_i \in X \setminus X_Q} P(X_i|pa(X_i))}$$

$\Rightarrow$ Berechnung dieser Wahrscheinlichkeit wird 'Updating' genannt

- Unterschied zum Updating in Credalnetzen?

Updating

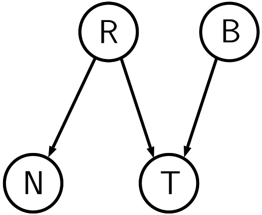
- Sei nun in einem Bayesnetz die Ausprägung einer oder mehrerer Variablen X_E bekannt (Es liegt Evidenz vor)
- Von Interesse sei nun die Eintrittswahrscheinlichkeit einer oder mehrerer weiteren Variablen X_Q gegeben dieser Zusatzinformation, also $P(X_Q|X_E)$

$$P(Q|E) = \frac{P(X_Q, X_E)}{P(X_E)} = \frac{\sum\limits_{X_i \in X \setminus \{X_Q \cup X_E\}} \prod\limits_{X_i \in X} P(X_i|pa(X_i))}{\sum\limits_{X_i \in X \setminus \{X_Q \cup X_E\}} \prod\limits_{X_i \in X \setminus X_Q} P(X_i|pa(X_i))}$$

$\Rightarrow$ Berechnung dieser Wahrscheinlichkeit wird 'Updating' genannt

- Unterschied zum Updating in Credalnetzen? $\Rightarrow$ Später!

Beispiel (I)



- $P(r) \in [0.4, 0.5]$
- $P(b) \in [0.4, 0.5]$
- $P(n|r) = 0.7 \quad P(n|r^c) = 0.2$
- $P(t|r,b) = 0.9 \quad P(t|r,b^c) = 0.3 \quad P(t|r^c,b) = 0.8 \quad P(t|r^c,b^c) = 0.1$

Beispiel (II)

- Es sei nun bekannt, dass $N = n$

Beispiel (II)

- Es sei nun bekannt, dass $N = n$
- Gesucht wird nach $P(t|n)$

Beispiel (II)

- Es sei nun bekannt, dass $N = n$
- Gesucht wird nach $P(t|n)$
- Wären alle Wahrscheinlichkeitsmaße präzise angegeben, so wäre aufgrund der Markov-Bedingung

$$P(t|n) = \frac{P(t,n)}{P(n)} = \frac{\sum_R \sum_B p(t|R,B) \cdot P(n|R) \cdot P(R) \cdot P(B)}{\sum_R \sum_B P(n|R) \cdot P(R) \cdot P(B)}$$

Beispiel (II)

- Es sei nun bekannt, dass $N = n$
- Gesucht wird nach $P(t|n)$
- Wären alle Wahrscheinlichkeitsmaße präzise angegeben, so wäre aufgrund der Markov-Bedingung

$$P(t|n) = \frac{P(t,n)}{P(n)} = \frac{\sum_R \sum_B p(t|R,B) \cdot P(n|R) \cdot P(R) \cdot P(B)}{\sum_R \sum_B P(n|R) \cdot P(R) \cdot P(B)}$$

- $P(R) \cdot P(B)$ soll hier die gemeinsame Wahrscheinlichkeitsverteilung von R und B ausdrücken $\Rightarrow$ ohne Weiteres auch umsetzbar bei impräzisen Wahrscheinlichkeiten?

Klassische Unabhängigkeitskonzepte

Zwei Variablen A und B gelten (für $P(A)$ und $P(B)$ präzise!) als stochastisch unabhängig voneinander, wenn ...

- $P(A, B) = P(A) \cdot P(B)$ (Faktorisierungs-Ansatz) oder

Klassische Unabhängigkeitskonzepte

Zwei Variablen A und B gelten (für $P(A)$ und $P(B)$ präzise!) als stochastisch unabhängig voneinander, wenn ...

- $P(A, B) = P(A) \cdot P(B)$ (Faktorisierungs-Ansatz) oder
- $P(A|B) := \frac{P(A,B)}{P(B)} = P(A)$ (Irrelevanz-Ansatz)

Klassische Unabhängigkeitskonzepte

Zwei Variablen A und B gelten (für $P(A)$ und $P(B)$ präzise!) als stochastisch unabhängig voneinander, wenn ...

- $P(A, B) = P(A) \cdot P(B)$ (Faktorisierungs-Ansatz) oder
- $P(A|B) := \frac{P(A,B)}{P(B)} = P(A)$ (Irrelevanz-Ansatz)

$\Rightarrow$ äquivalent, solange $P(B) \neq 0$ gilt!

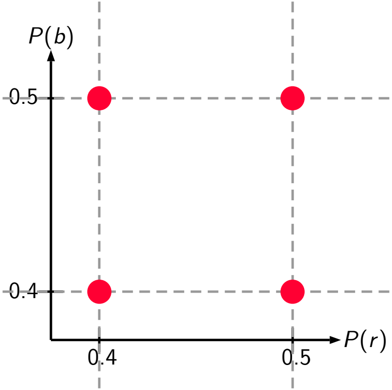
- Diese bzw. ähnliche Konzepte für den Fall der impräzisen Wahrscheinlichkeiten wünschenswert!

Starke Unabhängigkeit (I)

Zwei Variablen heißen stark unabhängig voneinander, wenn für sie unter den Extremalpunkten ihrer Eintrittswahrscheinlichkeiten klassische stochastische Unabhängigkeit gilt.

Starke Unabhängigkeit (I)

Zwei Variablen heißen stark unabhängig voneinander, wenn für sie unter den Extremalpunkten ihrer Eintrittswahrscheinlichkeiten klassische stochastische Unabhängigkeit gilt.



Starke Unabhängigkeit (II)

- Extremalpunkte von $(P(r); P(b))$:
  $(0.4, 0.4), (0.4, 0.5), (0.5, 0.4), (0.5, 0.5)$

Starke Unabhängigkeit (II)

- Extremalpunkte von $(P(r); P(b))$:
  $(0.4, 0.4), (0.4, 0.5), (0.5, 0.4), (0.5, 0.5)$
- Für jeden einzelnen dieser Punkte wird nun über $P(R) \cdot P(B)$ die gemeinsame Wahrscheinlichkeitsverteilung $P(R, B) = (P(r, b), P(r, b^c), P(r^c, b), P(r^c, b^c))$ berechnet (Strong extension)

$\Rightarrow$ Strong extension $\mathcal{M}(R, B)$:

Starke Unabhängigkeit (II)

- Extremalpunkte von $(P(r); P(b))$: $(0.4, 0.4), (0.4, 0.5), (0.5, 0.4), (0.5, 0.5)$
- Für jeden einzelnen dieser Punkte wird nun über $P(R) \cdot P(B)$ die gemeinsame Wahrscheinlichkeitsverteilung $P(R, B) = (P(r, b), P(r, b^c), P(r^c, b), P(r^c, b^c))$ berechnet (Strong extension)

$\Rightarrow$ Strong extension $\mathcal{M}(R, B)$: $((0.16, 0.24, 0.24, 0.36)$, $(0.2, 0.2, 0.3, 0.3)$, $(0.2, 0.3, 0.2, 0.3)$, $(0.25, 0.25, 0.25, 0.25))$

- Gegeben starker Unabhängigkeit kann die gemeinsame Wahrscheinlichkeitsverteilung jeden dieser Extremwerte und alle Konvexkombinationen annehmen

Berechnung der bedingten Wahrscheinlichkeit (I)

- Die vorhin gefragte bedingte Wahrscheinlichkeit $P(t|n)$ kann nun berechnet werden, aber ...

Berechnung der bedingten Wahrscheinlichkeit (I)

- Die vorhin gefragte bedingte Wahrscheinlichkeit $P(t|n)$ kann nun berechnet werden, aber ...
- ... diese wird wieder intervallwertig sein

Berechnung der bedingten Wahrscheinlichkeit (I)

- Die vorhin gefragte bedingte Wahrscheinlichkeit $P(t|n)$ kann nun berechnet werden, aber ...
- ... diese wird wieder intervallwertig sein
- Untere bedingte Wahrscheinlichkeit berechnet sich zu:

$$\underline{P}(t|n) = \min_{P(R,B)\in\mathcal{M}(R,B)} \frac{\sum_R \sum_B p(t|R,B)\cdot P(n|R)\cdot P(R,B)}{\sum_R \sum_B P(n|R)\cdot P(R,B)}$$

(obere bedingte Wahrscheinlichkeit $\overline{P}(t|n)$ analog mit Maximum)

Berechnung der bedingten Wahrscheinlichkeit (I)

- Die vorhin gefragte bedingte Wahrscheinlichkeit $P(t|n)$ kann nun berechnet werden, aber ...
- ... diese wird wieder intervallwertig sein
- Untere bedingte Wahrscheinlichkeit berechnet sich zu:

$$\underline{P}(t|n) = \min_{P(R,B)\in\mathcal{M}(R,B)} \frac{\sum\limits_R \sum\limits_B p(t|R,B) \cdot P(n|R) \cdot P(R,B)}{\sum\limits_R \sum\limits_B P(n|R) \cdot P(R,B)}$$

(obere bedingte Wahrscheinlichkeit $\overline{P}(t|n)$ analog mit Maximum)

$\Rightarrow$ $P(t|n) \in [\underline{P}(t|n), \overline{P}(t|n)]$

Berechnung der bedingten Wahrscheinlichkeit (I)

- Berechnung am Beispiel für $P(R, B) = (0.16, 0.24, 0.24, 0.36) := p_1$:

$$P(t|n, p_1) = \frac{0.9 \cdot 0.7 \cdot 0.16 + 0.3 \cdot 0.7 \cdot 0.24 + ... + ...}{0.7 \cdot 0.16 + 0.7 \cdot 0.24 + 0.2 \cdot 0.24 + 0.2 \cdot 0.36} = 0.492$$

Berechnung der bedingten Wahrscheinlichkeit (I)

- Berechnung am Beispiel für $P(R, B) = (0.16, 0.24, 0.24, 0.36) := p_1$:

$$P(t|n, p_1) = \frac{0.9 \cdot 0.7 \cdot 0.16 + 0.3 \cdot 0.7 \cdot 0.24 + ... + ...}{0.7 \cdot 0.16 + 0.7 \cdot 0.24 + 0.2 \cdot 0.24 + 0.2 \cdot 0.36} = 0.492$$

$\Rightarrow$ Mit Berechnung der anderen Werte für $P(R, B)$ ergibt sich $\underline{P}(t|n) = 0.492$ und $\overline{P}(t|n) = 0.5\overline{6}$

Epistemische Unabhängigkeit

- Versuch, das Konzept der Irrelevanz auf den Fall impräziser Wahrscheinlichkeiten zu übertragen

Epistemische Unabhängigkeit

- Versuch, das Konzept der Irrelevanz auf den Fall impräziser Wahrscheinlichkeiten zu übertragen
- Eine Variable A ist epistemisch irrelevant zu einer Variable B, falls gilt: $\underline{P}(B|A) = \underline{P}(B)$ und $\overline{P}(B|A) = \overline{P}(B)$
- Idee dahinter: Ist die Ausprägung der Variable A bekannt, ergibt sich dadurch kein zusätzliches Wissen über das Intervall der Eintrittswahrscheinlichkeit von B

Epistemische Unabhängigkeit

- Versuch, das Konzept der Irrelevanz auf den Fall impräziser Wahrscheinlichkeiten zu übertragen
- Eine Variable A ist epistemisch irrelevant zu einer Variable B, falls gilt: $\underline{P}(B|A) = \underline{P}(B)$ und $\overline{P}(B|A) = \overline{P}(B)$
- Idee dahinter: Ist die Ausprägung der Variable A bekannt, ergibt sich dadurch kein zusätzliches Wissen über das Intervall der Eintrittswahrscheinlichkeit von B
- Ist B auch epistemisch irrelevant zu A, so heißen A und B epistemisch unabhängig
- **Nicht** äquivalent zur starken Unabhängigkeit, aber:

Epistemische Unabhängigkeit

- Versuch, das Konzept der Irrelevanz auf den Fall impräziser Wahrscheinlichkeiten zu übertragen
- Eine Variable A ist epistemisch irrelevant zu einer Variable B, falls gilt: $\underline{P}(B|A) = \underline{P}(B)$ und $\overline{P}(B|A) = \overline{P}(B)$
- Idee dahinter: Ist die Ausprägung der Variable A bekannt, ergibt sich dadurch kein zusätzliches Wissen über das Intervall der Eintrittswahrscheinlichkeit von B
- Ist B auch epistemisch irrelevant zu A, so heißen A und B epistemisch unabhängig
- **Nicht** äquivalent zur starken Unabhängigkeit, aber: A und B stark unabhängig $\Rightarrow$ A und B epistemisch unabhängig

Updating unter epistemischer Unabhängigkeit

- Updating im Prinzip wie bei starker Unabhängigkeit, nur mit anderen Extrempunkten der gemeinsamen Wahrscheinlichkeitsverteilung mehrerer Variablen
- Berechnung dieser Extrempunkte (Natural Extension $\mathcal{N}(X)$) unter diesem Konzept meist computational

Updating unter epistemischer Unabhängigkeit

- Updating im Prinzip wie bei starker Unabhängigkeit, nur mit anderen Extrempunkten der gemeinsamen Wahrscheinlichkeitsverteilung mehrerer Variablen
- Berechnung dieser Extrempunkte (Natural Extension $\mathcal{N}(X)$) unter diesem Konzept meist computational

$\Rightarrow$ $\mathcal{N}(R, B)$ aus dem Beispiel besteht aus $\mathcal{M}(R, B)$ und den den Werten $(\frac{2}{9}, \frac{2}{9}, \frac{2}{9}, \frac{1}{3})$ und $(\frac{2}{11}, \frac{3}{11}, \frac{3}{11}, \frac{3}{11})$

$\Rightarrow$ Damit ergeben sich keine neuen Werte für $\underline{P}(t|n)$ und $\overline{P}(t|n)$

Fazit und Ausblick

- Impräzise Wahrscheinlichkeiten in Bayesnetzen verlangen bestimmtes Konzept der Unabhängigkeit

Fazit und Ausblick

- Impräzise Wahrscheinlichkeiten in Bayesnetzen verlangen bestimmtes Konzept der Unabhängigkeit
- Diese Konzepte können zu unterschiedlichen Resultaten führen

Fazit und Ausblick

- Impräzise Wahrscheinlichkeiten in Bayesnetzen verlangen bestimmtes Konzept der Unabhängigkeit
- Diese Konzepte können zu unterschiedlichen Resultaten führen
- Neben den vorgestellten Konzepten existieren noch weitere

Fazit und Ausblick

- Impräzise Wahrscheinlichkeiten in Bayesnetzen verlangen bestimmtes Konzept der Unabhängigkeit
- Diese Konzepte können zu unterschiedlichen Resultaten führen
- Neben den vorgestellten Konzepten existieren noch weitere
- Starke Unabhängigkeit ist das geläufigste Konzept, andere werden nach und nach näher erforscht

Literatur

- Thomas Augustin, Frank P.A. Coolen, Gert de Cooman, Matthias C.M. Troffaes: *Introduction to Imprecise Probabilities*
- Fabio G. Cozman: *Credal networks*
- Cassio P. de Campos, Fabio G. Cozman: Vortrag '*Credal networks*'
- Fabio G. Cozman: Separation Properties of Sets of Probability Measures