Probabilistic Relational Models

Seminarvortrag

TECHNISCHE UNIVERSITÄT DARMSTADT

Alexander Galitzki

galickis@gmail.com

Darmstadt, den 2. Dezember 2009

Ziele

TECHNISCHE UNIVERSITÄT DARMSTADT

Was wir in dieser Präsentation kennen lernen:

- den Formalismus und seine Motivation,
- verschiedene Arten von PRMs,
- die exakte und approximative Inferenz,
- die Lernalgorithmen.

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Versuch, Bayes'sche Netze um Elemente der Prädikatenlogik zu erweitern
- Wahrscheinlichkeitsrechnung auf beliebigen Aussagen schwierig:

$$P(\phi) = \sum_{M:\phi \text{ wahr in } M} \mu(M)$$

(ϕ – eine prädikatenlogische Aussage, μ – Wahrscheinlichkeit in einer Welt M)

- Benötigt ist eine Sprache, die:
  1. nur eine endliche Anzahl von Modellen zulässt und
  2. eine konsistente Wahrscheinlichkeitsverteilung ermöglicht.

Definition

Komponenten eines relationalen Schemas

- Klassenmengen $\mathcal{X} = \{X_1, \ldots, X_n\}$,
- Für jede Klasse $X \in \mathcal{X}$: Menge der Attribute $\mathcal{A}(X)$ mit Wertebereich $\mathcal{V}(X.A)$ für jedes Attribut $A \in \mathcal{A}(X)$,
- Für jede Klasse $X \in \mathcal{X}$: Menge der Referenz-Slots ρ_i mit Definitionsbereich X und Bildbereich $Y \in \mathcal{X}$.
- Ein inverser Slot ρ^{-1} ist die inverse Funktion von ρ.
- Eine Slot-Kette $\rho_1, \ldots, \rho_K$ ist eine Abfolge von (ggf. inversen) Slots, so dass für alle i gilt: ρ_i bildet auf den Definitionsbereich von ρ_{i+1} ab.
- Eine Instanz $\mathcal{I}$ gibt für jede Klasse X die Menge der Objekte der Klasse $\mathcal{I}(X)$.

Definition

Komponenten eines relationalen Schemas

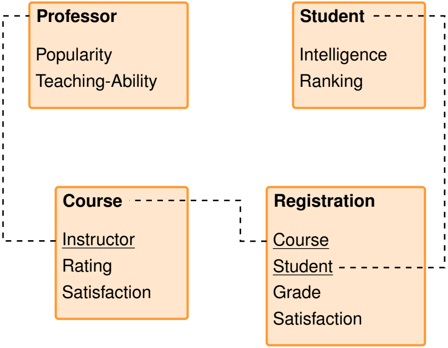
- Klassenmengen $\mathcal{X} = \{X_1, \dots, X_n\}$,
- Für jede Klasse $X \in \mathcal{X}$: Menge der Attribute $\mathcal{A}(X)$ mit Wertebereich $\mathcal{V}(X.A)$ für jedes Attribut $A \in \mathcal{A}(X)$,
- Für jede Klasse $X \in \mathcal{X}$: Menge der Referenz-Slots ρ_i mit Definitionsbereich X und Bildbereich $Y \in \mathcal{X}$.
- Ein inverser Slot ρ^{-1} ist die inverse Funktion von ρ.
- Eine Slot-Kette $\rho_1, \dots, \rho_k$ ist eine Abfolge von (ggf. inversen) Slots, so dass für alle i gilt: ρ_i bildet auf den Definitionsbereich von ρ_{i+1} ab.
- Eine Instanz $\mathcal{I}$ gibt für jede Klasse X die Menge der Objekte der Klasse $\mathcal{I}(X)$.

Definition

Komponenten eines relationalen Schemas

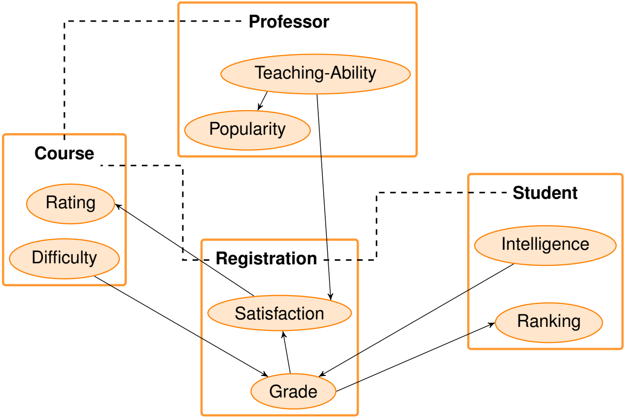
- Klassenmengen $\mathcal{X} = \{X_1, \dots, X_n\}$,
- Für jede Klasse $X \in \mathcal{X}$: Menge der Attribute $\mathcal{A}(X)$ mit Wertebereich $\mathcal{V}(X.A)$ für jedes Attribut $A \in \mathcal{A}(X)$,
- Für jede Klasse $X \in \mathcal{X}$: Menge der Referenz-Slots ρ_i mit Definitionsbereich X und Bildbereich $Y \in \mathcal{X}$.
- Ein inverser Slot ρ^{-1} ist die inverse Funktion von ρ.
- Eine Slot-Kette $\rho_1, \dots, \rho_k$ ist eine Abfolge von (ggf. inversen) Slots, so dass für alle i gilt: ρ_i bildet auf den Definitionsbereich von ρ_{i+1} ab.
- Eine Instanz $\mathcal{I}$ gibt für jede Klasse X die Menge der Objekte der Klasse $\mathcal{I}(X)$.

Beispielschema

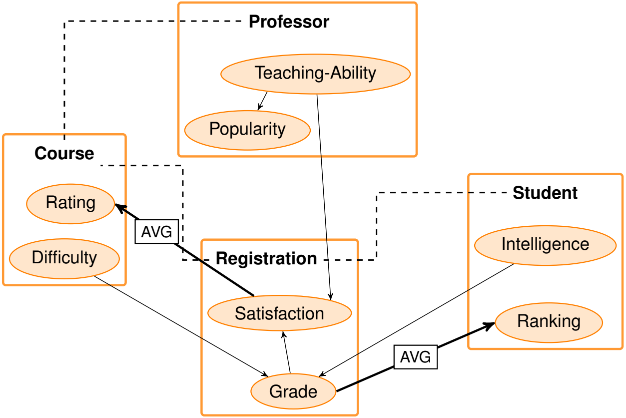


Abhängigkeiten im Beispielschema

TECHNISCHE UNIVERSITÄT DARMSTADT



Abhängigkeiten im Beispielschema



Definition

Probabilistisches Modell

Ein PRM Π für ein relationales Schema definiert sich wie folgt. Für jede Klasse $X \in \mathcal{X}$ und jedes beschreibende Attribut $A \in \mathcal{A}(X)$ gibt es:

- eine Menge der Eltern $\mathrm{Pa}(X.A) = \{U_1, \ldots, U_l\}$, wobei U_i die Form $X.B$ oder $\gamma(X.K.B)$ hat, K eine Slot-Kette und γ eine Aggregatfunktion von $X.K.B$ ist;
- eine gültige bedingte Wahrscheinlichkeitsverteilung $P(X.A | Pa(X.A))$.

Von PRMs zu Bayes'schen Netzen

TECHNISCHE UNIVERSITÄT DARMSTADT

Ein relationales Gerüst σ gibt die Menge der Objekte $\sigma(X_i)$ für jede Klasse an sowie die Relationen zwischen diesen Objekten, nicht jedoch die Attributwerte.

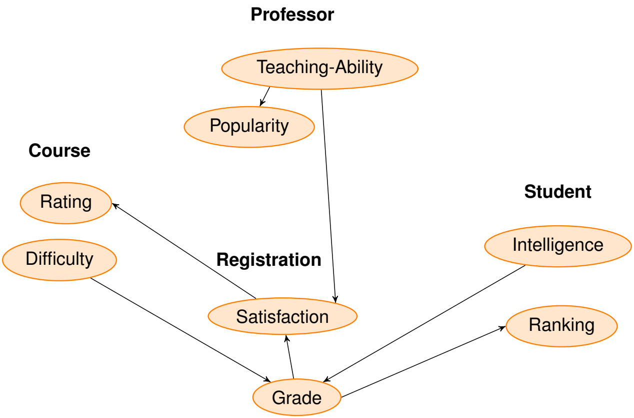
Ein PRM Π zusammen mit einem relationalen Gerüst σ definieren ein Bayes'sches Netz mit Knoten für alle Attribute bzw. Aggregate und Kanten für alle Abhängigkeiten.

$$\begin{aligned} P(\mathcal{I}|\sigma, \mathcal{S}, \theta_{\mathcal{S}}) &= \prod_{x \in \sigma} \prod_{A \in \mathcal{A}(x)} P(\mathcal{I}_{x.A}|\mathcal{I}_{Pa(x.A)}) \\ &= \prod_{X_i} \prod_{A \in \mathcal{A}(X_i)} \prod_{x \in \sigma(X_i)} P(\mathcal{I}_{x.A}|\mathcal{I}_{Pa(x.A)}) \end{aligned}$$

Gültigkeit des probabilistischen Modells

- Eine Instanz besteht aus endlich vielen Objekten (1)
- Aber: keine Garantie, dass $P(\mathcal{I}|\sigma, \mathcal{S}, \theta_{\mathcal{S}})$ eine Dichtefunktion ist. (2)
- Abhängigkeiten dürfen nicht zyklisch sein.
- Einführung von Abhängigkeitsgraphen:
  - für Instanzen (Zyklenfreiheit garantiert Kohärenz für *diese* Instanz)
  - für Klassenabhängigkeiten (Zyklenfreiheit garantiert Kohärenz für *alle* Instanzen)

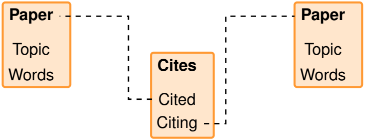
Graph der Klassenabhängigkeiten



Erweiterungen

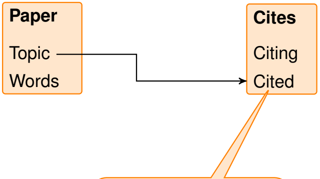
TECHNISCHE UNIVERSITÄT DARMSTADT

- PRM mit struktureller Unsicherheit:
  - Referenzielle Unsicherheit (Belegung der Referenz-Slots unbekannt)
  - Existenzielle Unsicherheit (Existenz einiger Objekte unbekannt)
- PRM mit Klassenhierarchie



Erweiterungen

- PRM mit struktureller Unsicherheit:
  - Referenzielle Unsicherheit (Belegung der Referenz-Slots unbekannt)
  - Existenzielle Unsicherheit (Existenz einiger Objekte unbekannt)



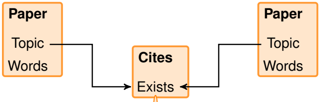
Topic	P1	P2
Theory	0.1	0.9
AI	0.99	0.01

- PRM mit Klassenhierarchie

Erweiterungen

TECHNISCHE UNIVERSITÄT DARMSTADT

- PRM mit struktureller Unsicherheit:
  - Referenzielle Unsicherheit (Belegung der Referenz-Slots unbekannt)
  - Existenzielle Unsicherheit (Existenz einiger Objekte unbekannt)

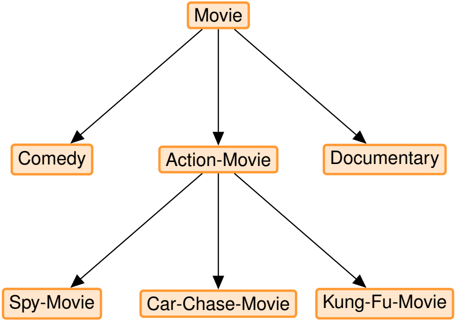


Citing.Topic	Cited.Topic	False	True
Theory	Theory	0.995	0.005
Theory	AI	0.999	0.001
AI	Theory	0.997	0.003
AI	AI	0.993	0.007

- PRM mit Klassenhierarchie

Erweiterungen

- PRM mit struktureller Unsicherheit:
  - Referenzielle Unsicherheit (Belegung der Referenz-Slots unbekannt)
  - Existenzielle Unsicherheit (Existenz einiger Objekte unbekannt)
- PRM mit Klassenhierarchie



Inferenz

TECHNISCHE UNIVERSITÄT DARMSTADT

- Exakt Ausnutzung der speziellen Struktur:
  - Dekomposition,
  - Wiederverwendung der Parameter.
- Approximativ Belief-Propagation-Algorithmus.
  - Idee: Benachbarte Knoten tauschen Informationen aus.
  - Konvergenz nur auf einfach zusammenhängeden Graphen
  - Sonst: kann konvergieren und dann gute Annäherung liefern

Parameterschätzung (I)

TECHNISCHE UNIVERSITÄT DARMSTADT

Log-Likelihood-Funktion gegeben ein Relationsgerüst σ, eine Abhängigkeitsstruktur $\mathcal{S}$, eine Instanz $\mathcal{I}$:

$$
\begin{aligned}
\log L(\theta_{\mathcal{S}} | \mathcal{I}, \sigma, \mathcal{S}) &= \log P(\mathcal{I} | \mathcal{S}, \sigma, \theta_{\mathcal{S}}) \\
&= \sum_{X_i} \sum_{A \in \mathcal{A}(X_i)} \left[\sum_{x \in \sigma(X_i)} \log P(\mathcal{I}_{x.A} | \mathcal{I}_{Pa(x.A)}) \right] \\
&= \sum_{X_i} \sum_{A \in \mathcal{A}(X_i)} \sum_{v \in \mathcal{V}(X.A)} \sum_{\mathbf{u} \in \mathcal{V}(Pa\, X.A)} C_{X.A}[v, \mathbf{u}] \cdot \log \theta_{v|\mathbf{u}},
\end{aligned}
$$

wobei $C_{X.A}[v, \mathbf{u}]$ die Anzahl der Beobachtungen von $\mathcal{I}_{x.A} = v$ und $\mathcal{I}_{Pa(x.A)} = \mathbf{u}$ ist.

Parameterschätzung (II)

TECHNISCHE UNIVERSITÄT DARMSTADT

$C_{X.A}[v, \mathbf{u}]$ können mittels SQL-Anfragen berechnet werden:

```
SELECT grade, intelligence, difficulty, COUNT(*)
FROM registration, student, course
GROUP BY grade, intelligence, difficulty
```

$C_{X.A}[v, \mathbf{u}]$ sind *suffiziente Statistiken* für die Parameter $\theta_{\mathcal{S}}$.

TECHNISCHE UNIVERSITÄT DARMSTADT

Strukturlernen

Wichtige Aspekte:

I. Prüfung der Abhängigkeitsstrukturen auf Zulässigkeit

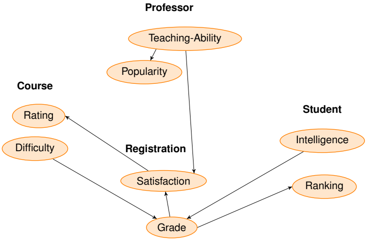
II. Bewertung der Kandidaten

III. Effizienz des Suchverfahrens

Strukturlernen (I)

Zulässige Strukturen

- Konstruktion eines Abhängigkeitsgraphen
- Nach dem Hinzufügen der Kante (u, v) prüfen, ob ein Weg von v nach u existiert
- Komplexität mit Tiefensuche: $O(|V| + |E|)$



Strukturlernen (II)

Strukturbewertung

TECHNISCHE UNIVERSITÄT DARMSTADT

Bewertung einer Struktur $\mathcal{S}$ gegeben eine Instanz $\mathcal{I}$ und ein Gerüst σ:

$$P(\mathcal{S}|\mathcal{I},\sigma) \propto P(\mathcal{I}|\mathcal{S},\sigma)P(\mathcal{S}|\sigma)$$

$P(\mathcal{S}|\sigma) = P(\mathcal{S})$, wobei lange Slot-Ketten penalisiert werden.

$P(\mathcal{I}|\mathcal{S},\sigma) = \int P(\mathcal{I}|\mathcal{S},\theta_{\mathcal{S}},\sigma)P(\theta_{\mathcal{S}}|\mathcal{S})\ d\theta_{\mathcal{S}}$ kann bei bestimmten Verteilungsnnahmen in geschlossener Form dargestellt werden.

Strukturlernen (III)

Struktursuche

TECHNISCHE UNIVERSITÄT DARMSTADT

- Lokale Suche im Raum der zulässigen Strukturen
- Lernen in Bayes-Netzen NP-schwer, daher kein effizientes Verfahren möglich
- Übergangsoperationen: *add edge, delete edge, reverse edge.*
- Verbesserungen: random restart, tabulist, simulated annealing etc.

Zusammenfassung

TECHNISCHE UNIVERSITÄT DARMSTADT

- PRMs erweitern Bayes'sche Netze um prädikatenlogische Elemente.
- PRMs erweitern relationale Schemata um probabilistische Elemente.
- Mit den Erweiterungen (strukturelle Unsicherheit, Klassenhierarchie) kann man weitere Zusammenhänge in den Daten modellieren.
- Ein PRM kann direkt aus einer relationalen Datenbank gelernt werden.