Struktur Vorlesung

- Kapitel 1: Einleitung
- Kapitel 2: Grundlagen
- Kapitel 3: Ausdrucksstärke und Modellkonstruktionen
- Kapitel 4: Tableau Algorithmen
- Kapitel 5: Komplexität
- Kapitel 6: ABoxen und Anfragebeantwortung
- Kapitel 7: Effiziente Beschreibungslogiken

Kapitel 7

Effiziente Beschreibungslogiken

Ziel des Kapitels

Für manche Anwendungen ist $\mathcal{ALC}$ zu komplex:

- Auch hoch-optimierte Reasoner können große Ontologien wie SNOMED CT nicht verarbeiten (oder nur nach intensivem Tuning)
- In der Anfragebeantwortung muss man oft mit sehr grossen Datenmengen umgehen und braucht schnelle Antworten (scalability)

Wir betrachten die Beschreibungslogik $\mathcal{EL}$:

- viel weniger ausdrucksstark als $\mathcal{ALC}$, Basisoperatoren $\sqcap$ und $\exists r.C$
- Erfüllbarkeit und Subsumption in Polyzeit entscheidbar
- Skalierbare Anfragebeantwortung mit Standard-SQL-Datenbanken möglich

EL

Definition 7.1

Ein $\mathcal{EL}$-*Konzept* ist ein $\mathcal{ALC}$-Konzept, in dem nur die Konstruktoren $\top$, $\sqcap$ und $\exists \boldsymbol{r}.\boldsymbol{A}$ verwendet werden.

Beliebt für biomedizinische Ontologien (groß, hoher Abstraktionsgrad):

$$\begin{aligned}
\mathsf{Perikardium} &\sqsubseteq \mathsf{Gewebe} \sqcap \exists \mathsf{teilVon}.\mathsf{Herz} \\
\mathsf{Perikarditis} &\equiv \mathsf{Entzündung} \sqcap \exists \mathsf{ort}.\mathsf{Perikardium} \\
\mathsf{Entzündung} &\sqsubseteq \mathsf{Krankheit} \sqcap \exists \mathsf{wirktAuf}.\mathsf{Gewebe}
\end{aligned}$$

SNOMED CT ist in unwesentlicher Erweiterung von EL formuliert

$\mathcal{EL}$ ist Grundlage des OWL EL Profiles von OWL2

Simulation

Intuitiv: $\mathcal{EL}$ ist “die Hälfte von $\mathcal{ALC}$”, entspricht “der Hälfte von Bisimulation”

Definition 7.2 (Simulation)

Seien $\mathcal{I}_1$ und $\mathcal{I}_2$ Interpretationen

Relation $\rho \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ ist *Simulation* von $\mathcal{I}_1$ nach $\mathcal{I}_2$ wenn

1. $d_1 \; \rho \; d_2$ und $d_1 \in A^{\mathcal{I}_1}$ impliziert $d_2 \in A^{\mathcal{I}_2}$, für alle $A \in \mathsf{N_C}$
2. $d_1 \; \rho \; d_2$ und $(d_1, d_1') \in r^{\mathcal{I}_1}$ impliziert die Existenz eines $d_2' \in \Delta^{\mathcal{I}_2}$ mit $d_1' \; \rho \; d_2'$ und $(d_2, d_2') \in r^{\mathcal{I}_1}$, für alle $r \in \mathsf{N_R}$

T7.1

Beachte: im Ggs. zu Bisimulationen sind Simulationen gerichtet!

Simulation

Seien $\mathcal{I}_1$ und $\mathcal{I}_2$ Interpretationen, $d_1 \in \Delta^{\mathcal{I}_1}$, $d_2 \in \Delta^{\mathcal{I}_2}$.

$(\mathcal{I}_1, d_1) \precsim (\mathcal{I}_2, d_2)$: es gibt Simulation ρ von $\mathcal{I}_1$ nach $\mathcal{I}_2$ mit

$d_1 \; \rho \; d_2$ (wir sagen: d_1 *simuliert* d_2).

Theorem 7.3.
Seien $\mathcal{I}_1, \mathcal{I}_2$ Interpretationen, $d_1 \in \Delta^{\mathcal{I}_1}$ und $d_2 \in \Delta^{\mathcal{I}_2}$.
Wenn $(\mathcal{I}_1, d_1) \precsim (\mathcal{I}_2, d_2)$, dann gilt für alle $\mathcal{EL}$-Konzepte C:
$d_1 \in C^{\mathcal{I}_1}$ impliziert $d_2 \in C^{\mathcal{I}_2}$.

Intuitiv: Wenn $(\mathcal{I}_1, d_1) \precsim (\mathcal{I}_2, d_2)$ und $(\mathcal{I}_2, d_2) \precsim (\mathcal{I}_1, d_1)$,
dann kann $\mathcal{EL}$ nicht zwischen d_1 und d_2 "unterscheiden".

Simulation

Natürlich sind Bisimulation und wechselseitige Simulation nicht dasselbe:

Lemma 7.4.

Es gibt $(\mathcal{I}_1, d_1)$ und $(\mathcal{I}, d_2)$ so dass

- $(\mathcal{I}_1, d_1) \precsim (\mathcal{I}_2, d_2)$ und $(\mathcal{I}_2, d_2) \precsim (\mathcal{I}_1, d_1)$
- $(\mathcal{I}_1, d_1) \not\sim (\mathcal{I}_2, d_2)$

T7.2

EL

In $\mathcal{EL}$ ist Erfüllbarkeit kein interessantes Schlussfolgerungsproblem:

Lemma 7.5

Jedes $\mathcal{EL}$-Konzept ist erfüllbar bzgl. jeder TBox. T7.3

Darum konzentrieren wir uns auf Subsumtion

Wir gehen in 2 Schritten vor:

- ohne TBox: kanonische Modelle für Konzepte
- mit TBox: kanonische Modelle für TBoxen

Subsumtion ohne TBox

Subsumtion ohne TBox

Definition 7.6.

Sei C ein $\mathcal{EL}$-Konzept. Wir setzen

$$\mathsf{ex}(C) = \{C\} \cup \{D \mid \exists r.D \in \mathsf{sub}(C)\}.$$

Das *kanonische Modell* $\mathcal{I}_C$ von C ist die folgende Interpretation:

- $\Delta^\mathcal{I} = \{d_D \mid D \in \mathsf{ex}(C)\}$;
- $A^\mathcal{I} = \{d_D \mid A$ ist ein Konjunkt von $D\}$
- $r^\mathcal{I} = \{(d_D, d_{D'}) \mid \exists r.D'$ ist ein Konjunkt von $D\}$ T7.4

Lemma 7.7.

Für jede Interpretation $(\mathcal{I}, d)$ und alle $D \in \mathsf{ex}(C)$ gilt:

$d \in D^\mathcal{I}$ gdw. $(\mathcal{I}_C, d_D) \precsim (\mathcal{I}, d)$. T7.5

Subsumtion ohne TBox

Da trivialerweise $(\mathcal{I}_C, d_C) \precsim (\mathcal{I}_C, d_C)$, folgt daraus:

Lemma 7.8.
$d_C \in C^{\mathcal{I}_C}$, also ist $\mathcal{I}_C$ Modell von C.

Lemma 7.9.
Für alle $\mathcal{EL}$-Konzepte C, D gilt: $C \sqsubseteq D$ gdw. $d_C \in D^{\mathcal{I}_C}$. T7.6

Theorem 7.10.
Subsumtion in $\mathcal{EL}$ kann in polynomieller Zeit entschieden werden:

- konstruiere $\mathcal{I}_C$ (wegen $|\Delta^{\mathcal{I}_C}| \leq |C|$ in polynomieller Zeit);
- überprüfe in polynomieller Zeit, ob $d_C \in D^{\mathcal{I}_C}$ (Lemma 4.1)

T7.7

Subsumtion mit TBox

Subsumtion mit TBox

TBoxen gehen in Konstruktion der kanonischen Modelle ein:

im Prinzip kanonisches Modell $\mathcal{I}_{C_0,\mathcal{T}}$ für Konzept C_0 und TBox $\mathcal{T}$.

Zwei Einschränkungen o.B.d.A.:

- nur Subsumtion von Konzeptnamen aus $\mathcal{T}$:

  $\mathcal{T} \models C \sqsubseteq D$ gdw. $\mathcal{T}' \models A_C \sqsubseteq A_D$

  wobei $\mathcal{T}' = \mathcal{T} \cup \{A_C \sqsubseteq C, D \sqsubseteq A_D\}$

- in $\mathcal{T}$ nur Existenzrestriktionen der Form $\exists r.A$, $A \in \mathsf{N_C}$:

  $$\exists r.C \quad \rightsquigarrow \quad \exists r.A_C \quad + A_C \equiv C$$

Dann: gemeinsames kanonisches Modell $\mathcal{I}_{\mathcal{T}}$ für alle Konzeptnamen in $\mathcal{T}$

Subsumtion mit TBox

Sei $\mathsf{N}_\mathsf{C}^\mathcal{T}$ die Menge der Konzeptnamen in $\mathcal{T}$

Algorithmus konstruiert Folge $\mathcal{I}_0, \mathcal{I}_1, \ldots$ von Interpretationen, deren Limit ist $\mathcal{I}_\mathcal{T}$.

$\mathcal{I}_0$ ist definiert wie folgt:

- $\Delta^{\mathcal{I}_0} = \{d_A \mid A \in \mathsf{N}_\mathsf{C}^\mathcal{T}\}$;
- $A^{\mathcal{I}_0} = \{d_A\}$ für alle $A \in \mathsf{N}_\mathsf{C}^\mathcal{T}$;
- $r^{\mathcal{I}_0} = \emptyset$ für alle $r \in \mathsf{N}_\mathsf{R}$.

Subsumtion mit TBox

$\mathcal{I}_{i+1}$ erhält man aus $\mathcal{I}_i$ durch einmaliges Anwenden einer Regel:

R1 Wenn $d \in C^{\mathcal{I}_i}$, $C \sqsubseteq D \in \mathcal{T}$, A Konjunkt von D und $d \notin A^{\mathcal{I}_i}$

definiere $\mathcal{I}_{i+1}$ wie $\mathcal{I}_i$, aber setze $A^{\mathcal{I}_{i+1}} = A^{\mathcal{I}_i} \cup \{d\}$

R2 Wenn $d \in C^{\mathcal{I}_i}$, $C \sqsubseteq D \in \mathcal{T}$, $\exists r.A$ Konjunkt von D und $(d, d_A) \notin r^{\mathcal{I}_i}$

definiere $\mathcal{I}_{i+1}$ wie $\mathcal{I}_i$, aber setze $r^{\mathcal{I}_{i+1}} = r^{\mathcal{I}_i} \cup \{(d, d_A)\}$

Man sieht leicht:

- Die Vorbedingungen können in Polyzeit geprüft werden (Lemma 4.1)
- Das Ergebnis ist unabhängig von der Reihenfolge der Regelanwendung
- Es sind max. $|\Delta^{\mathcal{I}_0}| \cdot |\mathsf{sub}(\mathcal{T})| \leq |\mathcal{T}|^2$ Regelanwendungen möglich. T7.8

Subsumtion mit TBox

Lemma 7.11.
Für jedes Modell $(\mathcal{I}, d)$ von $\mathcal{T}$ und alle $A \in \mathsf{N}_\mathsf{C}^\mathcal{T}$ gilt:
$d \in A^\mathcal{I}$ gdw. $(\mathcal{I}_\mathcal{T}, d_A) \precsim (\mathcal{I}, d)$. T7.9

Lemma 7.12.
$\mathcal{I}_\mathcal{T}$ ist Modell von $\mathcal{T}$. T7.10

Lemma 7.13.
Für alle $\mathcal{EL}$-Konzepte D und Konzeptnamen A, B in $\mathcal{T}$ gilt:
$\mathcal{T} \models A \sqsubseteq B$ gdw. $d_A \in B^{\mathcal{I}_\mathcal{T}}$. T7.11

Theorem 7.14.
Subsumtion in $\mathcal{EL}$ bzgl. TBoxen kann in polynomieller Zeit entschieden werden.

Erweiterungen von EL

Erweiterungen von EL

Der Algorithmus kann angepasst werden für $\mathcal{EL}$ erweitert mit:

- $\bot$
- Range Restrictions $\top \sqsubseteq \forall r.C$ und Range Restrictions $\top \sqsubseteq \forall r^-.C$
- allgemeine Rolleninklusionen $r_1 \circ \cdots r_n \sqsubseteq r$
- $\cdots$

Dies (und mehr) ist im OWL EL Profil von OWL2 realisiert.

Viele andere Erweiterungen lassen die Komplexität zurück auf ExpTime springen

Wir betrachten exemplarisch $\mathcal{ELU}$, die Erweiterung von $\mathcal{EL}$ mit $\sqcup$

und $\mathcal{ELU}_\bot$, die Erweiterung von $\mathcal{ELU}$ mit $\bot$

Erweiterungen von EL

Theorem 7.15.

Erfüllbarkeit in $\mathcal{ELU}_\bot$ bzgl. TBoxen ist ExpTime-vollständig.

Beweis: Reduktion von Erfüllbarkeit von Konzeptname A bzgl. $\mathcal{ALC}$-TBox $\mathcal{T}$

Schritt 1: Ersetze Werterestriktionen in $\mathcal{T}$ durch Existenzrestriktionen:

$$\forall r.C \qquad \text{wird} \qquad \neg\exists r.\neg C$$

Schritt 2: Modifiziere $\mathcal{T}$ so dass Negation nur vor Konzeptnamen auftritt:

$$A \sqsubseteq \exists s.(B' \sqcup \neg\exists r.B) \qquad \text{wird} \qquad A \sqsubseteq \exists s.(B' \sqcup \neg X)$$

$$X \doteq \exists r.B$$

(X neuer Konzeptname)

Erweiterungen von EL

Theorem 7.15.

Erfüllbarkeit in $\mathcal{ELU}_\bot$ bzgl. TBoxen ist ExpTime-vollständig.

Beweis: Reduktion von Erfüllbarkeit von Konzeptname A bzgl. $\mathcal{ALC}$-TBox $\mathcal{T}$

Schritt 3: Entferne Negation vollständig aus $\mathcal{T}$:

- Ersetze jedes $\neg X$ durch $\overline{X}$, $\overline{X}$ neuer Konzeptname
- Erzwinge korrektes Verhalten von $\overline{X}$:

$$\top \sqsubseteq X \sqcup \overline{X}$$

$$X \sqcap \overline{X} \sqsubseteq \bot$$

$\mathcal{T}'$ sei die resultierende $\mathcal{ELU}_\bot$-TBox.

T7.12

Lemma 7.16

Für alle Konzeptnamen A gilt: A erfüllbar bzgl. $\mathcal{T}$ gdw. A erfüllbar bzgl. $\mathcal{T}'$.

Erweiterungen von EL

Theorem 7.17.

Subsumtion in $\mathcal{ELU}$ bzgl. TBoxen ist ExpTime-vollständig.

Beweis: Reduktion von Erfüllbarkeit von Konzeptname A bzgl. $\mathcal{ELU}_\bot$-TBox $\mathcal{T}$

Konstruiere $\mathcal{ELU}$-TBox $\mathcal{T}'$:

- nimm o.B.d.A. an, dass $\bot$ nur in der Form $C \sqsubseteq \bot$ vorkommt, ersetze $\bot$ durch neuen Konzeptnamen L
- füge hinzu:

$$\exists r.L \sqsubseteq L \text{ für alle Rollennamen } r \text{ in } \mathcal{T}$$

Lemma 7.18

A unerfüllbar bzgl. $\mathcal{T}$ gdw. $\mathcal{T}' \models A \sqsubseteq L$.

T7.13

Erweiterungen von EL

$\mathcal{EL}^\forall$ ist $\mathcal{EL}$ erweitert um $\forall r.C$

Theorem 7.19. In $\mathcal{EL}^\forall$ ist Subsumtion bzgl. TBoxen ExpTime-vollständig.

Beweis: Reduktion von Subsumtion zwischen Knozeptnamen bzgl. $\mathcal{ELU}$-TBox $\mathcal{T}$

Wir können annehmen, das Disjunktion nur in den folgenden Formen vorkommt:

$$\underbrace{A_1 \sqcup A_2 \sqsubseteq A}_{} \qquad \text{und} \qquad A \sqsubseteq B_1 \sqcup B_2$$

$$= A_1 \sqsubseteq A,\ A_2 \sqsubseteq A$$

ersetze durch

$$A \sqcap \exists r.\top \sqsubseteq B_1$$
$$A \sqcap \forall r.X \sqsubseteq B_2$$

r, X neu

Erweiterungen von EL

$\mathcal{EL}^{\geq 2}$ ist $\mathcal{EL}$ erweitert um $(\geq 2\, r\, \top)$

Theorem 7.20. In $\mathcal{EL}^{\geq 2}$ ist Subsumtion bzgl. TBoxen ExpTime-vollständig.

Beweis: Reduktion von Subsumtion zwischen Knozeptnamen bzgl. $\mathcal{ELU}$-TBox $\mathcal{T}$

Wir können annehmen, das Disjunktion nur in den folgenden Formen vorkommt:

$$\underbrace{A_1 \sqcup A_2 \sqsubseteq A}_{= A_1 \sqsubseteq A,\ A_2 \sqsubseteq A} \qquad \text{und} \qquad A \sqsubseteq B_1 \sqcup B_2$$

ersetze durch

$$A \sqsubseteq \exists r.X \sqcap \exists r.Y$$

$$A \sqcap \exists r.(X \sqcap Y) \sqsubseteq B_1 \qquad r, X, Y \text{ neu}$$

$$A \sqcap (\geq 2\, r) \sqsubseteq B_2$$

Erweiterungen von EL

Erweiterung von $\mathcal{EL}$ ist convex wenn für alle TBoxen $\mathcal{T}$ und Konzepte C, D_1, D_2:

$$\mathcal{T} \models C \sqsubseteq D_1 \sqcup D_2 \quad \text{impliziert} \quad \mathcal{T} \models C \sqsubseteq D_1 \sqcup D_2 \text{ für ein } i \in \{1, 2\}$$

$\mathcal{EL} + \forall r.C$ ist nicht convex:

$$\top \models \exists r.\top \sqcup \forall r.X, \text{ aber } \top \not\models \exists r.\top \text{ und } \top \not\models \forall r.X$$

Die Existenz des kanonischen Modells $\mathcal{I}_\mathcal{T}$ impliziert Konvexität! T7.13

Unsere Beweise zeigen: jede nicht-konvexe Erweiterung von $\mathcal{EL}$ ist ExpTime-hart

Die Gegenrichtung ist leider falsch:
Zum Beispiel hat $\mathcal{EL}$ erweitert mit $\exists r^-.C$ kanonische Modelle,
ist aber ExpTime-vollständig.

Diskussion

Die $\mathcal{EL}$-Familie von BLen:

- Erlaubt Schlußfolgern in polynomieller Zeit
- Es gibt verscheidene Reasoner wie CEL, SNOROCKET und CB
- Skaliert auch auf große Terminologien wie SNOMED CT (> 400.100 Konzepte, wird in < 10 Min. klassifiziert)
- Die Beantwortung konjunktiver Anfragen ist NP-vollständig,...
- ...kann aber skalierbar mit normalen relationalen Datenbanksystemen implementiert werden