ĐẠI HỌC HUẾ
TRƯỜNG ĐẠI HỌC KHOA HỌC

BÁO CÁO ĐỀ TÀI CẤP CƠ SỞ ĐẠI HỌC HUẾ

HỌC KHÁI NIỆM ĐỐI VỚI CÁC CƠ SỞ TRI THỨC TRONG LOGIC MÔ TẢ DỰA VÀO MÔ PHỎNG HAI CHIỀU

Chủ nhiệm đề tài
ThS. TRẦN THANH LƯƠNG

Cán bộ phối hợp thực hiện
TS. HOÀNG THỊ LAN GIAO

Nội dung trình bày

1. Giới thiệu tổng quan
2. Logic mô tả và cơ sở tri thức
3. Mô phỏng hai chiều trong logic mô tả
4. Tính bất biến đối với mô phỏng hai chiều
5. Học khái niệm đối với các cơ sở tri thức trong logic mô tả sử dụng mô phỏng hai chiều
6. Kết luận

Giới thiệu tổng quan

- Logic mô tả có tầm quan trọng đặc biệt → cung cấp mô hình lý thuyết cho các hệ thống ngữ nghĩa và ontology.
- Tìm các khái niệm quan trọng và xây dựng được định nghĩa của các khái niệm đó → học khái niệm trong logic mô tả.
- Học khái niệm trong logic mô tả tương tự như phân lớp nhị phân trong học máy truyền thống.
- Điểm khác: các đối tượng không chỉ được đặc tả bằng các thuộc tính mà còn được đặc tả bằng các mối quan hệ giữa các đối tượng.
- Các mối quan hệ là một trong những yếu tố làm giàu thêm ngữ nghĩa của hệ thống huấn luyện → cần phải tận dụng được chúng như là một lợi thế của các phương pháp học.

Học khái niệm: Ba ngữ cảnh chính

Ngữ cảnh 1: Cho cơ sở tri thức $\mathcal{KB}$ trong logic mô tả L và các tập các cá thể E^+, E^-. Học khái niệm C trong L sao cho:

1. $\mathcal{KB} \models C(a)$ với mọi $a \in E^+$, và
2. $\mathcal{KB} \models \neg C(a)$ với mọi $a \in E^-$,

Ngữ cảnh 2: Ngữ cảnh này khác với ngữ cảnh đã đề cập ở trên là điều kiện thứ hai được thay bằng một điều kiện yếu hơn:

1. $\mathcal{KB} \models C(a)$ với mọi $a \in E^+$, và
2. $\mathcal{KB} \not\models C(a)$ với mọi $a \in E^-$.

Ngữ cảnh 3: Cho một diễn dịch $\mathcal{I}$ và các tập các cá thể E^+, E^-. Học khái niệm C trong logic mô tả L sao cho:

1. $\mathcal{I} \models C(a)$ với mọi $a \in E^+$, và
2. $\mathcal{I} \models \neg C(a)$ với mọi $a \in E^-$.

Chú ý rằng $\mathcal{I} \not\models C(a)$ tương đồng với $\mathcal{I} \models \neg C(a)$.

trong đó, tập E^+ chứa các mẫu dương và E^- chứa các mẫu âm của C.

Vấn đề đặt ra

Ngữ cảnh 2: Cho một diễn dịch $\mathcal{I}$ và các tập các cá thể E^+, E^-. Học khái niệm C trong logic mô tả L sao cho:

1. $\mathcal{KB} \models C(a)$ với mọi $a \in E^+$, và
2. $\mathcal{KB} \not\models C(a)$ với mọi $a \in E^-$.

trong đó, tập E^+ chứa các mẫu dương và E^- chứa các mẫu âm của C.

Mục tiêu chính

1. Nghiên cứu cú pháp, ngữ nghĩa đối với một lớp lớn các logic mô tả. Lớp các logic mô tả này phải có khả năng bao phủ những logic mô tả hữu ích như $\mathcal{SHOIQ}$, $\mathcal{SROIQ}$, ... Trên cơ sở đó xây dựng mô phỏng hai chiều cho lớp các logic mô tả này.
2. Xây dựng phương pháp làm mịn phân hoạch miền của các diễn dịch trong logic mô tả dựa trên mô phỏng hai chiều sử dụng các bộ chọn hợp lý cũng như các độ đo về gia lượng thông tin.
3. Đề xuất các thuật toán học khái niệm dựa trên mô phỏng hai chiều cho các cơ sở tri thức trong logic mô tả với ngữ cảnh (2).

Định nghĩa 1.4 (Ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$)

Cho Σ là bộ ký tự logic mô tả, Φ là tập các đặc trưng của logic mô tả và $\mathcal{L}$ đại diện cho $\mathcal{ALC}_{reg}$. Ngôn ngữ logic mô tả $\mathcal{L}_{\Sigma,\Phi}$ cho phép các *vai trò đối tượng* và các *khái niệm* được định nghĩa một cách đệ quy như sau:

- nếu $r \in \Sigma_{oR}$ thì r là một vai trò đối tượng của $\mathcal{L}_{\Sigma,\Phi}$,
- nếu $A \in \Sigma_C$ thì A là một khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
- ...
- nếu R và S là các vai trò đối tượng của $\mathcal{L}_{\Sigma,\Phi}$, C và D là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$, $r \in \Sigma_{oR}$, $\sigma \in \Sigma_{dR}$, $a \in \Sigma_I$ và n là một số tự nhiên thì
  - ε, $R \circ S$, $R \sqcup S$, R^* và $C?$ là các vai trò đối tượng của $\mathcal{L}_{\Sigma,\Phi}$,
  - $\top$, $\bot$, $\neg C$, $C \sqcap D$, $C \sqcup D$, $\exists R.C$ và $\forall R.C$ là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - ...
  - nếu $\mathrm{Self} \in \Phi$ thì $\exists r.\mathrm{Self}$ là một khái niệm của $\mathcal{L}_{\Sigma,\Phi}$. ■

Xem xét thêm: Thuộc tính, vai trò dữ liệu, tính chất hàm ($\mathcal{F}$), hạn chế số lượng không định tính ($\mathcal{N}$).

Định nghĩa (Cơ sở tri thức)

Một *cơ sở tri thức* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một bộ ba $\mathcal{KB} = \langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$, trong đó $\mathcal{R}$ là một RBox, $\mathcal{T}$ là một TBox và $\mathcal{A}$ là một ABox trong $\mathcal{L}_{\Sigma,\Phi}$. ■

Định nghĩa (Mô hình)

Một diễn dịch $\mathcal{I}$ là một *mô hình* của RBox $\mathcal{R}$ (tương ứng, TBox $\mathcal{T}$, ABox $\mathcal{A}$), ký hiệu là $\mathcal{I} \models \mathcal{R}$ (tương ứng, $\mathcal{I} \models \mathcal{T}$, $\mathcal{I} \models \mathcal{A}$), nếu $\mathcal{I}$ thỏa mãn tất cả các tiên đề vai trò trong $\mathcal{R}$ (tương ứng, tiên đề thuật ngữ trong $\mathcal{T}$, khẳng định cá thể trong $\mathcal{A}$). Một diễn dịch $\mathcal{I}$ là một *mô hình* của cơ sở tri thức $\mathcal{KB} = \langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$, ký hiệu là $\mathcal{I} \models \mathcal{KB}$, nếu nó là mô hình của cả $\mathcal{R}$, $\mathcal{T}$ và $\mathcal{A}$. ■

Suy luận trong logic mô tả

1. Thuật toán bao hàm theo cấu trúc
   - Thực hiện suy luận dựa trên việc so sánh các cấu trúc.
   - Áp dụng cho logic mô tả đơn giản như $\mathcal{FL}_0$, $\mathcal{FL}_\perp$, $\mathcal{ALN}$.
   - Không thực hiện được trên lớp các logic phức tạp như $\mathcal{ALC}$, $\mathcal{ALCI}$, $\mathcal{ALCIQ}$, $\mathcal{SHIQ}$, $\mathcal{SHOIQ}$.
2. Thuật toán tableaux
   - Thực hiện suy luận dựa trên việc triển khai các luật.
   - Áp dụng cho lớp các logic phức tạp như $\mathcal{ALC}$, $\mathcal{ALCI}$, $\mathcal{ALCIQ}$, $\mathcal{SHIQ}$, $\mathcal{SHOIQ}$.
   - Sử dụng trong các bộ suy luận thế hệ mới FaCT, FaCT^{++}, RACER, CEL và KAON 2.
   - Độ phức tạp cao, thường là hàm mũ.

Định nghĩa (Mô phỏng hai chiều)

Cho Σ và $\Sigma^\dagger$ là các bộ ký tự logic mô tả sao cho $\Sigma^\dagger \subseteq \Sigma$, Φ và $\Phi^\dagger$ là tập các đặc trưng của logic mô tả sao cho $\Phi^\dagger \subseteq \Phi$, $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch trong $\mathcal{L}_{\Sigma,\Phi}$. Một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-*mô phỏng hai chiều* giữa $\mathcal{I}$ và $\mathcal{I}'$ là một quan hệ nhị phân $Z \subseteq \Delta^\mathcal{I} \times \Delta^{\mathcal{I}'}$ thỏa các điều kiện sau với mọi $a \in \Sigma_I^\dagger$, $A \in \Sigma_C^\dagger$, $B \in \Sigma_A^\dagger \setminus \Sigma_C^\dagger$, $r \in \Sigma_{oR}^\dagger$, $\sigma \in \Sigma_{dR}^\dagger$, $d \in range(\sigma)$, $x, y \in \Delta^\mathcal{I}$, $x', y' \in \Delta^{\mathcal{I}'}$:

$$Z(a^\mathcal{I}, a^{\mathcal{I}'}) \tag{1}$$

$$Z(x, x') \Rightarrow [A^\mathcal{I}(x) \Leftrightarrow A^{\mathcal{I}'}(x')] \tag{2}$$

$$\dots \tag{3}$$

nếu Self $\in \Phi^\dagger$ thì

$$Z(x, x') \Rightarrow [r^\mathcal{I}(x, x) \Leftrightarrow r^{\mathcal{I}'}(x', x')], \tag{18}$$

trong đó $\#\Gamma$ ký hiệu cho lực lượng của tập hợp Γ. ■

Mô phỏng hai chiều trong logic mô tả

- Mô phỏng hai chiều là một quan hệ nhị phân đặc tả tính tương tự giữa hai trạng thái.
- Nguyen và Szałas đã nghiên cứu về mô phỏng hai chiều và tính không phân biệt được của các đối tượng để áp dụng vào việc học khái niệm trong logic mô tả ($\mathcal{ALC}_{reg} + \mathcal{I}$, $\mathcal{O}$, $\mathcal{Q}$, $\mathcal{U}$, Self).
- Tổng quát hóa và mở rộng các kết quả về mô phỏng hai chiều cho một lớp lớn các logic mô tả.
- Xem xét thêm một số vấn đề sau:
  - Xem xét các thuộc tính như là các phần tử cơ bản,
  - Xem xét vai trò dữ liệu,
  - Xem xét tính chất hàm ($\mathcal{F}$), hạn chế số lượng không định tính ($\mathcal{N}$).

Tính bất biến đối với mô phỏng hai chiều

Định nghĩa (Khái niệm bất biến)

Một khái niệm C được gọi là *bất biến* đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều nếu $Z(x, x')$ thỏa mãn thì $x \in C^{\mathcal{I}}$ khi và chỉ khi $x' \in C^{\mathcal{I}'}$ với mọi diễn dịch $\mathcal{I}$, $\mathcal{I}'$ trong $\mathcal{L}_{\Sigma,\Phi}$ thỏa $\Sigma^\dagger \subseteq \Sigma$, $\Phi^\dagger \subseteq \Phi$ và với mọi $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều Z giữa $\mathcal{I}$ và $\mathcal{I}'$. ■

Định lý

Tất cả các khái niệm của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ đều bất biến đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều. ■

Tính bất biến của khái niệm đối với mô phỏng hai chiều là một trong những tính chất quan trọng trong việc mô hình hóa tính không phân biệt được của các đối tượng.

Tự mô phỏng hai chiều

Định nghĩa (Tự mô phỏng hai chiều)

Cho $\mathcal{I}$ là một diễn dịch trong $\mathcal{L}_{\Sigma,\Phi}$. Một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-*tự mô phỏng hai chiều* của $\mathcal{I}$ là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và chính nó. Một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều Z của $\mathcal{I}$ được gọi là *lớn nhất* nếu với mọi $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều Z' của $\mathcal{I}$ thì $Z' \subseteq Z$. ■

- $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều lớn nhất của $\mathcal{I}$ ký hiệu là $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$.
- Quan hệ hai ngôi $\equiv_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$ trên $\Delta^\mathcal{I}$ là quan hệ thỏa mãn tính chất $x \equiv_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}} x'$ khi và chỉ khi x $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tương đương với x'.
- Nếu $\mathcal{I}$ là một phân nhánh hữu hạn đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ thì quan hệ $\equiv_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$ là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều lớn nhất của $\mathcal{I}$ (nghĩa là, quan hệ $\equiv_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$ và $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$ trùng khớp nhau).

Học khái niệm

Phân hoạch nhất quán

- Tập $Y \subseteq \Delta^{\mathcal{I}}$ bị *phân chia* bởi E nếu tồn tại $a \in E^+$ và $b \in E^-$ sao cho $\{a^{\mathcal{I}}, b^{\mathcal{I}}\} \subseteq Y$.
- Một phân hoạch $\mathbb{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ của $\Delta^{\mathcal{I}}$ được gọi là *nhất quán* với E nếu với mọi $1 \leq i \leq n$, Y_i không bị phân chia bởi E.

Nhận xét

Nếu A_d xác định được trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ bởi một khái niệm C, lúc đó:

- $C^{\mathcal{I}}$ phải là hợp của một số lớp tương đương trong phân hoạch $\mathbb{Y}$ của $\Delta^{\mathcal{I}}$ được phân hoạch thông qua $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$,
- $a^{\mathcal{I}} \in C^{\mathcal{I}}$ với mọi $a \in E^+$ và $a^{\mathcal{I}} \notin C^{\mathcal{I}}$ với mọi $a \in E^-$.

Bộ chọn cơ bản

Định nghĩa 3.4 (Bộ chọn cơ bản)

Một *bộ chọn cơ bản* trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ dùng để phân chia khối Y_{i_j} của phân hoạch $\mathbb{Y} = \{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}\}$ là một khái niệm thuộc một trong các dạng sau:

- A, trong đó $A \in \Sigma^\dagger_C$,
- $A = d$, trong đó $A \in \Sigma^\dagger_A \setminus \Sigma^\dagger_C$ và $d \in range(A)$,
- $\exists\sigma.\{d\}$, trong đó $\sigma \in \Sigma^\dagger_{dR}$ và $d \in range(\sigma)$,
- $\exists r.C_{i_t}$, trong đó $r \in \Sigma^\dagger_{oR}$ và $1 \leq t \leq k$,
- ...
- $\geq l\, r^-.C_{i_t}$ và $\leq m\, r^-.C_{i_t}$, nếu $\{\mathcal{Q},\mathcal{I}\} \subseteq \Phi^\dagger$, $r \in \Sigma^\dagger_{oR}$, $1 \leq t \leq k$, $0 < l \leq \#C^{\mathcal{I}}_{i_t}$ và $0 \leq m < \#C^{\mathcal{I}}_{i_t}$,
- $\exists r.\mathrm{Self}$, nếu $\mathrm{Self} \in \Phi^\dagger$ và $r \in \Sigma^\dagger_{oR}$. ∎

Độ đo dựa trên entropy

Cho $\mathcal{I}$ là một hệ thống thông tin, X và Y là các tập con của $\Delta^{\mathcal{I}}$, trong đó X đóng vai trò là tập các mẫu dương của khái niệm cần học, Y đóng vai trò là một khối của phân hoạch.

Định nghĩa 3.8 (Entropy)

Entropy của tập Y đối với tập X trong miền $\Delta^{\mathcal{I}}$ của hệ thống thông tin $\mathcal{I}$, ký hiệu là $E_{\Delta^{\mathcal{I}}}(Y/X)$, được xác định như sau:

$$E_{\Delta^{\mathcal{I}}}(Y/X) = \begin{cases} 0, \text{ nếu } Y \cap X = \emptyset \text{ hoặc } Y \subseteq X \\ -\dfrac{\#XY}{\#Y}\log_2\dfrac{\#XY}{\#Y} - \dfrac{\#\overline{X}Y}{\#Y}\log_2\dfrac{\#\overline{X}Y}{\#Y}, \text{ nếu ngược lại,} \end{cases} \quad (4)$$

trong đó XY đại diện cho tập $X \cap Y$ và $\overline{X}Y$ đại diện cho tập $\overline{X} \cap Y$. ■

Độ đo dựa trên entropy

Định nghĩa 3.9. (Độ đo gia lượng thông tin)

Gia lượng thông tin của bộ chọn D trong việc chia tập Y đối với tập X trong $\Delta^{\mathcal{I}}$ của hệ thống thông tin $\mathcal{I}$, ký hiệu là $IG_{\Delta^{\mathcal{I}}}(Y/X, D)$, được xác định như sau:

$$IG_{\Delta^{\mathcal{I}}}(Y/X, D) = E_{\Delta^{\mathcal{I}}}(Y/X) - \left(\frac{\#D^{\mathcal{I}}Y}{\#Y}E_{\Delta^{\mathcal{I}}}(D^{\mathcal{I}}Y/X) + \frac{\#\overline{D^{\mathcal{I}}}Y}{\#Y}E_{\Delta^{\mathcal{I}}}(\overline{D^{\mathcal{I}}}Y/X)\right) \quad (5)$$

trong đó $D^{\mathcal{I}}Y$ đại diện cho tập $D^{\mathcal{I}} \cap Y$ và $\overline{D^{\mathcal{I}}}Y$ đại diện cho tập $\overline{D^{\mathcal{I}}} \cap Y$. ■

Trong ngữ cảnh $\Delta^{\mathcal{I}}$ và X đã rõ ràng, chúng ta viết $E(Y)$ thay cho $E_{\Delta^{\mathcal{I}}}(Y/X)$ và $IG(Y, D)$ thay cho $IG_{\Delta^{\mathcal{I}}}(Y/X, D)$.

Phân hoạch miền của diễn dịch

- Cho $\mathcal{I}$ là một diễn dịch trong $\mathcal{L}_{\Sigma,\Phi}$ và $E = \langle E^+, E^- \rangle$, trong đó E^+ là tập các mẫu dương và E^- là tập các mẫu âm.
- Vấn đề đặt ra là phân hoạch miền $\Delta^\mathcal{I}$ của diễn dịch $\mathcal{I}$ sử dụng các bộ chọn trong ngôn ngữ $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, trong đó $\Sigma^\dagger \subseteq \Sigma$ và $\Phi^\dagger \subseteq \Phi$, để đạt được phân hoạch $\mathbb{Y}$ nhất quán với E.
- Khi diễn dịch $\mathcal{I}$ là mô hình của cơ sở tri thức $\mathcal{KB}$, chúng ta có:
  - với C là một khái niệm của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, lúc đó $C^\mathcal{I}$ là hợp của một số lớp tương đương của $\Delta^\mathcal{I}$ dựa trên $\equiv_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$,
  - $a^\mathcal{I} \in C^\mathcal{I}$ với mọi $a \in E^+$ và $a^\mathcal{I} \notin C^\mathcal{I}$ với mọi $a \in E^-$.

1. Bắt đầu từ phân hoạch $\{\Delta^{\mathcal{I}}\}$, thực hiện việc làm mịn để đạt được phân hoạch ương ứng với $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$.
   - Quá trình làm mịn dừng lại khi phân hoạch hiện tại nhất quán với E.
   - Thực hiện việc làm mịn bằng cách sử dụng các bộ chọn
   - Ký hiệu các khối được tạo ra trong quá trình làm mịn là $Y_1, Y_2, \ldots, Y_n$ với phân hoạch hiện tại là $\mathbb{Y} = \{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}\}$.
   - $s := n+1, \quad t := n+2, \quad n := n+2,$
     $Y_s := C_{i_j} \sqcap D, \quad Y_t := C_{i_j} \sqcap \neg D$
   - Phân hoạch mới của $\Delta^{\mathcal{I}}$ là $\mathbb{Y} = \{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}\} \setminus \{Y_{i_j}\} \cup \{Y_s, Y_t\}$. Chú ý, với mọi $1 \leq i \leq n$:
     - Y_i được biểu diễn thông qua một khái niệm C_i (sao cho $Y_i = C_i^{\mathcal{I}}$)
     - Ghi nhận lại thông tin của Y_i dù nó có bị chia bởi E hay không
2. Trả lại kết quả là phân hoạch $\mathbb{Y}$.

Algorithm 3.1: *Phân hoạch miền của diễn dịch trong logic mô tả*

Input: $\mathcal{I}$, $\Sigma^\dagger$, $\Phi^\dagger$, $E = \langle E^-, E^+\rangle$

Output: $\mathbb{Y} = \{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}\}$ sao cho $\mathbb{Y}$ nhất quán với E

$n := 1$; $Y_1 := \Delta^{\mathcal{I}}$; $\mathbb{Y} := \{Y_1\}$; $C_1 := \top$; $\mathbb{D} := \emptyset$;

Tạo và thêm các bộ chọn vào $\mathbb{D}$; `/* theo Định nghĩa 3.1 và Hình 3.1 */`

while ($\mathbb{Y}$ *không nhất quán với* E) **and** ($\mathbb{Y}$ *có thể phân hoạch*) **do**

- Chọn bộ chọn D_u trong $\mathbb{D}$ và khối Y_{i_j} trong $\mathbb{Y}$ sao cho D_u chia Y_{i_j} thành hai khối không rỗng;
- $s := n + 1$; $t := n + 2$; $n := n + 2$;
- $Y_s := Y_{i_j} \cap D_u^{\mathcal{I}}$; $\quad C_s := C_{i_j} \sqcap D_u$;
- $Y_t := Y_{i_j} \cap (\neg D_u)^{\mathcal{I}}$; $\quad C_t := C_{i_j} \sqcap \neg D_u$;
- $\mathbb{Y} := \mathbb{Y} \cup \{Y_s, Y_t\} \setminus \{Y_{i_j}\}$;
- Tạo và thêm các bộ chọn mới vào $\mathbb{D}$; `/* theo Định nghĩa 3.1 và Hình 3.1 */`

if ($\mathbb{Y}$ *nhất quán với* E) **then**

- **return** $\mathbb{Y}$;

else

- **return** failure;

Học khái niệm cho CSTT với Ngữ cảnh (2)

Ngữ cảnh (2)

Bài toán học khái niệm cho cơ sở tri thức trong logic mô tả đặt ra trong ngữ cảnh (2) là học khái niệm C như là một định nghĩa của A_d trong ngôn ngữ con cho trước $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, với $\Sigma^\dagger \subseteq \Sigma \setminus \{A_d\}$ và $\Phi^\dagger \subseteq \Phi$ sao cho:

- $\mathcal{KB} \models C(a)$ với mọi $a \in E^+$, và
- $\mathcal{KB} \not\models C(a)$, với mọi $a \in E^-$,

trong đó, tập E^+ chứa các mẫu dương và E^- chứa các mẫu âm của C.

- Sử dụng các mô hình của $\mathcal{KB}$ kết hợp với mô phỏng hai chiều trong mô hình đó $\rightarrow$ để mô hình hóa tính không phân biệt được.
- Sử dụng cây quyết định $\rightarrow$ để phân lớp dữ liệu phục vụ cho việc tìm kiếm khái niệm C.

Sử dụng các bộ chọn để làm mịn phân hoạch $\{\Delta^\mathcal{I}\}$ của diễn dịch $\mathcal{I}$ là mô hình của $\mathcal{KB}$ nhằm đạt được phân hoạch nhất quán với E.

Ngữ cảnh (2)-Ý tưởng chính của thuật toán BBCL2

- Xây dựng tập E_0^- và mở rộng nó sao cho E_0^- phủ càng lúc càng nhiều các cá thể trong E^-,
- Xây dựng tập $\mathbb{C}$ gồm các phần tử là các khái niệm D thỏa mãn điều kiện $\mathcal{KB} \models D(a)$ với mọi $a \in E^+$.
- Xây dựng tập $\mathbb{C}_0$: khi một khái niệm D không thỏa mãn điều kiện $\mathcal{KB} \models D(a)$ với mọi $a \in E^+ \to D$ được thêm vào $\mathbb{C}_0$. Sau này, khi cần, lấy hợp của các khái niệm trong $\mathbb{C}_0$ và kiểm tra xem nó có thỏa mãn điều kiện để thêm vào $\mathbb{C}$ hay không.

$\Rightarrow$ Như vậy ta luôn có:

- $\mathcal{KB} \models (\bigsqcap \mathbb{C})(a)\ \forall a \in E^+$, và
- $\mathcal{KB} \not\models (\bigsqcap \mathbb{C})(a)\ \forall a \in E_0^-$.

Mở rộng $\mathbb{C}$ sao cho $\mathcal{KB} \not\models (\bigsqcap \mathbb{C})(a)$ với càng lúc càng nhiều $a \in E^-$. Mở rộng $\mathbb{C} \to$ mở rộng E_0^-. Khi $E_0^- = E^-$ thuật toán trả về khái niệm $\bigsqcap \mathbb{C}$.

Algorithm 4.3: *BBCL2–Học khái niệm cho cơ sở tri thức trong logic mô tả với ngữ cảnh (2)*

Input: $\mathcal{KB}$, $\Sigma^\dagger$, $\Phi^\dagger$, $E = \langle E^+, E^- \rangle$, K

Output: Khái niệm C sao cho:

- $\mathcal{KB} \models C(a)$ với mọi $a \in E^+$, và
- $\mathcal{KB} \not\models C(a)$ for all $a \in E^-$.

$E_0^- := \emptyset$; $\mathbb{C} := \emptyset$; $\mathbb{C}_0 := \emptyset$;

while not (*too hard to extend* $\mathbb{C}$) and ($E_0^- \neq E^-$) **do**

  Xây dựng mô hình hữu hạn $\mathcal{I}$ của $\mathcal{KB}$ hoặc $\mathcal{I} = \mathcal{I}'_{|K}$;

  $\mathbb{Y}$:=`partition` ($\mathcal{I}$, $\Sigma^\dagger$, $\Phi^\dagger$, E); `/* phân hoạch` $\Delta^\mathcal{I}$ `theo Thuật toán 4.1 */`

  foreach $Y_{i_j} \in \mathbb{Y} \mid \exists a \in E^- : a^\mathcal{I} \in Y_{i_j}$ and $\forall a \in E^+ : a^\mathcal{I} \notin Y_{i_j}$ **do**

    if ($\mathcal{KB} \models \neg C_{i_j}(a), \forall a \in E^+$) **then**

      if ($\mathcal{KB} \not\models (\sqcap \mathbb{C} \sqsubseteq \neg C_{i_j})$) **then**

        $\mathbb{C} := \mathbb{C} \cup \{\neg C_{i_j}\}$;

        $E_0^- := E_0^- \cup \{a \in E^- \mid a^\mathcal{I} \in Y_{i_j}\}$;

    else

      $\mathbb{C}_0 := \mathbb{C}_0 \cup \{\neg C_{i_j}\}$;

while not (*too hard to extend* $\mathbb{C}$) and ($E_0^- \neq E^-$) **do**

  $D := D_1 \sqcup D_2 \sqcup \cdots \sqcup D_l$, với $D_1, D_2, \ldots, D_l$ được chọn ngẫu nhiên từ $\mathbb{C}_0$;

  if ($\mathcal{KB} \models D(a), \forall a \in E^+$) **then**

    if ($\mathcal{KB} \not\models (\sqcap \mathbb{C}) \sqsubseteq D$) and ($\exists a \in E^- \setminus E_0^- : \mathcal{KB} \not\models (\sqcap \mathbb{C} \sqcap D)(a)$) **then**

      $\mathbb{C} := \mathbb{C} \cup \{D\}$;

      $E_0^- := E_0^- \cup \{a \mid a \in E^- \setminus E_0^-, \mathcal{KB} \not\models (\sqcap \mathbb{C})(a)\}$;

if ($E_0^- = E^-$) **then**

  foreach $D \in \mathbb{C}$ **do**

    if $\mathcal{KB} \not\models \sqcap(\mathbb{C} \setminus \{D\})(a), \forall a \in E^-$ **then**

      $\mathbb{C} := \mathbb{C} \setminus \{D\}$;

  $C := \sqcap \mathbb{C}$;

  return C_{rs} := `simplify` (C); `/* rút gọn khái niệm` C `*/`

else

  return failure;

Mệnh đề 3.1 (Tính đúng đắn của thuật toán BBCL2)

Thuật toán BBCL2 là đúng đắn. Nghĩa là, nếu thuật toán BBCL2 trả về một khái niệm C_{rs} thì C_{rs} là một lời giải của bài toán học khái niệm cho cơ sở tri thức trong logic mô tả với ngữ cảnh (2). ∎

Độ phức tạp của thuật toán BBCL2

Học khái niệm cho cơ sở tri thức trong logic mô tả với ngữ cảnh (2) liên quan chặt chẽ với suy luận tự động trong logic mô tả. Đối với vấn đề suy luận tự động, độ phức tạp của bài toán này là ExpTime-khó ngay cả đối với logic mô tả cơ bản $\mathcal{ALC}$. Một cách tổng quát, bài toán kiểm tra tính thỏa trong logic mô tả thường là ExpTime-đầy đủ. Thuật toán BBCL và BBCL2 sử dụng một vòng lặp tuyến tính có giới hạn là lực lượng của $\mathbb{C}_0$ cho bài toán suy luận. Do đó, hai thuật toán này có độ phức tạp là hàm mũ (xét theo kích thước của $\mathcal{KB}$, E^+ và E^- với giả thiết là $\Sigma^\dagger$ cố định). ∎

Kết quả nghiên cứu của đề

1. Xây dựng ngôn ngữ logic mô tả $\mathcal{L}_{\Sigma,\Phi}$ dựa trên ngôn ngữ $\mathcal{ALC}_{reg}$ với tập các đặc trưng mở rộng gồm $\mathcal{I}$, $\mathcal{O}$, $\mathcal{N}$, $\mathcal{Q}$, $\mathcal{F}$, $\mathcal{U}$, Self. Ngoài ra ngôn ngữ được xây dựng còn cho phép sử dụng các thuộc tính như là các phần tử cơ bản của ngôn ngữ.

2. Xây dựng mô phỏng hai chiều trên lớp các logic mở rộng. Phát biểu và chứng minh các định lý, bổ đề, hệ quả liên quan đến mô phỏng hai chiều và tính bất biến đối với mô phỏng hai chiều.

3. Xây dựng thuật toán BBCL2 nhằm giải quyết các bài toán học khái niệm cho cơ sở tri thức trong logic mô tả với ngữ cảnh (2).

Các vấn đề cần quan tâm nghiên cứu thêm

1. Xây dựng các chiến lược học khác nhau thông qua các độ đo trong việc quyết định khối nào nên phân hoạch trước. So sánh các chiến lược học với nhau.
2. Xây dựng các module học khái niệm trong logic mô tả với các ngữ cảnh khác nhau như là một API cho phép tích hợp vào các hệ thống khác.
3. Nghiên cứu các thuật toán học nữa giám sát, học không giám sát, học theo xác suất cho các cở sở tri thức trong logic mô tả.
4. Nghiên cứu khả năng học chính xác khái niệm cho các logic mô tả khác nhau.

Danh mục các công trình đã công bố

1. Trần Thanh Lương, Hoàng Thị Lan Giao. Áp dụng độ đo entropy để phân hoạch khối cho các hệ thống thông tin dựa trên logic mô tả. *Kỷ yếu Hội thảo quốc gia lần thứ XV: Một số vấn đề chọn lọc của Công nghệ Thông tin và Truyền thông*, trang 11–18, Nhà xuất bản Khoa học và Kỹ thuật, 2013.
2. T.-L. Tran, Q.-T. Ha, T.-L.-G. Hoang, L. A. Nguyen, and H. S. Nguyen. Bisimulation-based concept learning in description logics. In *Proceedings of CS&P'2013*, pages 421–433. CEURWS.org, 2013.
3. T.-L. Tran, L. A. Nguyen, and T.-L.-G. Hoang. A domain partitioning method for bisimulationbased concept learning in description logics. In *Proceedings of ICCSAMA'2014, volume 282 of Advances in Intelligent Systems and Computing*, pages 297–312. Springer, 2014.
4. T.-L. Tran, Q.-T. Ha, T.-L.-G. Hoang, L. A. Nguyen, and H. S. Nguyen. Bisimulation-based concept learning in description logics. *Fundam. Inform.*, 133(2-3):287–303, 2014.
5. T.-L. Tran, T.-L.-G. Hoang. Entropy-based measures for partitioning the domain of an interpretation in description logics. *Journal of Science, Hue University*, volume 96, number 8: pages 87-101, 2014.
6. T.-L. Tran, L. A. Nguyen, and T.-L.-G. Hoang. Bisimulation-based concept learning for information systems in description logics. *Vietnam Journal of Computer Science*, Springer, 2015. (Online first).

Danh mục các sản phẩm đào tạo

Đào tạo thạc sĩ

- Họ tên: Nguyễn Công Ẩn, Khóa 2011–2013, Chuyên ngành KHMT
  Đề tài: Nghiên cứu một số thuật toán suy luận trong logic mô tả
  Năm thực hiện: 2013
  Người hướng dẫn: Hoàng Thị Lan Giao (thành viên đề tài)

Đào tạo cử nhân

- Họ tên: Nguyễn Hữu Hải, Khóa 33, Ngành Tin học
  Đề tài: Tìm hiểu về logic mô tả và ứng dụng trong xây dựng ontology
  Năm thực hiện: 2013
  Người hướng dẫn: Trần Thanh Lương (chủ nhiệm đề tài)

- Họ tên: Đặng Tuấn Anh, Khóa 34, Ngành Tin học
  Đề tài: Phân lớp dữ liệu với nhiều tập huấn luyện bằng cây quyết định
  Năm thực hiện: 2014
  Người hướng dẫn: Trần Thanh Lương (chủ nhiệm đề tài)

XIN CẢM ƠN QUÝ THẦY CÔ
VÀ CÁC ANH CHỊ ĐÃ LẮNG NGHE!