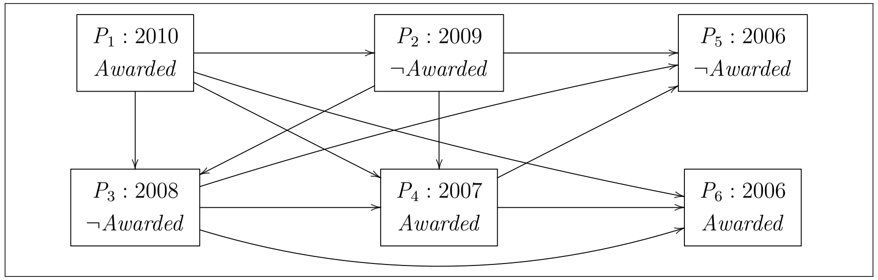
khái niệm C dựa trên cơ sở tri thức $\mathcal{KB}$ được ký hiệu là $\mathcal{KB} \not\models C(a)$. Khái niệm D được gọi là *bao hàm* khái niệm C dựa trên cơ sở tri thức $\mathcal{KB}$, ký hiệu là $\mathcal{KB} \models C \sqsubseteq D$, nếu với mọi diễn dịch $\mathcal{I}$ là mô hình của $\mathcal{KB}$ thì $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

Ví dụ 1.1. Ví dụ sau đây là một cơ sở tri thức đề cập về các ấn phẩm khoa học.

$$\Sigma_I = \{P_1, P_2, P_3, P_4, P_5, P_6\}, \qquad \Phi = \{\mathcal{I}, \mathcal{O}, \mathcal{N}, \mathcal{Q}\},$$
$$\Sigma_C = \{Pub, Awarded, UsefulPub, A_d\}, \qquad \Sigma_{dA} = \Sigma_C, \quad \Sigma_{nA} = \{Year\},$$
$$\Sigma_{oR} = \{cites, cited\_by\}, \quad \Sigma_{dR} = \emptyset,$$
$$\mathcal{R} = \{cites^- \sqsubseteq cited\_by, cited\_by^- \sqsubseteq cites, \mathtt{Irr}(cites)\},$$
$$\mathcal{T} = \{\top \sqsubseteq Pub, UsefulPub \equiv \exists cited\_by.\top\},$$
$$\begin{aligned}\mathcal{A}_0 = \{&Awarded(P_1), \neg Awarded(P_2), \neg Awarded(P_3), Awarded(P_4),\\ &\neg Awarded(P_5), Awarded(P_6), Year(P_1) = 2010, Year(P_2) = 2009,\\ &Year(P_3) = 2008, Year(P_4) = 2007, Year(P_5) = 2006, Year(P_6) = 2006,\\ &cites(P_1, P_2), cites(P_1, P_3), cites(P_1, P_4), cites(P_1, P_6), cites(P_2, P_3),\\ &cites(P_2, P_4), cites(P_2, P_5), cites(P_3, P_4), cites(P_3, P_5), cites(P_3, P_6),\\ &cites(P_4, P_5), cites(P_4, P_6)\}.\end{aligned}$$

Lúc đó $\mathcal{KB}_0 = \langle \mathcal{R}, \mathcal{T}, \mathcal{A}_0 \rangle$ là cơ sở tri thức trong $\mathcal{L}_{\Sigma,\Phi}$. Tiên đề $\top \sqsubseteq Pub$ để chỉ ra rằng miền của bất kỳ mô hình nào của $\mathcal{KB}_0$ đều chỉ gồm các ấn phẩm khoa học. Cơ sở tri thức $\mathcal{KB}_0$ được minh họa như trong Hình 1.3. Trong hình này, các nút ký hiệu cho các ấn phẩm và các cạnh ký hiệu cho các trích dẫn (khẳng định của vai trò *cites*). Hình này chỉ biểu diễn những thông tin về các khẳng định *Year*, *Awarded* và *cites*. ■



Hình 1.3: Một minh họa cho cơ sở tri thức của Ví dụ 1.1

1.5. Suy luận trong logic mô tả

1.5.1. Giới thiệu

Có nhiều bài toán suy luận được đặt ra trong các hệ thống biểu diễn tri thức dựa trên logic mô tả. Bài toán suy luận quan trọng nhất trong logic mô tả là bài toán kiểm

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC HUẾ
TRƯỜNG ĐẠI HỌC KHOA HỌC

TÓM TẮT BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CẤP CƠ SỞ

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

Mã số: DHH2013-01-41

Chủ nhiệm đề tài: ThS. TRẦN THANH LƯƠNG

Thừa Thiên Huế, 11/2014

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC HUẾ
TRƯỜNG ĐẠI HỌC KHOA HỌC

TÓM TẮT BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CẤP CƠ SỞ

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

Mã số: DHH2013-01-41

Xác nhận của cơ quan chủ trì đề tài **Chủ nhiệm đề tài**

ThS. TRẦN THANH LƯƠNG

Thừa Thiên Huế, 11/2014

1.4.3. Bộ khẳng định cá thể

Định nghĩa 1.10 (Khẳng định cá thể)**.** Một *khẳng định cá thể* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một biểu thức có dạng $C(a)$, $R(a,b)$, $\neg R(a,b)$, $a = b$, $a \neq b$, trong đó C là một khái niệm và R là một vai trò đối tượng của $\mathcal{L}_{\Sigma,\Phi}$. ■

Ý nghĩa của các khẳng định cá thể trong Định nghĩa 1.11 được hiểu như sau:

- $C(a)$ được gọi là một *khẳng định khái niệm*,
- $R(a,b)$ được gọi là một *khẳng định vai trò đối tượng dương*,
- $\neg R(a,b)$ được gọi là một *khẳng định vai trò đối tượng âm*,
- $a = b$ được gọi là một *khẳng định bằng nhau*,
- $a \neq b$ được gọi là một *khẳng định khác nhau.*

Ngữ nghĩa của các khẳng định cá thể được xác định thông qua diễn dịch $\mathcal{I}$ như sau:

$$\begin{array}{lll}
\mathcal{I} \models C(a) & \text{nếu} & a^{\mathcal{I}} \in C^{\mathcal{I}}, \\
\mathcal{I} \models R(a,b) & \text{nếu} & \langle a^{\mathcal{I}}, b^{\mathcal{I}} \rangle \in R^{\mathcal{I}}, \\
\mathcal{I} \models \neg R(a,b) & \text{nếu} & \langle a^{\mathcal{I}}, b^{\mathcal{I}} \rangle \notin R^{\mathcal{I}}. \\
\mathcal{I} \models a = b & \text{nếu} & a^{\mathcal{I}} = b^{\mathcal{I}}, \\
\mathcal{I} \models a \neq b & \text{nếu} & a^{\mathcal{I}} \neq b^{\mathcal{I}}.
\end{array}$$

Giả sử φ là một khẳng định cá thể. Chúng ta nói rằng $\mathcal{I}$ *thỏa mãn* φ nếu $\mathcal{I} \models \varphi$.

Định nghĩa 1.11 (Bộ khẳng định cá thể)**.** *Bộ khẳng định cá thể* (*ABox*) trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một tập hữu hạn các khẳng định cá thể trong $\mathcal{L}_{\Sigma,\Phi}$. ■

1.4.4. Cơ sở tri thức và mô hình của cơ sở tri thức

Định nghĩa 1.12 (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 1.13 (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 $\mathcal{A}$. ■

Cơ sở tri thức $\mathcal{KB}$ được gọi là *thỏa mãn* nếu $\mathcal{KB}$ có mô hình. Một cá thể a được gọi là *thể hiện* của một khái niệm C dựa trên cơ sở tri thức $\mathcal{KB}$, ký hiệu là $\mathcal{KB} \models C(a)$, nếu với mọi diễn dịch $\mathcal{I}$ là mô hình của $\mathcal{KB}$ thì $a^{\mathcal{I}} \in C^{\mathcal{I}}$. Cá thể a không phải thể hiện của vai trò hoặc một khẳng định vai trò trong $\mathcal{L}_{\Sigma,\Phi}$. ■

Ý nghĩa của các khẳng định vai trò trong Định nghĩa 1.6 được hiểu như sau:

- $\mathtt{Ref}(r)$ được gọi là một *khẳng định vai trò phản xạ*,
- $\mathtt{Irr}(r)$ được gọi là một *khẳng định vai trò không phản xạ*,
- $\mathtt{Sym}(r)$ được gọi là một *khẳng định vai trò đối xứng*,
- $\mathtt{Tra}(r)$ được gọi là một *khẳng định vai trò bắc cầu*,
- $\mathtt{Dis}(R,S)$ được gọi là một *khẳng định vai trò không giao nhau.*

Ngữ nghĩa của các tiên đề vai trò được xác định thông qua diễn dịch $\mathcal{I}$ như sau:

$\mathcal{I} \models \varepsilon \sqsubseteq r$	nếu	$\varepsilon^{\mathcal{I}} \subseteq r^{\mathcal{I}}$,
$\mathcal{I} \models R_1 \circ R_2 \circ \cdots \circ R_k \sqsubseteq r$	nếu	$R_1^{\mathcal{I}} \circ R_2^{\mathcal{I}} \circ \cdots \circ R_k^{\mathcal{I}} \sqsubseteq r^{\mathcal{I}}$,
$\mathcal{I} \models \mathtt{Ref}(r)$	nếu	$r^{\mathcal{I}}$ phản xạ,
$\mathcal{I} \models \mathtt{Irr}(r)$	nếu	$r^{\mathcal{I}}$ không phản xạ,
$\mathcal{I} \models \mathtt{Sym}(r)$	nếu	$r^{\mathcal{I}}$ đối xứng,
$\mathcal{I} \models \mathtt{Tra}(r)$	nếu	$r^{\mathcal{I}}$ bắc cầu,
$\mathcal{I} \models \mathtt{Dis}(R,S)$	nếu	$R^{\mathcal{I}}$ và $S^{\mathcal{I}}$ không giao nhau.

Giả sử φ là một tiên đề vai trò. Chúng ta nói rằng $\mathcal{I}$ *thỏa mãn* φ nếu $\mathcal{I} \models \varphi$.

Định nghĩa 1.7 (Bộ tiên đề vai trò)**.** *Bộ tiên đề vai trò* (*RBox*) trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một tập hữu hạn các tiên đề vai trò trong $\mathcal{L}_{\Sigma,\Phi}$. ■

1.4.2. Bộ tiên đề thuật ngữ

Định nghĩa 1.8 (Tiên đề thuật ngữ)**.** Một *tiên đề bao hàm khái niệm tổng quát* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một biểu thức có dạng $C \sqsubseteq D$, trong đó C và D là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$. Một *tiên đề tương đương khái niệm* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một biểu thức có dạng $C \equiv D$, trong đó C và D là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$. Một *tiên đề thuật ngữ* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một tiên đề bao hàm khái niệm tổng quát hoặc một tiên đề tương đương khái niệm trong $\mathcal{L}_{\Sigma,\Phi}$. ■

Ngữ nghĩa của các tiên đề thuật ngữ được xác định thông qua diễn dịch $\mathcal{I}$ là $\mathcal{I} \models C \sqsubseteq D$ nếu $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ và $\mathcal{I} \models C \equiv D$ nếu $C^{\mathcal{I}} = D^{\mathcal{I}}$.

Giả sử φ là một tiên đề thuật ngữ. Chúng ta nói rằng $\mathcal{I}$ *thỏa mãn* φ nếu $\mathcal{I} \models \varphi$.

Định nghĩa 1.9 (Bộ tiên đề thuật ngữ)**.** *Bộ tiên đề thuật ngữ* (*TBox*) trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một tập hữu hạn các tiên đề thuật ngữ trong $\mathcal{L}_{\Sigma,\Phi}$. ■

DANH SÁCH THÀNH VIÊN THAM GIA NGHIÊN CỨU:

• **TS. Hoàng Thị Lan Giao**,
Khoa Công nghệ Thông tin, Trường Đại học Khoa học, Đại học Huế
77 Nguyễn Huệ, Thành phố Huế, Tỉnh Thừa Thiên Huế, Việt Nam

ĐƠN VỊ PHỐI HỢP NGHIÊN CỨU:

• **Viện Tin học**,
Khoa Toán - Tin học và Cơ học, Trường Đại học Tổng hợp Vác-xa-va, Ba Lan
Banacha 2, 02-097 Warsaw, Poland

MỤC LỤC

$$
\begin{array}{lll}
(R \circ S)^{\mathcal{I}} = R^{\mathcal{I}} \circ S^{\mathcal{I}} & (R^*)^{\mathcal{I}} = (R^{\mathcal{I}})^* & (C?)^{\mathcal{I}} = \{\langle x,x\rangle \mid C^{\mathcal{I}}(x)\} \\
(R \sqcup S)^{\mathcal{I}} = R^{\mathcal{I}} \cup S^{\mathcal{I}} & (R^-)^{\mathcal{I}} = (R^{\mathcal{I}})^{-1} & \varepsilon^{\mathcal{I}} = \{\langle x,x\rangle \mid x \in \Delta^{\mathcal{I}}\} \\
U^{\mathcal{I}} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} & \top^{\mathcal{I}} = \Delta^{\mathcal{I}} \quad \bot^{\mathcal{I}} = \emptyset & (\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}} & (C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}} & \{a\}^{\mathcal{I}} = \{a^{\mathcal{I}}\}
\end{array}
$$

$$(A \le d)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid A^{\mathcal{I}}(x) \text{ xác định và } A^{\mathcal{I}}(x) \le d\}$$

$$(A \ge d)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid A^{\mathcal{I}}(x) \text{ xác định và } A^{\mathcal{I}}(x) \ge d\}$$

$$
\begin{array}{ll}
(A = d)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid A^{\mathcal{I}}(x) = d\} & (A \ne d)^{\mathcal{I}} = (\neg(A = d))^{\mathcal{I}} \\
(A < d)^{\mathcal{I}} = ((A \le d) \sqcap (A \ne d))^{\mathcal{I}} & (A > d)^{\mathcal{I}} = ((A \ge d) \sqcap (A \ne d))^{\mathcal{I}} \\
(\forall R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \forall y\,[R^{\mathcal{I}}(x,y) \Rightarrow C^{\mathcal{I}}(y)]\} & (\exists r.\mathsf{Self})^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(x,x)\} \\
(\exists R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y\,[R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)]\} & (\exists \sigma.\{d\})^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \sigma^{\mathcal{I}}(x,d)\} \\
(\ge n\, R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \#\{y \mid R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)\} \ge n\} & (\ge n\, R)^{\mathcal{I}} = (\ge n\, R.\top)^{\mathcal{I}} \\
(\le n\, R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \#\{y \mid R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)\} \le n\} & (\le n\, R)^{\mathcal{I}} = (\le n\, R.\top)^{\mathcal{I}}
\end{array}
$$

Hình 1.2: Diễn dịch của các vai trò phức và khái niệm phức.

1.3. Các dạng chuẩn

1.3.1. Dạng chuẩn phủ định của khái niệm

Khái niệm C được gọi là ở *dạng chuẩn phủ định* nếu toán tử phủ định chỉ xuất hiện trước các tên khái niệm xuất hiện trong C.

1.3.2. Dạng chuẩn nghịch đảo của vai trò

Vai trò đối tượng R được gọi là một vai trò ở *dạng chuẩn nghịch đảo* (*Converse Normal Form - CNF*) nếu tạo tử nghịch đảo chỉ áp dụng cho các tên vai trò đối tượng xuất hiện trong R (không xét đến vai trò đối tượng phổ quát U)

Đặt $\Sigma_{oR}^{\pm} = \Sigma_{oR} \cup \{r^- \mid r \in \Sigma_{oR}\}$. Một *vai trò đối tượng cơ bản* là một phần tử thuộc $\Sigma_{oR}^{\pm}$ nếu ngôn ngữ được xem xét cho phép vai trò nghịch đảo hoặc một phần tử thuộc Σ_{oR} nếu ngôn ngữ được xem xét không cho phép vai trò nghịch đảo.

1.4. Cơ sở tri thức trong logic mô tả

1.4.1. Bộ tiên đề vai trò

Định nghĩa 1.6 (Tiên đề vai trò)**.** Một *tiên đề bao hàm vai trò* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một biểu thức có dạng $\varepsilon \sqsubseteq r$ hoặc $R_1 \circ R_2 \circ \cdots \circ R_k \sqsubseteq r$, trong đó $k \ge 1$, $r \in \Sigma_{oR}$ và $R_1, R_2, \ldots, R_k$ là các vai trò đối tượng cơ bản của $\mathcal{L}_{\Sigma,\Phi}$ khác với vai trò phổ quát U. Một *khẳng định vai trò* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một biểu thức có dạng $\mathtt{Ref}(r)$, $\mathtt{Irr}(r)$, $\mathtt{Sym}(r)$, $\mathtt{Tra}(r)$ hoặc $\mathtt{Dis}(R,S)$, trong đó $r \in \Sigma_{oR}$ và R, S là các vai trò đối tượng của $\mathcal{L}_{\Sigma,\Phi}$ khác với vai trò phổ quát U. Một *tiên đề vai trò* trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ là một tiên đề bao hàm

- 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 $d \in \mathit{range}(\sigma)$ thì $\exists\sigma.\{d\}$ là một khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\mathcal{I} \in \Phi$ thì R^- là một vai trò đối tượng của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\mathcal{O} \in \Phi$ thì $\{a\}$ là một khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\mathcal{F} \in \Phi$ thì $\leq 1\, r$ là một khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\{\mathcal{F}, \mathcal{I}\} \subseteq \Phi$ thì $\leq 1\, r^-$ là một khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\mathcal{N} \in \Phi$ thì $\geq n\, r$ và $\leq n\, r$ là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\{\mathcal{N}, \mathcal{I}\} \subseteq \Phi$ thì $\geq n\, r^-$ và $\leq n\, r^-$ là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\mathcal{Q} \in \Phi$ thì $\geq n\, r.C$ và $\leq n\, r.C$ là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\{\mathcal{Q}, \mathcal{I}\} \subseteq \Phi$ thì $\geq n\, r^-.C$ và $\leq n\, r^-.C$ là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\mathcal{U} \in \Phi$ thì U là một vai trò đối tượng của $\mathcal{L}_{\Sigma,\Phi}$,
  - nếu $\mathsf{Self} \in \Phi$ thì $\exists r.\mathsf{Self}$ là một khái niệm của $\mathcal{L}_{\Sigma,\Phi}$. ■

Định nghĩa 1.5 (Ngữ nghĩa của $\mathcal{L}_{\Sigma,\Phi}$)**.** Một *diễn dịch* trong $\mathcal{L}_{\Sigma,\Phi}$ là một bộ $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$, trong đó $\Delta^\mathcal{I}$ là một tập không rỗng được gọi là *miền* của $\mathcal{I}$ và $\cdot^\mathcal{I}$ là một ánh xạ được gọi là *hàm diễn dịch* của $\mathcal{I}$ cho phép ánh xạ mỗi cá thể $a \in \Sigma_I$ thành một phần tử $a^\mathcal{I} \in \Delta^\mathcal{I}$, mỗi tên khái niệm $A \in \Sigma_C$ thành một tập $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, mỗi thuộc tính $A \in \Sigma_A \setminus \Sigma_C$ thành một hàm từng phần $A^\mathcal{I} : \Delta^\mathcal{I} \to \mathit{range}(A)$, mỗi tên vai trò đối tượng $r \in \Sigma_{oR}$ thành một quan hệ nhị phân $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$ và mỗi vai trò dữ liệu $\sigma \in \Sigma_{dR}$ thành một quan hệ nhị phân $\sigma^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \mathit{range}(\sigma)$. Hàm $\cdot^\mathcal{I}$ được mở rộng cho các vai trò đối tượng phức và các khái niệm phức như trong Hình 1.2, trong đó $\#\Gamma$ ký hiệu cho lực lượng của tập Γ. ■

Cho diễn dịch $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$. Chúng ta nói rằng đối tượng $x \in \Delta^\mathcal{I}$ có *độ sâu* là k nếu k là số tự nhiên lớn nhất sao cho tồn tại các đối tượng $x_0, x_1, \ldots, x_k \in \Delta^\mathcal{I}$ khác nhau từng đôi một thỏa mãn:

- $x_k = x$ và $x_0 = a^\mathcal{I}$ với $a \in \Sigma_I$,
- $x_i \neq b^\mathcal{I}$ với mọi $1 \leq i \leq k$ và với mọi $b \in \Sigma_I$,
- với mỗi $1 \leq i \leq k$ tồn tại một vai trò đối tượng R_i của $\mathcal{L}_{\Sigma,\Phi}$ sao cho $R_i^\mathcal{I}(x_{i-1}, x_i)$ thỏa mãn.

Chúng ta ký hiệu $\mathcal{I}_{|k}$ là diễn dịch thu được từ diễn dịch $\mathcal{I}$ bằng cách hạn chế miền $\Delta^\mathcal{I}$ của diễn dịch $\mathcal{I}$ chỉ bao gồm tập các đối tượng có độ sâu không lớn hơn k và hàm diễn dịch $\cdot^\mathcal{I}$ được hạn chế một cách tương ứng.

THÔNG TIN KẾT QUẢ NGHIÊN CỨU

1. Thông tin chung:

- Tên đề tài: **HỌC KHÁI NIỆM ĐỐI VỚI CÁC CƠ SỞ TRI THỨC TRONG LOGIC MÔ TẢ DỰA VÀO MÔ PHỎNG HAI CHIỀU**
- Mã số: DHH-2013-01-41
- Chủ nhiệm đề tài: ThS. Trần Thanh Lương
  Điện thoại: `091 4145414` E-mail: `ttluong@hueuni.edu.vn`
- Cơ quan chủ trì đề tài: Trường Đại học Khoa học, Đại học Huế
- Cơ quan và cá nhân phối hợp thực hiện:

  - TS. Hoàng Thị Lan Giao,
  Khoa Công nghệ Thông tin Trường Đại học Khoa học, Đại học Huế
  77 Nguyễn Huệ, Thành phố Huế, Tỉnh Thừa Thiên Huế, Việt Nam

  - Viện Tin học,
  Khoa Toán, Tin học và Cơ học, Trường Đại học Tổng hợp Vác-xa-va, Ba Lan
  Banacha 2, 02-097 Warsaw, Poland
- Thời gian thực hiện: Từ tháng 01 năm 2013 đến tháng 12 năm 2013

2. Mục tiêu

Mở rộng lý thuyết mô phỏng hai chiều và phương pháp học khái niệm cho cơ sở tri thức trong logic mô tả với một số điều kiện cho trước. Đề xuất thuật toán học khái niệm dựa trên mô phỏng hai chiều để giải quyết bài toán phân lớp các đối tượng trong logic mô tả.

3. Tính mới và sáng tạo

Xây dựng mô phỏng hai chiều để mô hình hóa tính không phân biệt của các đối tượng được trên một lớp lớn các logic mô tả. Từ đó đề xuất thuật toán phân hoạch miền và học khái niệm cho cơ sở tri thức trong logic mô tả sử dụng mô phỏng hai chiều.

4. Kết quả nghiên cứu

- Xây dựng ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ dựa trên logic mô tả $\mathcal{ALC}_{reg}$ với tập các đặc trưng logic mô tả 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 (mỗi thuộc tính có thể là rời rạc hoặc số) như là các phần tử cơ bản của ngôn ngữ. Cách tiếp cận này rất phù hợp đối với các hệ thống thông tin trong thực tế.
- Xây dựng mô phỏng hai chiều trên lớp các logic mở rộng đang nghiên cứu. Các định lý, bổ đề, hệ quả, mệnh đề liên quan đến mô phỏng hai chiều và tính bất biến đối

các phần tử của Σ_R được gọi là *vai trò nguyên tố*. *Logic mô tả động* $\mathcal{ALC}_{reg}$ cho phép các khái niệm và các vai trò được định nghĩa một cách đệ quy như sau:

- nếu $r \in \Sigma_R$ thì r là một vai trò của $\mathcal{ALC}_{reg}$,
- nếu $A \in \Sigma_C$ thì A là một khái niệm của $\mathcal{ALC}_{reg}$,
- nếu C, D là các khái niệm và R, S là các vai trò thì
  - ε, $R \circ S$, $R \sqcup S$, R^*, $?C$ là các vai trò của $\mathcal{ALC}_{reg}$,
  - $\top$, $\bot$, $\neg C$, $C \sqcap D$, $C \sqcup D$, $\exists R.C$ và $\forall R.C$ là các khái niệm của $\mathcal{ALC}_{reg}$. ∎

1.2.2. Ngôn ngữ logic mô tả $\mathcal{L}_{\Sigma,\Phi}$

Một *bộ ký tự logic mô tả* là một tập hữu hạn $\Sigma = \Sigma_I \cup \Sigma_{dA} \cup \Sigma_{nA} \cup \Sigma_{oR} \cup \Sigma_{dR}$, trong đó Σ_I là tập các *cá thể*, Σ_{dA} là tập các *thuộc tính rời rạc*, Σ_{nA} là tập các *thuộc tính số*, Σ_{oR} là tập các *tên vai trò đối tượng* và Σ_{dR} là tập các *vai trò dữ liệu*.[1] Tất cả các tập Σ_I, Σ_{dA}, Σ_{nA}, Σ_{oR} và Σ_{dR} rời nhau từng đôi một.

Đặt $\Sigma_A = \Sigma_{dA} \cup \Sigma_{nA}$. Khi đó mỗi thuộc tính $A \in \Sigma_A$ có một miền giá trị là $range(A)$. Miền $range(A)$ là một tập không rỗng đếm được nếu A là thuộc tính rời rạc và có thứ tự "$\leq$" nếu A là thuộc tính liên tục.[2] (Để đơn giản, chúng ta không ghi ký hiệu "$\leq$" kèm theo thuộc tính A.) Một thuộc tính rời rạc được gọi là *thuộc tính Bool* nếu $range(A) = \{\mathsf{true}, \mathsf{false}\}$. Chúng ta xem các thuộc tính Bool như là các tên khái niệm. Gọi Σ_C là tập các tên khái niệm của Σ, lúc đó ta có $\Sigma_C \subseteq \Sigma_{dA}$.

Xét các *đặc trưng của logic mô tả* gồm $\mathcal{I}$ (*nghịch đảo vai trò*), $\mathcal{O}$ (*định danh*), $\mathcal{F}$ (*tính chất hàm*), $\mathcal{N}$ (*hạn chế số lượng không định tính*), $\mathcal{Q}$ (*hạn chế số lượng có định tính*), $\mathcal{U}$ (*vai trò phổ quát*), Self (*tính phản xạ cục bộ của vai trò*). *Tập các đặc trưng của logic mô tả* Φ là một tập rỗng hoặc tập chứa một số các đặc trưng nêu trên.

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 $A \in \Sigma_A \setminus \Sigma_C$ và $d \in range(A)$ thì $A = d$ và $A \neq d$ là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,
- nếu $A \in \Sigma_{nA}$ và $d \in range(A)$ thì $A \leq d$, $A < d$, $A \geq d$ và $A > d$ là các khái niệm của $\mathcal{L}_{\Sigma,\Phi}$,

[1] Các tên vai trò đối tượng là các vai trò đối tượng nguyên tố.
[2] Có thể giả sử rằng nếu A là một thuộc tính số thì $range(A)$ là tập các số thực và "$\leq$" là một quan hệ thứ tự tuyến giữa các số thực.

1.1.3.6. Phân cấp vai trò

Tạo tử *phân cấp vai trò* (*role hierarchive*), ký hiệu là $\mathcal{H}$, cho phép người sử dụng biểu diễn mối quan hệ giữa các vai trò theo phương cách cụ thể hóa hoặc theo phương cách tổng quát hóa. Vai trò r là cụ thể hóa của vai trò s (hay nói cách khác, vai trò s là tổng quát hóa của vai trò r) và được viết là $r \sqsubseteq s$.

1.1.3.7. Bao hàm vai trò phức

Tạo tử *bao hàm vai trò phức* (*complex role inclusion*), ký hiệu là $\mathcal{R}$, cho phép người sử dụng biểu diễn các tiên đề bao hàm dạng $r \circ s \sqsubseteq r$ (hoặc $r \circ s \sqsubseteq s$).

1.2. Cú pháp và ngữ nghĩa của logic mô tả

1.2.1. Ngôn ngữ logic mô tả $\mathcal{ALC}$

Định nghĩa 1.1 (Cú pháp của $\mathcal{ALC}$)**.** Cho Σ_C là tập các *tên khái niệm* và Σ_R là tập các *tên vai trò* ($\Sigma_C \cap \Sigma_R = \emptyset$). Các phần tử của Σ_C được gọi là *khái niệm nguyên tố. Logic mô tả* $\mathcal{ALC}$ cho phép các khái niệm được định nghĩa một cách đệ quy như sau:

- nếu $A \in \Sigma_C$ thì A là một khái niệm của $\mathcal{ALC}$,
- nếu C, D là các khái niệm và $r \in \Sigma_R$ là một vai trò thì $\top$, $\bot$, $\neg C$, $C \sqcap D$, $C \sqcup D$, $\exists r.C$ và $\forall r.C$ cũng là các khái niệm của $\mathcal{ALC}$. ∎

Định nghĩa 1.2 (Ngữ nghĩa của $\mathcal{ALC}$)**.** Một *diễn dịch* trong logic mô tả $\mathcal{ALC}$ là một bộ $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$, trong đó $\Delta^\mathcal{I}$ là một tập không rỗng được gọi là *miền* của $\mathcal{I}$ và $\cdot^\mathcal{I}$ là một ánh xạ, được gọi là *hàm diễn dịch* của $\mathcal{I}$, cho phép ánh xạ mỗi cá thể $a \in \Sigma_I$ thành một phần tử $a^\mathcal{I} \in \Delta^\mathcal{I}$, mỗi tên khái niệm $A \in \Sigma_C$ thành một tập $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ và mỗi tên vai trò $r \in \Sigma_R$ thành một quan hệ nhị phân $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. Diễn dịch của các khái niệm phức được xác định như sau:

$$
\begin{aligned}
\top^\mathcal{I} &= \Delta^\mathcal{I},\\
\bot^\mathcal{I} &= \emptyset,\\
(\neg C)^\mathcal{I} &= \Delta^\mathcal{I} \setminus C^\mathcal{I},\\
(C \sqcap D)^\mathcal{I} &= C^\mathcal{I} \cap D^\mathcal{I},\\
(C \sqcup D)^\mathcal{I} &= C^\mathcal{I} \cup D^\mathcal{I},\\
(\exists r.C)^\mathcal{I} &= \{x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I}\, [r^\mathcal{I}(x,y) \wedge C^\mathcal{I}(y)]\},\\
(\forall r.C)^\mathcal{I} &= \{x \in \Delta^\mathcal{I} \mid \forall y \in \Delta^\mathcal{I}\, [r^\mathcal{I}(x,y) \Rightarrow C^\mathcal{I}(y)]\}.
\end{aligned}
$$
∎

Định nghĩa sau đây trình bày logic mô tả $\mathcal{ALC}$ tương ứng với logic động mệnh đề, được gọi là *logic mô tả động* và được ký hiệu là $\mathcal{ALC}_{reg}$.

Định nghĩa 1.3 (Cú pháp của $\mathcal{ALC}_{reg}$)**.** Cho Σ_C là tập các *tên khái niệm* và Σ_R là tập các *tên vai trò* ($\Sigma_C \cap \Sigma_R = \emptyset$). Các phần tử của Σ_C được gọi là *khái niệm nguyên tố* và

với mô phỏng hai chiều cũng được phát triển và chứng minh trên lớp các logic mở rộng này.

- Dựa vào mô phỏng hai chiều, xây dựng thuật toán để phân hoạch miền của mô hình của cơ sở tri thức và thuật toán BBCL2 để học khái niệm cho cơ sở tri thức trong logic mô tả.

5. Sản phẩm

- Hướng dẫn 01 luận văn Thạc sĩ Khoa học chuyên ngành Khoa học Máy tính (Người hướng dẫn: TS. Hoàng Thị Lan Giao, thành viên đề tài).
- Hướng dẫn 02 khóa luận Tốt nghiệp Đại học ngành Tin học (Người hướng dẫn: ThS. Trần Thanh Lương, chủ trì đề tài).
- Công bố 04 bài báo trên các tạp chí/hội thảo khoa học trong nước và quốc tế.
- Báo cáo đề tài.

6. Hiệu quả, phương thức chuyển giao kết quả nghiên cứu và khả năng áp dụng

- Phương pháp học khái niệm cho cơ sở tri thức trong logic mô tả sử dụng mô phỏng hai chiều có thể áp dụng để tìm kiếm, xây dựng các định nghĩa khái của các niệm phù hợp cho hệ thống ngữ nghĩa nói chung và Web ngữ nghĩa nói riêng.
- Báo cáo làm tài liệu tham khảo cho sinh viên đại học, học viên cao học và những người nghiên cứu trong chuyên ngành Khoa học Máy tính nói chung cũng như logic mô tả và học máy nói riêng.
- Địa chỉ ứng dụng: Khoa Công nghệ Thông tin, Khoa Tin học của các trường Đại học trong cả nước.

Ngày 25 tháng 11 năm 2014

Cơ quan chủ trì **Chủ nhiệm đề tài**

ThS. TRẦN THANH LƯƠNG

INFORMATION ON RESEARCH RESULTS

1. General information

- Project title: **CONCEPT LEARNING FOR KNOWLEDGE BASES IN DESCRIPTION LOGIC USING BISIMULATION**
- Code number: DHH-2013-01-41
- Coordinator: Msc. Tran Thanh Luong
- Implementing institution: College of Sciences, Hue University
- Cooperating institution(s):

  - Dr. Hoang Thi Lan Giao,
  Department of Information Technology, College of Sciences, Hue University
  77 Nguyen Hue, Hue City, Thua Thien Hue Province, Vietnam

  - Institution of Informatics,
  Faculty of Mathematics, Informatics and Mechanics, Warsaw University, Poland Banacha 2, 02-097 Warsaw, Poland
- Duration: from January 2014 to December 2015

2. Objective(s)

- Extend the theory of bisimulation and method of concept learning for knowledge bases in description logics (DLs) using given conditions.
- Propose a bisimulation-based concept learning algorithm for classifying objects in DLs.

3. Creativeness and innovativeness

We built bisimulation to model indiscernibility of objects in a large class of DLs. We also proposed algorithms for partitioning and leaning concepts for knowledge bases in DLs using bisimulation.

4. Research results

A report consists of information about:

- Consider the language $\mathcal{L}_{\Sigma,\Phi}$, where $\mathcal{L}$ stands for $\mathcal{ALC}_{reg}$, with the set of DL-features, including $\mathcal{I}$, $\mathcal{O}$, $\mathcal{N}$, $\mathcal{Q}$, $\mathcal{F}$, $\mathcal{U}$, Self. In addition, this language allows to use attributes as basic elements (each attribute may be discrete or numeric). This approach is suitable for practical information systems based on DLs.

tri thức đã biết thông qua các khẳng định về các cá thể bao gồm khẳng định khái niệm, khẳng định vai trò (vai trò dương tính và vai trò âm tính), khẳng định đẳng thức, khẳng định bất đẳng thức, . . .

• **Hệ thống suy luận (*Inference System - IS*)**: Hệ thống suy luận cho phép trích rút ra những tri thức tiềm ẩn từ những tri thức đã có được thể hiện trong RBox, TBox và ABox.

• **Giao diện người dùng (*User Interface - UI*)**: Giao diện người dùng được sử dụng để giao tiếp với người sử dụng. Giao diện người dùng được thiết kế tùy thuộc vào từng ứng dụng cụ thể.

1.1.3. Khả năng biểu diễn của logic mô tả

1.1.3.1. Hạn chế số lượng

- *Hạn chế số lượng có định tính* (*qualified number restrictions*), ký hiệu là $\mathcal{Q}$, là hạn chế số lượng trên các vai trò có chỉ ra tính chất của các đối tượng cần hạn chế.
- *Hạn chế số lượng không định tính* (*unqualified number restrictions*), ký hiệu là $\mathcal{N}$, là hạn chế số lượng trên các vai trò nhưng không chỉ ra tính chất của các đối tượng cần hạn chế. Đây là một dạng đặc biệt của hạn chế số lượng có định tính bằng cách thay khái niệm thể hiện tính chất cần định tính bằng khái niệm đỉnh.

1.1.3.2. Tính chất hàm

Ràng buộc *tính chất hàm* (*functionality*), ký hiệu là $\mathcal{F}$, cho phép chỉ ra tính chất hàm cục bộ của các vai trò, nghĩa là các thể hiện của các khái niệm có quan hệ tối đa với một cá thể khác thông qua vai trò được chỉ định.

1.1.3.3. Định danh

Tạo tử *định danh* (*nominal*), ký hiệu là $\mathcal{O}$, cho phép xây dựng khái niệm dạng $\{a\}$ từ một cá thể đơn lẻ a. Khái niệm này biểu diễn cho tập có thể hiện chỉ là một cá thể.

1.1.3.4. Nghịch đảo vai trò

Một logic mô tả với *vai trò nghịch đảo* (*inverse role*), ký hiệu là $\mathcal{I}$, cho phép người sử dụng định nghĩa các vai trò là nghịch đảo của nhau nhằm tăng sự ràng buộc đối với các đối tượng trong miền biểu diễn. Nghịch đảo của vai trò r được viết là r^-.

1.1.3.5. Vai trò bắc cầu

Tạo tử *vai trò bắc cầu* (*transitive role*), ký hiệu là $\mathcal{S}$, được đưa vào logic mô tả nhằm tăng khả năng biểu diễn của logic mô tả đó. Một vai trò r được gọi là bắc cầu nếu $r \circ r \sqsubseteq r$.

Chương 1.

LOGIC MÔ TẢ VÀ CƠ SỞ TRI THỨC

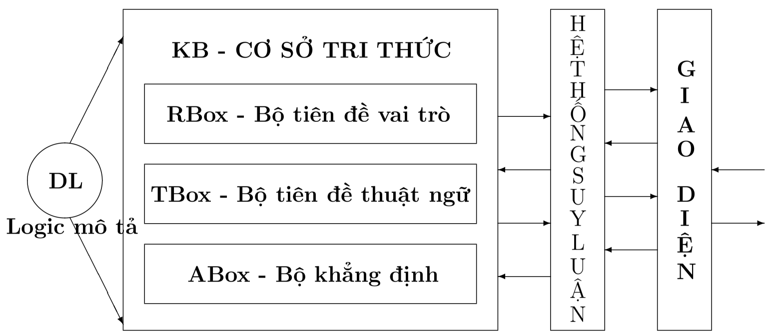
1.1. Giới thiệu về logic mô tả

1.1.1. Tổng quan về logic mô tả

Logic mô tả được xây dựng dựa vào ba thành phần cơ bản gồm tập các *cá thể*, tập các *khái niệm nguyên tố* và tập các *vai trò nguyên tố*. Các logic mô tả khác nhau được đặc trưng bởi tập các *tạo tử khái niệm* và *tạo tử vai trò* mà nó được phép sử dụng để xây dựng các *khái niệm phức*, *vai trò phức* từ các khái niệm nguyên tố (còn được gọi là *tên khái niệm*) và vai trò nguyên tố (còn được gọi là *tên vai trò*).

1.1.2. Biểu diễn tri thức trong logic mô tả

Từ các cá thể, các khái niệm và các vai trò, người ta có thể xây dựng một hệ thống để biểu diễn và suy luận tri thức dựa trên logic mô tả. Thông thường, một hệ thống biểu diễn và suy luận tri thức gồm có các thành phần sau:



Hình 1.1: Kiến trúc của một hệ cơ sở tri thức trong logic mô tả

• **Bộ tiên đề vai trò (*Role Box - RBox*)**: Bộ tiên đề vai trò chứa các tiên đề về vai trò bao gồm các tiên đề bao hàm vai trò và các khẳng định vai trò.

• **Bộ tiên đề thuật ngữ (*Terminology Box - TBox*)**: Bộ tiên đề thuật ngữ chứa các tiên đề về thuật ngữ, nó cho phép xây dựng các khái niệm phức từ những khái niệm nguyên tố và vai trò nguyên tố, đồng thời bộ tiên đề thuật ngữ cho biết mối quan hệ giữa các khái niệm thông qua các tiên đề bao hàm tổng quát.

• **Bộ khẳng định (*Assertion Box - ABox*)**: Bộ khẳng định dùng để chứa những

- Study bisimulation for the considered class of DLs. Theorems, lemmas, corollaries, propositions related to bisimulations and invariant results for bisimulation are developed and proved for the considered class of DLs.
- Develop bisimulation-based algorithms to partition the domain of model of knowledge bases and learn concepts for knowledge bases in DLs (Algorithm BBCL2).

5. Products

- 01 Master thesis, major: Computer Science
  (Supervisor: Dr. Hoang Thi Lan Giao, a co-applicant)
- 02 Bachelor theses, major: Informatics
  (Supervisor: Msc. Tran Thanh Luong, the coordinator)
- 04 papers published national/international journals/conferences
- A report of project.

6. Effects, transfer alternatives of research results and applicability

Bisimulation-based concept learning method for knowledge bases in DLs can be applied to find and build definitions of suitable concepts for ontologies and Semantic Webs.

This project is a material for students and researchers in the computer science major in general as well as DLs and machine learning in particular.

Application address: Department of Information Technology, Department of Informatics and Department of Computer Science in Universities of Vietnam.

MỞ ĐẦU

Logic mô tả (*Description Logics*) là một họ các ngôn ngữ hình thức rất thích hợp cho việc biểu diễn và suy luận tri thức trong một miền quan tâm cụ thể. Nó có tầm quan trọng đặc biệt trong việc cung cấp mô hình lý thuyết cho các hệ thống ngữ nghĩa và ontology. Trong logic mô tả, miền quan tâm được mô tả thông qua các thuật ngữ về cá thể, khái niệm và vai trò. Một cá thể đại diện cho một đối tượng, một khái niệm đại diện cho một tập các đối tượng và một vai trò đại diện cho một quan hệ hai ngôi giữa các đối tượng. Các khái niệm phức được xây dựng từ các tên khái niệm, tên vai trò và tên cá thể bằng cách kết hợp với các tạo tử.

Đặc tả các khái niệm phù hợp cho các hệ thống ngữ nghĩa là một trong những vấn đề rất được quan tâm. Do vậy, vấn đề đặt ra là cần tìm được các khái niệm quan trọng và xây dựng được định nghĩa của các khái niệm đó. Học khái niệm trong logic mô tả nhằm mục đích kiểm tra, suy luận và tìm ra được các khái niệm này phục vụ cho các ứng dụng cụ thể.

Học khái niệm trong logic mô tả tương tự như việc phân lớp nhị phân trong học máy truyền thống. Tuy nhiên, việc học khái niệm trong ngữ cảnh logic mô tả khác với học máy truyền thống ở chỗ, 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 này là một trong những yếu tố làm giàu thêm ngữ nghĩa của hệ thống huấn luyện. Do đó các phương pháp học khái niệm trong logic mô tả cần phải tận dụng được chúng như là một lợi thế.

Vấn đề học khái niệm trong logic mô tả được đặt ra theo ba ngữ cảnh chính như sau:

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 $a \in E^-$,

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

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 $\mathcal{KB} \not\models C(a)$ 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 $\mathcal{I} \models \neg C(a)$.

Chúng tôi tiến hành nghiên cứu bài toán học khái niệm cho các cơ sở tri thức trong logic mô tả theo ngữ cảnh (2), gọi tắt là *học khái niệm cho các cơ sở tri thức trong logic mô tả*. Từ các khảo sát như đã trình bày ở trên trên, mục tiêu chính của đề tài đặt ra là:

- Nghiên cứu cú pháp và ngữ nghĩa đối với một lớp lớn các logic mô tả, trong đó có những logic mô tả hữu ích như $\mathcal{SHOIQ}$, $\mathcal{SROIQ}$,... và xây dựng mô phỏng hai chiều cho lớp các logic mô tả đó.
- Xây dựng phương pháp làm mịn phân hoạch miền của các diễn dịch trong logic mô tả sử dụng mô phỏng hai chiều và các độ đo dựa trên entropy.
- 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) sử dụng mô phỏng hai chiều.

tra tính *thỏa mãn/không thỏa mãn* của một khái niệm trong một cơ sở tri thức.

1.5.2. Các thuật toán suy luận

1.5.2.1. Thuật toán bao hàm theo cấu trúc

Thuật toán bao hàm theo cấu trúc thực hiện quá trình suy luận dựa trên việc so sánh cấu trúc cú pháp của các khái niệm (thường đã được chuyển về ở dạng chuẩn phủ định). Thuật toán này tỏ ra hiệu quả đối với các ngôn ngữ logic mô tả đơn giản có khả năng biểu diễn yếu như $\mathcal{FL}_0$, $\mathcal{FL}_\perp$, $\mathcal{ALN}$. Với lớp ngôn ngữ logic mô tả rộng hơn, chẳng hạn như $\mathcal{ALC}$, $\mathcal{ALCI}$, $\mathcal{ALCIQ}$, $\mathcal{SHIQ}$, $\mathcal{SHOIQ}$, thuật toán bao hàm theo cấu trúc không thể giải quyết được các bài toán suy luận cơ bản.

1.5.2.2. Thuật toán tableaux

Để khắc phục những nhược điểm của thuật toán bao hàm theo cấu trúc, năm 1991, Schmidt-Schauß và Smolka đề xuất thuật toán tableaux để kiểm tra tính thỏa mãn của một khái niệm trong logic mô tả $\mathcal{ALC}$. Hướng tiếp cận này sau đó đã được áp dụng trên một lớp lớn các logic mô tả là logic mở rộng của $\mathcal{ALC}$ và được áp dụng để cài đặt các bộ suy luận FaCT, FaCT^{++}, RACER, CEL và KAON 2.

Chương 2.

MÔ PHỎNG HAI CHIỀU TRONG LOGIC MÔ TẢ VÀ TÍNH BẤT BIẾN ĐỐI VỚI MÔ PHỎNG HAI CHIỀU

2.1. Giới thiệu

Mô phỏng hai chiều được J. van Benthem giới thiệu lần đầu dưới tên gọi *p-quan hệ* (*p-relation*) và *quan hệ zig-zag* (*zig-zag relation*). Nó được phát triển trong logic hình thái (*modal logic*) và trong các hệ thống chuyển trạng thái (*state transition systems*). Mô phỏng hai chiều là một quan hệ nhị phân cho phép đặc tả tính tương tự giữa hai trạng thái cũng như tính tương tự giữa các mô hình Kripke.

2.2. Mô phỏng hai chiều trong logic mô tả

2.2.1. Mô phỏng hai chiều

Định nghĩa 2.1 (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 $a \in \Sigma_I^\dagger$, $A \in \Sigma_C^\dagger$, $B \in \Sigma_A^\dagger \setminus \Sigma_C^\dagger$, $r \in \Sigma_{oR}^\dagger$, $\sigma \in \Sigma_{dR}^\dagger$, $d \in range(\sigma)$, $x, y \in \Delta^{\mathcal{I}}$, $x', y' \in \Delta^{\mathcal{I}'}$:

$$Z(a^{\mathcal{I}}, a^{\mathcal{I}'}) \tag{2.1}$$

$$Z(x,x') \Rightarrow [A^{\mathcal{I}}(x) \Leftrightarrow A^{\mathcal{I}'}(x')] \tag{2.2}$$

$$Z(x,x') \Rightarrow [B^{\mathcal{I}}(x) = B^{\mathcal{I}'}(x') \text{ hoặc đều không xác định}] \tag{2.3}$$

$$[Z(x,x') \wedge r^{\mathcal{I}}(x,y)] \Rightarrow \exists y' \in \Delta^{\mathcal{I}'} \mid [Z(y,y') \wedge r^{\mathcal{I}'}(x',y')] \tag{2.4}$$

$$[Z(x,x') \wedge r^{\mathcal{I}'}(x',y')] \Rightarrow \exists y \in \Delta^{\mathcal{I}} \mid [Z(y,y') \wedge r^{\mathcal{I}}(x,y)] \tag{2.5}$$

$$Z(x,x') \Rightarrow [\sigma^{\mathcal{I}}(x,d) \Leftrightarrow \sigma^{\mathcal{I}'}(x',d)], \tag{2.6}$$

nếu $\mathcal{I} \in \Phi^\dagger$ thì

$$[Z(x,x') \wedge r^{\mathcal{I}}(y,x)] \Rightarrow \exists y' \in \Delta^{\mathcal{I}'} \mid [Z(y,y') \wedge r^{\mathcal{I}'}(y',x')] \tag{2.7}$$

$$[Z(x,x') \wedge r^{\mathcal{I}'}(y',x')] \Rightarrow \exists y \in \Delta^{\mathcal{I}} \mid [Z(y,y') \wedge r^{\mathcal{I}}(y,x)], \tag{2.8}$$

nếu $\mathcal{O} \in \Phi^\dagger$ thì

$$Z(x,x') \Rightarrow [x = a^{\mathcal{I}} \Leftrightarrow x' = a^{\mathcal{I}'}], \tag{2.9}$$

nếu $\mathcal{N} \in \Phi^\dagger$ thì

$$Z(x,x') \Rightarrow \#\{y \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(x,y)\} = \#\{y' \in \Delta^{\mathcal{I}'} \mid r^{\mathcal{I}'}(x',y')\}, \tag{2.10}$$

nếu $\{\mathcal{N}, \mathcal{I}\} \subseteq \Phi^\dagger$ thì

$$Z(x,x') \Rightarrow \#\{y \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(y,x)\} = \#\{y' \in \Delta^{\mathcal{I}'} \mid r^{\mathcal{I}'}(y',x')\}, \tag{2.11}$$

nếu $\mathcal{F} \in \Phi^\dagger$ thì

$$Z(x,x') \Rightarrow [\#\{y \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(x,y)\} \leq 1 \Leftrightarrow \#\{y' \in \Delta^{\mathcal{I}'} \mid r^{\mathcal{I}'}(x',y')\} \leq 1], \tag{2.12}$$

nếu $\{\mathcal{F}, \mathcal{I}\} \subseteq \Phi^\dagger$ thì

$$Z(x,x') \Rightarrow [\#\{y \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(y,x)\} \leq 1 \Leftrightarrow \#\{y' \in \Delta^{\mathcal{I}'} \mid r^{\mathcal{I}'}(y',x')\} \leq 1], \tag{2.13}$$

nếu $\mathcal{Q} \in \Phi^\dagger$ thì

$$\begin{array}{l}\text{nếu } Z(x,x') \text{ thỏa mãn thì với mọi } r \in \Sigma^\dagger_{oR}, \text{ tồn tại một song ánh} \\ h : \{y \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(x,y)\} \to \{y' \in \Delta^{\mathcal{I}'} \mid r^{\mathcal{I}'}(x',y')\} \text{ sao cho } h \subseteq Z,\end{array} \tag{2.14}$$

nếu $\{\mathcal{Q}, \mathcal{I}\} \subseteq \Phi^\dagger$ thì

$$\begin{array}{l}\text{nếu } Z(x,x') \text{ thỏa mãn thì với mọi } r \in \Sigma^\dagger_{oR}, \text{ tồn tại một song ánh} \\ h : \{y \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(y,x)\} \to \{y' \in \Delta^{\mathcal{I}'} \mid r^{\mathcal{I}'}(y',x')\} \text{ sao cho } h \subseteq Z,\end{array} \tag{2.15}$$

nếu $\mathcal{U} \in \Phi^\dagger$ thì

$$\forall x \in \Delta^{\mathcal{I}},\ \exists x' \in \Delta^{\mathcal{I}'}, Z(x,x') \tag{2.16}$$
$$\forall x' \in \Delta^{\mathcal{I}'},\ \exists x \in \Delta^{\mathcal{I}}, Z(x,x'), \tag{2.17}$$

nếu $\mathsf{Self} \in \Phi^\dagger$ thì

$$Z(x,x') \Rightarrow [r^{\mathcal{I}}(x,x) \Leftrightarrow r^{\mathcal{I}'}(x',x')], \tag{2.18}$$

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

Bổ đề 2.1.

1. *Quan hệ $\{\langle x,x\rangle \mid x \in \Delta^{\mathcal{I}}\}$ là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}$.*
2. *Nếu Z là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$ thì Z^{-1} cũng là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}'$ và $\mathcal{I}$.*
3. *Nếu Z_1 là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}_0$ và $\mathcal{I}_1$, Z_2 là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}_1$ và $\mathcal{I}_2$ thì $Z_1 \circ Z_2$ là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}_0$ và $\mathcal{I}_2$.*
4. *Nếu $\mathcal{Z}$ là một tập các $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$ thì $\bigcup \mathcal{Z}$ là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$.* ■

2.2.2. Quan hệ tương tự hai chiều và quan hệ tương đương

Định nghĩa 2.2. Cho $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$. Ta nói rằng $\mathcal{I}$ $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-*tương tự hai chiều* với $\mathcal{I}'$ nếu tồn tại một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$. ■

Định nghĩa 2.3. Cho $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$, $x \in \Delta^{\mathcal{I}}$ và $x' \in \Delta^{\mathcal{I}'}$. Ta nói rằng x $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-*tương tự hai chiều* với x' nếu tồn tại một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$ sao cho $Z(x, x')$ thỏa mãn. ■

Định nghĩa 2.4. Cho $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$, $x \in \Delta^{\mathcal{I}}$ và $x' \in \Delta^{\mathcal{I}'}$. Ta nói rằng x $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-*tương đương* với x' nếu với mọi khái niệm C của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, $x \in C^{\mathcal{I}}$ khi và chỉ khi $x' \in C^{\mathcal{I}'}$. ■

Theo Bổ đề 2.1, chúng ta thấy rằng quan hệ tương tự hai chiều giữa các diễn dịch là một quan hệ tương đương và quan hệ tương tự hai chiều giữa các phần tử trong diễn dịch cũng là một quan hệ tương đương.

2.3. Tính bất biến đối với mô phỏng hai chiều

2.3.1. Quan hệ giữa mô phỏng hai chiều với các khái niệm và vai trò

Bổ đề 2.2. *Cho $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$, Z là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$. Lúc đó, với mọi khái niệm C của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, mọi vai trò đối tượng R của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, mọi đối tượng $x, y \in \Delta^{\mathcal{I}}$, $x', y' \in \Delta^{\mathcal{I}'}$ và mọi cá thể $a \in \Sigma_I^\dagger$, các điều kiện sau sẽ được thỏa mãn:*

$$Z(x, x') \Rightarrow [C^{\mathcal{I}}(x) \Leftrightarrow C^{\mathcal{I}'}(x')] \tag{2.19}$$

$$[Z(x, x') \wedge R^{\mathcal{I}}(x, y)] \Rightarrow \exists y' \in \Delta^{\mathcal{I}'} \mid [Z(y, y') \wedge R^{\mathcal{I}'}(x', y')] \tag{2.20}$$

$$[Z(x, x') \wedge R^{\mathcal{I}'}(x', y')] \Rightarrow \exists y \in \Delta^{\mathcal{I}} \mid [Z(y, y') \wedge R^{\mathcal{I}}(x, y)], \tag{2.21}$$

nếu $\mathcal{O} \in \Phi^\dagger$ thì:

$$Z(x, x') \Rightarrow [R^{\mathcal{I}}(x, a^{\mathcal{I}}) \Leftrightarrow R^{\mathcal{I}'}(x', a^{\mathcal{I}'})]. \quad ■ \tag{2.22}$$

2.3.2. Tính bất biến của khái niệm

Định nghĩa 2.5 (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 ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ thỏa $\Sigma^\dagger \subseteq \Sigma$, $\Phi^\dagger \subseteq \Sigma$ và với mọi $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều Z giữa $\mathcal{I}$ và $\mathcal{I}'$. ■

Định lý 2.1. *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.* ■

Định lý này cho phép mô hình hóa tính không phân biệt được của các đối tượng thông qua ngôn ngữ con $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$. Tính không phân biệt của các đối tượng là một trong những đặc trưng cơ bản trong quá trình phân lớp dữ liệu. Điều này có nghĩa là chúng ta có thể sử dụng ngôn ngữ con $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ cho các bài toán học máy trong logic mô tả.

2.3.3. Tính bất biến của cơ sở tri thức

Định nghĩa 2.6. Một TBox $\mathcal{T}$ (tương ứng, ABox $\mathcal{A}$) trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ được gọi là *bất biến đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều* nếu với mọi diễn dịch $\mathcal{I}$ và $\mathcal{I}'$ trong $\mathcal{L}_{\Sigma,\Phi}$ tồn tại một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$ sao cho $\mathcal{I}$ là mô hình của $\mathcal{T}$ (tương ứng, $\mathcal{A}$) khi và chỉ khi $\mathcal{I}'$ là mô hình của $\mathcal{T}$ (tương ứng, $\mathcal{A}$). ∎

Hệ quả 2.1. *Nếu $\mathcal{U} \in \Phi^\dagger$ thì tất cả các TBox trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ đều bất biến đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều.* ∎

Một diễn dịch $\mathcal{I}$ trong $\mathcal{L}_{\Sigma,\Phi}$ được gọi là *kết nối đối tượng được đối với* $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ nếu với mọi đối tượng $x \in \Delta^\mathcal{I}$ tồn tại cá thể $a \in \Sigma_I^\dagger$, các đối tượng $x_0, x_1, \ldots, x_k \in \Delta^\mathcal{I}$ và các vai trò đối tượng cơ bản $R_1, R_2, \ldots, R_k$ của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ với $k \geq 0$ sao cho $x_0 = a^\mathcal{I}$, $x_k = x$ và $R_i^\mathcal{I}(x_{i-1}, x_i)$ thỏa mãn với mọi $1 \leq i \leq k$.

Định lý 2.2. *Cho $\mathcal{T}$ là một TBox trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch trong $\mathcal{L}_{\Sigma,\Phi}$ thỏa điều kiện kết nối đối tượng được đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ sao cho tồn tại một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$. Lúc đó $\mathcal{I}$ là mô hình của $\mathcal{T}$ khi và chỉ khi $\mathcal{I}'$ là mô hình của $\mathcal{T}$.* ∎

Định lý 2.3. *Cho $\mathcal{A}$ là một ABox trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$. Nếu $\mathcal{O} \in \Phi^\dagger$ hoặc $\mathcal{A}$ chỉ chứa các khẳng định dạng $C(a)$ thì $\mathcal{A}$ bất biến đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều.* ∎

Hệ quả 2.2. *Cho cơ sở tri thức $\mathcal{KB} = \langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$ trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ sao cho $\mathcal{R} = \emptyset$ và giả thiết $\mathcal{O} \in \Phi^\dagger$ hoặc $\mathcal{A}$ chỉ chứa các khẳng định có dạng $C(a)$, $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch kết nối đối tượng được trong $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ sao cho tồn tại một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$. Lúc đó $\mathcal{I}$ là mô hình của $\mathcal{KB}$ khi và chỉ khi $\mathcal{I}'$ là mô hình của $\mathcal{KB}$.* ∎

2.4. Tính chất Hennessy-Milner đối với mô phỏng hai chiều

Định nghĩa 2.7. Một diễn dịch $\mathcal{I}$ trong $\mathcal{L}_{\Sigma,\Phi}$ được gọi là *phân nhánh hữu hạn* (hay *hữu hạn ảnh*) đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ nếu với mọi $x \in \Delta^\mathcal{I}$ và với mọi vai trò $r \in \Sigma_{oR}^\dagger$ thì:

- tập $\{y \in \Delta^\mathcal{I} \mid r^\mathcal{I}(x, y)\}$ là hữu hạn,
- nếu $\mathcal{I} \in \Phi^\dagger$ thì tập $\{y \in \Delta^\mathcal{I} \mid r^\mathcal{I}(y, x)\}$ là hữu hạn. ∎

Định lý 2.4 (Tính chất Hennessy-Milner)**.** *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 $\mathcal{L}_{\Sigma,\Phi}$ thỏa mãn điều kiện phân nhánh hữu hạn đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, sao cho với mọi $a \in \Sigma_I^\dagger$, $a^\mathcal{I}$ $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tương đương với $a^{\mathcal{I}'}$. Giả thiết rằng $\mathcal{U} \notin \Phi^\dagger$ hoặc $\Sigma_I^\dagger \neq \emptyset$. Lúc đó, $x \in \Delta^\mathcal{I}$ $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tương đương với $x' \in \Delta^{\mathcal{I}'}$ khi và chỉ khi tồn tại một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều Z giữa $\mathcal{I}$ và $\mathcal{I}'$ sao cho $Z(x, x')$ thỏa mãn.* ∎

Hệ quả 2.3. *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 $\mathcal{L}_{\Sigma,\Phi}$ thỏa điều* *kiện phân nhánh hữu hạn đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$. Giả thiết rằng $\Sigma_I^\dagger \neq \emptyset$ và với mọi $a \in \Sigma_I^\dagger$, $a^\mathcal{I}$ $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tương đương với $a^{\mathcal{I}'}$. Lúc đó, quan hệ $\{\langle x, x'\rangle \in \Delta^\mathcal{I} \times \Delta^{\mathcal{I}'} \mid x\ \mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tương đương với $x'\}$ là một $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều giữa $\mathcal{I}$ và $\mathcal{I}'$.* ■

2.5. Tự mô phỏng hai chiều

Định nghĩa 2.8 (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 $Z' \subseteq Z$. ■

Cho $\mathcal{I}$ là một diễn dịch trong $\mathcal{L}_{\Sigma,\Phi}$, chúng ta ký hiệu $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều lớn nhất của $\mathcal{I}$ là $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$, và ký hiệu quan hệ nhị phân $\equiv_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$ 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'.

Định lý 2.5. *Cho Σ và $\Sigma^\dagger$ là các bộ ký tự của logic mô tả sao cho $\Sigma^\dagger \subseteq \Sigma$, Φ và $\Phi^\dagger$ là tập các đặc trưng của logic mô tả sao cho $\Phi^\dagger \subseteq \Phi$, $\mathcal{I}$ là một diễn dịch trong $\mathcal{L}_{\Sigma,\Phi}$. Lúc đó:*

1. *$\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều lớn nhất của $\mathcal{I}$ tồn tại và nó là một quan hệ tương đương,*
2. *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).* ■

Chúng ta nói rằng tập Y *bị phân chia* bởi tập X nếu $Y \setminus X \neq \emptyset$ và $Y \cap X \neq \emptyset$. Như vậy, tập Y không bị phân chia bởi tập X nếu hoặc $Y \subseteq X$ hoặc $Y \cap X = \emptyset$. Phân hoạch $\mathbb{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ được gọi là *nhất quán* với tập X nếu với mọi $1 \leq i \leq n$, Y_i không bị phân chia bởi X.

Định lý 2.6. *Cho Σ và $\Sigma^\dagger$ là các bộ ký tự của logic mô tả sao cho $\Sigma^\dagger \subseteq \Sigma$, Φ và $\Phi^\dagger$ là tập các đặc trưng của logic mô tả sao cho $\Phi^\dagger \subseteq \Phi$, $\mathcal{I}$ là một diễn dịch hữu hạn trong $\mathcal{L}_{\Sigma,\Phi}$ và $X \subseteq \Delta^\mathcal{I}$. Gọi $\mathbb{Y}$ là phân hoạch của $\Delta^\mathcal{I}$ thông qua quan hệ $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$. Lúc đó:*

1. *nếu tồn tại khái niệm C của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ sao cho $C^\mathcal{I} = X$ thì phân hoạch $\mathbb{Y}$ nhất quán với tập X,*
2. *nếu phân hoạch $\mathbb{Y}$ nhất quán với tập X thì tồn tại khái niệm C của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ sao cho $C^\mathcal{I} = X$.* ■

KẾT LUẬN

Kết luận

Kết quả nghiên cứu của đề tài được tóm tắt như sau:

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 (bao gồm thuộc tính rời rạc và liên tục) như là các phần tử cơ bản của ngôn ngữ nhằm mô tả các hệ thống thông tin phù hợp với thực tế hơn.
2. Xây dựng mô phỏng hai chiều trên lớp các logic mở rộng đang nghiên cứu.
3. Xây dựng thuật toán phân hoạch miền của mô hình của cơ sở tri thức và xây dựng thuật toán BBCL2 để học khái niệm trong logic mô tả cho cơ sở tri thức với ngữ cảnh (2).

Những vấn đề cần tiếp tục nghiên cứu

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 khả năng học chính xác khái niệm cho các logic mô tả khác nhau.

cách tổng quát, bài toán kiểm tra tính thỏa trong logic mô tả thường là ExpTime*-đầy đủ. Thuật toán BBCL2 sử dụng một vòng lặp tuyến tính có giới hạn là lực lượng của C_0 cho bài toán suy luận. Do đó 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).*

3.3.3. Ví dụ minh họa

Ví dụ 3.2. Xét cơ sở tri thức $\mathcal{KB}_0 = \langle \mathcal{R}, \mathcal{T}, \mathcal{A}_0 \rangle$ như đã cho trong Ví dụ 1.1 và $E = \langle E^+, E^- \rangle$ đối với $E^+ = \{P_4, P_6\}$, $E^- = \{P_1, P_2, P_3, P_5\}$, $\Sigma^\dagger = \{Awarded, cited\_by\}$ và $\Phi^\dagger = \emptyset$. Học định nghĩa cho khái niệm A_d với $\mathcal{KB} = \langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$, trong đó $\mathcal{A} = \mathcal{A}_0 \cup \{A_d(a) \mid a \in E^+\} \cup \{\neg A_d(a) \mid a \in E^-\}$. Thuật toán BBCL2 thực hiện các bước như sau:

1. $\mathbb{C} := \emptyset$, $\mathbb{C}_0 := \emptyset$, $E_0^- := \emptyset$.
2. $\mathcal{KB}$ có nhiều mô hình, trong đó mô hình $\mathcal{I}$ được đặc tả trong Ví dụ 3.1 như sau:

$$\begin{array}{rcl} \Delta^\mathcal{I} & = & \{P_1, P_2, P_3, P_4, P_5, P_6\}, \qquad x^\mathcal{I} = x \text{ với } x \in \{P_1, P_2, P_3, P_4, P_5, P_6\}, \\ Pub^\mathcal{I} & = & \Delta^\mathcal{I}, \quad Awarded^\mathcal{I} = \{P_1, P_4, P_6\}, \quad UsefulPub^\mathcal{I} = \{P_2, P_3, P_4, P_5, P_6\}, \\ cites^\mathcal{I} & = & \{\langle P_1, P_2\rangle, \langle P_1, P_3\rangle, \langle P_1, P_4\rangle, \langle P_1, P_6\rangle, \langle P_2, P_3\rangle, \langle P_2, P_4\rangle, \\ & & \langle P_2, P_5\rangle, \langle P_3, P_4\rangle, \langle P_3, P_5\rangle, \langle P_3, P_6\rangle, \langle P_4, P_5\rangle, \langle P_4, P_6\rangle\}, \\ cited\_by^\mathcal{I} & = & (cites^\mathcal{I})^{-1}, \text{ hàm từng phần } Year^\mathcal{I} \text{ được đặc tả theo từng cá thể.} \end{array}$$

3. Áp dụng Thuật toán 3.1 để làm mịn phân hoạch $\{\Delta^\mathcal{I}\}$ của $\mathcal{I}$, chúng ta thu được phân hoạch $\mathbb{Y} = \{Y_3, Y_4, Y_5\}$ nhất quán với E tương ứng với các khái niệm đặc trưng C_3, C_4, C_5, trong đó $Y_3 = \{P_2, P_3, P_5\}$, $Y_4 = \{P_4, P_6\}$, $Y_5 = \{P_1\}$ và $C_3 \equiv \neg Awarded$, $C_4 \equiv Awarded \sqcap \exists cited\_by.\top$, $C_5 \equiv \neg Awarded \sqcap \exists cited\_by.\top$. (Xem quá trình phân hoạch ở Ví dụ 3.1)
4. Vì $Y_3 \subseteq E^-$ nên ta tiến hành xem xét đối với $C_3 \equiv \neg Awarded$. Vì $\mathcal{KB} \models \neg C_3(a)$ với mọi $a \in E^+$ nên ta thêm $\neg C_3$ vào $\mathbb{C}$ và thêm các phần tử của Y_3 vào E_0^-. Do đó, ta có $\mathbb{C} = \{C_3\}$ và $E_0^- = \{P_2, P_3, P_5\}$.
5. Vì $Y_5 \subseteq E^-$ nên ta tiến hành xem xét đối với $C_5 \equiv Awarded \sqcap \neg\exists cited\_by.\top$. Vì $\mathcal{KB} \models \neg C_5(a)$ với mọi $a \in E^+$ và $\bigsqcap \mathbb{C}$ không bị bao hàm bởi C_5 dựa trên $\mathcal{KB}$ nên ta thêm $\neg C_5$ vào $\mathbb{C}$. Do đó, ta có $\mathbb{C} = \{\neg C_3, \neg C_5\}$, $\bigsqcap \mathbb{C} \equiv \neg C_3 \sqcap \neg C_5 \equiv \neg\neg Awarded \sqcap \neg(Awarded \sqcap \neg\exists cited\_by.\top)$ và $E_0^- = \{P_1, P_2, P_3, P_5\}$.
6. Vì $E_0^- = E^-$ nên ta có $C \equiv \bigsqcap \mathbb{C} \equiv \neg\neg Awarded \sqcap \neg(Awarded \sqcap \neg\exists cited\_by.\top)$. Rút gọn C ta được kết quả trả về là $C_{rs} \equiv Awarded \sqcap \exists cited\_by.\top$. ■

Chương 3.

HỌC KHÁI NIỆM TRONG LOGIC MÔ TẢ SỬ DỤNG MÔ PHỎNG HAI CHIỀU

3.1. Giới thiệu

Trong những năm gần đây, logic mô tả đã và đang được ứng dụng trong nhiều lĩnh vực của Web ngữ nghĩa. Nó được xem là một thành phần quan trọng trong các hệ thống ngữ nghĩa và được sử dụng để biểu diễn cũng như suy luận tri thức trong các hệ thống này. Đặc tả các khái niệm phù hợp cho các hệ thống ngữ nghĩa là một trong những vấn đề rất được quan tâm. Do vậy, vấn đề đặt ra là cần tìm được các khái niệm quan trọng và xây dựng được định nghĩa của các khái niệm đó. Học khái niệm trong logic mô tả nhằm mục đích kiểm tra, suy luận và tìm ra được các khái niệm này để phục vụ cho các ứng dụng khác nhau như: tin sinh học, tin học trong y tế, quản trị tri thức, kỹ nghệ phần mềm, . . .

Vấn đề học khái niệm cho cơ sở tri thức trong logic mô tả của đề tài này được đặt ra theo ngữ cảnh như sau:

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

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

3.2. Phân hoạch miền của diễn dịch trong logic mô tả

Cho diễn dịch $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ trong ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$. Ý tưởng cơ bản của việc làm mịn phân hoạch miền $\Delta^\mathcal{I}$ của diễn dịch $\mathcal{I}$ là dựa trên phương pháp học khái niệm sử dụng mô phỏng hai chiều. Bắt đầu từ phân hoạch $\{\Delta^\mathcal{I}\}$, chúng ta thực hiện làm mịn liên tục $\{\Delta^\mathcal{I}\}$ để đạt được phân hoạch tương ứng với quan hệ $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$ với các kỹ thuật cụ thể đặt ra như sau:

- Quá trình làm mịn có thể dừng lại khi một số điều kiện đặt ra được thỏa mãn.
- Trong quá trình làm mịn phân hoạch $\{\Delta^\mathcal{I}\}$, các khối được tạo ra ở tất cả các bước được ký hiệu là $Y_1, Y_2, \ldots, Y_n$. Để thực hiện được điều này, khi mỗi khối được tạo ra, chúng ta sử dụng một chỉ số mới gán cho khối đó bằng cách tăng giá trị của n.

Ta gọi phân hoạch hiện thời là $\mathbb{Y} = \{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}\} \subseteq \{Y_1, Y_2, \ldots, Y_n\}$. Như vậy, $\mathbb{Y}$ là một tập con của các khối được tạo ra ở trên. Chúng ta thiết lập thông tin nhằm ghi nhận lại khái niệm C_i đặc trưng cho khối Y_i sao cho $C_i^{\mathcal{I}} = Y_i$.

- Chúng ta có thể sử dụng các chiến lược khác nhau cũng như các hàm tính điểm để tối ưu hóa quá trình làm mịn.

3.2.1. Bộ chọn cơ bản

Định nghĩa 3.1 (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 $A \in \Sigma_C^\dagger$,
- $A = d$, trong đó $A \in \Sigma_A^\dagger \setminus \Sigma_C^\dagger$ và $d \in range(A)$,
- $\exists\sigma.\{d\}$, trong đó $\sigma \in \Sigma_{dR}^\dagger$ và $d \in range(\sigma)$,
- $\exists r.C_{i_t}$, trong đó $r \in \Sigma_{oR}^\dagger$ và $1 \leq t \leq k$,
- $\exists r^-.C_{i_t}$, nếu $\mathcal{I} \in \Phi^\dagger$, $r \in \Sigma_{oR}^\dagger$ và $1 \leq t \leq k$,
- $\{a\}$, nếu $\mathcal{O} \in \Phi^\dagger$ và $a \in \Sigma_I^\dagger$,
- $\leq 1\, r$, nếu $\mathcal{F} \in \Phi^\dagger$ và $r \in \Sigma_{oR}^\dagger$,
- $\leq 1\, r^-$, nếu $\{\mathcal{F}, \mathcal{I}\} \subseteq \Phi^\dagger$ và $r \in \Sigma_{oR}^\dagger$,
- $\geq l\, r$ và $\leq m\, r$, nếu $\mathcal{N} \in \Phi^\dagger$, $r \in \Sigma_{oR}^\dagger$, $0 < l \leq \#\Delta^{\mathcal{I}}$ và $0 \leq m < \#\Delta^{\mathcal{I}}$,
- $\geq l\, r^-$ và $\leq m\, r^-$, nếu $\{\mathcal{N}, \mathcal{I}\} \subseteq \Phi^\dagger$, $r \in \Sigma_{oR}^\dagger$, $0 < l \leq \#\Delta^{\mathcal{I}}$ và $0 \leq m < \#\Delta^{\mathcal{I}}$,
- $\geq l\, r.C_{i_t}$ và $\leq m\, r.C_{i_t}$, nếu $\mathcal{Q} \in \Phi^\dagger$, $r \in \Sigma_{oR}^\dagger$, $1 \leq t \leq k$, $0 < l \leq \#C_{i_t}$ và $0 \leq m < \#C_{i_t}$,
- $\geq l\, r^-.C_{i_t}$ và $\leq m\, r^-.C_{i_t}$, nếu $\{\mathcal{Q}, \mathcal{I}\} \subseteq \Phi^\dagger$, $r \in \Sigma_{oR}^\dagger$, $1 \leq t \leq k$, $0 < l \leq \#C_{i_t}$ và $0 \leq m < \#C_{i_t}$,
- $\exists r.\mathsf{Self}$, nếu $\mathsf{Self} \in \Phi^\dagger$ và $r \in \Sigma_{oR}^\dagger$. ■

Định lý 3.1. *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 logic mô tả sao cho $\Phi^\dagger \subseteq \Phi$, $\mathcal{I}$ là một diễn dịch hữu hạn trong $\mathcal{L}_{\Sigma,\Phi}$. Xuất phát từ phân hoạch $\{\Delta^{\mathcal{I}}\}$ và thực hiện việc làm mịn liên tục nó bằng các bộ chọn cơ bản ta sẽ nhận được một phân hoạch tương ứng với quan hệ tương đương $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$.* ■

Trong thực tế, để quá trình phân hoạch đạt hiệu quả cao, chúng ta có thể xem xét sử dụng các bộ chọn được trình bày trong Hình 3.1.

Algorithm 3.2: *BBCL2–Học khái niệm đối với cơ sở tri thức trong logic mô tả*

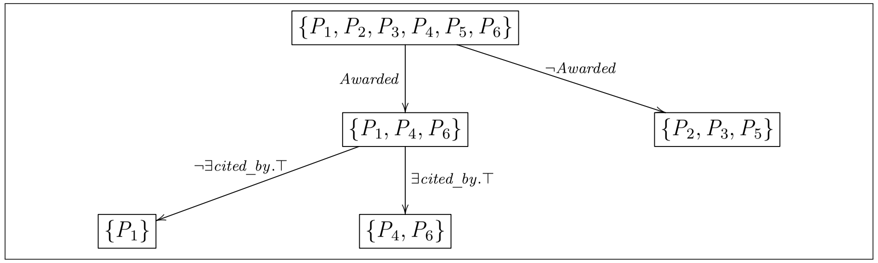
Input: $\mathcal{KB}$, $\Sigma^\dagger$, $\Phi^\dagger$, $E = \langle E^+, E^- \rangle$, K

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 miền của $\mathcal{I}$ */
  foreach $Y_{i_j} \in \mathbb{Y}$, $\exists a \in E^- : a^{\mathcal{I}} \in Y_{i_j}$ 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 ($\mathcal{KB} \not\models (\bigsqcap \mathbb{C} \sqsubseteq \neg C_{i_j})$) **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 ($\mathcal{KB} \not\models (\bigsqcap \mathbb{C}) \sqsubseteq D$) and ($\exists a \in E^- \setminus E_0^- : \mathcal{KB} \not\models (\bigsqcap \mathbb{C})(a)$) **then**
      $\mathbb{C} := \mathbb{C} \cup \{D\}$;
      $E_0^- := E_0^- \cup \{a \mid a \in E^- \setminus E_0^-, \mathcal{KB} \not\models (\bigsqcap \mathbb{C})(a)\}$;
if ($E_0^- = E^-$) **then**
  foreach $D \in \mathbb{C}$ **do**
    if $\mathcal{KB} \not\models \bigsqcap(\mathbb{C} \setminus \{D\})(a), \forall a \in E^-$ **then**
      $\mathbb{C} := \mathbb{C} \setminus \{D\}$;
  $C := \bigsqcap \mathbb{C}$;
  return C_{rs} := `Simplify` (C);   /* rút gọn khái niệm C */
else
  return failure;

trong logic mô tả liên quan chặt chẽ với vấn đề 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 $\mathcal{ALC}$. Đến nay, các công trình nghiên cứu vẫn chưa chỉ ra được dấu hiệu nào cho thấy có thể hy vọng làm giảm độ phức tạp của bài toán suy luận. Một*



Hình 3.2: Cây quyết định minh họa quá trình làm mịn phân hoạch của Ví dụ 3.1

3.3. Học khái niệm trong logic mô tả

3.3.1. Thuật toán BBCL2

Ý tưởng chính của thuật toán BBCL2 để giải quyết bài toán này là sử dụng các mô hình của $\mathcal{KB}$ kết hợp với mô phỏng hai chiều trong mô hình đó (để mô hình hóa tính không phân biệt được) và cây quyết định (để phân lớp dữ liệu) cho việc tìm kiếm khái niệm C. Thuật toán này sử dụng Thuật toán 3.1 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 $E = \langle E^+, E^- \rangle$.

Thuật toán tiến hành xây dựng tập E_0^- và mở rộng nó sao cho E_0^- phủ càng lúc càng nhiều cá thể trong E^-, xây tập $\mathbb{C}$ gồm các phần tử là các khái niệm D sao cho $\mathcal{KB} \models D(a)$ với mọi $a \in E^+$ và xây dựng tập $\mathbb{C}_0$ gồm khái niệm để trợ giúp cho việc xây dựng khái niệm C. Khi một khái niệm D không thỏa mãn điều kiện $\mathcal{KB} \models \neg D(a)$ với mọi $a \in E^-$ nhưng nó là một khái niệm ứng viên "tốt" thì khái niệm D được đưa vào $\mathbb{C}_0$. Sau này, khi cần thiết, các khái niệm trong $\mathbb{C}_0$ được lấy ra, thực hiện phép hợp và kiểm tra xem nó có thỏa mãn điều kiện để thêm vào $\mathbb{C}$ hay không. Như vậy, trong quá trình học chúng ta luôn có:

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

Thuật toán mở rộng $\mathbb{C}$ sao cho $\mathcal{KB} \not\models (\bigsqcap \mathbb{C})(a)$ với càng lúc càng nhiều $a \in E^-$. Như vậy, mở rộng $\mathbb{C}$ đồng nghĩa với việc mở rộng E_0^-. Khi $E_0^- = E^-$ thuật toán trả về khái niệm $\bigsqcap \mathbb{C}$ sau khi đã thực hiện việc chuẩn hóa và đơn giản hóa.

3.3.2. Tính đúng của thuật toán BBCL2

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).* ∎

Ghi chú 3.2 (Độ phức tạp của thuật toán BBCL2)**.** *Học khái niệm cho cơ sở tri thức*

- $A < d$, $A \le d$, trong đó $A \in \Sigma_{nA}^{\dagger}$, $d \in range(A)$ và d không phải là giá trị nhỏ nhất của $range(A)$,
- $A > d$, $A \ge d$, trong đó $A \in \Sigma_{nA}^{\dagger}$, $d \in range(A)$ và d không phải là giá trị lớn nhất của $range(A)$,
- $\exists r.\top$, $\exists r.C_i$ và $\forall r.C_i$, trong đó $r \in \Sigma_{oR}^{\dagger}$ và $1 \le i \le n$,
- $\exists r^-.\top$, $\exists r^-.C_i$, $\forall r^-.C_i$, nếu $\mathcal{I} \in \Phi^{\dagger}$, $r \in \Sigma_{oR}^{\dagger}$ và $1 \le i \le n$,
- $\ge l\, r.C_i$ và $\le m\, r.C_i$, nếu $\mathcal{Q} \in \Phi^{\dagger}$, $r \in \Sigma_{oR}^{\dagger}$, $1 \le i \le n$, $0 < l \le \#C_i$ và $0 \le m < \#C_i$,
- $\ge l\, r^-.C_i$ và $\le m\, r^-.C_i$, nếu $\{\mathcal{Q},\mathcal{I}\} \subseteq \Phi^{\dagger}$, $r \in \Sigma_{oR}^{\dagger}$, $1 \le i \le n$, $0 < l \le \#C_i$ và $0 \le m < \#C_i$.

Hình 3.1: Các bộ chọn được sử dụng thêm trong thực tế

3.2.2. Gia lượng thông tin trong việc phân hoạch miền

Cho diễn dịch $\mathcal{I}$ 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 3.2. *Entropy* của tập Y đối với tập X trong miền $\Delta^{\mathcal{I}}$ của diễn dịch $\mathcal{I}$, ký hiệu là $E_{\Delta^{\mathcal{I}}}(Y/X)$, được xác định như sau:

$$E_{\Delta^{\mathcal{I}}}(Y/X) = \begin{cases} 0, \text{ nếu } Y \cap X = \emptyset \text{ hoặc } Y \subseteq X \\ -\dfrac{\#XY}{\#Y}\log_2\dfrac{\#XY}{\#Y} - \dfrac{\#\overline{X}Y}{\#Y}\log_2\dfrac{\#\overline{X}Y}{\#Y}, \text{nếu ngược lại,} \end{cases} \tag{3.1}$$

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

Ghi chú 3.1. *Theo phương trình* (3.1)*, chúng ta thấy rằng $E_{\Delta^{\mathcal{I}}}(Y/X) = 0$ khi và chỉ khi tập Y không bị phân chia bởi tập X.* ∎

Định nghĩa 3.3. *Gia lượng thông tin* của bộ chọn D trong việc chia tập Y đối với tập X trong $\Delta^{\mathcal{I}}$ của diễn dịch $\mathcal{I}$, ký hiệu là $IG_{\Delta^{\mathcal{I}}}(Y/X, D)$, được xác định như sau:

$$IG_{\Delta^{\mathcal{I}}}(Y/X, D) = E_{\Delta^{\mathcal{I}}}(Y/X) - \left(\frac{\#D^{\mathcal{I}}Y}{\#Y}E_{\Delta^{\mathcal{I}}}(D^{\mathcal{I}}Y/X) + \frac{\#\overline{D^{\mathcal{I}}}Y}{\#Y}E_{\Delta^{\mathcal{I}}}(\overline{D^{\mathcal{I}}}Y/X)\right) \tag{3.2}$$

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 $IG_{\Delta^{\mathcal{I}}}(Y/X, D)$.

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 $D_u \in \mathbb{D}$ và $Y_{i_j} \in \mathbb{Y}$ sao cho D_u chia Y_{i_j} thành hai khối không rỗng;
  $s := n + 1$; $t := n + 2$; $n := n + 2$;
  $Y_s := Y_{i_j} \cap D_u^\mathcal{I}$; $C_s := C_{i_j} \sqcap D_u$;
  $Y_t := Y_{i_j} \cap (\neg D_u)^\mathcal{I}$; $C_t := C_{i_j} \sqcap \neg D_u$;
  $\mathbb{Y} := \mathbb{Y} \cup \{Y_s, Y_t\} \setminus \{Y_{i_j}\}$;
  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;

3.2.3. Phân hoạch miền của diễn dịch

Ta nói rằng 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.

Cho diễn dịch $\mathcal{I}$ là một hệ thống thông tin huấn luyện trong $\mathcal{L}_{\Sigma,\Phi}$. Gọi $A_d \in \Sigma_C$ là một khái niệm đại diện cho "thuộc tính quyết định", $E = \langle E^+, E^- \rangle$ với $E^+ = \{a \mid a^\mathcal{I} \in A_d^\mathcal{I}\}$ và $E^- = \{a \mid a^\mathcal{I} \in (\neg A_d^\mathcal{I})\}$ tương ứng là tập các mẫu dương và mẫu âm của A_d trong $\mathcal{I}$. Giả sử rằng A_d có thể được biểu diễn bởi một khái niệm C trong ngôn ngữ con $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, trong đó $\Sigma^\dagger \subseteq \Sigma \setminus \{A_d\}$ và $\Phi^\dagger \subseteq \Phi$. Vấn đề đặt ra là phân hoạch miền $\Delta^\mathcal{I}$ của $\mathcal{I}$ sử dụng các khái niệm của $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$ để thu được phân hoạch $\mathbb{Y}$ nhất quán với E.

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

- theo khẳng định thứ nhất của Định lý 2.6, $C^\mathcal{I}$ 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 $a \in E^-$, nghĩa là phân hoạch $\mathbb{Y}$ nhất quán với E.

Với các nhận xét trên, chúng tôi thiết kế Thuật toán 3.1 để phân hoạch miền của một diễn dịch trong logic mô tả. Giả sử chúng ta có phân hoạch hiện thời là $\mathbb{Y} = \{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}\}$ và tập các bộ chọn hiện thời là $\mathbb{D} = \{D_1, D_2, \ldots, D_h\}$ (được xây dựng

bằng cách sử dụng các luật trong Định nghĩa 3.1 và Hình 3.1). Với mỗi khối $Y_{i_j} \in \mathbb{Y}$, gọi S_{i_j} là bộ chọn đơn giản nhất có được từ $\arg\max_{D_u \in \mathbb{D}}\{IG(Y_{i_j}, D_u)\}$. Bộ chọn S_{i_j} được chọn để phân hoạch khối Y_{i_j}. Tiếp theo, chúng ta chọn khối $Y_{i_j} \in \arg\max_{Y_{i_j} \in \mathbb{Y}}\{IG(Y_{i_j}, S_{i_j})\}$ để phân chia trước. Quá trình lựa chọn này dựa trên độ đo gia lượng thông tin.

Ví dụ 3.1. Xét cơ sở tri thức $\mathcal{KB}$ như đã cho trong Ví dụ 1.1 và diễn dịch $\mathcal{I}$ là mô hình của $\mathcal{KB}$ như sau:

$$\begin{aligned}
\Delta^\mathcal{I} &= \{P_1, P_2, P_3, P_4, P_5, P_6\}, \qquad x^\mathcal{I} = x \text{ với } x \in \{P_1, P_2, P_3, P_4, P_5, P_6\},\\
\mathit{Pub}^\mathcal{I} &= \Delta^\mathcal{I}, \quad \mathit{Awarded}^\mathcal{I} = \{P_1, P_4, P_6\}, \quad \mathit{UsefulPub}^\mathcal{I} = \{P_2, P_3, P_4, P_5, P_6\},\\
\mathit{cites}^\mathcal{I} &= \{\langle P_1, P_2\rangle, \langle P_1, P_3\rangle, \langle P_1, P_4\rangle, \langle P_1, P_6\rangle, \langle P_2, P_3\rangle, \langle P_2, P_4\rangle,\\
&\qquad \langle P_2, P_5\rangle, \langle P_3, P_4\rangle, \langle P_3, P_5\rangle, \langle P_3, P_6\rangle, \langle P_4, P_5\rangle, \langle P_4, P_6\rangle\},\\
\mathit{cited_by}^\mathcal{I} &= (\mathit{cites}^\mathcal{I})^{-1}, \text{ hàm từng phần } \mathit{Year}^\mathcal{I} \text{ được đặc tả theo từng cá thể.}
\end{aligned}$$

Cho $E = \langle E^+, E^- \rangle$ với $E^+ = \{P_4, P_6\}$ và $E^- = \{P_1, P_2, P_3, P_5\}$, ngôn ngữ con $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$, trong đó $\Sigma^\dagger = \{\mathit{Awarded}, \mathit{cited\_by}\}$ và $\Phi^\dagger = \emptyset$. Các bước làm mịn phân hoạch $\{\Delta^\mathcal{I}\}$ của diễn dịch $\mathcal{I}$ được mô tả như sau:

1. $Y_1 := \Delta^\mathcal{I}$, $C_1 := \top$, $\mathbb{Y} := \{Y_1\}$
2. Theo độ đo gia lượng thông tin, bộ chọn tốt nhất để phân chia Y_1 là *Awarded*. Phân chia khối Y_1 bởi *Awarded* chúng ta thu được:
   - $Y_2 := \{P_1, P_4, P_6\}$, $C_2 := \mathit{Awarded}$
   - $Y_3 := \{P_2, P_3, P_5\}$, $C_3 := \neg\mathit{Awarded}$
   - $\mathbb{Y} := \{Y_2, Y_3\}$
3. Theo độ đo gia lượng thông tin, các bộ chọn tốt nhất để phân chia khối Y_2 là $\exists \mathit{cited\_by}.\top$, $\exists \mathit{cited\_by}.C_2$ và $\exists \mathit{cited\_by}.C_3$. Chúng ta sử dụng bộ chọn đơn giản nhất $\exists \mathit{cited\_by}.\top$ để phân chia Y_2 và thu được:
   - $Y_4 := \{P_4, P_6\}$, $C_4 := C_2 \sqcap \exists \mathit{cited\_by}.\top$
   - $Y_5 := \{P_1\}$, $C_5 := C_2 \sqcap \neg\exists \mathit{cited\_by}.\top$
   - $\mathbb{Y} := \{Y_3, Y_4, Y_5\}$

Phân hoạch đạt được là $\mathbb{Y} = \{Y_3, Y_4, Y_5\}$ nhất quán với E, gồm khối Y_4 chứa P_4, P_6 với $P_4, P_6 \in E^+$ và các khối Y_3, Y_5 không chứa cá thể nào của E^+ nên ta có kết quả trả về là $\mathbb{Y} = \{Y_3, Y_4, Y_5\}$ (phân hoạch này không tương ứng với quan hệ $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$). ■

Quá trình làm mịn phân hoạch của Ví dụ 3.1 được minh họa thông qua cây quyết định như trong Hình 3.2.