ĐẠI HỌC HUẾ
TRƯỜNG ĐẠI HỌC KHOA HỌC

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
Thời gian: từ tháng 01/2013 đến tháng 12/2014

Chủ trì 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 quát
2. Tình hình nghiên cứu trong và ngoài nước
3. Mục tiêu đề tài
4. Nội dung nghiên cứu
   - Logic mô tả và cơ sở tri thức
   - Mô phỏng hai chiều trong logic mô tả
   - Học khái niệm trong logic mô tả sử dụng mô tả hai chiều
5. Kết quả nghiên cứu
6. Sản phẩm của đề tài

Giới thiệu tổng quát

- Học máy là một lĩnh vực của trí tuệ nhân tạo → nghiên cứu và xây dựng các kĩ thuật cho phép các hệ thống "*học*" tự động từ dữ liệu để giải quyết những vấn đề cụ thể.
- Học khái niệm trong logic mô tả tương tự như việc phân lớp nhị phân trong học máy truyền thống.
- Điểm khác là trong logic mô tả 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.

Giới thiệu tổng quát

Ba ngữ cảnh chính về học khái niệm trong logic mô tả:

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

E^+ chứa các mẫu dương và E^- chứa các mẫu âm của C.

(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 $a \in E^-$.

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

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

Tình hình nghiên cứu

Học khái niệm trong logic mô tả đã được nhiều nhà khoa học quan tâm nghiên cứu:

- Quinlan (1990): học các định nghĩa của mệnh đề Horn từ các dữ liệu được biểu diễn thông qua các quan hệ.
- Cohen và Hirsh (1994): lý thuyết về khả năng học PAC của logic mô tả CLASSIC và một logic con của nó là C-CLASSIC; đề xuất thuật toán học khái niệm LCSLearn dựa trên các "bao hàm chung nhỏ nhất".
- Lambrix và Larocchia (1998): thuật toán học khái niệm đơn giản dựa trên việc chuẩn hóa khái niệm và lựa chọn khái niệm thông qua các thể hiện của dạng chuẩn hóa.

Tình hình nghiên cứu

- Badea và Nienhuys-Cheng (Badea2000): toán tử làm mịn trên xuống (*downward refinement operators*) của logic mô tả $\mathcal{ALER}$ được thiết kế cho thuật toán học trên-xuống.
- Fanizzi và các cộng sự (2004, 2008): toán tử làm mịn trên xuống trong logic mô tả $\mathcal{ALN}$ và xây dựng hệ thống DL-FOIL; khai thác dữ liệu không gán nhãn như trong học bán giám sát.
- Iannone cùng cộng sự (2007): phương pháp bán tự động, xây dựng thuật toán suy luận trong logic mô tả $\mathcal{ALC}$; tìm và loại bỏ các phần của khái niệm có chứa lỗi phân loại.
- Lisi và Straccia (2013): học các tiên đề bao hàm tổng quát mờ từ các khẳng định chính xác $\rightarrow$ thuật toán FOIL-$\mathcal{DL}$.

Tình hình nghiên cứu

- Revoredo cùng các cộng sự (2010): học các khái niệm trong logic mô tả có xác suất từ dữ liệu quan hệ $\rightarrow$ thuật toán học trong $\textsc{cr}\mathcal{ALC}$ sử dụng các toán tử làm mịn + hàm tính điểm.
- Nguyen và Szałas (Nguyen2013): áp dụng mô phỏng hai chiều (*bisimulation*) trong logic mô tả để mô hình hóa tính không phân biệt được của các đối tượng:
  - tiên phong trong việc sử dụng mô phỏng hai chiều
  - xấp xỉ khái niệm với lý thuyết tập thô của Pawlak.
  - logic mô tả $\mathcal{ALC}$ với các tạo tử vai trò I (*vai trò nghịch đảo*), U (*vai trò phổ quát*) và các tạo tử khái niệm O (*định danh*), Q (*hạn chế số lượng có định tính*), Self (*tính phản xạ cục bộ của vai trò*).

Tình hình nghiên cứu

- Chúng tôi (2012) nghiên cứu học khái niệm trong logic mô tả:
  - tổng quát hóa và mở rộng phương pháp học khái niệm cho các hệ thống thông tin dựa trên logic mô tả sử dụng mô phỏng hai chiều.
  - xem các thuộc tính như là các yếu tố cơ bản của ngôn ngữ.
  - xem hệ thống thông tin trong logic mô tả là một diễn dịch hữu hạn trong logic đó.
- Công trình của Nguyen và chúng tôi đều giải quyết bài toán sử dụng ngữ cảnh (3) và cùng sử dụng mô phỏng hai chiều làm nền tảng cho quá trình tìm khái niệm kết quả.
- Mô phỏng hai chiều được van Benthem giới thiệu lần đầu dưới tên gọi *p-quan hệ* và *quan hệ zig-zag*.
- Nó được phát triển trong logic hình thái và trong các hệ thống chuyển trạng thái.

Mục tiêu của đề tài

- Xây dựng mô phỏng hai chiều cho một lớp lớn các logic mô tả, trong đó có $\mathcal{SHOIQ}$, $\mathcal{SROIQ}$ là những logic mô tả làm cơ sở cho OWL và OWL 2.
- Xây dựng phương pháp làm mịn phân hoạch miền của các diễn dịch trong logic mô tả dựa trên mô phỏng hai chiều.
- Phát triển các thuật toán học khái niệm dựa trên mô phỏng hai chiều cho các hệ thống thông tin dựa trên logic mô tả với ngữ cảnh (2).

Nội dung nghiên cứu

- Logic mô tả và cơ sở tri thức
- Mô phỏng hai chiều trong logic mô tả
- Học khái niệm trong logic mô tả sử dụng mô tả hai chiều

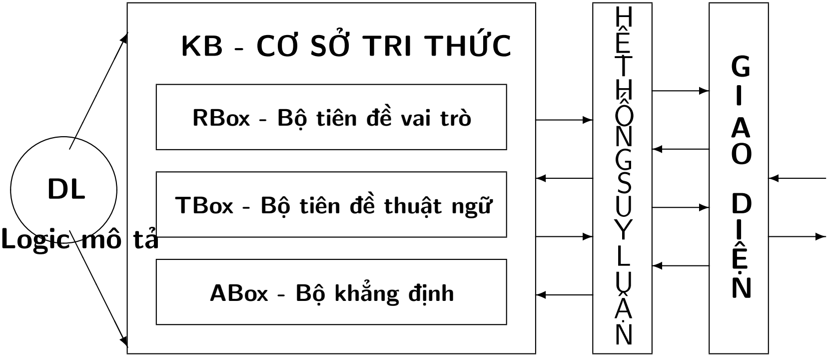


Figure: Kiến trúc của một hệ cơ sở tri thức trong logic mô tả

Bộ tiên đề vai trò - RBox

$hasParent \equiv hasChild^-$, $\quad hasChild \sqsubseteq hasDescendant$,

$\mathtt{Irr}(hasChild)$, $\quad \mathtt{Sym}(marriedTo)$.

$hasDescendant \circ hasDescendant \sqsubseteq hasDescendant$,

Bộ tiên đề thuật ngữ - TBox

$Human \equiv \top$, $\quad Parent \equiv Human \sqcap \exists hasChild.Human$,

$Male \equiv \neg Female$, $\quad Husband \equiv Male \sqcap \exists marriedTo.Female$,

$Male \sqcap Female \equiv \bot$, $\quad Husband \sqsubseteq \forall marriedTo.Female$.

Bộ khẳng định - ABox

$Human(\mathrm{LAN})$, $\quad Male(\mathrm{HUNG})$,

$Husband(\mathrm{HAI})$, $\quad hasChild(\mathrm{LAN}, \mathrm{HUNG})$,

$(\neg Female \sqcap Rich)(\mathrm{HUNG})$.

Definition (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 $\mathcal{ALC}$. ■

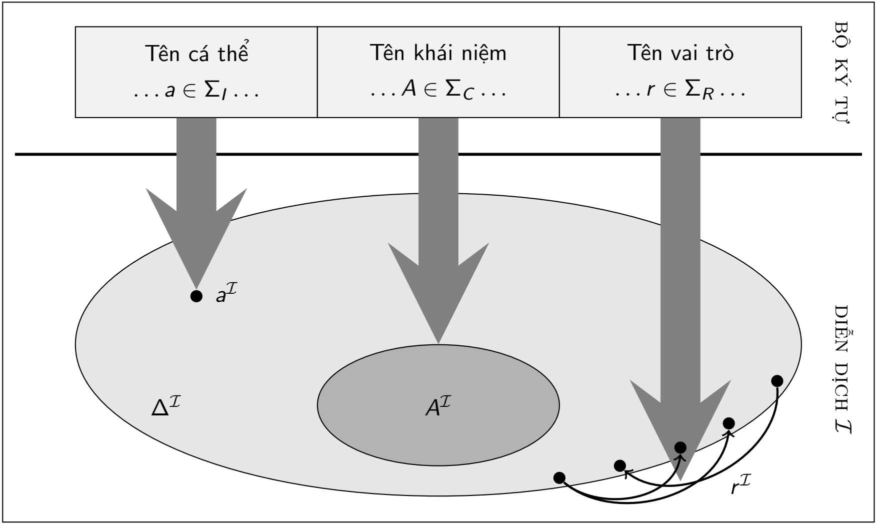
$$C, D \quad \longrightarrow \quad A \mid \top \mid \bot \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \exists r.C \mid \forall r.C$$

Definition (Ngữ nghĩa của $\mathcal{ALC}$)

Một *diễn dịch* trong logic $\mathcal{ALC}$ 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:

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

Logic mô tả $\mathcal{ALC}$



Definition (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 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 $A \in \Sigma_C$ 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
  - $R \circ S$, $R \sqcup S$, 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 $\mathcal{ALC}_{reg}$. ■

Cú pháp $\mathcal{ALC}_{reg}$ có thể mô tả một cách vắn tắt bằng các luật sau:

$$R, S \longrightarrow r \mid R \circ S \mid R \sqcup S \mid R^* \mid C?$$

$$C, D \longrightarrow A \mid \top \mid \bot \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \exists R.C \mid \forall R.C$$

- 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*.
- Đặt $\Sigma_A = \Sigma_{dA} \cup \Sigma_{nA}$.
- A là thuộc tính Bool nếu $dom(A) = \{\mathrm{true}, \mathrm{false}\}$.
- Các thuộc tính Bool như là các tên khái niệm.
- Gọi Σ_C là tập các tên khái niệm của Σ.

Ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ và ngữ nghĩa

Các *đặc trưng của logic mô tả* gồm:

- $\mathcal{I}$ *vai trò nghịch đảo,*
- $\mathcal{U}$ *vai trò phổ quát*
- $\mathcal{O}$ *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,*
- Self *tính phản xạ cục bộ của vai trò.*

Tập các đặc trưng của logic mô tả Φ là một tập chứa không hoặc một số các đặc trưng trên.

$\Phi = \{\mathcal{I}, \mathcal{O}, \mathcal{Q}\}$ để chỉ tập đặc trưng của logic mô tả gồm: nghịch đảo vai trò, định danh và hạn chế số lượng có định tính.

Ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ và ngữ nghĩa

- 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 thông qua các quy tắc đệ quy với các tính chất tương ứng với tập đặc trưng Φ.

Ngôn ngữ $\mathcal{L}_{\Sigma,\Phi}$ và ngữ nghĩa

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

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

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

$$(A = d)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid A^{\mathcal{I}}(x) = d\} \qquad (A \neq d)^{\mathcal{I}} = (\neg(A = d))^{\mathcal{I}}$$

$$(A < d)^{\mathcal{I}} = ((A \leq d) \sqcap (A \neq d))^{\mathcal{I}} \qquad (A > d)^{\mathcal{I}} = ((A \geq d) \sqcap (A \neq d))^{\mathcal{I}}$$

$$(\forall R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \forall y\, [R^{\mathcal{I}}(x,y) \Rightarrow C^{\mathcal{I}}(y)]\} \qquad (\exists r.\mathrm{Self})^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid r^{\mathcal{I}}(x,x)\}$$

$$(\exists R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y\, [R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)]\} \qquad (\exists \sigma.\{d\})^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \sigma^{\mathcal{I}}(x,d)\}$$

$$(\geq n\, R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \#\{y \mid R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)\} \geq n\} \qquad (\geq n\, R)^{\mathcal{I}} = (\geq n\, R.\top)^{\mathcal{I}}$$

$$(\leq n\, R.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \#\{y \mid R^{\mathcal{I}}(x,y) \wedge C^{\mathcal{I}}(y)\} \leq n\} \qquad (\leq n\, R)^{\mathcal{I}} = (\leq n\, R.\top)^{\mathcal{I}}$$

Cơ sở tri thức

Definition (Cơ sở tri thức)

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

Definition (Mô hình)

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 $\mathcal{A}$. ■

- Một cơ sở tri thức được gọi là *thỏa mãn* nếu nó 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 khái niệm C dựa trên cơ sở tri thức $\mathcal{KB}$ được ký hiệu là $\mathcal{KB} \not\models C(a)$.

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.
- Divroodi và Nguyen đã nghiên cứu mô phỏng hai chiều cho một số logic mô tả cụ thể.
- Nguyen và Szalas 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.
- 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:
  - các thuộc tính như là các phần tử cơ bản,
  - vai trò dữ liệu,
  - tính chất hàm,
  - hạn chế số lượng không định tính.

Mô phỏng hai chiều trong logic mô tả

Lemma

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* $\mathcal{I}'$. ∎

Lemma

Cho $\mathcal{I}$ và $\mathcal{I}'$ là các diễn dịch trong $\mathcal{L}_{\Sigma,\Phi}$ và 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 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:

$$Z(x,x') \Rightarrow [C^{\mathcal{I}}(x) \Leftrightarrow C^{\mathcal{I}'}(x')] \tag{1}$$

$$[Z(x,x') \wedge R^{\mathcal{I}}(x,y)] \Rightarrow \exists y' \in \Delta^{\mathcal{I}'} \mid [Z(y,y') \wedge R^{\mathcal{I}'}(x',y')] \tag{2}$$

$$[Z(x,x') \wedge R^{\mathcal{I}'}(x',y')] \Rightarrow \exists y \in \Delta^{\mathcal{I}} \mid [Z(y,y') \wedge R^{\mathcal{I}}(x,y)], \tag{3}$$

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

$$Z(x,x') \Rightarrow [R^{\mathcal{I}}(x,a^{\mathcal{I}}) \Leftrightarrow R^{\mathcal{I}'}(x',a^{\mathcal{I}'})]. \blacksquare \tag{4}$$

Tính bất biến đối với mô phỏng hai chiều

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 $\Sigma^\dagger \subseteq \Sigma$, $\Phi^\dagger \subseteq \Sigma$ và với mọi $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều Z giữa $\mathcal{I}$ và $\mathcal{I}'$.

Theorem

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. ■

Mô phỏng hai chiều - Tính bất biến

Các định lý về tính bất biến đối với mô phỏng hai chiều:

- Điều kiện để một ABox $\mathcal{A}$ bất biến đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều,
- Điều kiện để một TBox $\mathcal{T}$ bất biến đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều,
- Điều kiện để một cơ sở tri thức $\mathcal{KB} = \langle \mathcal{R}, \mathcal{T}, \mathcal{A} \rangle$ bất biến đối với $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-mô phỏng hai chiều.

Definition (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 $Z' \subseteq Z$. ■

- Ký hiệu $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều lớn nhất của $\mathcal{I}$ là $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$,
- 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'.

Tự mô phỏng hai chiều

- Ký hiệu $\mathcal{L}_{\Sigma^\dagger,\Phi^\dagger}$-tự mô phỏng hai chiều lớn nhất của $\mathcal{I}$ là $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$,
- 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'.

Theorem

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 nhau).* ■

Tự mô phỏng hai chiều

- Ta nói rằng tập Y *bị phân chia* bởi tập X nếu $Y \setminus X \neq \emptyset$ và $Y \cap X \neq \emptyset$. $\Leftrightarrow$ tập Y không bị phân chia bởi tập X nếu hoặc $Y \subseteq X$ hoặc $Y \cap X = \emptyset$.

- Một phân hoạch $\mathbb{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ *nhất quán* với tập X nếu với mọi $1 \leq i \leq n$, Y_i không bị phân chia bởi X.

Theorem

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 hệ thống thông tin 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 $X = C^{\mathcal{I}}$ 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 $C^{\mathcal{I}} = X$.* ∎

Vấn đề học khái niệm cho cơ sở tri thức

Bài toán học khái niệm

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:

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

E^+ chứa các mẫu dương và E^- chứa các mẫu âm của C.

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 $\mathcal{I}$.
- Bắt đầu từ phân hoạch $\{\Delta^{\mathcal{I}}\}$, làm mịn liên tục $\{\Delta^{\mathcal{I}}\} \rightarrow$ phân hoạch tương ứng với quan hệ $\sim_{\Sigma^{\dagger},\Phi^{\dagger},\mathcal{I}}$ với các kỹ thuật:
  - 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 $Y_1, Y_2, \ldots, Y_n$.
  - Ta gọi phân hoạch hiện thời là $\mathbb{Y} = \{Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}\} \subseteq \{Y_1, Y_2, \ldots, Y_n\}$.
  - 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.

Bộ chọn cơ bản

Definition (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 $1 \leq t \leq k$,
- $\ldots$
- $\geq l\, r^-.C_{i_t}$ và $\leq m\, r^-.C_{i_t}$, nếu $\{\mathcal{Q}, \mathcal{I}\} \subseteq \Phi^\dagger$, $r \in \Sigma^\dagger_{oR}$, $1 \leq t \leq k$, $0 < l \leq \#C_{i_t}$ và $0 \leq m < \#C_{i_t}$,
- $\exists r.\mathrm{Self}$, nếu $\mathrm{Self} \in \Phi^\dagger$ và $r \in \Sigma^\dagger_{oR}$. ■

Bộ chọn cơ bản

Theorem

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* $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$. ∎

Theo đó, các bộ chọn cơ bản là điều kiện đủ để khi làm mịn liên tục phân hoạch $\{\Delta^\mathcal{I}\}$ chúng ta được đạt phân hoạch tương ứng với quan hệ tương đương $\sim_{\Sigma^\dagger,\Phi^\dagger,\mathcal{I}}$.

Các bộ chọn khác

Trong thực tế, chúng ta có thể xem xét sử dụng các bộ chọn khác.

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

Figure: Các bộ chọn được sử dụng thêm trong thực tế

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 := Δ^I; 𝕐 := {Y_1}; C_1 := ⊤; 𝔻 := ∅;
Tạo và thêm các bộ chọn vào 𝔻 ;          /* theo Định nghĩa 3.1 và Hình 3.1 */
while (𝕐 không nhất quán với E) and (𝕐 có thể phân hoạch) do
    Chọn bộ chọn D_u trong 𝔻 và khối Y_{i_j} trong 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} ∩ D_u^I;        C_s := C_{i_j} ⊓ D_u;
    Y_t := Y_{i_j} ∩ (¬D_u)^I;     C_t := C_{i_j} ⊓ ¬D_u;
    𝕐 := 𝕐 ∪ {Y_s, Y_t} \ {Y_{i_j}};
    Tạo và thêm các bộ chọn mới vào 𝔻;  /* theo Định nghĩa 3.1 và Hình 3.1 */
if (𝕐 nhất quán với E) then
    return 𝕐;
else
    return failure;
```

Học khái niệm trong logic mô tả

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 ($\mathcal{KB} \not\models (\sqcap \mathbb{C} \sqsubseteq \neg C_{i_j})$) **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 ($\exists a \in E^- \setminus E_0^- : \mathcal{KB} \not\models (\sqcap \mathbb{C})(a)$) **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;

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

- Đối với vấn đề suy luận tự động, độ phức tạp của bài toán này là $\textsc{ExpTime}$-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 $\textsc{ExpTime}$-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).

Kết luận

- 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.
- Xây dựng mô phỏng hai chiều trên lớp các logic mở rộng đang nghiên cứu. Các định lý, bổ đề, hệ quả liên quan đến mô phỏng hai chiều và tính bất biến đối 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 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).

Sản phẩm

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

EM XIN CẢM ƠN SỰ THAM DỰ CỦA THẦY CÔ