

Các Phương pháp Lập luận và Tích hợp Ontology Mờ

Trương Hải Bằng

Trường Đại học Công nghệ Thông tin, Đại học quốc gia TP.HCM

bangth@uit.edu.vn

Tóm tắt. Ontology có vai trò quan trọng trong việc tổ chức và quản lý tri thức ở các lĩnh vực nghiên cứu và ứng dụng khác nhau. Tri thức Ontology đã thu hút sự quan tâm nghiên cứu của các nhà khoa học trên thế giới. Tuy nhiên, khái niệm ontology truyền thống còn thiếu khả năng biểu diễn các thông tin mờ trong lĩnh vực tri thức không chắc chắn. Mặt khác bài toán lập luận và tích hợp các ontology mờ hiện vẫn còn là vấn đề đang có nhiều trao đổi và nghiên cứu về cả lý thuyết lẫn ứng dụng. Bài báo đề xuất một mô hình về ontology mờ nhằm hướng đến giải quyết bài toán tích hợp ontology mờ dựa trên lý thuyết đồng thuận trong quá trình giải quyết mâu thuẫn giữa các ontology mờ và một số kết quả nghiên cứu gần đây.

1 Giới thiệu

Ontology đóng vai trò thiết yếu trong quá trình trao đổi tri thức giữa các hệ thống thông tin phân tán. Để thực hiện điều này một cách hiệu quả, các ontology phân tán cần phải được tích hợp lại. Tuy nhiên tích hợp ontology là quá trình rất phức tạp, vì tính chất đa dạng về cấu trúc của nó. Do vậy tích hợp ontology là bài toán được nhiều nhà nghiên cứu quan tâm theo các hướng tiếp cận khác nhau [13]. Với mục tiêu tích hợp ontology, định nghĩa sau đây được sử dụng. Ontology là một bộ tứ được xác định bởi các thành phần như sau [19]:

$O = (C, I, R, Z)$, trong đó

- C là tập hợp các khái niệm (concepts);
- I là tập hợp các thực thể (instances);
- R là tập hợp các quan hệ hai ngôi được định nghĩa trên C;
- Z là tập hợp các luật biểu diễn ràng buộc toàn vẹn hoặc các mối quan hệ giữa các thực thể và các khái niệm không thể hiện được bởi những quan hệ trong bộ R.

Ontology dựa trên logic mô tả truyền thống, không đủ khả năng để mô tả thông tin mờ, không thể đại diện đầy đủ và xử lý tri thức không chắc chắn, thông tin không chính xác trong các miền ứng dụng khác nhau. Năm 2006, Straccia [25] dựa vào nền tảng của logic mô tả và lý thuyết tập mờ của Zadeh [26] đã đưa ra logic mô tả mờ (Fuzzy Description Logic) nhằm phục vụ cho việc xử lý tri thức không chắc chắn trên Web ngữ nghĩa. Từ đó việc nghiên cứu và phát triển logic mô tả mờ như là một cơ sở cho việc biểu diễn tri thức và lập luận được đặt ra. Trong các nghiên cứu của mình, Calegari và Ciucci [23], [24] đã đề cập đến các phương pháp tích hợp logic mô tả mờ

vào ontology truyền thống để mở rộng ontology phù hợp hơn cho việc giải quyết các vấn đề suy luận không chắc chắn và các vấn đề cần thiết để xây dựng một ontology mờ cho Web ngữ nghĩa.

Sự khác biệt giữa ontology và ontology mờ là cho phép phân biệt ở các mức độ khác nhau về sự mô tả các khái niệm và quan hệ giữa chúng trong thế giới thực. Các nghiên cứu về ontology mờ có thể phân thành hai nhóm công trình. Nhóm thứ nhất bao gồm các phương pháp tiếp cận dựa trên logic, cụ thể là logic mô tả mờ [16], [20]. Nhóm thứ hai sử dụng tiếp cận không-logic [14]. Nhóm này đề xuất một mô hình ontology mờ trong đó một mô tả khái niệm được mờ hóa mức độ của giá trị các thuộc tính bằng cách sử dụng một hàm thành viên mờ [27]-[29]. Trong nhóm này Blanco [15] đề xuất một mô hình ontology linh hoạt cho phép lưu trữ thông tin mờ trong cơ sở dữ liệu. Đề xuất này cho phép người dùng quản lý thông tin không chính xác.

Trong phần tiếp theo, chúng tôi trình bày các nghiên cứu liên quan đến các phương pháp lập luận và tích hợp tri thức ontology mờ. Trong phần 3 trình bày các vấn đề cơ bản liên quan phương pháp lập luận và tích hợp ontology mờ. Phần 4 đề xuất phương pháp sử dụng lý thuyết đồng thuận để tích hợp ontology mờ và một số kết quả đã được nhóm tác giả công bố gần đây [31]-[34]. Cuối cùng là các nghiên cứu trong thời gian tiếp theo.

2 Các Công trình Liên quan

Các công trình mới nhất về tích hợp ontology từ năm 2009 đến nay đã được trình bày [1]-[12]. Các nghiên cứu tập trung đề cập vấn đề tích hợp ontology theo các phương pháp khác nhau về so khớp, ánh xạ và trộn ontology tùy thuộc vào ứng dụng cụ thể chủ yếu liên quan đến web ngữ nghĩa, truy vấn thông tin. Các phương pháp tiếp cận để tích hợp ontology được trình bày không thể hiện các tiêu chí một cách rõ ràng cho quá trình tích hợp, các thuật toán tích hợp chỉ liên quan đến các nghiên cứu và ứng dụng thực tế cho một vấn đề cụ thể.

Hiện nay có rất nhiều định nghĩa về tích hợp ontology trong đó định nghĩa được nhiều nghiên cứu tham khảo [19] được phát biểu như sau: Cho n ontologies O_1, \dots, O_n ta cần xác định một ontology O^* tốt nhất đại diện cho các ontology O_1, \dots, O_n . Để thực hiện được điều này tùy thuộc vào ứng dụng và mục tiêu nghiên cứu, cần tiến hành những kỹ thuật sau:

So khớp ontology: là tìm sự tương ứng của các khái niệm và quan hệ giữa các ontology. Quá trình này được sử dụng khi chúng ta cần giao tiếp trao đổi giữa hai hệ thống hoặc muốn sử dụng thông tin của hệ thống thứ hai này cho người dùng của hệ thống thứ nhất. Được sử dụng trong nhiều công việc khác như trộn ontology, trả lời truy vấn, dịch dữ liệu, hoặc duyệt web ngữ nghĩa [1], [3], [4], [8], [10].

Ánh xạ ontology: biểu diễn một sự tương ứng giữa các các khái niệm và quan hệ hai ontology là quá trình một chiều, chuyển các khái niệm và quan hệ từ một ontology này sang một ontology khác. Định nghĩa này phù hợp với định nghĩa ánh xạ trong toán học, là sự tương ứng của các khái niệm và quan hệ của ontology nguồn và ontology đích [6], [7], [9], [12].

Trộn ontology: là quá trình tạo ra một ontology mới từ các ontology khác. Được sử dụng khi cần tích hợp một số hệ thống để tạo ra một hệ thống mới [5], [11], [21], [22].

Xung đột ontology: trong quá trình tích hợp ontology thông thường cần phải giải quyết sự xung đột (không thống nhất) giữa các thực thể, khái niệm và quan hệ giữa các ontology [16], [17], [19], [21], [22], [30]-[34].

3 Các Khái niệm Liên quan

3.1 Định nghĩa Ontology Mờ

Cho (\mathbf{A}, \mathbf{V}) là một thể giới thực, trong đó \mathbf{A} là tập hữu hạn các thuộc tính, \mathbf{V} miền giá trị của \mathbf{A} , $V^c = \bigcup_{a \in \mathbf{A}} V_a$, V_a là miền giá trị của thuộc tính a . Ontology mờ được định nghĩa như sau [30], [32], [34]: *Fuzzy ontology* = $(\mathbf{C}, \mathbf{R}, \mathbf{Z})$, trong đó

- \mathbf{C} là tập hữu hạn các khái niệm. Một khái niệm của ontology mờ được định nghĩa là một bộ tứ: (c, A^c, V^c, f^c) , với c là tên duy nhất của khái niệm, $A^c \subseteq \mathbf{A}$ là tập các thuộc tính mô tả khái niệm, $V^c \subseteq \mathbf{V}$ là miền giá trị của thuộc tính: $V^c = \bigcup_{a \in \mathbf{A}} V_a$ và f^c là hàm thành viên mờ: $f^c: A^c \rightarrow [0,1]$ biểu diễn mức thuộc tính mô tả khái niệm c . Bộ (A^c, V^c, f^c) được gọi là cấu trúc mờ của c .
- \mathbf{R} là tập các quan hệ mờ giữa các khái niệm, $\mathbf{R} = \{R_1, R_2, \dots, R_m\}$, $R_i \subseteq \mathbf{C} \times \mathbf{C} \times (0,1]$, $i = 1, 2, \dots, m$. Một quan hệ là một tập bao gồm một cặp khái niệm và giá trị mờ biểu diễn mức độ quan hệ giữa chúng. Mỗi quan hệ R_i giữa hai khái niệm trong ontology chỉ được biểu diễn bằng một giá trị mờ duy nhất, nghĩa là nếu $(c, c', v) \in R_i$ và $(c, c', v') \in R_i$ thì $v = v'$.
- \mathbf{Z} là tập hợp các tiên đề, có thể được hiểu là ràng buộc toàn vẹn hoặc mối quan hệ giữa các khái niệm và là tập hợp các hạn chế hay điều kiện (cần và đủ) để xác định các khái niệm trong \mathbf{C} .

3.2 Các Phương pháp Tích hợp Ontology

Trong quá trình thực hiện so khớp, ánh xạ, trộn và xung đột ontology cần phải giải quyết hai vấn đề chính sau:

- Các ontology đã cho đồng dạng (Similarity) với nhau đến mức độ nào?

- Những mâu thuẫn thường xuất hiện giữa các ontology như thế nào?

Các phương pháp xác định mức độ đồng dạng giữa các ontology

Các phương pháp xác định mức độ đồng dạng giữa các ontology bao gồm các kỹ thuật cơ bản như sau:

- Kỹ thuật dựa trên chuỗi;
- Kỹ thuật dựa trên ngôn ngữ;
- Kỹ thuật dựa trên cấu trúc;
- Kỹ thuật dựa trên ngữ nghĩa.

Một cách tiếp cận đơn giản để liên kết hai ontology là mỗi khái niệm thuộc ontology này phải so khớp với tất cả các khái niệm ở ontology khác nhằm đưa ra quyết định có hay không một cặp khái niệm đồng dạng với nhau. Thuật toán này có độ phức tạp $O(n^2)$, trong đó n là số lượng các khái niệm trong mỗi ontology. Để khắc phục hạn chế này trong phép toán so khớp ontology mà chúng tôi đề xuất thuật toán dựa trên khái niệm những phần chung tiềm năng viết tắt là PCP (potentially common parts) có độ phức tạp tính toán $O(n \log(n))$ [34].

Một số khái niệm:

Khoảng cách PCP. Giả sử $P_1 = \{c_1, c_2, \dots, c_n\}$ và $P'_2 = \{c'_1, c'_2, \dots, c'_n\}$ là tập các khái niệm của các PCP thuộc hai ontologies tương ứng O_1 và O_2 . Khoảng cách nhỏ nhất cho một cặp khái niệm tương ứng theo tiếp cận PCP sẽ được định nghĩa như sau:

$$D_0(P_1, P_2) = \sum_{c_i \in P_1} \sum_{c_j \in K} |CI(c_i) - CI(c_j)| \quad (1)$$

Với $K = \{c_j \in P_2 | d(c_i, c_j) \leq \text{threshold}, c_i \in P_1\}$, trong đó:

$$\text{Độ quan trọng của khái niệm} : CI(c) = AI(c) + RI(c) + \text{bias} \quad (2)$$

$$\text{Độ quan trọng của thuộc tính} : AI(c) = \sum_{a \in A^c} f_c(a) \quad (3)$$

$$\text{Độ quan trọng của quan hệ} : RI(c) = \sum_{R \in R^c} f_c(R) * CI(c') \quad (4)$$

Phương pháp so khớp:

Nếu $D(P_1, P_2)$ là nhỏ nhất, suy ra cặp khái niệm (c_i, c_j) có khả năng tương đương, cần kiểm tra sự tương đương cho các cặp khái niệm (c_i, c_j) trong P_1 và P_2 .

Để tìm ra khái niệm có khả năng tương đương, chúng ta sử dụng công thức đệ quy sau:

$$D_{k+1}(P_1, P_2) = D_k(P_1, P_2) + E(P_1, P_2) \quad (5)$$

trong đó

- $D_{k+1} - D_0 > \text{threshold}$ (vượt một ngưỡng cho trước)
- $E(P_1, P_2)$ là trọng số tăng bởi sự so khớp lại P_1 và P_2 .

Quy tắc tính sự tương đương theo tiếp cận PCP:

Quy tắc 1: Tính sự tương đương của cặp khái niệm có cả 2 đường được đánh dấu trong bảng. Nếu được so khớp, loại khỏi bảng, tính lại D_0 và cập nhật D_{k+1}

Quy tắc 2: Nếu chỉ một trong 2 đường được đánh dấu, chúng ta tính độ tương đương cho cặp khái niệm theo khoảng cách nhỏ nhất. Nếu được so khớp loại khỏi dòng hoặc cột tương ứng, cập nhật lại bảng, tính lại D_0, D_{k+1} .

Quy tắc 3: Nếu $D_{k+1} - D_0 > \text{Threshold}$ hoặc tất cả các cặp khái niệm được so khớp.

Ý tưởng thuật toán:

Bắt đầu từ một cặp khái niệm được so khớp duyệt các khái niệm theo mối quan hệ hoặc các trong kiến trúc ontology theo một mức cụ thể để xác định được phần chung tiềm năng ban đầu. Quá trình này lan truyền đến cặp khái niệm lân cận và kết quả là một tập hợp của sự sắp xếp giữa các khái niệm trong tập hợp kết nối ngữ nghĩa của hai đồ thị gọi là phân đoạn. Để so sánh sự giống nhau giữa các cặp khái niệm của các

ontology láng giềng quá trình này lặp đi lặp lại cho đến khi thuật toán thỏa mãn "hoặc là tất cả các khái niệm thu thập được liên kết tìm thấy, hoặc không có cặp liên kết mới".

Thuật toán PCP:

input: Hai ontology O_1 và O_2 , hai Medoids, và một mức DepthLevel

output: Tập các khái niệm so khớp $C^* = \bigcup_{i=1}^n C_i$, $C_i \in O_1$ hoặc O_2 ;

For mỗi cặp $(C_i^1, C_j^2) \in \text{Medoids}$ hoặc C^* do

Nếu (C_i^1, C_j^2) là medoids hoặc $\text{Similarity}(C_i^1, C_j^2) \geq \text{Threshold} * \text{then}$

RealMatch = (C_i^1, C_j^2) , $i=1..n$; $j=1..m$

CommonPart 1 = $\bigcup c_k^1$; CommonPart 2 = $\bigcup c_h^2$;

Tính CI CommonPart 1 và CommonPart 2 theo công thức (2,3,4);

For với mỗi $C_k^1 \in \text{CommonPart 1}$, $C_h^2 \in \text{CommonPart 2}$ do

Table \leftarrow Distance (C_k^1, C_h^2) ;

End for

Tính D_0 ;

$D_k = D_0$;

while thỏa quy tắc(3): $D_k - D_0 \leq \text{Threshold}$ do

Medoids \leftarrow Xác định cặp khái niệm so khớp theo quy tắc (1) và (2)

Loại bỏ các khái niệm tương đương ra khỏi Medoids từ C^* ;

End while

}

End

End for

Return (PossibleMatch)

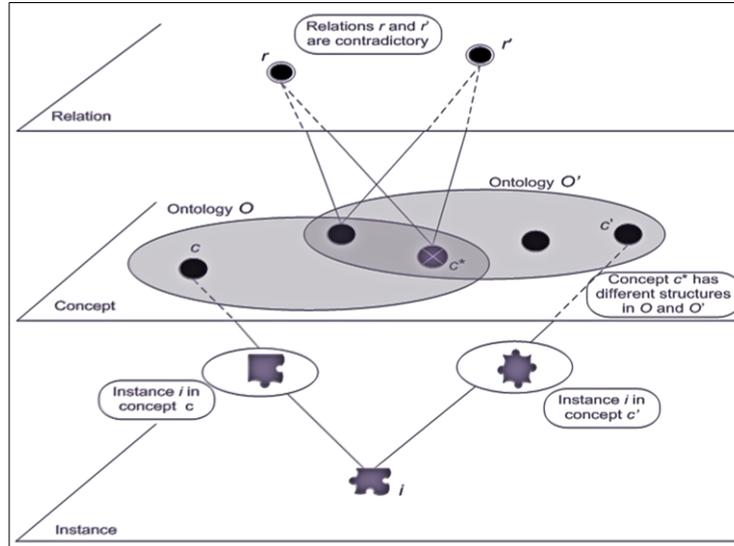
Các phương pháp giải quyết mâu thuẫn giữa các ontology

Tích hợp ontology là một trường hợp đặc biệt của quá trình tích hợp tri thức, quá trình tích hợp cần thỏa mãn các tiêu chí sau:

- *Bảo toàn thông tin:* Dữ liệu của Ontology nguồn phải ở trong kết quả tích hợp.
- *Giải quyết mâu thuẫn:* Tất cả các xung đột xuất hiện trong các yếu tố thành phần sau khi được tích hợp phải được giải quyết.
- *Bảo toàn cấu trúc:* Các kiểu cấu trúc của kết quả tích hợp phải giống cấu trúc của các yếu tố thành phần ban đầu.

Có ba mức mâu thuẫn ontology:

- *Mâu thuẫn ở mức thực thể:* cùng thực thể của các ontology khác nhau có mô tả mâu thuẫn với nhau.
- *Mâu thuẫn ở mức khái niệm:* cùng khái niệm giống nhau nhưng có cấu trúc khác nhau trong các ontology khác nhau.
- *Mâu thuẫn ở mức quan hệ:* có mâu thuẫn về quan hệ giữa hai khái niệm trong các ontology khác nhau.



Hình 1. Các mức mâu thuẫn ontology [19]

3.3 Lý thuyết Đồng thuận

Giới thiệu

Phương pháp đồng thuận rất hữu ích trong việc giải quyết các xung đột hay mâu thuẫn tri thức trong thế giới thực đặc biệt liên quan đến vấn đề tích hợp tri thức ontology. Phương pháp đồng thuận là một cách tiếp cận hiệu quả trong quá trình giải quyết xung đột tri thức nói chung và ontology nói riêng.

Xét bài toán đồng thuận như sau: Cho một tập hợp của các đối tượng trong vũ trụ, cần phải xác định một đối tượng tốt nhất đại diện cho các đối tượng này. Mục tiêu của vấn đề giải quyết xung đột là xác định được lời giải thích hợp cho bài toán đặt ra. Có hai trường hợp xảy ra trong quá trình này:

- i. Các giải pháp thích hợp độc lập với ý kiến của các người tham gia xung đột. Ví dụ cho trường hợp này là việc tính GDP hàng năm của một quốc gia được tạo ra bởi các chuyên gia tài chính khác nhau. Vấn đề xác định GDP được chính xác khi kết thúc năm, giá trị này là độc lập của dự báo đã cho. Trường hợp này được gọi là xung đột độc lập.
- ii. Giải pháp phụ thuộc vào ý kiến của những người tham gia cuộc xung đột, gọi là xung đột phụ thuộc. Ví dụ cho trường hợp này là quá trình bỏ phiếu trong các cuộc bầu cử.

Các khái niệm cơ bản

Cho U là một tập hợp hữu hạn của các đối tượng đại diện cho ý kiến tiềm năng cho các đối tượng xung đột

Định nghĩa 1. (Hàm khoảng cách)

Cho tập U chúng ta định nghĩa hàm khoảng cách d như sau: $\delta: Y \times Y \rightarrow [0, 1]$, thỏa các điều kiện:

Không âm: $\forall \xi, \psi \in Y: \delta(\xi, \psi) \geq 0$,

Phân xạ: $\forall \xi, \psi \in Y: \delta(\xi, \psi) = 0$ nếu $\xi = \psi$,

Đối xứng: $\forall \xi, \psi \in Y: \delta(\xi, \psi) = \delta(\psi, \xi)$

Định nghĩa 2. (Profile) Profile $X = \{r_i \in \text{TUPLE}(T_i): T_i \subseteq A \text{ for } i = 1, 2, \dots, n\}$, A là tập các thuộc tính, $\text{TUPLE}(T)$ là tập hợp các bộ kiểu T_i .

Định nghĩa 3. (xung đột Profile)

U : là tập hữu hạn biểu diễn các đối tượng cho sự xung đột.

$\Pi_k(U)$: là tất cả các tập con k -phần tử của U for $k \in \mathbb{N}$ (\mathbb{N} là số tự nhiên), Ký hiệu: $\Pi(U) = \bigcup_{k \in \mathbb{N}} \Pi_k(U)$, $\Pi(U)$ là tập hợp của tất cả các tập con khác rỗng của U , khi đó mỗi phần tử của $\Pi(U)$ được gọi là một xung đột Profile.

Ví dụ: Trong các profile sau, mức độ xung đột (conflict) tăng dần:

$X = \{\text{Yes}, \text{No}, \text{Neutral}\}$, $X' = \{\text{Yes}, \text{Yes}, \text{No}\}$, $X'' = \{\text{Yes}, \text{Yes}, \text{Neutral}\}$.

Định nghĩa 4. (Sự đồng thuận)

Cho U , d được cho trong định nghĩa 1

Cho Ánh xạ $C: \Pi(U) \rightarrow 2^U$. Với mỗi *conflict profile* $X \in \Pi(U)$, tập $C(X)$ được gọi là sự đồng thuận của X , và một phần tử của $C(X)$ được gọi là một đồng thuận của profile X .

Bài toán tích hợp tri thức phát biểu như sau: cho một tập các profile

$X = \{r_i \in \text{TUPLE}(T_i): r_i \subseteq A \text{ for } i = 1, 2, \dots, n\}$, cần xác định một bộ r^* là đại diện tốt nhất cho các bộ r_1, \dots, r_n . Bộ r^* được gọi là tích hợp của các profile.

Các tiêu chí để lựa chọn phương pháp đồng thuận

Có nhiều tiêu chuẩn liên quan đến bài toán xác định sự đồng thuận [19], hai tiêu chuẩn quan trọng thường được sử dụng trong bài toán tích hợp tri thức là:

- Tính tối ưu O_1 : [Kemeny, 1959]:

Nếu và chỉ nếu $(x \in C(X)) \Rightarrow (d(x, X) = \min_{y \in U} d(y, X)), \forall X \in \Pi(U)$.

(tổng khoảng cách từ một đồng thuận đến các phần tử của profile $P(U)$ là nhỏ nhất)

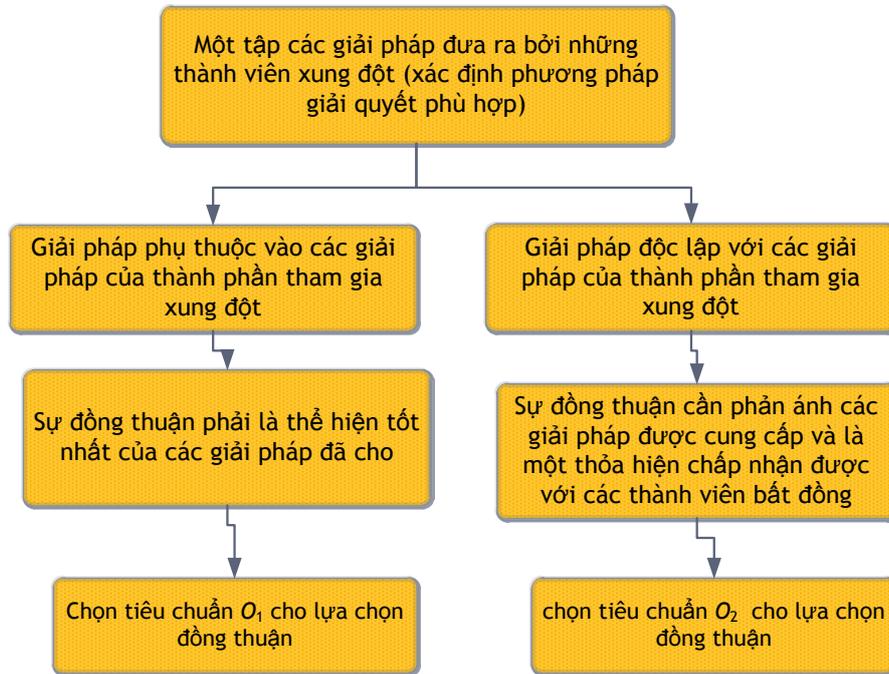
- Tính tối ưu O_2 :

Nếu và chỉ nếu $(x \in C(X)) \Rightarrow (d^2(x, X) = \min_{y \in U} d^2(y, X)), \forall X \in \Pi(U)$.

(tổng bình phương khoảng cách từ một đồng thuận đến các phần tử của profile $P(U)$ là nhỏ nhất)

Tóm lại, lược đồ của phương pháp đồng thuận bao gồm các bước như sau:

- Xác định tập các phiên bản tiềm năng của dữ liệu
- Định nghĩa hàm đo khoảng cách giữa các phiên bản
- Chọn tiêu chí lựa chọn phương pháp đồng thuận
- Thực hiện thuật toán để lựa chọn phương pháp đồng thuận



Hình 2. Sơ đồ sử dụng của các chức năng đồng thuận [19]

4 Một số Kết quả Nghiên cứu: Tích hợp Ontology Mờ Dựa trên Lý thuyết Đồng thuận

4.1 Tích hợp Ontology Mờ Mức Khái niệm

Cho hai ontology mờ O_1 và O_2 , khái niệm c thuộc O_1 là concepts $(c, A^{c1}, V^{c1}, f1)$ và thuộc O_2 là $(c, A^{c2}, V^{c2}, f2)$. Ta nói rằng sự mâu thuẫn khái niệm xảy ra nếu $A^{c1} \neq A^{c2}$ hoặc $V^{c1} \neq V^{c2}$ hoặc $f1 \neq f2$

Bài toán 1:

Cho tập cấu trúc mờ của các khái niệm $X = \{(A^i, V^i, f^i) | (A^i, V^i, f^i)$ cấu trúc mờ của khái niệm c trong ontology $O_i, i=1, \dots, n\}$, Cần xác định bộ ba: $c^* = (A^*, V^*, f^*)$ là biểu diễn tốt nhất thỏa tiêu chuẩn O_2 của lý thuyết đồng thuận cho các cấu trúc trên.

Thuật toán 1. Tích hợp mức khái niệm

Input: cho cấu trúc mờ của khái niệm c trong n ontology

$X = \{(A^i, V^i, f^i) : (A^i, V^i, f^i)$ là cấu trúc mờ của khái niệm c trong n ontology $O_i, i=1, \dots, n\}$

Output: Triple $c^* = (A^*, V^*, f^*)$ biểu diễn tốt nhất từ X thỏa tiêu chuẩn đồng thuận O_2

Procedure:

BEGIN

Set $A^* = \bigcup_{i=1}^n A_i;$

Set $V^* = \cup_{i=1}^n V_i$;
 For each $a \in A^*$ do
 Begin
 xác định $X_a = \{f^i(a) : \text{if } f^i(a) \text{ tồn tại, } i = 1, \dots, n\}$;
 tính toán $f^*(a) = \frac{1}{\text{card}(X_a)} \sum_{v \in X_a} v$;
 End.

END.

Chứng minh: Theo Định lý 8.1. [19] như sau:

Cho profile $X = \{x^{(i)} = x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)} : i = 1, 2, \dots, m\}$ vector $x = (x_1, x_2, \dots, x_m)$ thỏa tiêu chuẩn đồng thuận O_2 khi và chỉ khi: $x_j = \frac{1}{n} \sum_{i=1}^n x_j^{(i)}$, $j = 1, 2, \dots, m$

4.2 Tích hợp Ontology Mờ Mức Quan hệ

Cho hai ontology mờ O_1 và O_2 . $c \in O_1$, $c' \in O_2$. chúng ta nói rằng sự mâu thuẫn mức quan hệ xảy ra nếu $R_{i1}(c, c') \neq R_{i2}(c, c')$, $i \in \{1, \dots, m\}$.

Xét trường hợp sau: trong bảng 1, cho hai ontology O_1 và O_2 có các quan hệ R_1, R_2, R_3 và các khái niệm a, b, c , các mâu thuẫn mức quan hệ xảy ra trong ví dụ này là: $R_{11} \neq R_{12}$, $R_{21} \neq R_{22}$, $R_{31} \neq R_{32}$

Bảng 1. Mâu thuẫn mức quan hệ

	R_1	R_2	R_3
O_1	$\langle a, b, 0.5 \rangle$ $\langle a, c, 0.3 \rangle$	$\langle a, c, 0.3 \rangle$ $\langle b, c, 0.8 \rangle$	$\langle c, a, 0.7 \rangle$ $\langle b, c, 0.7 \rangle$
O_2	$\langle a, c, 0.3 \rangle$ $\langle a, b, 0.7 \rangle$	$\langle a, c, 0.4 \rangle$ $\langle b, c, 0.8 \rangle$	$\langle c, a, 0.7 \rangle$

Bài toán 2:

Cho $i \in \{1, \dots, m\}$ và tập các quan hệ $X = \{R_{ij}(c, c') : i=1, \dots, m; j = 1, \dots, n\}$ giữa 2 khái niệm c và c' trong n ontologies, cần xác định $R_i(c, c')$ - quan hệ tốt nhất giữa c và c' trong tập các quan hệ đã cho thỏa tiêu chuẩn đồng thuận O_1 .

Thuật toán 2: Tích hợp mức quan hệ

Input: Cho tập các quan hệ giữa 2 khái niệm c và c' trong n ontologies $X = \{R_{ij}(c, c') : j = 1, \dots, n\}$

Output: Quan hệ $R_i(c, c') = (c, c', v)$ tốt nhất trong tập các quan hệ X thỏa O_1 - consensus.

Procedure:

BEGIN

Thiết lập thứ tự $X = \{x_1, x_2, \dots, x_n\}$;

Thiết lập khoảng $\frac{X_{n+1}}{2}, \frac{X_{n+2}}{2}$

Xác định giá trị v trong khoảng $\frac{X_{n+1}}{2}, \frac{X_{n+2}}{2}$

END.

Chứng minh: Theo Định lý 8.2. [19]:

Cho profile $X = \{x^{(i)} = x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)} : i = 1, 2, \dots, m\}$

vector $x = (x_1, x_2, \dots, x_m)$ thỏa tiêu chuẩn đồng thuận O_1 nếu không tồn tại bất kỳ vector y nào sao cho: $d(y, x) > d(y, x^{(i)})$, $\forall i = 1, 2, \dots, m$

Bài toán 3:

Cho $i \in \{1, \dots, m\}$ và tập các quan hệ giữa 2 khái niệm c và c' trong n ontology: $X = \{R_{ij} \subseteq C \times C \times (0, 1] : j = 1, \dots, n\}$. Cần xác định $R_i(c, c')$ quan hệ $R_{ij} \subseteq C \times C \times (0, 1]$ tốt nhất thỏa tiêu chuẩn O_1 của lý thuyết đồng thuận trong tập các quan hệ X đã cho.

Loại mâu thuẫn quan hệ thứ 2 ở mức độ phức tạp hơn, giải quyết cho trường hợp quan hệ có tính bắc cầu. (c, c', v_1) của R_i , (c', c'', v_2) của R_i , trong trường hợp này trọng số v_1 của quan hệ $\langle c, c'' \rangle$ trong sẽ là bao nhiêu.

Thuật toán 3:

Input: - Tập quan hệ cùng loại giữa các khái niệm trong n ontologies $X = \{R_{ij} \subseteq C \times C \times (0, 1] : j = 1, \dots, n\}$

- Quan hệ có tính bắc cầu.

Output: Quan hệ $R_i \subseteq C \times C \times (0, 1]$ tốt nhất của X thỏa tiêu chuẩn đồng thuận O_1

Procedure:

BEGIN

Set $R_i = \emptyset$;

For each pair $(c, c') \in C \times C$ do

Begin

Xác định tập $X_{(c,c')} = \{v : \langle c, c', v \rangle \in R_{ij} \text{ for } j = 1, \dots, n\}$;

Thiết lập thứ tự $X_{(c,c')}$ tăng dần $X = \{x_1, x_2, \dots, x_k\}$;

thiết lập khoảng $(\frac{x_{k+1}}{2}, \frac{x_{k+2}}{2})$

Lấy một giá trị v trong khoảng trên ;

đặt $R_i := R_i \cup \{\langle c, c', v \rangle\}$

End;

For each $(c, c', c'') \in C \times C \times C$ do

Begin

If $\langle c, c', v_1 \rangle \in R_i$, $\langle c, c', v_2 \rangle \in R_i$ and $\langle c, c', v_3 \rangle \in R_i$ then

$v_3 = \min\{v_1, v_2\}$;

If only $\langle c, c', v_1 \rangle \in R_i$ and $\langle c, c', v_2 \rangle \in R_i$ then set $R_i := R_i \setminus \langle c, c', v_3 \rangle$

where $v_3 = \min\{v_1, v_2\}$;

End

END.

Chứng minh: Tương tự thuật toán 2

4.3 Mô hình Đa Thuộc tính và Đa Giá trị cho Bài toán Tích hợp Ontology Mở ở Mức Thực thể

Cho A là tập các thuộc tính của một ontology. Mỗi thuộc tính $a \in A$ có miền giá V_a , V_a được gọi là tập các giá trị cơ bản (elementary values). Một giá của thuộc tính a là

một tập con của V_a . Tập 2^{V_a} được gọi là siêu miền (super domain) của thuộc tính a . giả sử có tập B là tập con của A : $B \subseteq A$, lấy $V_b = \bigcup_{b \in B} V_b$ và $2^{-B} = \bigcup_{b \in B} 2^{V_b}$

Định nghĩa: Một thực thể mờ của khái niệm c được mô tả bởi các thuộc tính của tập A^c có các giá trị thuộc tập 2^{V_x} ($X = A^c$) là một cặp (i, v) , với i là định danh của thực thể, v là giá trị của thực thể là một bộ có kiểu A^c được biểu diễn là một hàm như sau $v: A \rightarrow 2^{-A^c}$, $v(a) \in 2^{V_a}, \forall a \in A^c$.

Cho hai ontology O_1 và O_2 , khái niệm c thuộc O_1 là khái niệm (c, A^{c1}, V^{c1}, f^1) và c' thuộc O_2 là $(c', A^{c2}, V^{c2}, f^2)$. Cho $(i, v) \in (O_1, c)$ và $(i, v') \in (O_2, c')$. Ta nói rằng sự mâu thuẫn thực thể xảy ra nếu $v(a) \neq v'(a), a \in A^c \cap A^{c'}$.

Bài toán 4:

Cho tập hợp các thực thể $X = \{(i, v_1), \dots, (i, v_n)\}$, với $v_i \in A_i \subseteq A$, $v_i: A_i \rightarrow V_i$ $i = 1, \dots, n$ và $V_i = \bigcup_{i \in A_i} V_a$, cần xác định (i, v) tốt nhất trong tập X thỏa tiêu chuẩn O_1 của lý thuyết đồng thuận.

Ví dụ về mâu thuẫn giữa các thực thể của các ontology về thời tiết

Bảng 2. Thông tin thời tiết được thu thập từ sáu trạm khí tượng (tương ứng với sáu ontology)

Thực thể \ Thuộc tính	I ₁	I ₂	I ₃	I ₄	I ₅	I ₆
Áp suất	990–995		990–997	992–999	993–997	992–998
Hướng gió	{W, W–N}	{E, E–N}	{S, W–S}	{S, W–S}	{W–N}	
Vận tốc gió	10–12	5–10	20–30	40–50	0–10	
Nhiệt độ	15–25	20–24	12–21	12–21	22–24	17–20
Mưa	không	có	không	không	có	có
Nắng	không	không	không	không	có	không
Tuyết	có	không				có

Trong bảng 1 ta thấy các thực thể có thể có các giá trị thuộc tính khác nhau, xảy ra mâu thuẫn.

Thuật toán 4:

Input: tập các mô tả của các thực thể: $X = \{r_i \in \text{TUPLE}(T_i): T_i \subseteq A, i = 1, 2, \dots, n\}$

và hàm khoảng cách d_a cho các thuộc tính $a \in A$,

d_a là hàm được xác định như sau $d_a: 2^{V_a} \times 2^{V_a} \rightarrow [0,1]$

Output: bộ $t^* \in T^* \subseteq A$ là đại diện tốt nhất được tích hợp từ các bộ của X .

Procedure:

BEGIN

1. $A = \bigcup_{i=1}^n T_i$;

2. For each $a \in A$ xác định một tập $X_a = \{t_{ia}: t_i \in X \text{ for } i = 1, 2, \dots, n\}$;

3. For each $a \in A$ sử dụng hàm khoảng cách d_a xác định một giá trị $v_a \subseteq V_a$ sao cho

$$\sum_{r_{ia} \in X_a} d_a(v_a, r_{ia}) = \min_{v'_a \in V_a} \sum_{r_{ia} \in X_a} d_a(v'_a, r_{ia})$$

4. khởi tạo bộ t^* bao gồm các giá trị $v_a, a \in A$;
END.

5 Kết luận và Hướng Phát triển

Bài báo trình bày một số khái niệm cơ bản của tích hợp ontology, đề xuất một mô hình ontology mờ cho bài toán tích hợp ontology mờ. Một đóng góp quan trọng của nghiên cứu là sử dụng cách tiếp cận dựa trên lý thuyết đồng thuận cho các phương pháp lập luận & tích hợp trên Cơ sở tri thức không chắc chắn Ontology Mờ.

Các công việc trong tương lai liên quan đến thử nghiệm các thuật toán tích hợp và thực hiện thử nghiệm trên dữ liệu ontology mờ thời tiết được cài đặt dựa trên các ontology thời tiết đã có.

Tài liệu tham khảo

1. Md. Hanif Seddiqui, Masaki Aono., An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size, *Web Semantics: Science, Services and Agents on the World Wide Web*, (Elsevier 2009) 344–3567
2. Deryle Lonsdale a, David W. Embley a, Yihong Ding a, Li Xub, Martin Hepp, 2010], Reusing ontologies and language components for ontology generation, *Data & Knowledge Engineering Journal* (2010) 318–330
3. Watson Wei Khong Chua, Jung-jae Kim, BOAT: Automatic alignment of biomedical ontologies using term informativeness and candidate selection, *Journal of Biomedical Informatics* 45 (2012) 337–349
4. Jürgen Bock a, Jan Hettenhausen, Discrete particle swarm optimisation for ontology alignment, *Information Sciences* (2012) 152–173
5. Asad Masood Khattak, Zeeshan Pervez, Khalid Latif, Sungyoung, Knowledge accumulation through automatic merging of ontologies, *Knowledge-Based Systems* (2012)
6. Adolfo Guzmán-Arenas, Alma-Delia Cuevas. Time efficient reconciliation of mappings in dynamic web ontologies, *Expert Systems with Applications* (2010) 1991–2005
7. Ming Mao, Yefei Peng, Michael Spring, An adaptive ontology mapping approach with neural network based constraint satisfaction, *Web Semantics: Science, Services and Agents on the World Wide Web* (2010) 14–25
8. Hacene Belhadef, A new bidirectional method for ontologies matching, *Procedia Engineering* (2011) 558 – 564
9. Mohamed Bakillah, A Fuzzy Logic Semantic Mapping Approach for Fuzzy Geospatial Ontologies, *The Fifth International Conference on Advances in Semantic Processing*, 2011
10. Hacene Belhadef, A new bidirectional method for ontologies matching, *Procedia Engineering* 23 (2011) 558 – 564
11. Rung-Ching Chen, Cho-Tscan Bau, Chun-Ju Yeh, Merging domain ontologies based on the WordNet system and Fuzzy Formal Concept Analysis techniques, *Applied Soft*

Computing 11 (2011) 1908–1923

12. Rujuan Wang, Lei Wang and Lei Liu, Gang Chen and Qiushuang Wang, Combination of the Improved Method for Ontology Mapping, *Physics Procedia* 25 (2012) 2167 – 2172
13. Chitra Ramesh and Aghila Gnanasekaran, Methodology based survey on ontology management, *International Journal of Computer Science & Engineering Survey (IJCSES)*, 2010)
14. Abulaish, M., Dey, A.: A Fuzzy Ontology Generation Framework for Handling Uncertainties and Non-uniformity in Domain Knowledge Description. In: *Proceedings of the International Conference on Computing: Theory and Applications*, pp. 287–293. IEEE (2007)
15. Blanco, I.J., Vila, M.A., Martinez-Cruz, C.: The Use of Ontologies for Representing Database Schemas of Fuzzy Information. *International Journal of Intelligent Systems* 23(4), 419–445 (2008)
16. Duong, T.H., Nguyen, N.T., Jo, G.S.: A Method for Integrating Multiple Ontologies. *Cybernetics and Systems* 40(2), 123–145 (2009)
17. Fernandez-Breis, J.T., Martinez-Bejar, R.: A Cooperative Framework for Integrating Ontologies. *Int. J. Human-Computer Studies* 56, 665–720 (2002)
18. Lu, J., Li, Y., Zhou, B., Kang, D., Zhang, Y.: *Distributed Reasoning with Fuzzy Description*
19. Nguyen, N.T.: *Advanced methods for inconsistent knowledge management*. Springer, London (2008)
20. Nguyen, N.T.: Inconsistency of Knowledge and Collective Intelligence. *Cybernetics and Systems* 39(6), 542–562 (2008)
21. Noy, N.F., Musen, M.A.: SMART: Automated Support for Ontology Merging and Alignment. In: *Proc. of the 12th Workshop on Knowledge Acquisition, Modelling and Management (KAW 1999)*, Banff, Canada, pp. 1–20 (1999)
22. Pinto, H.S., Martins, J.P.: A Methodology for Ontology Integration. In: *Proceedings of the First International Conference on Knowledge Capture*, pp. 131–138. ACM Press (2001)
23. Calegari, S., Ciucci, D.: Fuzzy Ontology, Fuzzy Description Logics and Fuzzy-OWL. In: *Proceedings of WILF 2007*. Volume 4578 of LNCS. (2007) In printing.
24. Calegari, S., Ciucci, D.: Fuzzy Ontology and Fuzzy-OWL in the KAON Project. In: *FUZZIEEE 2007. IEEE International Conference on Fuzzy Systems* (2007).
25. Straccia U., A Fuzzy Description Logic for the semantic Web, *Proc. in Capturing Intelligence: Fuzzy logic and the semantic Web*, Elie Sanchez ed., Elsevier, 2006.
26. Zadeh, L. A. Fuzzy sets. *Information and control*, Elsevier, 1965, pp. 338-358.
27. Thanh Tho Quan, Siu Cheung Hui and Tru Hoang Cao.: Automatic fuzzy ontology generation for semantic Web, *Knowledge and Data Engineering Journals IEEE*, Volume: 18, Issue: 6, pp: 842 - 856 , Product Type: Journals & Magazines, 2006.
28. Silvia, Calegari and Elie Sanchez.: A Fuzzy Ontology-Approach to improve Semantic Information Retrieval. *Proceedings of the Third ISWCWorkshop on Uncertainty Reasoning for the Semantic Web - URSW'07*, 2007.
29. Hua Mao Gu.: Educational Game for Middle & Primary School Students- EGMPSS. *Proceedings of the 18th Conference of International Maritime Lecturers' Association*, 2008.
30. Hai Bang Truong, Ngoc Thanh Nguyen: A Multi-attribute and Multi-valued Model for Fuzzy Ontology Integration Instance Level. *ACIIDS The 4nd Asian Conference on Intelligent Information and Database Systems*, Springer Verlag in a volume LNAI, 2012: 187-197
31. Trong Hai Duong, Hai Bang Truong, Ngoc Thanh Nguyen: Local Neighbor Enrichment for Ontology Integration. *ACIIDS The 4nd Asian Conference on Intelligent Information and Database Systems*, Springer Verlag in a volume LNAI, 2012: 156-166

32. Hai Bang Truong, Ngoc Thanh Nguyen, Phi Khu Nguyen: Fuzzy Ontology Building and Integration for Fuzzy Inference Systems in Weather Forecast Domain. ACIIDS-2011 The 3rd Asian Conference on Intelligent Information and Database Systems, Springer Verlag in a volume LNAI, in Daegu city, Korea: 517-527
33. Hai Bang Truong, Ngoc Thanh Nguyen: A framework of an effective fuzzy ontology alignment technique. International Conference on Systems, Man and Cybernetics, Anchorage, Alaska, USA, IEEE 2011, ISBN 978-1-4577-0652-3: 931-935
34. Ngoc Thanh Nguyen, Hai Bang Truong: A Consensus-Based Method for Fuzzy Ontology Integration. The Second International Conference on Computational Collective Intelligenc, Springer Verlag in a volume LNAI, ICCCI10, Kaohsiung city, Taiwan: 480-489.