

## MỘT PHƯƠNG PHÁP CẤU TRÚC MẠNG NGỮ NGHĨA

LÊ TỰ THÀNH

Viện Điện tử và Tin học Hà Nội

### 1. ĐẶT VẤN ĐỀ

Mạng ngữ nghĩa là một kiểu ngôn ngữ quan hệ dùng để biểu diễn mối liên hệ giữa một khái niệm với các khái niệm liên quan. Mạng ngữ nghĩa lần đầu được Ross Quillian trình bày năm 1966 cho một đề án xây dựng hệ hiểu ngôn ngữ tiếng Anh. Từ đó tới nay, mạng ngữ nghĩa trở nên một ngôn ngữ biểu diễn tri thức với nhiều mô hình khác nhau, như cách tiếp cận của Woods, Anderson, Cohen, McCathy ...

Trong một mạng gồm có các nốt và các quan hệ. Nói một cách tổng quát, nốt dùng để biểu diễn các đối tượng của tư duy, ví dụ như một từ trong ngôn ngữ, một bộ phận trong cấu trúc, ... Những đối tượng như vậy có thể là một đối tượng tồn tại khách quan, nhưng có những trường hợp chúng chỉ là khái niệm do tư duy đặt ra (xem [4]). Bên cạnh các nốt còn có các quan hệ cho biết mối liên quan giữa hai đối tượng được biểu diễn bởi các nốt. Trong nhiều trường hợp, những quan hệ được dùng để định nghĩa một khái niệm qua các khái niệm khác. Một tổ hợp các nốt và các quan hệ như vậy tạo nên một mạng. Để biểu diễn tri thức, rõ ràng phải tồn tại quan hệ giữa các nốt với nhau thông qua các quan hệ và các nốt trung gian. Đó là những mạng liên thông. Dưới đây, khi nói đến mạng tức là nói mạng liên thông; khi có những bộ phận tách rời nhau thì được xem đó là những mạng khác nhau.

Một cách hình thức, nếu gọi tập các nốt là  $X = \{Z_1, Z_2, \dots, Z_n\}$  và tập các quan hệ là  $Q = \{R_1, R_2, \dots, R_m\}$  thì ta có định nghĩa (xem [2]):

**Định nghĩa 1.** Một mạng ngữ nghĩa  $L$  gồm tập các nốt  $N$  và tập các quan hệ  $FN$  được ký hiệu là  $L = (N, FN)$ , với

- a)  $N = \{Z_i | Z_i \in X, i = 1, \dots, n\}$ ;
- b)  $FN = \{FZ_1, FZ_2, \dots, FZ_n\}$  với  $FZ_i = \{R_k Z_j | \exists Z_i R_k Z_j : Z_i, Z_j \in N, R_k \in Q, j \neq i, i = 1, \dots, n\}$ .

### 2. CẤU TRÚC CỦA MẠNG

Những mạng được định nghĩa như trên đủ thuận tiện cho việc biểu diễn tri thức. Tuy vậy chúng khá cồng kềnh khi lưu trữ. Do đó ta đưa ra định nghĩa mạng cơ sở:

**Định nghĩa 2.** Một mạng ngữ nghĩa  $L_i = (N, FN)$  được gọi là mạng cơ sở nếu

- a)  $N = Z_i \cup \{Z_j | \exists R_k Z_j \in FN\}$ ;
- b)  $FN = FZ_i \neq \emptyset (FZ_j = \emptyset \forall j \neq i)$ .

Từ định nghĩa này dễ dàng thấy rằng, mạng cơ sở chỉ bao gồm các nốt có quan hệ trực tiếp với nốt  $Z$  và bản thân nốt đó. Đó là những tri thức cơ sở để tư duy, còn trong mạng, đó là những thành phần cơ sở xây dựng nên mạng. Những thành phần cơ sở đó chỉ có thể tạo nên một mạng nếu chúng có liên hệ với nhau.

**Định nghĩa 3.** Một tập các mạng cơ sở được gọi là tập liên hệ và được ký hiệu là  $SL = \{L_i = (N_i, FN_i)\}$  nếu  $\forall L_i \in SL, \exists L_j \in SL$  với  $i \neq j$  sao cho :  $\exists Z_k$  mà  $Z_k \in L_i \wedge Z_k \in L_j$ .

**Định lý 1.** Mọi mạng ngữ nghĩa đều phân rã được thành một tập liên hệ.

*Chứng minh:* Trước hết cần xây dựng một thuật toán  $DEC(L)$  để phân rã một mạng ngữ nghĩa  $L$  bất kỳ. Thuật toán nhận dữ liệu vào là mạng  $L = (N, FN)$  và phải đưa ra tập liên hệ  $SL$ . Các bước như sau:

- a - Lấy lần lượt các tập  $FZ_i \in L$ ; nếu hết thì đưa ra  $SL$ , kết thúc thuật toán.
- b - Tạo một mạng cơ sở  $L_i$  từ tập  $FZ_i$  với
 
$$N_i = Z_i \cup \{Z_j \mid \exists R_k Z_j \in FZ_i\},$$

$$FN_i = FZ_i.$$
- c - Bổ sung  $L_i$  vào  $SL$ . Quay lại bước a.

Do qui ước liên thông, dễ dàng thấy rằng  $SL$  là một tập liên hệ.

**Hệ quả 1.** Nếu  $SL = \{L_i = (N_i, FN_i)\}$  là tập liên hệ nhận được từ phân rã của mạng  $L = (N, FN)$  theo thuật toán  $DEC(L)$  thì

- a)  $N = \bigcup_{i=1}^n N_i,$
- b)  $FN = \bigcup_{i=1}^n FN_i.$

*Chứng minh:* Do cách tạo các mạng cơ sở ở bước a và b, ta có:

$$FN_i = FZ_i \text{ với } \forall i \Rightarrow FN = \{FZ_1, FZ_2, \dots, FZ_n\} = \bigcup_{i=1}^n FN_i.$$

Để chứng minh a), ta giả sử có  $Z_k \in N$ , nhưng  $Z_k \notin N_j$  với  $\forall j$ . Theo bước b ta suy ra  $Z_k$  không có quan hệ với  $\forall Z_m \in L_j$  với  $\forall j$ . Điều này làm cho mạng không liên thông. Vậy  $N = \bigcup_{i=1}^n N_i.$

Từ những mạng cơ sở, cần phải tạo nên được các mạng ngữ nghĩa. Ta đưa ra hai phương thức phát triển từ một mạng ban đầu  $L$  với các mạng cơ sở như sau:

- Nối tiến :  $L_0 = L_0 \cup L_i$  khi  $\exists Z_k : Z_k \in L_0 \wedge Z_k \in L_i, k \neq i$ ;
- Nối lùi :  $L_0 = L_0 \cup L_i$  khi  $Z_i \in L_0$ .

**Định lý 2.** Phương thức nối tiến và nối lùi là đủ để phát triển một mạng ngữ nghĩa từ tập liên hệ nhận được từ phân rã của chính nó.

*Chứng minh:* Thuật toán phát triển mạng  $CONS(SL)$ , với dữ liệu vào là tập liên hệ  $SL = \{L_i = (N_i, FN_i)\}$  gồm các bước sau:

- a - Lấy  $L_0 = L_i$  bất kỳ.
- b - Phát triển mạng bằng phương thức nối tiến và nối lùi.
- c - Kiểm tra  $N_0 = \bigcup_{i=1}^n N_i$ ? Nếu chưa, quay lại bước b. Nếu bằng, đưa ra mạng  $L_0$ , kết

thức thuật toán.

Trước hết phải chứng minh có thể đạt được điều kiện ra của thuật toán. Giả sử sau  $k$  lần lặp,  $N_0^{(k)} = N_0^{(k-1)}$  (với  $N_0^{(k)}$  là tập nốt của  $L_0$  sau lần lặp thứ  $k$ ) và  $N \setminus N_0^{(k)} \neq \emptyset$ . Lấy  $Z_m \in N \setminus L_0^{(k)}$ . Giả sử  $Z_m \in L_j$ , khi đó  $L_j \notin L_0$ , vì nếu  $L_j \in L_0$  sau lần lặp  $i < k$  thì theo hai phương thức nối  $Z_m \in N_j \subset N_0^{(k)}$ . Nhưng  $SL$  là tập liên hệ, dễ dàng suy ra  $\forall L_i \in SL$  đều không thuộc vào  $L_0$ , điều này mâu thuẫn với bước a.

Nếu mạng ban đầu là  $L = (N, FN)$ , thì ta đã có  $N_0 = \bigcup_{i=1}^n N_i$ , nên theo hệ quả 1 của phần a),  $N_0 = N$ . Theo thuật toán  $CONS(SL)$ , tất cả các mạng cơ sở đều đã được dùng để phát triển mạng  $L_0$ , nên  $FN_0 = \bigcup_{i=1}^n FN_i$ . Theo hệ quả 1, phần b), ta có  $FN_0 = FN$ .

**Hệ quả 2.** Mạng được phát triển bằng thuật toán  $CONS(SL)$  từ một tập liên hệ  $SL$  cho trước là duy nhất.

Ta dễ dàng chứng minh hệ quả này.

**Hệ quả 3.** Các phép hợp, giao và lấy phần bù giữa các mạng có thể được thực hiện bởi các thuật toán  $DEC(L)$  và  $CONS(SL)$  như sau:

- Hợp  $L = \bigcup_{i=1}^m L_i$  với  $L_i = (N_i, FN_i) : L = CONS(\bigcup_{i=1}^m DEC(L_i))$ .
- Giao  $L = \bigcap_{i=1}^m L_i$  với  $L_i = (N_i, FN_i) : L = CONS(\bigcap_{i=1}^m DEC(L_i))$ .
- Phần bù  $L = L_1 \setminus L_2$  với  $L_1 = (N_1, FN_1)$  và  $L_2 = (N_2, FN_2) : L = CONNS(DEC(L_1) \setminus DEC(L_2))$ .

Ta dễ dàng chứng minh hệ quả này.

### 3. TẬP BIÊN

Khi biểu diễn mạng bằng đồ thị, với quan hệ là các cung có hướng xuất phát từ một nốt tới một nốt liên hệ khác, thì mạng ngữ nghĩa trở thành một đồ thị có hướng. Trên đồ thị đó, có những nốt chỉ có các cung đi và có những nốt chỉ có các cung đến, chúng ta gọi chúng tương ứng là các nốt gốc và các nốt biên. Chúng biểu diễn những khái niệm cuối cùng hoặc đầu tiên để suy luận. Ngoài những nốt biên, những nốt còn lại kể cả những nốt gốc được gọi là những nốt trong.

Trên đồ thị biểu diễn một mạng ngữ nghĩa, mọi nốt trong đều có liên hệ với các nốt biên thông qua các nhánh đồ thị đi khỏi nốt trong đó. Ta đưa ra định nghĩa sau.

**Định nghĩa 4.** Tập tất cả các nốt biên có quan hệ với một nốt trong  $Z_i$  qua các nhánh đồ thị xuất phát từ nốt đó làm thành tập biên  $B(Z_i)$  của nốt trong  $Z_i$ .

**Thuật toán BOUN(Z):** Xác định tập biên của nốt trong  $Z$ .

Vào : Tập liên hệ  $SL$  nhận được từ phân rã mạng  $L$ .

Ra : Tập biên  $B(Z)$ .

Phương pháp:

a) Đưa  $Z$  vào tập nốt quan hệ  $R(Z)$  và tập nốt đang xét  $C(Z)$ .

b)  $\forall Z_i \in C(Z)$  tìm  $L_i = (N_i, FN_i) \in SL$ .

- Nếu không tìm thấy, bổ sung  $Z_i$  vào tập biên  $B(Z)$ ;  
 Nếu tìm thấy, bổ sung tập  $\{Z_j \mid \exists R_k Z_j \in FN_i\}$  vào tập  $C'(Z)$ .
- c) Kiểm tra tập  $C'(Z)$   
 Nếu  $C'(Z) = \emptyset$  thì đưa ra  $B(Z)$ , kết thúc thuật toán;  
 Nếu  $Z_j \in C'(Z)$  mà  $Z_j \in R(Z)$  thì loại bỏ  $Z_j$  khỏi  $C'(Z)$ .
- d) Bổ sung tập  $C'(Z)$  vào  $R(Z)$ , thay  $C(Z)$  bằng  $C'(Z)$ , xóa  $C'(Z)$ . Quay lại bước b).

**Bổ đề 1.** Thuật toán BOUN( $Z$ ) xác định đúng tập biên của nốt trong  $Z$ .

*Chứng minh:* Trước hết cần chứng tỏ là thuật toán này dừng. Do bước kiểm tra c),  $C'(Z)$  chỉ chứa các nốt mới chưa được xét qua bước b); và do tập  $SL$  hữu hạn nên BOUN( $Z$ ) dừng.

Theo bước b), mọi nốt được đưa vào  $B(Z)$  đều là nốt biên và có quan hệ với  $Z$ . Ta còn phải chứng tỏ  $B(Z)$  chứa mọi nốt biên liên quan với  $Z$ . Giả sử có một nốt biên  $Z_{i,m}$ . Trên đồ thị sẽ có một cung đi từ  $Z$  qua  $Z_{i_1}, Z_{i_2}, \dots, Z_{i,m}$ . Theo bước b), Nếu  $Z_{i,m}$  không được đưa vào  $B(Z)$  thì  $Z_{i,m-1}$  phải không thuộc  $C'(Z)$ . Cuối cùng suy ra  $Z$  không thuộc vào  $C'(Z)$ . Điều đó mâu thuẫn với bước a). Vậy  $B(Z)$  xác định bởi thuật toán BOUN( $Z$ ) chính là tập biên của  $Z$ .

#### 4. CÁC THỂ HIỆN CỦA MỘT MẠNG

Một khái niệm, ngoài cách định nghĩa thông qua các khái niệm khác còn có thể được định nghĩa thông qua danh sách các thể hiện. Ví dụ như khái niệm *xe*, ngoài định nghĩa qua các khái niệm *bánh xe*, *tay lái*,... còn có thể liệt kê các thể hiện như *xe đạp*, *xe máy*,... Định nghĩa kiểu liệt kê như vậy có dạng

$$Z = \{Z_1, Z_2, \dots, Z_n\} \text{ và}$$

$$R = \{R_1, R_2, \dots, R_m\}.$$

Trong mạng ngữ nghĩa, các nốt thể hiện bao giờ cũng được kế thừa các tính chất của nốt biểu diễn khái niệm tổng quát hơn. Từ đó, nếu các nốt khái niệm lập nên một mạng cơ sở thì các nốt thể hiện chúng cũng sẽ lập nên mạng thể hiện có cấu trúc tương đương.

**Định nghĩa 5.** Mạng  $L' = (N', FN')$  là mạng thể hiện có cấu trúc tương đương của mạng  $L = (N, FN)$  với các định nghĩa liệt kê:

$$Z_i = \{z_{i_1}, z_{i_2}, \dots, z_{i_n}\} \quad i = 1, 2, \dots, n \text{ và}$$

$$R_k = \{r_{k_1}, r_{k_2}, \dots, r_{k_m}\} \quad k = 1, 2, \dots, m \text{ nếu}$$

$$a) N' = \{z_i \mid \exists Z_i \in N : z_i \in Z_i \quad i = 1, 2, \dots, n\};$$

$$b) FN' = \{r_k z_j \mid \exists R_k Z_j \in FN : r_k \in R_k \quad k = 1, 2, \dots, m\}.$$

Mạng mức trên và mạng thể hiện tương ứng được Cluicop gọi là mạng macro và micro. Cách gọi như vậy là tương đối so với nhau. Tuy nhiên, ở đây ta chỉ định nghĩa cho mạng cơ sở. Để tạo mạng thể hiện bất kỳ, có thể áp dụng thuật toán CONS. Ta không định nghĩa cho mạng bất kỳ, bởi vì định nghĩa như vậy chỉ đúng khi mọi nốt đều có định nghĩa liệt kê.

#### 5. KẾT LUẬN

Chúng ta đã xây dựng một phương thức để khảo sát cấu trúc mạng ngữ nghĩa. Nền tảng của phương thức này là mạng cơ sở, được lắp ghép hoặc phân rã nhờ các thuật toán CONS và

DEC. Những phép hợp, giao và lấy phần bù giữa các mạng đều có thể thực hiện qua chúng.

Việc khảo sát như vậy nhằm tìm ra một cách biểu diễn và lưu trữ tri thức bằng mạng ngữ nghĩa. Dĩ nhiên việc này còn đòi hỏi những công việc nghiên cứu tiếp tục, mà chưa được trình bày ở đây.

Trong những phần cuối, chúng ta đã đưa ra khái niệm tập biên với điều kiện biểu diễn mạng trên đồ thị và khái niệm mạng thể hiện cho mạng cơ sở. Chúng sẽ là những phương tiện để thực hiện suy luận trên các mạng biểu diễn tri thức cụ thể.

Thực chất những kết quả đã nêu là những nghiên cứu về cấu trúc hình thức của mạng nhằm tạo một công cụ khi biểu diễn tri thức trên mạng; do vậy đã không đi sâu được vào những vấn đề về ngữ nghĩa hiện đang được tranh luận. Để đạt được một số kết quả về cấu trúc, ta đã hạn chế phần nào tính đa dạng của ngữ nghĩa. Điều đó thật đáng tiếc nhưng cũng khó tránh khỏi.

*Nhận ngày 20-7-1990*

### TÀI LIỆU THAM KHẢO

1. Bạch Hưng Khang, Hoàng Kiếm - Trí tuệ nhân tạo các chuyên ngành và ứng dụng. NXB Khoa học và Kỹ thuật, Hà Nội, 1989, 216 trang.
2. Клыкoв, Ю. И., Горков, К. Л. - Банки данных для принятия решений. Москва. Советское Радио. 1980, 208 стр.
3. Hayes Roth F., Waterman D. A., Lenat D. B. (Eds.) - Building Expert Systems, London, Addison-Wesley, 1983, 444 pp.
4. Maida A. S., Shapiro S. C. - Intensional Concept in Propositional Semantic Networks In: Readings in Knowledge Representation, California, Morgan Kaufman Publishers, 1985, pp. 169-189.
5. Ringland G. A., Duce D. A. - Approaches to Knowledge Representation - An Introduction. Research Studies Press LTD, Lechtworth, Hertfordshire, England and Jonh Wiley & Sons, New-York, 1988, 260 pp.
6. Swamy M. N. S., Thulasiraman K. - Graph Networks and Algorithms. New-York, Jonh Wiley & Sons, 1981, 304 pp.
7. Woods W. A. - What's in a Link: Foundations for Semantic Networks In: Readings in knowledge Representation, California, Morgan Kaufman Publishers, 1985, pp. 217 - 241.

### ABSTRACT

This paper deals with the construction of Semantic Networks, one of the knowledge Representation Languages. The definitions of the Basic Networks, of the Related Set of Basic Networks are introduced to decompose and construct any Semantic Network. There are the algorithms DEC and CONS to do these works, and some operations on Networks may be expressed with them. The Boundary Set of an internal node is introduced, which shows the relationship between nodes on a graph representing Semantic Network. In last chapter, the concept of installed Basic Networks is considered based on the list definitions.