

Bài 2: Biểu diễn tri thức và lập luận

Các Hệ cơ sở tri thức

Trường Đại học Công nghệ Thông tin

Source: Giáo trình các hệ Cơ sở tri thức

Hoàng Kiếm, Đỗ Phúc, Đỗ Văn Nhon

Tri thức, biểu diễn tri thức.

Các loại tri thức: được chia thành 5 loại

- 1. Tri thức thủ tục: mô tả cách thức giải quyết một vấn đề. Loại tri thức này đưa ra giải pháp để thực hiện một công việc nào đó. Các dạng tri thức thủ tục tiêu biểu thường là các luật, chiến lược, lịch trình và thủ tục.
- 2. Tri thức khai báo: cho biết một vấn đề được thấy như thế nào. Loại tri thức này bao gồm các phát biểu đơn giản, dưới dạng các khẳng định logic đúng hoặc sai. Tri thức khai báo cũng có thể là một danh sách các khẳng định nhằm mô tả đầy đủ hơn về đối tượng hay một khái niệm nào đó.

Các loại tri thức

- 3. Siêu tri thức (meta-knowledge): mô tả *tri thức về tri thức*. Loại tri thức này giúp lựa chọn tri thức thích hợp nhất trong số các tri thức khi giải quyết một vấn đề. Các chuyên gia sử dụng tri thức này để điều chỉnh hiệu quả giải quyết vấn đề bằng cách hướng các lập luận về miền tri thức có khả năng hơn ca.
- 4. Tri thức heuristic: mô tả các *"kinh nghiệm"* để dẫn dắt tiến trình lập luận. Tri thức heuristic là *tri thức* không đảm bảo hoàn toàn 100% chính xác về kết quả giải quyết vấn đề. Các chuyên gia thường dùng các tri thức khoa học như sự kiện, luật, ... sau đó chuyển chúng thành các tri thức heuristic để thuận tiện hơn trong việc giải quyết một số bài toán.
- 5. Tri thức có cấu trúc: mô tả tri thức theo cấu trúc. Loại tri thức này mô tả mô hình tổng quan hệ thống theo quan điểm của chuyên gia, bao gồm khái niệm, khái niệm con, và các đối tượng; diễn tả chức năng và mối liên hệ giữa các tri thức dựa theo cấu trúc xác định.

Phân loại tri thức

- 1. Hà nội là thủ đô của Việt Nam.
- 2. Bạn Lan thích đá banh.
- 3. Modus Ponens.
- 4. Modus Tollens.
- 5. Thuật toán tìm kiếm BFS
- 6. Thuật toán tìm kiếm DFS
- 7. Thuật toán AKT
- 8. Thuật giải Greedy

Phân loại tri thức

- 9. Một số cách chiếu tướng trong khi chơi cờ tướng.
- 10. Hệ thống các khái niệm trong hình học.
- 11. Cách tập viết chữ đẹp.
- 12. Tóm tắt quyển sách về Khai phá dữ liệu.
- 13. Chọn loại cổ phiếu để mua bán cổ phiếu.
- 14. Chọn các thông tin về các loại cổ phiếu trên thị trường.

Các kỹ thuật biểu diễn tri thức

- Bộ ba Đối tượng-Thuộc tính-Giá trị
- Các luật dẫn
- Mạng ngữ nghĩa
- Frames
- Logic

Bộ ba Đối tượng-Thuộc tính-Giá trị

- Một sự kiện có thể được dùng để xác nhận giá trị của một thuộc tính xác định của một vài đối tượng.
- Ví dụ, mệnh đề "quả bóng màu đỏ" xác nhận "đỏ" là giá trị thuộc tính "màu" của đối tượng "quả bóng".
- Kiểu sự kiện này được gọi là bộ ba Đối tượng-Thuộc tính-Giá trị (O-A-V – Object-Attribute-Value).

Bộ ba Đối tượng-Thuộc tính-Giá trị

- Trong các sự kiện O-A-V, một đối tượng có thể có nhiều thuộc tính với các kiểu giá trị khác nhau. Hơn nữa một thuộc tính cũng có thể có một hay nhiều giá trị. Chúng được gọi là các sự kiện *đơn trị* (single-valued) hoặc *đa trị* (multi-valued). Điều này cho phép các hệ tri thức linh động trong việc biểu diễn các tri thức cần thiết.
- Các sự kiện không phải lúc nào cũng bảo đảm là đúng hay sai với độ chắc chắn hoàn toàn. Vì thế, khi xem xét các sự kiện, ta còn dùng khái niệm là *độ tin cậy*.
- Phương pháp truyền thống để quản lý thông tin không chắc chắn là sử dụng nhân tố chắc chắn CF (certainly factor).
- Khái niệm này bắt đầu từ hệ thống MYCIN (khoảng năm 1975) dùng trong y khoa, để diễn đạt các thông tin suy luận không chắc chắn.
- Khi đó, trong sự kiện O-A-V sẽ có thêm một giá trị xác định độ tin cậy của nó là CF.

Các luật dẫn

- Luật là cấu trúc tri thức dùng để liên kết thông tin đã biết với các thông tin khác giúp đưa ra các suy luận, kết luận từ những thông tin đã biết.
- Trong hệ thống dựa trên luật, ta thu thập các tri thức lĩnh vực trong một tập và lưu chúng trong cơ sở tri thức của hệ thống.
- Việc xử lý các luật trong hệ thống dựa trên các luật được quản lý bằng một module gọi là *bộ suy diễn*.

Các luật dẫn(tiếp)

- **Các dạng luật cơ bản: 7 dạng**
- **1. Quan hệ:**
 - IF Bình điện hóng THEN Xe sẽ không khởi động được
- **2. Lời khuyên:**
 - IF Xe không khởi động được THEN Đi bộ
- **3. Hướng dẫn**
 - IF Xe không khởi động được AND Hệ thống nhiên liệu tốt THEN Kiểm tra hệ thống điện

Các luật dẫn(tiếp)

- **4. Chiến lược**
 - IF Xe không khởi động được THEN Đầu tiên hãy kiểm tra hệ thống nhiên liệu, sau đó kiểm tra hệ thống điện
- **5. Diễn giải**
 - IF Xe nổ AND tiếng giòn THEN Động cơ hoạt động bình thường
- **6. Chẩn đoán**
 - IF Sốt cao AND hay ho AND Họng đỏ THEN Viêm họng
- **7. Thiết kế**
 - IF Là nữ AND Da sáng THEN Nên chọn Xe Honda Lead AND Chọn màu sáng

Các luật dẫn(tiếp)

- Mở rộng cho các luật Trong một số áp dụng cần thực hiện cùng một phép toán trên một tập hay các đối tượng giống nhau. Lúc đó cần các *luật có biến*.
- Ví dụ: IF X là nhân viên AND Tuổi của X > 65 THEN X có thể nghỉ hưu
- Khi mệnh đề phát biểu về sự kiện, hay bản thân sự kiện có thể không chắc chắn, người ta dùng hệ số chắc chắn CF. Luật thiết lập quan hệ không chính xác giữa các sự kiện giả thiết và kết luận được gọi là *luật không chắc chắn*.
- Ví dụ: IF Lạm phát CAO THEN Hầu như chắc chắn lãi suất sẽ CAO
- Luật này được viết lại với giá trị CF có thể như sau:
- IF Lạm phát cao THEN Lãi suất cao, CF = 0.8
- Dạng luật tiếp theo là *siêu luật (meta-rule)*:
- Một luật với mô tả cách thức dùng các luật khác. Siêu luật sẽ đưa ra chiến lược sử dụng các luật theo lĩnh vực chuyên dụng, thay vì đưa ra thông tin mới.
- Ví dụ: IF Xe không khởi động AND Hệ thống điện làm việc bình thường THEN Có thể sử dụng các luật liên quan đến hệ thống điện

Mạng ngữ nghĩa

Mạng ngữ nghĩa là một phương pháp biểu diễn tri thức dùng đồ thị trong đó nút biểu diễn đối tượng và cung biểu diễn quan hệ giữa các đối tượng.

Logic mệnh đề

1. Logic mệnh đề

IF Xe không khởi động được (A)

AND Khoảng cách từ nhà đến chỗ làm là xa (B)

THEN Sẽ trễ giờ làm (C)

Luật trên có thể biểu diễn lại như sau: $A \wedge B \Rightarrow C$

2. Logic vị từ

Logic vị từ, cũng giống như logic mệnh đề, dùng các ký hiệu để thể hiện tri thức. Những ký hiệu này gồm **hằng số**, **vị từ**, **biến** và **hàm**.

SUY DIỄN DỮ LIỆU

Modus ponens

- 1. E1
- 2. E1 E2
- 3. E2
- Nếu có tiền đề khác, có dạng **E2 E3** thì **E3** được đưa vào danh sách.

2. Modus tollens

- 1. E2
- 2. E1 E2
- 3. E1

Ví dụ suy diễn tiến

- Luật 1.** IF Bệnh nhân rất họng AND Nghi viêm nhiễm
- THEN** Tin rằng bệnh nhân viêm họng, đi chữa họng.
- Luật 2.** IF Nhiệt độ bệnh nhân quá 37 độ
- THEN** Bệnh nhân bị sốt
- Luật 3.** IF Bệnh nhân ốm trên 1 tuần AND Bệnh nhân sốt
- THEN** Nghi bệnh nhân viêm nhiễm.
- Thông tin từ bệnh nhân là:
 - Bệnh nhân có nhiệt độ 39 độ
 - Bệnh nhân đã ốm hai tuần
 - Bệnh nhân họng rất
- Khi hệ thống thấy giả thiết của luật khớp với thông tin trong bộ nhớ, câu kết luận của luật được bổ sung vào bộ nhớ.
- Mình họa Ví dụ suy diễn lùi**

Cơ chế suy diễn

Suy diễn với logic mệnh đề:

1. Thuật toán suy diễn tiến

Input: - Tập luật Rule = {r₁, r₂, ..., r_m}
 - GT, KL
 Output: Thông báo "thành công" nếu GT KL
 Ngược lại, thông báo "không thành công"

Method:
 TD=GT;
 T=Loc(Rule, TD);
 While (KL TD) AND (T) Do
 {
 r = Get(T);
 TD=TD (q); // r:left q
 Rule = Rule \ {r};
 T=Loc(Rule, TD);
 }
 If KL TD THEN Return "True"
 else Return "False"

Ví dụ: Rule = {r₁:a c, r₂:b d, r₃:a e, r₄:a d e, r₅:b c f, r₆:e f g}
 Hỏi a b g? a g?

Thuật toán suy diễn lùi

```

If KL GT THEN Return "True"
Else (TDich = ; Vét = ; First=1; Quaylui= False;)
For Each q KL DO TDich=TDich {(q,0)};
Repeat
  first++;
  ((f,i)=Get(TDich);
  If (f GT) THEN
  {
    j = TimLuật(f,i,Rule); // ri: Leftj f
    If (Tim có ri) THEN
    { Vet = Vet {(f,j)};
      For Each t ((Leftj/GT) DO TDich = TDich {(t,0)};
    }
    else
    { Quaylui=True;
      While (f KL) AND Quaylui DO
      { Repeat { (g,k)=Get(Vét);
        TDich = TDich \ Leftk;
        Until f Leftk;
        j=TimLuật(g,k,Rule);
      }
    }
  }

```

Thuật toán suy diễn lùi

```

If KL GT THEN Return "True"
Else (TDich= ; Vét = ; First=1; Quaylui= False;)
For Each q KL DO TDich=TDich {(q,0)};
Repeat
  first ++;
  ((f,i)=Get(TDich);
  If (f GT) THEN
  {
    j = Timluật(f,i,Rule); // ri; Lefti f
    If (Tìm có ri) THEN
      { Vét = Vét {(f,j)};
        For Each t (LeftiGT) DO TDich = TDich {(t,0)};
      }
    else
      { Quaylui=True;
        While (f KL) AND Quaylui DO
        {
          Repeat { (g,k)=Get(Vét);
                  TDich = TDich \ Leftk;
          Until f Leftk;
          I=Timluật(g,k,Rule);
        }
      }
  }

```

Thuật toán suy diễn lùi

```

If (Tìm có ri) THEN
{ TDich = TDich \ Lefti ;
  For Each t (Lefti\GT) DO
    TDich = TDich {(t,0)};
    Vét = Vét {(g,i)};
    Quaylui = False;
  }
  else f=g;
}
}
Until (TDich = ) OR ((f KL) and (First>2));
If (f KL) then Return False else Return True;
Ví dụ: Rule = {r1:a c, r2:b d, r3:a e, r4:a d e, r5:b c f, r6:e f g}
Hỏi a b g? , a g?

```

Ưu điểm

- * **Suy diễn tiến**
- Ưu điểm chính của suy diễn tiến là làm việc tốt khi bài toán về bản chất đi thu thập thông tin rồi thấy điều cần suy diễn.
- Suy diễn tiến cho ra khối lượng lớn các thông tin từ một số thông tin ban đầu. Nó sinh ra nhiều thông tin mới. * Suy diễn tiến là tiếp cận lý tưởng đối với loại bài toán cần giải quyết các nhiệm vụ như lập kế hoạch, điều hành điều khiển và diễn dịch.
- * **Suy diễn lùi**
- Một trong các ưu điểm chính của suy diễn lùi là phù hợp với bài toán đưa ra giả thuyết rồi xem hiệu quả giả thiết đó có đúng không.
- Suy diễn lùi tập trung vào đích đã cho. Nó tạo ra một loạt câu hỏi chỉ liên quan đến vấn đề đang xét, đến hoàn cảnh thuận tiện đối với người dùng.
- Khi suy diễn lùi muốn suy diễn cái gì đó từ các thông tin đã biết, nó chỉ tìm trên một phần của cơ sở tri thức thích đáng đối với bài toán đang xét.

Nhược điểm

- * **Suy diễn tiến**
- Một nhược điểm chính của hệ thống suy diễn tiến là không cảm nhận được rằng chỉ một vài thông tin là quan trọng. Hệ thống hỏi các câu hỏi có thể hỏi mà không biết rằng chỉ một ít câu đã đi đến kết luận được.
- Hệ thống có thể hỏi cả câu không liên quan. Có thể các câu trả lời cũng quan trọng, nhưng làm người dùng lúng túng khi phải trả lời các câu không dính đến chủ đề.
- * **Suy diễn lùi**
- Nhược điểm cơ bản của suy diễn này là nó thường tiếp theo dòng suy diễn, thay vì đúng ra phải đúng ở đó mà sang nhánh khác. Tuy nhiên có thể dùng nhân tố tin cậy và các luật meta để khắc phục.