

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN

BỘ MÔN CÔNG NGHỆ TRI THỨC

-----oOo-----

LÊ TRỌNG NGỌC

9912622

HỆ CBR CHẨN ĐOÁN BỆNH TIM

GIÁO VIÊN HƯỚNG DẪN

TH.S LÊ HOÀNG THÁI

TP. HCM 7/ 2003

MỤC LỤC

CHƯƠNG 1: SUY DIỄN DỰA TÌNH HUỐNG	4
1.1 Giới thiệu :.....	4
1.2 Nội dung của luận văn :	5
1.3 Suy diễn dựa tình huống :.....	5
1.3.1 Suy diễn dựa tình huống là gì :	5
1.3.2 Các kiểu CBR :.....	8
1.3.2.1 CBR giải thích :.....	9
1.3.2.2 CBR giải quyết vấn đề :.....	10
1.3.3 Tại sao lại dùng CBR :.....	11
1.4 CBR và các kỹ thuật khác :	14
1.4.1 CBR và kỹ thuật truy tìm thông tin :.....	14
1.4.2 CBR và các hệ trên cơ sở luật :.....	15
1.4.3 CBR và phương pháp máy học :.....	16
1.4.4 CBR và mạng neural :.....	16
1.5 CBR và các tiếp cận liên quan :.....	17
1.6 Lịch sử và các ứng dụng CBR :.....	19
CHƯƠNG 2: CÁC THÀNH PHẦN VÀ CÁC TIẾN TRÌNH CỦA HỆ CBR	22
2.1 Các tiến trình :	22
2.1.1 Tiến trình nhớ :.....	22
2.1.2 Tiến trình sửa đổi :	24
2.1.3 Tiến trình học :	26
2.2 Tình huống :	27
2.2.1 Tình huống là gì :	27

2.2.2 Các kiểu tình huống :	28
2.2.3 Tình huống trong cơ sở tình huống :	28
2.2.4 Tích hợp tiến trình đánh giá tương tự và thích nghi trong tình huống :	29
2.2.5 Các thành phần cơ bản của một tình huống :	30
2.2.6 Tính chất của tình huống :	31
2.3 Biểu diễn tính chất của tình huống :	32
2.3.1 Biểu diễn cặp tính chất-giá trị :	32
2.3.2 Trọng số của tính chất :	33
2.3.3 Biểu diễn định tính :	33
2.3.4 Biểu diễn định lượng :	34
2.3.5 Biểu diễn bằng kỹ thuật mờ :	35
2.4 Một số phương pháp đánh giá tương tự :	37
2.4.1 Đánh giá tương tự dựa trên khoảng cách metric :	37
2.4.2 Dùng cây phân loại :	38
2.4.3 Tiếp cận của Vargas & Bourne :	38
2.4.4 Tiếp cận của Werner Dubitzky :	39
2.5 Cơ sở tình huống :	40
2.5.1 Cơ sở tình huống là gì :	40
2.5.2 Một số vấn đề cần chú ý khi xây dựng cơ sở tình huống :	42
2.5.2.1 Tích hợp tri thức cơ bản và tri thức đặc biệt :	42
2.5.2.2 Biểu diễn tri thức không đầy đủ và không chắc chắn :	42
2.5.2.3 Vấn đề chỉ mục :	43
2.5.2.4 Ngữ cảnh :	44
2.5.2.5 Vấn đề thu thập tri thức :	45
2.6 Một số mô hình cơ sở tình huống :	45
2.6.1 Tiếp cận cơ sở MOP:	46

2.6.2 Mô hình phân loại các mẫu :	47
2.6.3 Mô hình tình huống trừu tượng :	49
2.6.4 Tiếp cận dùng kỹ thuật mờ :	50
2.6.5 Mô hình PERCEPT :	53
CHƯƠNG 3: XÂY DỰNG HỆ CBR CHẨN ĐOÁN BỆNH TIM	54
3.1 CBR trong y khoa	54
3.1.1 Các đặc trưng của y khoa	54
3.1.2 Thuận lợi của CBR trong y khoa	56
3.1.3 Một số hệ CBR trong y khoa	58
3.2 CBR chẩn đoán bệnh tim	60
3.2.1 Giới thiệu	60
3.2.2 Nguồn dữ liệu	61
3.2.3 Biểu diễn tình huống	61
3.2.4 Đánh giá tương tự giữa các tình huống	62
3.2.5 Phân loại tình huống	66
3.2.6 Truy tìm tình huống	66
3.2.7 Thích nghi tình huống	68
3.2.8 Tiến trình học	69
CHƯƠNG 4: ĐÁNH GIÁ KẾT QUẢ VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỀ TÀI	70
4.1 Đánh giá kết quả	70
4.2 Các hướng phát triển	73
4.2.1 Đối với ứng dụng	73
4.2.2 Đối với CBR	73

CHƯƠNG 1: SUY DIỄN DỰA TÌNH HUỐNG

1.1 Giới thiệu :

Các hệ cơ sở tri thức hay hệ chuyên gia thường được dùng trong môi trường thế giới thực và môi trường nghiên cứu để nhắm đến các bài toán mở. Các bài toán mở thường liên quan tới các lĩnh vực mà nền tảng lý thuyết yếu.

Trong các lĩnh vực mà nền tảng lý thuyết yếu, tri thức cơ bản không đủ mạnh để mô tả tất cả các hiện tượng trong lĩnh vực. Đặc biệt tri thức cơ bản không đầy đủ có thể quá hẹp để cho phép phát triển các lời giải đúng đắn cho tất cả các vấn đề xuất hiện trong lĩnh vực. Nền tảng lý thuyết yếu bắt nguồn từ những quan hệ không chắc chắn giữa các khái niệm của lĩnh vực, lĩnh vực càng yếu quan hệ càng không chắc chắn. Điển hình cho loại lĩnh vực này là chẩn đoán y khoa.

Theo truyền thống các hệ cơ sở tri thức cho các lĩnh vực này thường dùng các luật Heuristic để biểu diễn tri thức. Tiếp cận này đã bộc lộ những giới hạn của nó. Gần đây suy diễn và tri thức lĩnh vực yếu được biểu diễn xung quanh các tình huống (case) quá khứ, tiếp cận này được biết đến như là suy diễn dựa tình huống (CBR) hay suy diễn dựa trên kinh nghiệm. Trong suy diễn dựa tình huống, nguồn tri thức chủ yếu được biểu diễn bởi bộ nhớ các tình huống. Các tình huống này ghi lại các tình tiết đặc biệt trước đó, và các lời giải mới được tạo bằng cách truy tìm các tình huống phù hợp nhất từ

bộ nhớ và làm thích nghi chúng cho vừa với các tình huống mới . Thuận lợi chính của các hệ CBR là tính đơn giản và hiệu quả của chúng .

Luận văn này sẽ nghiên cứu về suy diễn dựa tình huống và ứng dụng nó trong việc xây dựng hệ CBR chẩn đoán bệnh tim.

1.2 Nội dung của luận văn :

Chương 1 sẽ nêu tổng quát về CBR và những ưu điểm của nó trong việc xây dựng hệ chuyên gia. Chương này cũng so sánh CBR với các kỹ thuật khác của trí tuệ nhân tạo, so sánh CBR với các tiếp cận gần gũi với CBR . Cuối cùng chương này nêu sự hình thành và những thành công mà lĩnh vực CBR đạt được .

Chương 2 sẽ phân tích kỹ hơn về hệ CBR, biểu diễn tri thức tình huống và đánh giá tương tự giữa chúng, tổ chức cơ sở tình huống và các tiến trình của chu trình suy diễn CBR.

Chương 3 sẽ trình bày những đặc điểm của y khoa và thuận lợi mà CBR cung cấp cho việc xây dựng hệ chuyên gia chẩn đoán y khoa.Sau đó sẽ triển khai xây dựng hệ CBR chẩn đoán bệnh tim

Chương 4 sẽ nêu những kết quả đạt được và những điều chưa đạt được trong luận văn này.Chương này cũng phân tích những xu hướng phát triển của hệ CBR

1.3 Suy diễn dựa tình huống :

1.3.1 Suy diễn dựa tình huống là gì :

Suy diễn dựa tình huống (CBR-case based reasoning) là phương pháp trí tuệ nhân tạo khá mới mẻ.Nó giải quyết vấn đề khác về cơ bản so với các

phương pháp trí tuệ nhân tạo khác. Thay vì giải quyết vấn đề trên cơ sở tri thức tổng quát hay trên cơ sở các tri thức được tổng quát hóa, hệ CBR giải quyết vấn đề dựa trên những tri thức đặc biệt của các tình huống đã bắt gặp trong quá khứ.

Một cách hình thức hơn ta mô tả CBR như sau :

Suy diễn dựa tình huống có nghĩa là suy diễn từ các tình huống có sẵn. Tình huống có sẵn ở đây là các tình tiết, các trường hợp hay các kinh nghiệm trong quá khứ. Hệ suy diễn dựa tình huống sẽ dùng những tình huống này để đưa ra giải pháp cho tình huống mới. Giải pháp có thể là một lời giải hoàn chỉnh, một phương pháp thích nghi, một lời cảnh báo, một sự phân loại tình huống,...

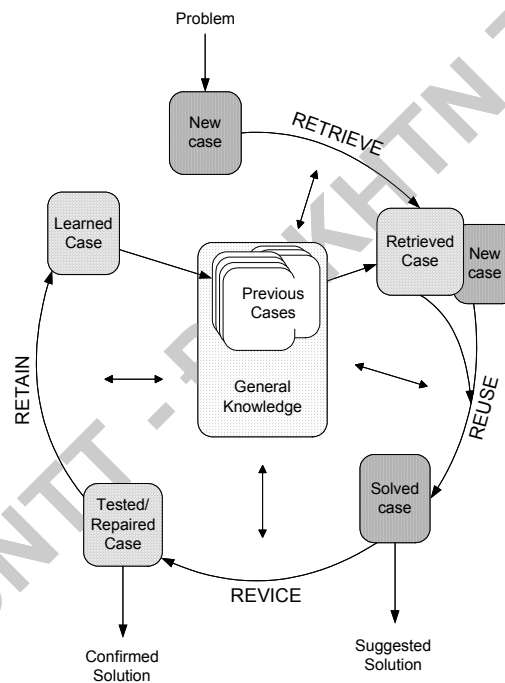
Theo Kolodner và Leak thì niềm tin của phương pháp suy diễn dựa tình huống dựa vào 4 giả định sau :

1. Các hành động tiến hành trong điều kiện giống nhau hay tương tự nhau thường dẫn đến các kết quả giống nhau hay tương tự nhau.
2. Các sự kiện có xu hướng lặp lại. Như vậy các kinh nghiệm trong hệ CBR thường hữu ích trong tương lai.
3. Những thay đổi nhỏ trong thế giới chỉ yêu cầu những thay đổi nhỏ trong nhận thức của chúng ta về thế giới, và những thay đổi nhỏ trong cách mà chúng ta thích nghi những tình tiết thay đổi này.
4. Mặc dù các tình huống hiếm khi lặp lại một cách chính xác nhưng sự khác biệt là thường rất nhỏ và những khác biệt này là dễ dàng bù đắp.

Amodt và Plaza đã mô tả CBR như là chu trình gồm 4 bước như sau :

1. Truy tìm tình huống tương tự nhất từ cơ sở tình huống.
2. Tái sử dụng tình huống này để đề nghị lời giải.
3. Xem xét lại lời giải này.
4. Ghi nhận lại các thông tin về tình huống mới sau khi được giải quyết.

Chu trình này được thể hiện trực quan hơn qua hình vẽ sau :

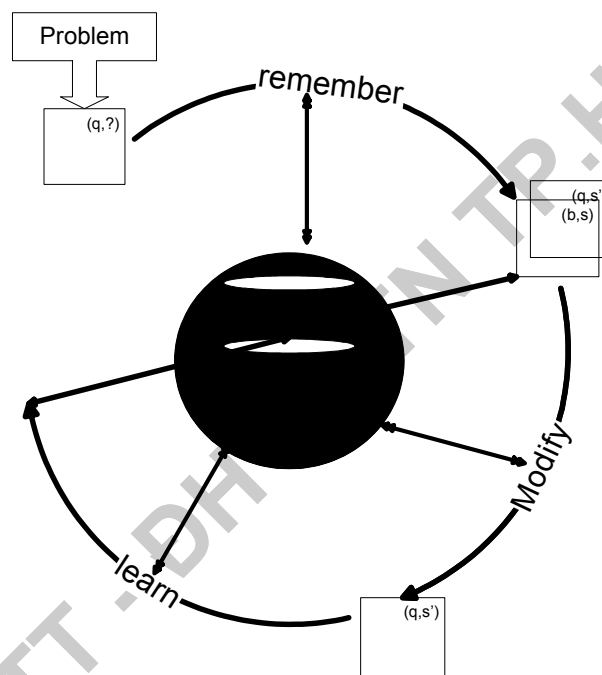


Tuy nhiên xét về mặt bản chất Werner Dubitzky mô tả lại chu trình này thành 3 bước : nhớ, sửa đổi và học

Hệ CBR sẽ dùng những mô tả tình huống bài toán mới để truy tìm tình huống tương tự nhất trong cơ sở tình huống (tiến trình nhớ). Tình huống này

sẽ được sửa đổi cho phù hợp với ngữ cảnh bài toán mới và từ đó đề nghị lời giải cho nó. Thông tin về tình huống mới này sẽ được ghi nhận lại trong cơ sở tình huống (học) để sử dụng trong tương lai .

Ta thể hiện chu trình này trên hình vẽ như sau :



Ta nhận thấy rằng trong mô hình suy diễn dựa tình huống học được xem như là phân tích hợp của chu trình suy diễn.

1.3.2 Các kiểu CBR :

Thông thường CBR chia làm hai loại là CBR giải thích và CBR giải quyết vấn đề .CBR giải thích sẽ dùng các tình huống trước đó để phân loại hay đặc tả các tình huống mới ,trong khi CBR giải quyết vấn đề dùng các tình

huống trước đó đề đề xuất lời giải mới cho tình huống mới. Trong thực tế sự phân biệt này là không rõ ràng và cả hai kiểu có thể xuất hiện cùng nhau .

1.3.2.1 CBR giải thích :

Như ta nói trong CBR giải thích , hệ suy diễn có xu hướng tập trung vào mô tả tình huống .Cụ thể công việc này bao gồm đưa ra một phán quyết hay một sự phân loại tình huống mới . Điều này được làm bằng cách so sánh tình huống mới với tình huống có sẵn đã được phân loại trong quá khứ .

CBR giải thích mà ta thường bắt gặp nhất là việc đưa ra chứng cứ hay căn nguyên cho tính đúng đắn của luận điểm nào đó .Kiểu suy diễn này thường được các luật sư sử dụng , họ thường trích dẫn những tình huống phù hợp đã xảy ra trước đó .

CBR giải thích tiếp theo là sự phân lớp , nó sẽ đặt tình huống mới trong ngữ cảnh đặt biệt . Điều này thường yêu cầu quyết định tình huống sẽ thuộc tập nào trong các tập tình huống đã định sẵn .

CBR giải thích cuối cùng là việc dự báo, nó sẽ cố gắng dự đoán ảnh hưởng của các quyết định hay giải pháp hiện hành .

Một cách tổng quát các tiên trình giả thích có đầu vào mô tả tình huống hay lời giải được đề nghị cho bài toán và đầu ra của nó là sự phân lớp tình huống ,hỗ trợ tranh luận cho sự phân lớp hay lời giải ,hỗ trợ cho phán quyết ...

Tổng quát tiên trình CBR giải thích gồm 4 bước :

- 1.Đánh giá tình huống
- 2.Dùng kết quả của bước đánh giá tình huống ,hệ suy diễn truy tìm các tình huống trước đó phù hợp với tình huống hiện tại

3. Hệ suy diễn xác định giải thích phù hợp nhất chứa trong các tình huống được truy tìm và cố gắng ứng dụng vào tình huống hiện tại.
4. Hệ suy diễn nhớ tình huống hiện tại cùng với giải thích mới .

1.3.2.2 CBR giải quyết vấn đề :

Mục tiêu của CBR giải quyết vấn đề là áp dụng lời giải của quá khứ để tạo lời giải cho bài toán mới . Trong xây dựng các lời giải cho bài toán mới , các tình huống thi hành hai chức năng chính : chúng đề nghị các lời giải hầu như đúng cho bài toán mới mà sau đó sẽ được sửa cho phù hợp với bài toán mới và chúng cảnh báo tiềm ẩn của sự thất bại . Trong lĩnh vực giải quyết vấn đề, nắm bắt kinh nghiệm trong quá khứ cho phép hệ suy diễn trở nên hiệu quả hơn đúng đắn hơn qua thời gian. Các lĩnh vực áp dụng CBR giải quyết vấn đề thường là thiết kế , lập lịch, diễn tả và chẩn đoán .

Thiết kế : Hệ suy diễn phải tìm ra các lời giải cho các bài toán , chúng được định nghĩa như là tập các ràng buộc. Các bài toán thiết kế điển hình bao gồm : thiết kế tòa nhà, máy, mạch điện tử. Nhớ các tình huống thiết kế trong quá khứ để tạo các ràng buộc tương tự cho bài toán hiện tại có thể giúp người thiết kế xây dựng một lời giải mới thỏa các ràng buộc trong bài toán thiết kế mới .

Lập lịch : hệ suy diễn phải tạo chuỗi các bước hay các hành động hay các lịch trình cho một công việc nào đó . Thường công việc lập lịch áp dụng cho các lĩnh vực : giao thông , công thức nấu ăn ,....

Chẩn đoán : Hệ suy diễn cố gắng diễn tả một tập các triệu chứng hay đặc trưng . Khi có một lượng lớn các diễn tả hay chẩn đoán có thể thì chẩn đoán trở thành bài toán giải thích . Nếu số lượng là nhỏ thì chẩn đoán trở thành bài toán phân lớp . Các công việc chẩn đoán thường dùng là chẩn đoán

bệnh hay chẩn đoán lỗi . Dùng các chẩn đoán hay các diễn tả trước đó có thể giúp để chẩn đoán một tập các triệu chứng mới và để cảnh báo một diễn tả không phù hợp trong quá khứ .

Chất lượng lời giải được đề nghị bởi hệ CBR trong các lĩnh vực mà hiểu biết không đầy đủ cũng cao hơn được đề nghị bởi hệ trên cơ sở luật do nó phản ánh một tình huống thực sự xảy ra hay không xảy ra trong tập các tình huống đã cho chứ không phải là những tình huống dự đoán theo mô hình lý thuyết

Cuối cùng một hệ có được chấp nhận hay không phụ thuộc vào nó có thuyết phục được người dùng rằng kết luận được suy diễn một cách hợp lý.CBR cung cấp một cơ chế suy diễn thuyết phục được người dùng .

1.3.3 Tại sao lại dùng CBR :

Ta thấy rằng mục tiêu cuối cùng của trí tuệ nhân tạo là xây dựng một hệ thống trí tuệ nhân tạo hiệu quả,tin cậy mô phỏng con người trong các hoạt động ra quyết định và giải quyết vấn đề.

Quan sát thấy rằng con người dùng suy diễn dựa tình huống trong nhiều ngữ cảnh công việc khác nhau cả để ra quyết định và giải quyết vấn đề ,đã khuyến khích các nhà nghiên cứu trí tuệ nhân tạo dùng như là framework cho các hệ thống máy tính thông minh.Con người là một hệ giải quyết vấn đề mạnh mẽ;họ giải quyết các vấn đề khó bất chấp các tri thức không chắc chắn và không đầy đủ và năng lực giải quyết vấn đề của họ cải tiến với kinh nghiệm.Những phẩm chất này là mong ước của các hệ máy tính thông minh

hoạt động trong thế giới thực. Các nhà nghiên cứu trí tuệ nhân tạo đã nhận diện ra 6 vấn đề chính trong trí tuệ nhân tạo mà CBR có khả năng hỗ trợ.

Vấn đề thu thập tri thức

Trong các hệ trí tuệ nhân tạo xây dựng trên cơ sở luật thì luôn gặp phải khó khăn là vấn đề tạo luật. Tạo luật tức là từ những mô tả, phân tích tri thức mà chuyên gia cung cấp xây dựng thành các luật để biểu diễn trong máy tính. Tuy nhiên các chuyên gia thì rất giỏi về chuyên môn nhưng không giỏi trong lĩnh vực máy tính, còn những người xây dựng hệ thống thì ngược lại do đó việc tạo luật rất nặng nề và cho kết quả có độ tin cậy không cao. Đôi lúc các luật không được sự đồng thuận của các chuyên gia. Trong nhiều trường hợp việc tạo luật thậm chí rất khó và số lượng luật đôi khi rất lớn không thể quản lý được. Các hệ xây dựng trên cơ sở suy diễn dựa tình huống suy diễn từ các tình tiết đặc biệt do đó tránh được vấn đề tạo luật này. Nhiều lĩnh vực CBR phù hợp một cách tự nhiên bởi vì các tình huống là một phần trong thủ tục giải quyết vấn đề.

Linh động trong việc biểu diễn tri thức

Suy diễn trên cơ sở luật còn giới hạn các kiểu tri thức được biểu diễn trong khi CBR có thể khai thác nhiều kiểu tri thức. Tri thức trong các hệ CBR không những chỉ biểu diễn trong bản thân tình huống mà còn trong sơ đồ chỉ mục, trong cơ chế đánh giá tương tự, và trong chiến lược thích nghi tình huống. Điều này cho phép các nhà phát triển hệ thống linh động trong việc chọn lựa các khả năng biểu diễn tri thức.

Vấn đề bảo trì tri thức

Bên cạnh khó khăn trong việc tạo luật và hạn chế trong việc biểu diễn tri thức các hệ trên cơ sở luật còn gặp một khó khăn khác mang tính sống còn là bảo trì tri thức. Trong thực tế việc hiểu biết bất kỳ một lĩnh vực nào đó là không hoàn chỉnh và luôn được bổ sung thêm. Điều này dẫn đến các tri thức mà chuyên gia cung cấp luôn phải được bổ sung và sửa đổi dẫn đến hệ luật trong cơ sở tri thức cũng phải được bổ sung và sửa đổi. Khi bổ sung một luật vào hệ luật hay sửa đổi một luật có sẵn trong hệ luật có thể dẫn đến mâu thuẫn trong hệ luật và làm đổ vỡ toàn bộ cơ sở tri thức. Trong hệ CBR việc cập nhật cơ sở tri thức là một phần trong toàn bộ tiến trình suy diễn và việc cập nhật này được thực hiện mà không cần sự giúp đỡ của chuyên gia. Bên cạnh đó hệ CBR suy diễn từ những tình huống cụ thể do đó chỉ cần nắm bắt những tình huống thực sự xuất hiện trong thực tế chứ không cần phải là toàn bộ những tình huống theo nguyên lý.

Tăng hiệu quả giải quyết vấn đề

Các hệ CBR còn làm tăng hiệu quả giải quyết vấn đề nhờ việc tái sử dụng. Thay vì lặp lại toàn bộ suy diễn như trong hệ trên cơ sở luật hệ CBR sử dụng lại những suy diễn cũ. Hơn thế các hệ CBR lưu trữ cả thất bại lẫn thành công nên nó có khả năng cảnh báo những vấn đề tiềm ẩn để tránh

Tăng chất lượng của lời giải

Chất lượng lời giải được đề nghị bởi hệ CBR trong các lĩnh vực mà hiểu biết không đầy đủ cũng cao hơn được đề nghị bởi hệ trên cơ sở luật do nó phản ánh một tình huống thực sự xảy ra hay không xảy ra trong tập các

tình huống đã cho chứ không phải là những tình huống dự đoán theo mô hình lý thuyết

Sự chấp nhận của người dùng

Cuối cùng một hệ có được chấp nhận hay không phụ thuộc vào nó có thuyết phục được người dùng rằng kết luận được suy diễn một cách hợp lý. CBR cung cấp một cơ chế suy diễn thuyết phục được người dùng.

1.4 CBR và các kỹ thuật khác :

1.4.1 CBR và kỹ thuật truy tìm thông tin :

CBR và kỹ thuật truy tìm thông tin (IR-information Retrieval) đều tập trung vào việc truy tìm thông tin từ cơ sở dữ liệu (CSDL), cho phép truy vấn CSDL một cách linh động, và trả về một tập các số khớp phù hợp nhưng không chính xác.

Tuy nhiên chúng cũng có khác nhau :

Các phương pháp IR hoạt động trên dữ liệu văn bản trong khi CBR làm việc trên nhiều dữ liệu hỗn hợp

Các phương pháp IR có thể nắm bắt một lượng khổng lồ dữ liệu và có thể tìm kiếm thông qua hàng ngàn tài liệu nhưng CBR tương đối ít hơn

Các hệ IR làm việc độc lập với công việc giải quyết vấn đề của người dùng. IR cung cấp một động cơ chỉ mục và truy tìm chung cho một miền rộng các bài toán. Trong khi CBR dùng ngay tri thức về tiến trình giải quyết công việc để xây dựng chỉ mục hiệu quả và cải tiến độ chính xác của việc truy tìm.

Các tool của hệ CBR biểu diễn rõ ràng tri thức mà chúng dùng ,trong khi các hệ IR thì không.

Trong các ứng dụng có cấu trúc phức tạp yêu cầu tích hợp nhiều phương pháp học và giải quyết vấn đề giàu tri thức khác nhau,sự khác biệt giữa các hệ CBR và IR trở nên rất rõ ràng.

1.4.2 CBR và các hệ trên cơ sở luật :

Phát triển các hệ chuyên gia trên cơ sở luật để giải quyết các vấn đề thế giới thực phức tạp là công việc rất khó khăn.Một trong những khó khăn chính là các luật phải được cung cấp bởi những chuyên gia trong khi những chuyên gia này mặc dù giải quyết các vấn đề thực tế thuộc lĩnh vực của họ rất tốt nhưng việc diễn tả những kinh nghiệm của họ trong việc giải quyết bài toán dưới dạng luật là rất khó khăn và không đáng tin cậy.CBR cung cấp một phương pháp tốt hơn trong việc này,thay vì phải diễn tả những kinh nghiệm của mình dưới dạng luật logic ,các chuyên gia chỉ cần cung cấp những ví dụ cụ thể.Điều này cho phép những người xây dựng hệ CBR không phải quá phụ thuộc vào chuyên gia và các chuyên gia cũng dễ dàng hơn trong việc cung cấp tri thức.

CBR rất hữu ích trong những lĩnh vực mà sự hiểu biết không đầy đủ và quá nhiều ngoại lệ đối với tập luật đã biết.

Một vấn đề khá quan trọng trong hệ chuyên gia là việc bảo trì tri thức.Theo thời gian các tri thức cho việc giải quyết vấn đề sẽ được cập nhật ngày càng nhiều,tuy nhiên việc đưa những tri thức cập nhật vào hệ chuyên gia trên cơ sở luật là rất khó khăn thậm chí có thể làm đổ vỡ luôn hệ cơ sở luật.Hệ CBR cho phép cập nhật kinh nghiệm một cách dễ dàng,điều này làm cho cơ sở tình huống ngày càng tốt hơn và hệ CBR trở nên mạnh hơn theo thời gian .

1.4.3 CBR và phương pháp máy học :

Máy học phân chia rõ ràng giữa học và giải quyết vấn đề .Học là việc phân tích các mẫu huấn luyện để trích chiết ra các hàm hay luật ;giải quyết vấn đề là việc ứng dụng những hàm hay luật này vào vấn đề mới.CBR không phân chia hai khái niệm này.

Trong máy học thuật toán cho việc học được chú ý hơn là khía cạnh giải quyết vấn đề của hệ thống .

Một sự khác biệt quan trọng là tiến trình học trong máy học là tiến trình biên dịch trong khi học trong CBR là tiến trình thông dịch,có nghĩa là học tại thời điểm suy diễn.

1.4.4 CBR và mạng neural :

Mạng neural thi hành tốt hơn CBR trong các môi trường nghèo tri thức khi dữ liệu không thể được biểu diễn dưới dạng symbolic ví dụ như nhận dạng tín hiệu radar.Mạng neural cũng mở rộng lĩnh vực đến nhận dạng mẫu ở đó có nhiều dữ liệu thô như trong xử lý hình ảnh,tiếng nói,thị giác.

Mạng neural không phù hợp khi tri thức lĩnh vực nền phải được đưa vào hệ thống.Mạng neural không thể đương đầu với các cấu trúc phức tạp và để thực hiện tốt thông tin của lĩnh vực phải hiểu biết toàn diện trong pha học.

Mạng neural làm việc như hộp đen vì vậy chúng chịu thiệt thòi vì thiếu tính trong sáng.Giá trị của hệ thống là quyết định không thể được phán quyết bởi tính tự nhiên của các công việc bên trong,đầu ra của mạng là hàm các vector trọng số mà phụ thuộc vào kiến trúc của mạng và phương pháp học được dùng.

1.5 CBR và các tiếp cận liên quan :

Ta nhận thấy rằng CBR và các phương pháp suy diễn khác như: suy diễn tương tự (analogical reasoning), suy diễn theo mẫu (exemplar-based reasoning), suy diễn dựa theo ví dụ (instance-based reasoning), suy diễn trên cơ sở nhớ (memory-based reasoning) có chung tiến trình như nhau đó là đều dựa trên tình huống tương tự trước đó để suy diễn ra thông tin cho tình huống mới. Tuy nhiên chúng vẫn là các phương pháp khác nhau do đó phần này sẽ so sánh CBR với các phương pháp này để hiểu một cách chính xác hơn về CBR.

Suy diễn tương tự thường được đặc tả bởi giải quyết vấn đề thông qua những tương tự liên lĩnh vực, trong khi các phương pháp CBR thường tập trung vào những tương tự bên trong một lĩnh vực nào đó. Theo truyền thống suy diễn tương tự nghiên cứu tập trung vào bài toán ánh xạ và bài toán tương quan. Bài toán ánh xạ liên quan tới việc xác định các tính chất nào của tình huống mới nên được so khớp với các tính chất trong tình huống đã biết (về tổng quát các đặc trưng với vai trò chức năng giống nhau nên được so khớp). Bởi vì các hệ CBR thường hoạt động trong một lĩnh vực nên bài toán ánh xạ thường không rắc rối. Nó quan tâm hơn tới những xem xét thực tế liên quan tới toàn bộ tiến trình truy tìm, ánh xạ, chọn lựa, thích nghi, lưu trữ và hữu ích của các chi mục, các lời giải và các tình huống.

Trong suy diễn theo mẫu, giải quyết vấn đề có nghĩa là phân lớp một sự kiện, tình huống, trường hợp hay đối tượng mới được biết như là các mẫu. Tìm một mẫu tương tự nhất với tình huống mới xác định lớp mà tình huống đó thuộc về. Tập các lớp đã biết tạo thành không gian các lời giải cho suy diễn trên cơ sở mẫu. Suy diễn trên cơ sở mẫu xuất phát từ ba quan điểm khoa học nhận thức chủ yếu: quan điểm lớp, quan điểm mẫu và quan điểm nguyên mẫu. Trong quan điểm mẫu giả định quan trọng là các lớp được biểu

diễn thông qua các mẫu hay các lớp khác hơn là thông qua việc mô tả tính chất. CBR và suy diễn trên cơ sở mẫu liên quan xung quanh việc truy tìm các thực thể phù hợp từ cơ sở tri thức. Như vậy hai tiếp cận chia sẻ những vấn đề tổ chức bộ nhớ và làm chắc rằng các thực thể phù hợp được truy tìm tại những thời điểm phù hợp (chi mục). Nhưng suy diễn dựa tình huống khác với suy diễn trên cơ sở mẫu bởi vì tiến trình suy diễn dựa tình huống còn bao hàm việc suy diễn trên cơ sở thông tin được chứa trong tình huống khớp nhất và làm thích nghi nó cho vừa với tình tiết mới

Về cơ bản **suy diễn dựa theo ví dụ** là tiếp cận liên quan tới việc học hay tương tự về mặt khái niệm với các phương pháp máy học đệ quy. Nhưng không như các phương pháp máy học đệ quy truyền thống, học theo ví dụ ghi nhớ những ví dụ hay mẫu đặc biệt sau đó tổng quát hoá từ các thể hiện này. Nhân mạnh trong suy diễn theo ví dụ là học với ít hay không có sự hướng dẫn từ tri thức nền tổng quát; sự thiếu sót tri thức nền tổng quát này được bù đắp bởi số lượng lớn các thể hiện được lưu trữ trong hệ thống. Suy diễn theo ví dụ cũng tập trung vào việc học tự động nghĩa là học không cần sự can thiệp của người dùng. Tương tự các phương pháp trên cơ sở mẫu, mục tiêu của học theo ví dụ thường là sự phân lớp các thể hiện mới. Suy diễn theo ví dụ được đặc tả bởi những biểu diễn vector đặc trưng đơn giản về mặt quan hệ, trong khi CBR, suy diễn trên cơ sở mẫu, suy diễn tương tự liên quan xung quanh việc biểu diễn giàu tri thức hơn.

Cuối cùng tiếp cận **suy diễn trên cơ sở nhớ** xem tập hợp các tình huống, các thể hiện, các trường hợp, hay các mẫu như là bộ nhớ lớn và chu trình suy diễn như là một tiến trình tìm kiếm trong bộ nhớ đó. Suy diễn trên cơ sở nhớ tương tự với CBR trong nhiều trường hợp. Về truyền thống các phương pháp trên cơ sở nhớ tập trung vào các kỹ thuật song song.

1.6 Lịch sử và các ứng dụng CBR :

Nguồn gốc của suy diễn dựa tình huống trong AI được phát hiện trong các công việc của Roger Schank về bộ nhớ động và vai trò trung tâm là sự nhớ lại các tình huống trước đó và các mẫu tình huống có trong giải quyết vấn đề và học. Những con đường khác vào lĩnh vực CBR đến từ nghiên cứu suy diễn tương tự, và sau nữa là từ lý thuyết khái niệm hình thức, giải quyết vấn đề và học kinh nghiệm trong tâm sinh lý. Cho ví dụ Wittgenstein đã quan sát rằng các 'khái niệm tự nhiên', có nghĩa là các khái niệm mà là một phần của thế giới tự nhiên-như chim, cam, ghế, xe hơi,.. là đa dạng. Điều đó có nghĩa là những trường hợp cụ thể của chúng có thể được phân loại bằng nhiều cách khác nhau và nó không thể đưa ra sự định nghĩa lớp đầy đủ, bằng ngôn ngữ tập hợp các tính chất đủ và cần thiết như các khái niệm. Trả lời cho bài toán này là để biểu diễn khái niệm một cách mở rộng, được định nghĩa bởi tập các trường hợp cụ thể của nó-hay tình huống.

Hệ đầu tiên mà có thể được gọi là hệ suy diễn dựa tình huống là hệ CYRUS được phát triển bởi Janet Kolodner tại đại học Yale (nhóm của Schank). CYRUS được dựa trên mô hình bộ nhớ động của Schank và lý thuyết MOP về giải quyết vấn đề và học. Về cơ bản nó là một hệ hỏi đáp với tri thức của nhiều chuyên đi và cuộc gặp của nguyên bộ trưởng Cyrus Vance của Mỹ. Mô hình bộ nhớ tình huống được phát triển cho hệ thống này sau đó được phục vụ như là cơ sở cho một số hệ suy diễn dựa tình huống khác (bao gồm MEDIATOR, PERSUADER, CHEF, JULIA, CASEY).

Cơ sở khác cho CBR và tập các mô hình khác được phát triển bởi Bruce Porter và nhóm của ông ta tại đại học Texas, Austin. Ban đầu nó nhắm đến vấn đề máy học của việc học khái niệm cho các công việc phân lớp. Điều

này dẫn đến sự phát triển của hệ PROTOS mà nhấn mạnh sự tích hợp tri thức tổng quát của chuyên ngành và tri thức đặc biệt của tình huống vào một cấu trúc biểu diễn hợp nhất. Sự kết hợp các tình huống với tri thức tổng quát của chuyên ngành được đưa vào trong GREBE, một ứng dụng trong chuyên ngành luật. Sự đóng góp có ý nghĩa khác cho CBR là công việc của Edwina Rissland và nhóm của cô ta ở đại học Massachusetts, Amherst. Với một số nhà khoa học về luật trong nhóm, họ quan tâm tới vai trò của suy diễn trước đó của sự phán xét hợp lệ. Các tình huống ở đây không được dùng để đưa ra một câu trả lời đơn mà để thông dịch một tình huống của tòa và để dẫn ra và đánh giá các tranh luận của cả hai bên. Điều này đã có kết quả trong hệ HYPO, và sau đó đã kết hợp hệ suy diễn dựa tình huống và hệ trên cơ sở luật CABARET. Phyllis Koton ở MIT đã nghiên cứu việc sử dụng suy diễn dựa tình huống để tối ưu hiệu quả trong các hệ cơ sở tri thức hiện hữu, ở đó lĩnh vực (chứng liệt tim) được mô tả bởi mô hình nhân quả, sâu sắc. Điều này đã có kết quả trong hệ CASEY mà suy diễn theo mô hình sâu và dựa tình huống được kết hợp.

Ở châu Âu nghiên cứu CBR được khởi xướng trễ hơn một chút so với ở Mỹ. Các công việc CBR dường như được gắn mạnh hơn đến sự phát triển của hệ chuyên gia và nghiên cứu thu thập tri thức hơn ở Mỹ. Trong số những kết quả sớm nhất là CBR cho chẩn đoán kỹ thuật phức tạp trong hệ MOLTKE được làm bởi Michael Richter cùng với Klaus Dieter Althoff và những người khác ở đại học Kaiserslautern. Điều này dẫn đến hệ PATDEX, với Stefan Wess như người phát triển chính, và sau đó là một số phương pháp và hệ thống khác [III.A ở Blane, Enric Plaza và Ramon Lopez de Mantaras đã phát triển hệ học dựa tình huống cho chẩn đoán y khoa, và Beatrice Lopez đã nghiên cứu việc sử dụng các phương pháp dựa tình huống cho suy diễn ở

mức chiến lược. Ở Aberdeen, nhóm của Derek Sleeman đã nghiên cứu việc sử dụng các tình huống cho việc tinh chế cơ sở tri thức. Kết quả ban đầu là hệ REFINER được phát triển bởi Sunil Sharma. Kết quả khác là hệ IULIAN cho sự xem lại lý thuyết. Tại đại học Tondheim, Agnar Aamodt và các đồng nghiệp tại Sintef đã nghiên cứu khía cạnh học của CBR trong ngữ cảnh của việc rhu thập tri thức về tổng quát, và đặc biệt là bảo trì tri thức. Về giải quyết vấn đề, kết hợp việc dùng các tình huống và tri thức tổng quát của lĩnh vực được chú ý. Điều này dẫn đến sự phát triển của hệ CREEK và một khung tích hợp, và tiếp tục phát triển suy diễn dựa tình huống giàu tri thức. Về mặt khoa học nhận thức, công việc đầu tiên được thực hiện trên suy diễn tương tự bởi Mark Keane ở trường Trinity, Dublin, một nhóm mà đã được phát triển vào môi trường mạnh cho kiểu CBR này. Trong nhóm của Gerhard Strube ở đại học Freiburg, vai trò của tri thức rời rạc trong các mô hình nhận thức được tích hợp trong dự án EVENTS [Strube-90], mà dẫn đến tiêu sử của nghiên cứu hiện tại của nhóm về khoa học nhận thức và CBR.

Hiện tại hoạt động CBR ở Mỹ cũng như ở châu Âu đã lan rộng (ví dụ [DARPA-91],[IEEE-92],[EWCBR-93],[Allemagne-93], và tăng nhanh chóng số bài báo về CBR trong hầu hết các tạp chí AI). Đức dường như dẫn đầu trong số các nhà nghiên cứu và một số nhóm ở mức hoạt động và độ lớn có ý nghĩ đã được thiết lập gần đây. Từ Nhật Bản và các quốc gia châu Á khác cũng có những điểm tích cực, cho ví dụ ở Ấn Độ [Venkatamaran-93]. Ở Nhật Bản quan tâm là sự mở rộng tập trung đến tiếp cận tính toán song song của CBR [Kitano-93].

CHƯƠNG 2: CÁC THÀNH PHẦN VÀ CÁC TIẾN TRÌNH CỦA HỆ CBR

2.1 Các tiến trình :

2.1.1 Tiến trình nhớ :

Như ta thấy ở trên đầu vào của tiến trình nhớ là tình huống vấn đề mới và đầu ra là tình huống tương tự nhất với tình huống bài toán được lấy từ cơ sở tình huống. Tiến trình này được thực hiện bởi sự hỗ trợ của tri thức tổng quát.

Cơ sở tình huống là một thư viện các tình huống, đó là những kinh nghiệm mà hệ thống tham khảo khi thực hiện suy diễn. Cơ sở tình huống này có thể được tạo một cách trực tiếp bằng cách phỏng vấn các chuyên gia và thu thập các kinh nghiệm của họ; hoặc gián tiếp bằng các thủ tục tự động hay bán tự động để xây dựng từ các cơ sở dữ liệu có sẵn.

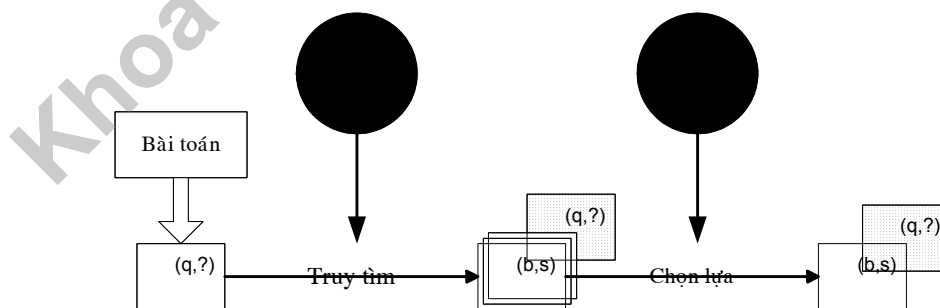
Một bài toán trong thực tế thường được mô tả không đơn giản, khi đưa vào máy tính thì những mô tả này khó có thể được xử lý tốt. Do vậy việc mô tả lại bài toán cho phù hợp với máy tính mà không làm mất đi bản chất của bài toán là cần thiết. Công việc này thực chất là lựa chọn những tính chất nào để mô tả bài toán. Các tính chất này thường là các tính chất nổi bật, quan

trọng. Và việc lựa chọn các tính chất có thể được thực hiện một cách thủ công hay tự động và được hỗ trợ bởi tri thức tổng quát đó là tri thức cho việc mô tả tình huống. Kết quả của việc lựa chọn là tập các tính chất hay tập các chỉ mục q . Các tình huống trong cơ sở tình huống cũng đã được đánh chỉ mục trước khi lưu trữ.

Hệ CBR sẽ dùng tập chỉ mục này để truy tìm tình huống, việc truy tìm được thực hiện bằng cách so khớp các chỉ mục. Thường thì việc truy tìm này được hỗ trợ bởi tri thức tổng quát.

Trên thực tế cơ sở tình huống thường rất lớn và như vậy việc truy tìm ra ngay một tình huống khớp nhất với tình huống bài toán sẽ rất khó khăn, thậm chí đôi khi không đáp ứng được về vấn đề thời gian cũng như ngữ nghĩa. Để giải quyết vấn đề này người ta thường chia thành hai tiến trình riêng biệt và liên tiếp nhau. Tiến trình đầu tiên sẽ truy tìm một tập các tình huống làm ứng viên, việc truy tìm này sẽ sử dụng các kỹ thuật so khớp đơn giản, và tri thức tổng quát hỗ trợ cho tiến trình này là tri thức hỗ trợ việc so khớp chỉ mục.

Tập các ứng viên này chính là đầu vào của tiến trình tiếp theo: tiến trình chọn lựa. Tiến trình này sẽ đánh giá một cách tinh vi sự tương tự giữa tình huống bài toán và các tình huống trong tập ứng viên để chọn ra một tình huống tương tự nhất (b,s) .



2.1.2 Tiến trình sửa đổi :

Như ta đã nói ở trên các tình huống hiếm khi lặp lại một cách chính xác, do vậy tình huống được nhớ lại trong tiến trình trên hầu như không bao giờ khớp một cách chính xác với tình huống bài toán mới. Vì thế hệ CBR thường phải thích nghi lời giải cũ để ứng dụng vào tình huống mới. Thích nghi ở đây có nghĩa là bằng cách nào đó biến đổi lời giải s của tình huống khớp nhất thành lời giải s' phù hợp với tình huống bài toán mới. Công việc này được sự hỗ trợ của tri thức tổng quát hỗ trợ cho sự thích nghi.

Sự thích nghi được áp dụng sau khi bài toán với lời giải cũ được đưa ra, hay trong suốt việc hình thành lời giải. Sự thích nghi có thể đơn giản là xoá một vài thứ từ hay thay thế một vài thứ trong lời giải cũ, hay phức tạp hơn là các phần tử mới được thêm vào lời giải cũ hay lời giải cũ được truyền vào hoàn toàn hay một phần.

Theo khảo sát của Kolodner thì có ít nhất 10 phương pháp thích nghi khác nhau được đề cập; các phương pháp này có thể được nhóm thành 4 nhóm:

Nhóm 1: Các phương pháp thay thế:

Phương pháp này sẽ thay thế các giá trị trong tình huống bài toán cũ thành các giá trị khác phù hợp hơn với tình huống bài toán mới. Việc thay thế này có thể là thiết lập lời giải của tình huống cũ vào tương tự tình huống mới; có thể dùng heuristic để điều chỉnh các thông số lời giải của tình huống cũ dựa trên sự khác biệt của hai tình huống; có thể thay thế trên cơ sở đề nghị của các tình huống khác; v.v...

Nhóm 2: Các phương pháp biến đổi

Các phương pháp này sẽ biến đổi lời giải của tình huống cũ vào tình huống mới; Việc biến đổi có thể dùng heuristic thông minh trong việc thay thế, xóa, hay thêm các thành phần vào lời giải cũ; hoặc dựa theo mô hình nhân quả để biến đổi.

Nhóm 3: Các phương pháp thích nghi với mục đích đặc biệt

Phương pháp này dùng tri thức chuyên ngành để tiến hành thích nghi mà có thể thay đổi toàn bộ cấu trúc của lời giải.

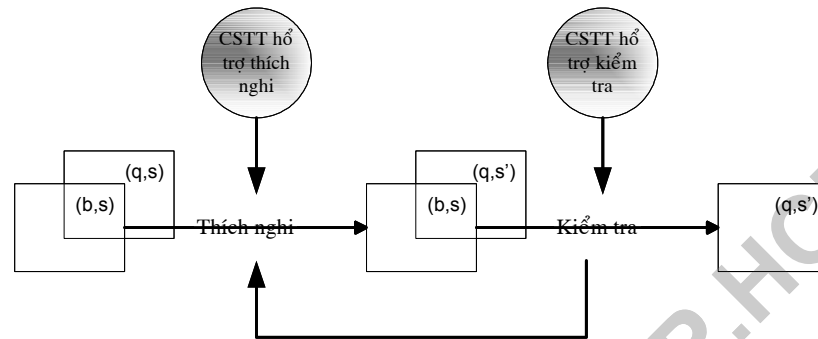
Nhóm 4: Các phương pháp lặp lại

Các phương pháp này dùng lại các bước trong tiến trình dẫn ra lời giải tình huống cũ để dẫn ra lời giải tình huống mới.

Sau khi thích nghi ta có một lời giải cho tình huống mới nhưng lời giải này không hẳn là hoàn hảo do đó ta cần phải kiểm tra lại lời giải. Việc kiểm tra lại lời giải cho phép hệ thống loại bỏ những lời giải nghèo nàn và dò tìm các vấn đề tiềm ẩn khác chứa trong nó.

Việc kiểm tra lại lời giải đòi hỏi nhiều tri thức của lĩnh vực và nỗ lực suy diễn lớn do đó việc thực hiện công đoạn này trong chu trình CBR khá khó khăn. Thông thường việc kiểm tra thường được thực hiện bằng cách dự đoán ảnh hưởng và kết quả của lời giải mới. Việc dự đoán này có thể được thực hiện bằng cách áp dụng trên môi trường của chương trình mô phỏng máy tính, hay bằng cách hỏi các chuyên gia hoặc áp dụng trực tiếp vào thế giới

thực.kết quả sau khi kiểm tra có thể được thích nghi một lần nữa.Tri thức tổng quát hỗ trợ cho tiến trình này là cơ sở tri thức hỗ trợ cho việc kiểm tra



2.1.3 Tiến trình học :

Học là tiến trình cuối cùng trong chu trình suy diễn dựa tình huống, tiến trình này ghi nhận lại những kinh nghiệm tại vị trí phù hợp trong cơ sở tình huống sau khi suy diễn ,và kinh nghiệm này sẽ được áp dụng trong tương lai.Như vậy thông qua tiến trình học này mà hệ CBR sẽ hiệu quả hơn và có năng lực hơn trong tương lai.

Học truyền thống trong trí tuệ nhân tạo thường dựa trên quy nạp hoặc dựa trên giải thích.Ý tưởng của học dựa trên quy nạp là tổng quát hóa các quan sát đặc biệt.Cụ thể là khi quan sát một số tình huống của một lớp ta thấy rằng các tình huống này đều sở hữu tính chất p,điều này cho ta kết luận rằng mọi tình huống trong lớp đều sở hữu tính chất p.Ý tưởng của việc học dựa trên giải thích là trên cơ sở các quan sát cụ thể để xây dựng một diễn tả cho lớp sau đó dùng tri thức chuyên ngành để xây dựng một định nghĩa cho lớp.

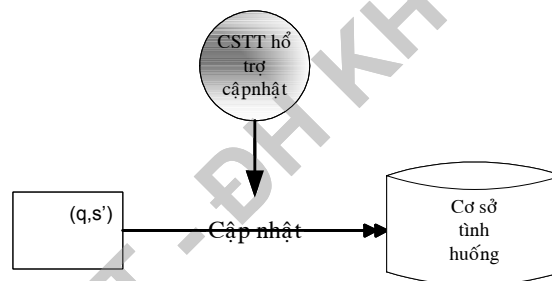
Trong hệ CBR hầu hết việc học được thực hiện theo hai cách khác nhau: tích lũy các tình huống mới và sửa đổi cấu trúc chỉ mục.

Tích lũy các tình huống mới nghĩa là cập nhật các tình huống bài toán mới được giải quyết vào cơ sở tình huống. Các tình huống bài toán mới này cung cấp thêm ngữ cảnh giải quyết vấn đề hay diễn dịch tình huống cho hệ suy diễn, và như vậy hệ suy diễn sẽ cải tiến năng lực của nó.

Sửa đổi cấu trúc chỉ mục tức là cập nhật thêm các chỉ mục cho phép hệ suy diễn cải tiến chất lượng của tiến trình nhớ, và như vậy các tình huống mà nó nhớ sẽ phù hợp với tình huống bài toán mới hơn.

Học trong hệ CBR cũng thể hiện ở việc sửa đổi cấu trúc tri thức hỗ trợ cho các tiến trình chọn lựa, truy tìm, thích nghi,...

Theo Smyth và Keane thì học trong CBR còn thể hiện ở việc xóa các tình huống nghèo nàn trong cơ sở tình huống.



2.2 Tình huống :

2.2.1 Tình huống là gì :

Các tình huống hình thành tri thức lõi trong suy diễn dựa tình huống. Thông thường một tình huống biểu diễn một kinh nghiệm đặc biệt và cụ thể đã bắt gặp trong quá khứ. Kinh nghiệm này có thể là một sự kiện, một đối tượng hay tình tiết. Những kinh nghiệm này là những thông tin rất cụ thể chứ không phải là những thông tin được tổng quát hóa. Các tình huống có thể

tiến hóa theo thời gian ví dụ như tình huống diễn tiến bệnh của bệnh nhân ,các tình huống có thể phản ánh thông tin mà khía cạnh thời gian không quan trọng hay không đáng kể,các tình huống có thể biểu diễn một tình tiết giải quyết vấn đề ví dụ như một dự án lập lịch.

Tuy nhiên không phải mọi thông tin đều tạo nên một tình huống hữu ích.

2.2.2 Các kiểu tình huống :

Thông thường các hệ CBR thường có 3 kiểu tình huống:tình huống truy vấn,tình huống cơ sở và tình huống mẫu

Tình huống truy vấn là tình huống bài toán mới,các thông tin trong tình huống này thường chưa được chuẩn hoá.Đặc biệt thông tin về kết luận và lời giải chưa có,những thông tin này chỉ đầy đủ sau khi đi qua tiến trình CBR

Tình huống cơ sở là tình huống đã được ghi nhận trong cơ sở tình huống.Các tình huống này đã được cung cấp đủ các thông tin cần thiết của một tình huống.Chúng là những mẫu tri thức đặc biệt cấu thành tri thức lõi của hệ CBR

Trong các mô hình phân loại các mẫu,thường các tình huống thường được phân loại trên cơ sở đánh giá tương tự của nó với các tình huống đặc biệt gọi là tình huống mẫu.Mỗi loại sẽ có một tình huống mẫu,tình huống này được xem là thuộc hoàn toàn về mẫu đó

2.2.3 Tình huống trong cơ sở tình huống :

Để phục vụ cho việc truy tìm ,cơ sở tình huống của các hệ CBR thường được tổ chức dưới dạng phân lớp.Điều này có nghĩa là các tình huống tương tự nhau theo một ý nghĩa nào đó sẽ được lưu trữ trong cùng một lớp trong cơ

sở tình huống. Ví dụ như trong chẩn đoán đông y thì các chứng bệnh phát nhiệt có triệu chứng lâm sàng tương tự nhau, như vậy các tình huống của các chứng bệnh này sẽ được lưu trữ trong cùng một lớp của cơ sở tình huống. Lúc này thay vì tập trung tới việc lấy tình huống nào trong cơ sở tình huống để tái sử dụng cho tình huống mới, tiến trình CBR quan tâm tới việc lưu trữ tình huống mới ở đâu trong cơ sở tình huống tức là trong lớp nào của cơ sở tình huống

Do việc phân lớp trong cơ sở tình huống, các tiến trình truy tìm và cập nhật trong chu trình CBR được thực hiện qua hai công đoạn. Trước hết tìm lớp sẽ được truy tìm hay cập nhật, sau đó thực hiện truy tìm hay cập nhật trong lớp này.

Vấn đề đặt ra ngay giờ ở đây là quan hệ giữa các lớp, quan hệ giữa các tình huống trong cùng một lớp, quan hệ giữa tình huống với lớp được thể hiện như thế nào trong cơ sở tình huống để việc truy tìm và cập nhật hiệu quả. Trong khuôn khổ của chương là biểu diễn và đánh giá tương tự giữa các tình huống ta chỉ quan tâm tới quan hệ giữa các tình huống trong cùng lớp và quan hệ giữa tình huống với lớp. Quan hệ giữa các lớp và chi tiết quan hệ giữa lớp và tình huống sẽ được xem xét trong phần xây dựng cơ sở tình huống.

2.2.4 Tích hợp tiến trình đánh giá tương tự và thích nghi trong tình huống :

Ta nhận thấy rằng mục tiêu của việc tổ chức tình huống là tạo cơ chế cho việc biểu diễn các tri thức của tình huống và dễ dàng hơn cho việc đánh giá tương tự cũng như thực hiện tiến trình thích nghi. Cơ chế hướng đối tượng cho phép ta tích hợp các tiến trình này như những phương thức của đối tượng.

Thông thường tương tự giữa hai tình huống được tính dựa theo sự tương tự của mỗi tính chất trong tập thuộc tính. Sự tính toán này đơn giản hay phức tạp, hiệu quả hay không hiệu quả phụ thuộc vào việc ta biểu diễn các tính chất tình huống như thế nào

Sự thích nghi thì khác hơn một chút, ngoài việc phụ thuộc vào biểu diễn tri thức tình huống; nó phụ thuộc mạnh mẽ vào phương pháp thích nghi và tri thức hỗ trợ cho việc thích nghi. Trong chương này ta không bàn sâu vào sự thích nghi

2.2.5 Các thành phần cơ bản của một tình huống :

Theo Kolodner thì tình huống được mô tả bởi 3 thành phần: mô tả tình huống, lời giải và kết luận

Như ta đã trình bày trong chương trước thì sau khi xử lý mô tả bài toán ta được một tập chỉ mục đó chính là mô tả tình huống

Lời giải tình huống là phương pháp giải quyết vấn đề, nó có thể là một mô tả ngắn kết quả của bài toán, hoặc có thể là mô tả các bước hay các hành động dẫn đến lời giải. Những mô tả này có thể được thể hiện dưới dạng text hay một đoạn video

Với hai thành phần này, mô tả tình huống và lời giải, hệ suy diễn dựa tình huống có thể giải quyết vấn đề mới bằng cách tìm tình huống phù hợp và sau đó thích nghi lời giải cũ của tình huống cho vừa với các tình tiết của bài toán mới.

Tuy nhiên với chỉ hai thành phần này thì không đủ để đảm bảo rằng hệ sẽ cho một lời giải tốt. Xác suất xuất hiện lời giải tốt và lời giải nghèo là như nhau. Do đó để hệ thống hiệu quả hơn và phù hợp hơn theo thời gian những lời giải của tình huống cần phải được đánh giá chất lượng trước khi lưu trữ

tình huống vào cơ sở tình huống. Thông tin đánh giá chất lượng này được biểu diễn bởi thành phần kết luận

2.2.6 Tính chất của tình huống :

Các tính chất của tình huống là một trong 3 phần hình thành nên tình huống .Ngoài việc được dùng để mô tả tình huống tập các tính chất này còn được dùng để đánh giá tương tự giữa các tình huống và thích nghi tình huống. Do đó việc biểu diễn các tính chất này như thế nào có ảnh hưởng rất lớn đến hiệu quả của hệ CBR

Tình huống có thể được mô tả bởi các tính chất chung, hoặc bởi các tính chất riêng. Ví dụ để mô tả một chiếc xe bus ta có thể dùng các tính chất chung là : phương tiện vận chuyển hành khách và đời của xe; hoặc chúng ta có thể dùng các tính chất riêng là: có động cơ, số cửa chính, số cửa phụ,...

Trong nhiều trường hợp các tình huống cũng có thể được mô tả bởi những tính chất mô tả hoặc bởi các tính chất trừu tượng. Các tính chất mô tả có thể được xác định một cách dễ dàng bởi bất cứ người nào. Tuy nhiên việc xác định các tính chất trừu tượng đòi hỏi rất nhiều tri thức chuyên ngành và thời gian xử lý. Việc xác định này thường chỉ được xác định bởi các chuyên gia của lĩnh vực.

Quan hệ giữa các tính chất cũng nên được xem xét biểu diễn. Trong nhiều nghiên cứu thực nghiệm người ta đã nhận thấy rằng con người thường dùng các quan hệ giữa các tính chất để phân loại các đối tượng. Quan hệ giữa các tính chất thường được đặc tả bởi sự phụ thuộc. Ví dụ các loài chim hay hót thường có kích thước nhỏ

Quan hệ giữa các tính chất cũng còn được phản ánh qua việc mô tả tính chất này thông qua tính chất khác; điều này có nghĩa là mô tả bởi tập các tính

chất khác. Việc biểu diễn các tính chất này đòi hỏi một phương tiện phức tạp và tinh vi hơn.

Một số tính chất có thể được biểu diễn trong nhiều tình huống. Các tính chất này thường được dùng để biểu diễn phân cấp tình huống, tức là từ những tính chất chung này tạo một tình huống trừu tượng khác. Cách biểu diễn này tạo một phương tiện kinh tế cho việc truy tìm tình huống. Tuy nhiên một tính chất không nên chung cho quá nhiều tình huống

Các tính chất riêng cho mỗi tình huống được dùng cho việc phân loại tình huống. Vì vậy các tính chất này chỉ nên dùng khi nó đóng góp đáng kể vào tiến trình phân loại.

Cuối cùng việc biểu diễn tính chất tình huống cũng nên chú ý tới vấn đề ngữ cảnh. Các giá trị tính chất của các thể hiện khác nhau của tình huống sẽ có ý nghĩa khác nhau

Nhận thấy rằng ngữ cảnh đóng vai trò quan trọng trong các tiến trình đánh giá tương tự và thích nghi

2.3 Biểu diễn tính chất của tình huống :

2.3.1 Biểu diễn cặp tính chất-giá trị :

Các tính chất mô tả tình huống thường được biểu diễn bằng cặp tính chất-giá trị. Những tính chất là những từ vựng được định nghĩa trước cùng với kiểu dữ liệu, miền dữ liệu của tính chất đó. Giá trị của tính chất sẽ được gắn với tính chất tại thời điểm suy diễn nếu tình huống đang xét là tình huống truy vấn. Còn tình huống là tình huống cơ sở thì nó đã có giá trị sẵn.

Có 3 dạng giá trị mà tính chất tình huống có thể nhận đó là dạng số, dạng symbolic, và dạng mờ. Dạng số dùng để biểu diễn các thông tin chính

xác, có thể xác định được giá trị rõ ràng. Dạng giá trị mờ biểu diễn các thông tin không đầy đủ và không chắc chắn.

Thường thì mỗi tính chất chỉ nhận một dạng dữ liệu, tuy nhiên có một số hệ như PERCEPT mỗi tính chất có thể nhận cả 3 dạng dữ liệu

2.3.2 Trọng số của tính chất :

Các tính chất biểu diễn tình huống là những tính chất nổi bật và đặc trưng cho tình huống. Tuy nhiên không phải tất cả các tính chất đều quan trọng như nhau đối với tình huống. Mức quan trọng của mỗi tính chất đối với tình huống được biểu diễn bằng trọng số của nó. Việc xác định trọng số này sẽ tùy thuộc vào quan điểm của từng người và từng ứng dụng. Thông thường việc xác định này mang tính thực nghiệm.

Một cách xác định trọng số là xác suất xuất hiện của tính chất đối với lớp tình huống. Ví dụ như khoảng 9/10 loài chim biết bay, như vậy ta có thể biểu diễn trọng số tính chất bay của loài chim là 0.9

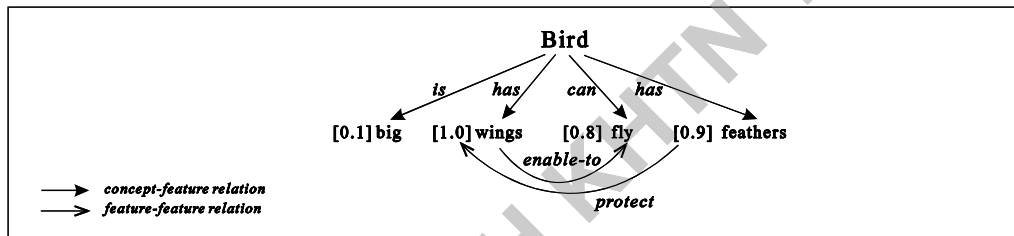
2.3.3 Biểu diễn định tính :

Biểu diễn định tính là dùng những tính chất nổi bật để mô tả lớp tình huống ví dụ như lớp chim sẽ được mô tả bởi các tính chất : có cánh, có thể bay, có lông vũ, là lớn. Các lớp tình huống sẽ được phân biệt dựa vào các tính chất lớp này có mà lớp kia không có. Thuận lợi của cách biểu diễn này là ở tính đơn giản của nó.

Khuyết điểm của cách biểu diễn này là không thể diễn tả và nắm bắt sự biến đổi về lượng của tính chất. Ví dụ như khi biểu diễn tính chất của con gà là lớn hệ không biết được trọng lượng của nó là bao nhiêu

Biểu diễn định tính thường dùng danh sách các tính chất để mô tả lớp tình huống. Tuy nhiên cấu trúc này không đủ để đặc tả tri thức lớp tình huống. Theo các nhà khoa học về nhận thức thì khi con người nghĩ về một đối tượng nào đó, ngoài việc liên tưởng tới các tính chất của nó, họ còn quan tâm tới quan hệ giữa các tính chất đó và dùng những tính chất này để phân lớp đối tượng. Những người thiết kế hệ CBR cũng thường dùng ý tưởng này kết hợp với cấu trúc danh sách tình huống để đặc tả lớp tình huống

Xem ví dụ sau :



Ta thấy rằng ngoài việc dùng các tính chất là lớn, có cánh, có thể bay, có lông vũ; các quan hệ cho phép giữa tính chất có cánh và có thể bay, quan hệ bảo vệ giữa có lông vũ và có cánh vũ cũng được dùng để mô tả lớp loài chim

Việc biểu diễn định tính không cho phép nắm bắt sự biến đổi về mặt ngữ cảnh

2.3.4 Biểu diễn định lượng :

Phương pháp biểu diễn này được dùng nhằm lượng hóa các tính chất mô tả lớp tình huống. Ý tưởng cơ bản của phương pháp là mỗi lớp tình huống được xem như một điểm trong không gian metric nhiều chiều. Mỗi chiều biểu

diễn cho một tính chất lớp tình huống. Tọa độ của lớp tình huống được xác định là tọa độ trọng tâm của các tình huống thuộc lớp đó

Thuận lợi của phương pháp này là khả năng nắm bắt sự biến đổi về lượng của tính chất. Phương pháp này cũng cho phép chúng ta biểu diễn định tính bằng cách giới hạn giá trị của tính chất trong miền giá trị $[a, b]$ và quy ước nếu giá trị tính chất đó bằng a thì lớp tình huống được xem như không có tính chất đó, nếu giá trị tính chất đó có giá trị bằng b thì xem như tính chất đó tồn tại.

Nhược điểm của phương pháp này là không có khả năng hỗ trợ rõ ràng quan hệ giữa các tính chất. Phương pháp này cũng không hỗ trợ việc nắm bắt những biến đổi ngữ cảnh.

2.3.5 Biểu diễn bằng kỹ thuật mờ :

Lý thuyết tập mờ :

Các chuyên gia cũng như những người không phải là chuyên gia thường mô tả tình huống một cách tự nhiên. Những tri thức mô tả này thường không rõ ràng và khó diễn đạt bằng các hệ thống logic truyền thống. Để diễn đạt sự không rõ ràng này, năm 1965 Lofti Zadeh đã đề xuất một phương tiện toán học là lý thuyết tập mờ.

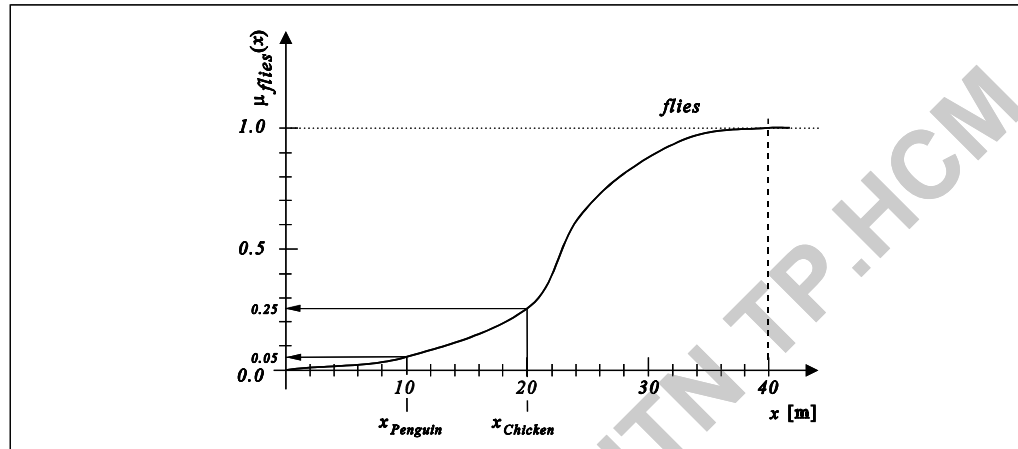
Cho một tập hợp A , theo quan điểm của Cantor thì một phần tử x hoặc sẽ thuộc về hoặc không thuộc về tập này. Theo quan điểm của tập mờ thì khác, khái niệm thuộc về được mở rộng nhằm phản ánh mức độ x là phần tử của tập A

Một tập mờ A được đặc trưng bằng hàm thành viên μ và cho x là một phần tử thì $\mu(x)$ phản ánh mức độ x thuộc về A . Hàm thành viên này được định nghĩa như sau

$$\mu_A(x):U_A \rightarrow [0,1]$$

Ví dụ:

Xem xét tính chất *bay* của lớp chim .Hàm thành viên của nó biểu diễn bởi đồ thị sau:



$$\mu_{Bay}(Gà)=0.25$$

$$\mu_{Bay}(Gà \text{ tây})=0.05$$

Theo đồ thị hàm thành viên thì những loài chim có khoảng cách bay >40m mới có mức thành viên đầy đủ trong tập mờ bay.

Biểu diễn tính chất bằng kỹ thuật mờ :

Phương pháp biểu diễn này xem mỗi tính chất của lớp tình huống như là tập mờ.Như vậy việc biểu diễn tính chất phải kèm theo hàm thành viên cho mỗi giá trị của tính chất.Tính chất của tình huống sẽ được cung cấp giá trị thông qua hàm thành viên

Ta nhận thấy rằng phương pháp biểu diễn này có thể nắm bắt được việc biểu diễn định tính và biểu diễn định lượng

Để thấy các giá trị của tính chất được định lượng bởi hàm thành viên của nó, các giá trị thu được là một con số cụ thể trong đoạn $[0,1]$. Và như vậy nó nắm bắt được sự thay đổi về lượng của giá trị tính chất

Xem lại hàm thành viên của tính chất bay ở trên. Ta có thể biểu diễn sự có mặt hay không có mặt của tính chất bay như sau

$$\mu(\text{có})=1$$

$$\mu(\text{không})=0$$

Điều này thể hiện khả năng biểu diễn định tính của phương pháp biểu diễn bằng kỹ thuật mờ

2.4 Một số phương pháp đánh giá tương tự :

Tương tự giữa các tình huống sẽ phụ thuộc vào tương tự của các tính chất do đó phần này ta chỉ quan tâm tới việc đánh giá tương tự của các tính chất

2.4.1 Đánh giá tương tự dựa trên khoảng cách metric :

Ta nhận thấy rằng các tính chất có khoảng cách metric càng nhỏ thì độ tương tự giữa chúng càng lớn. Trong CBR việc đánh giá tương tự được dùng để so sánh và tìm tình huống tương tự nhất. Do đó thay vì đi so sánh độ tương tự ta có thể đi so sánh khoảng cách metric

Trường hợp tính chất nhận giá trị số ta thường dùng các khoảng cách metric sau

Cho v_1, v_2 là hai giá trị số , khoảng cách Metric thường dùng là :

$$d(v_1, v_2) = \frac{|v_1 - v_2|}{\text{Max}(v_1, v_2)}$$

$$d(v_1, v_2) = \frac{|v_1 - v_2|}{|u|}$$

$$d(v_1, v_2) = \frac{v_1 - v_2}{v_1 + v_2}$$

Nếu giá trị tính chất là các giá trị symbolic rời rạc, ta có khoảng cách được định nghĩa như sau

$$d_{x,x} = 0$$

$$d_{x,y;x \neq y} = 1$$

2.4.2 Dùng cây phân loại :

Phương pháp này thường dùng cho việc biểu diễn các tính chất nhận giá trị symbolic. Để định nghĩa tập giá trị của thuộc tính ta dùng cây phân loại. Cây phân loại là cây n nhánh trong đó các node biểu diễn các giá trị symbolic. Cây phân loại không chỉ đơn giản là biểu diễn giá trị thuộc tính mà nó còn biểu diễn quan hệ giữa giá trị ở các node thông qua vị trí của chúng trong cây phân loại. Quan hệ này sẽ diễn tả tri thức về tương tự giữa các symbol trong cây phân loại.

2.4.3 Tiếp cận của Vargas & Bourne :

Trong khi xây dựng hệ CBR chẩn đoán thiết bị, Vargas và Bourne đã phát triển cái gọi là fuzzy scale (tạm dịch là thực thể mờ). Mỗi thực thể mờ được kết hợp với một tính chất tương ứng. Thực thể mờ cho ta nhiều thuận tiện trong việc tinh chỉnh tri thức, tuy nhiên ở đây ta chỉ quan tâm đến việc đánh giá tương tự giữa chúng.

Cũng như các phương pháp biểu diễn bằng kỹ thuật mờ khác thực thể mờ ở đây cũng được thể hiện bằng hàm thành viên $\mu(x)$. Việc đánh giá tương tự giữa hai giá trị a,b được thực hiện như sau :

(1) A, B là Symbolic

Nếu A, B có thể so khớp được thì

$$\text{Sim}(A,B) = 1 \text{ nếu } A = B$$

$$\text{Sim}(A,B) = 1 - (K/u) \text{ nếu } A \neq B$$

Ở đây n : số đoạn trong thực thể mờ

K : số đoạn phân chia A bởi B

Nếu A,B không thể so khớp được thì

$$\text{Sim}(A,B) = 1 \text{ nếu } A = Bm$$

$$\text{Sim}(A,B) = 0 \text{ nếu } A \neq B$$

(2) a , b là giá trị số

$$\text{Sim}(a,b) = 1 - |\mu_A(a) - \mu_A(b)|$$

A : Symbolic

(3) A là Symbolic và b là số

$$\text{Sim}(A,b) = 1 - |1 - \mu_A(b)| = \mu_A(b)$$

2.4.4 Tiếp cận của Werner Dubitzky :

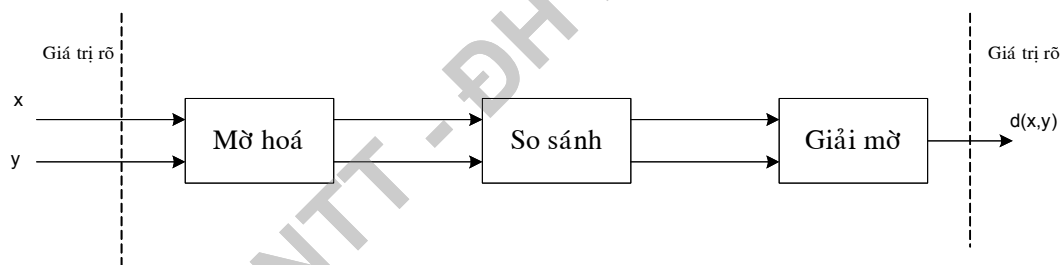
Tiếp cận này dựa trên khái niệm về tính chất đa hình. Tính chất đa hình được Werner Dubitzky (Đại học Ulster, Vương quốc Anh) đề nghị trong khi xây dựng mô hình suy diễn dựa tình huống PERCEPT. Tính chất đa hình bao gồm tính chất nguyên và tính chất phức. Tính chất nguyên được biểu diễn như ý nghĩa mà tri thức trong cơ sở tình huống được diễn tả và xử lý. Tính chất phức được xây dựng từ các tính chất đơn và bản thân các tính chất phức

Tính chất đa hình (tính chất nguyên) cho phép biểu diễn 3 dạng giá trị tính chất đó là dạng số, dạng linguistic, dạng dự báo mờ. Điều này cung cấp cho người dùng một cơ chế linh động để diễn tả các mức không chắc chắn khác nhau. Tính chất đa hình còn cho phép tạo một cơ chế đánh giá tương tự giữa các dạng giá trị khác nhau

Tính chất phức cung cấp cơ chế nhóm các tính chất lại với nhau và nắm bắt sự tương quan giữa các tính chất. Việc kết hợp các tính chất lại với nhau thành tính chất phức làm cho việc thu thập tri thức trở nên đơn giản hơn bởi vì nó giảm số đơn vị tri thức cần được nhận thức

Tính chất phức được xem như là tính chất trừu tượng ,nó trừu tượng hóa các tính chất tạo thành tính chất phức. Điều này có nghĩa là nó đóng gói tập các tính chất đơn giản hơn.

Đánh giá tương tự được thực hiện theo quy trình như hình vẽ sau



Các giá trị sẽ được mờ hoá để so sánh, sau đó kết quả sẽ được làm rõ trở lại. Phương pháp này có nhiều ưu điểm nhưng quá rắc rối nên không trình bày ở đây.

2.5 Cơ sở tình huống :

2.5.1 Cơ sở tình huống là gì :

Cơ sở tình huống là thư viện các tình huống được tổ chức và lưu trữ theo một tiêu chí nào đó. Cơ sở tình huống biểu diễn khối lượng tri thức của hệ suy diễn dựa tình huống. Tri thức này sẽ tiến hoá trong suốt quá trình hoạt động của hệ CBR.

Trong các hệ chuyên gia, thông thường cơ sở tri thức được tổ chức phù hợp với cơ chế suy diễn. Trong hệ CBR, cơ sở tri thức của nó chính là cơ sở tình huống, nó phải được tổ chức phù hợp với cơ chế suy diễn CBR đó là hai tiến trình truy tìm và cập nhật. Tiến trình truy tìm sẽ tìm trong cơ sở tình huống một tình huống phù hợp nhất với tình huống hiện tại, còn tiến trình cập nhật sẽ tìm vị trí phù hợp nhất trong cơ sở tình huống để lưu trữ tình huống bài toán mới. Trong các cơ sở tình huống nhỏ thì việc truy tìm và cập nhật khá đơn giản, nhưng với cơ sở tình huống lớn thì để hỗ trợ tốt cho hai tiến trình này thì cơ sở tình huống đòi hỏi phải được tổ chức tinh vi hơn.

Bên cạnh hỗ trợ cơ chế suy diễn, cơ sở tình huống còn phải được tổ chức làm sao để hỗ trợ các loại tri thức cần thiết cho lĩnh vực đó là tri thức cụ thể của mọi tình huống, tri thức tổng quát của lĩnh vực tri thức ngữ cảnh, tri thức suy diễn,....

Việc tích hợp toàn bộ các kiểu tri thức này làm cho việc tổ chức cơ sở tình huống trở nên nặng nề và dường như mâu thuẫn với mục đích hỗ trợ cơ chế suy diễn. Hai mục tiêu này không thể đạt được đồng thời, cơ sở tri thức càng phức tạp thì việc suy diễn càng khó khăn và ngược lại.

Mặc dù hai mục tiêu của việc tổ chức cơ sở tình huống trái ngược nhau nhưng vai trò của chúng là quan trọng như nhau trong hệ CBR. Việc điều phối hai mục tiêu này phụ thuộc mạnh mẽ vào lĩnh vực mà hệ CBR phục vụ. Thông thường khi xây dựng hệ CBR thì tiến trình truy tìm và cập nhật được tích hợp vào cơ chế biểu diễn tri thức.

2.5.2 Một số vấn đề cần chú ý khi xây dựng cơ sở tình huống :

2.5.2.1 Tích hợp tri thức cơ bản và tri thức đặc biệt :

Mỗi lĩnh vực đều được xây dựng trên một cơ sở lý thuyết nào đó cơ sở lý thuyết này mang tri thức chung , tổng quát cho lĩnh vực đó và tri thức này ít thay đổi theo thời gian .Ta gọi tri thức này là tri thức cơ bản của lĩnh vực .

Ta cũng biết rằng hệ CBR hoạt động dựa vào việc thích nghi những tình huống đặc biệt trong quá khứ. Các tình huống này chứa những tri thức đặc biệt ,riêng lẽ. Chỉ dựa trên những tri thức riêng lẽ và đặc biệt này CBR cũng đã đạt được nhiều thành tựu. Tuy nhiên theo Aamodt và Plaza thì hệ CBR nên tích hợp tri thức cơ bản .Việc này cũng phù hợp với những nghiên cứu về cách giải quyết vấn đề của con người .Con người luôn tích hợp tri thức cơ bản với kinh nghiệm của mình khi gặp một vấn đề gì đó cần giải quyết .Ta sẽ đưa ra một số ưu điểm khi tích hợp các tri thức cơ bản trong CBR .

Biểu diễn tri thức cơ bản của lĩnh vực trong hệ CBR có thể được dùng để cải tiến chất lượng và sự linh động của nhiều tiến trình con . Đó là đánh giá tình huống , truy tìm tình huống ,so khớp và chọn lựa , thích nghi, kiểm tra ,sửa đổi và học .

Tri thức cơ bản cũng hướng dẫn chọn lựa các chỉ mục và tổ chức cơ sở tình huống , đồng thời cũng hỗ trợ tiến trình thu thập tri thức.

Như vậy việc tích hợp tri thức cơ bản vào hệ CBR (tức là tích hợp với tri thức đặc biệt) là vấn đề nên được quan tâm khi xây dựng cơ sở tình huống .

2.5.2.2 Biểu diễn tri thức không đầy đủ và không chắc chắn :

Trong các hệ CBR , tri thức không chắc chắn xuất hiện khắp tất cả các thành phần của nó trong các đặc trưng được dùng để mô tả tình huống, trong các quan hệ giữa các tình huống và các đặc trưng của nó , trong các phương pháp đánh giá tương tự và trong cách mà cơ sở tình huống được tổ chức và truy cập. Những tri thức không chắc chắn này xuất hiện bởi những lý do sau :

Thông tin không thể định lượng ví dụ như tuổi của bệnh nhân dễ dàng xác định nhưng điều kiện của họ không thể định lượng

Thông tin không đầy đủ do đo lường không chính xác , ví dụ như huyết áp thường được cung cấp dưới dạng “khoảng 145 mmHg” chứ không phải là “145mmHg”

Thông tin khó kiểm , đó là những thông tin cần phải có chi phí quá cao để kiểm nó

Toán học đã cung cấp hai phương tiện để biểu diễn những tri thức này đó là lý thuyết xác suất và lý thuyết tập mờ . Tuy nhiên để biểu diễn điều này trong hệ CBR là một vấn đề không giản và đây là một vấn đề nên nhắm đến khi xây dựng hệ CBR.

2.5.2.3 Vấn đề chỉ mục :

Chỉ mục là tập các tính chất mô tả tình huống , những tính chất này được xác định trước .Các tính chất mô tả mỗi tình huống là tập con của tập chỉ mục và nó phân biệt (không phải tất cả) với các tình huống khác .

Tập chỉ mục cũng được dùng để so khớp tình huống trong các tiến trình đánh giá tương tự, truy tìm và cập nhật. Các tình huống sẽ được đánh chỉ mục như thế nào để truy tìm hiệu quả nhất , đó là vấn đề lớn của CBR.

Những người phát triển hệ CBR đã nhận ra rằng tập chỉ mục tối thiểu phải thoả các điều liên sau :

Các chỉ mục phải biểu diễn kèm theo mức quan trọng của nó trong tình huống và như vậy những chỉ mục quan trọng hơn sẽ có ảnh hưởng hơn trong so khớp tình huống .

Đảm bảo rằng việc so khớp tình huống phải được mô tả bởi tập chỉ mục cần và đủ để xác định nó .

Các tình huống được chọn sẽ dựa trên sự so khớp chỉ mục này

Qua đó ta thấy rằng vấn đề chỉ mục liên quan đến toàn bộ việc thiết lập CBR, bao gồm cả tổ chức cơ sở tình huống và các tiến trình truy cập.

Ngoài những điều kiện trên để được xem là tập chỉ mục tốt nó cần phải thêm các điều kiện sau :

Tập chỉ mục phải đủ trừu tượng để truy tìm tình huống phù hợp trong nhiều trường hợp khác. Tuy nhiên quá trừu tượng, chỉ mục sẽ có sự hướng truy tìm quá nhiều tình huống .

Tập chỉ mục phải đủ cụ thể để nhận ra các tình huống trong tương lai. Tuy nhiên quá cụ thể sẽ trở nên ít có ích trong nhiều trường hợp, bởi nó hiếm khi dẫn đến tình huống .

Tóm lại để được một tập chỉ mục tốt đòi hỏi quá nhiều nỗ lực, điều này luôn làm đau đầu các nhà thiết kế hệ CBR. Tuy nhiên, nó luôn là mối quan tâm đầu tiên khi thiết kế cơ sở tình huống .

2.5.2.4 Ngữ cảnh :

Ngữ cảnh được dùng khá thường xuyên trong các hệ trí tuệ nhân tạo, chúng được dùng để phân biệt hay phân rã các mục dựa trên các tập mô tả đã cho, hay nhấn mạnh vào các đặc trưng hay các kết hợp đặc trưng nhất định trong mô tả. Tổ chức tri thức xung quanh ngữ cảnh có thể hữu ích cho việc chọn lựa tri thức phù hợp và giải thích tri thức . Trong CBR tri thức ngữ cảnh

nên được đưa vào khi xác định sự phù hợp giữa các tình huống đã được lưu trữ trong hệ thống . Biểu diễn rõ ràng thông tin ngữ cảnh tập trung cho việc tích lũy tình huống , đánh giá tương tự , thực hiện thích nghi các tình huống .

Trong CBR ngữ cảnh sẽ được sử dụng chủ yếu trong việc làm tăng hiệu quả các tiến trình truy cập . Ngữ cảnh cũng làm tăng tính linh động của việc biểu diễn tri thức và tăng tính chính xác của việc truy cập tình huống .

Thực vậy ta nhận thấy rằng một người châu Á cao 1.8m gọi là rất cao nhưng người châu Âu chỉ được xem là cao . Nếu ta biểu diễn tri thức về người có phân biệt ngữ cảnh này thì sẽ rõ ràng hơn .

Nói tóm lại ngữ cảnh là mối quan tâm trong việc xây dựng hệ cơ sở tình huống.

2.5.2.5 Vấn đề thu thập tri thức :

Thu thập tri thức luôn là vấn đề của các hệ trí tuệ nhân tạo . Công việc này phụ thuộc vào chuyên gia, mức độ phụ thuộc tùy thuộc vào ta đang dùng phương pháp trí tuệ nhân tạo gì.

Hệ cơ sở tình huống tạo thuận lợi cho việc thu thập tri thức vì chúng chỉ là mô tả những tình huống cụ thể đặc biệt. Tuy nhiên công việc này cũng nên được xét đến khi xây dựng cơ sở tình huống ,việc hỗ trợ thu thập tri thức sẽ làm cho hệ thành công hơn.

Đó chỉ là những quan tâm cơ bản nhất cần quan tâm khi thiết kế cơ sở tình huống .Sẽ còn rất nhiều điều cần làm để hệ CBR có thể đạt được tính hiệu quả .Những vấn đề này sẽ xuất hiện trong suốt toàn bộ luận văn này.

2.6 Một số mô hình cơ sở tình huống :

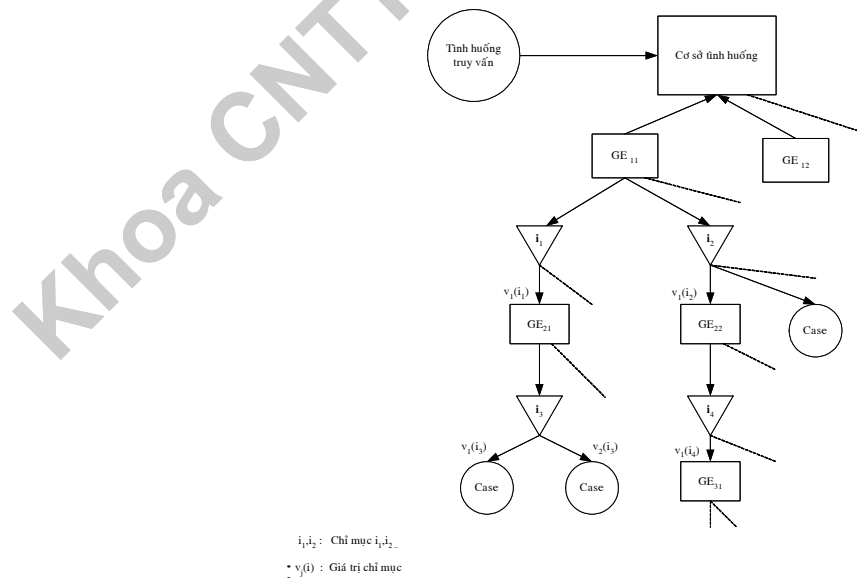
2.6.1 Tiếp cận cơ sở MOP:

MOP là viết tắt của cụm từ tiếng Anh Memory Organisation Package , nghĩa là gói tổ chức bộ nhớ. Khái niệm này được Schank (Đại học Cambridge , Vương quốc Anh) giới thiệu trong khi mô hình việc học và suy diễn xung quanh bộ nhớ động. Nhiều mô hình đã được xây dựng và phát triển trên cơ sở khái niệm này với các tên khác nhau E-MOP, TOP, T-MOP, i-MOP, S-MOP.....

Ý tưởng cơ bản của MOP là xây dựng một cơ sở tình huống phân cấp , cho phép cùng một thuật toán chung cho việc truy tìm và cập nhật , cung cấp một cơ chế thay đổi động một phần của cấu trúc tổ chức cơ sở tình huống.

Một cơ sở tình huống trên cơ sở MOP sẽ gồm 3 kiểu đối tượng : *norms*, *case* và *indices* . *Case* là các tình huống cơ sở của lĩnh vực bài toán . *Norms* biểu diễn những tính chất chung của các tình huống được tổ chức dưới một cấu trúc tổng quát hơn là GE . *indices* là các tính chất của tình huống cùng một GE dùng để phân biệt nhau. Ta nhận thấy rằng GE được hình thành từ tình huống và GE khác .

Mô hình này được minh họa bởi hình vẽ sau



Nhận thấy rằng cấu trúc của cơ sở tình huống dựa trên MOP thể hiện một cây phân rã. Việc truy tìm và cập nhật sẽ dựa vào việc duyệt cây phân rã này. Mô tả tình huống bài toán mới được dùng để so khớp với tập các tính chất của GE hay tình huống để tìm GE hay tình huống khớp nhất. Nếu kết quả là GE thì những tính chất còn lại của tình huống bài toán mới (tính chất không phản ánh trên GE) sẽ được so khớp để tìm tình huống hay GE bên dưới nó. Cứ như vậy ta sẽ tìm được tình huống phù hợp nhất (khớp nhất) cần tìm.

Tương tự tiến trình truy tìm, tiến trình cập nhật sẽ tìm tình huống phù hợp nhất. Tình huống sẽ được kết hợp với tình huống bài toán mới để tạo ra một GE mới, lúc này hai tình huống này sẽ được chỉ mục bên dưới GE mới.

Ưu điểm của phương pháp tổ chức này là:

- + Cơ sở tình huống được cập nhật tự động
- + Dùng chung một thuật toán truy tìm và cập nhật.
- + Tri thức được tổng quát trong GE
- + Tích hợp cơ chế biểu diễn tri thức vào tiến trình truy tìm và cập nhật

Nhược điểm của phương pháp tổ chức này là:

- + Việc cập nhật tự động là một ưu điểm nhưng nó cũng tạo nhược điểm là cơ sở tình huống tăng trưởng nhanh và dư thừa chỉ mục.

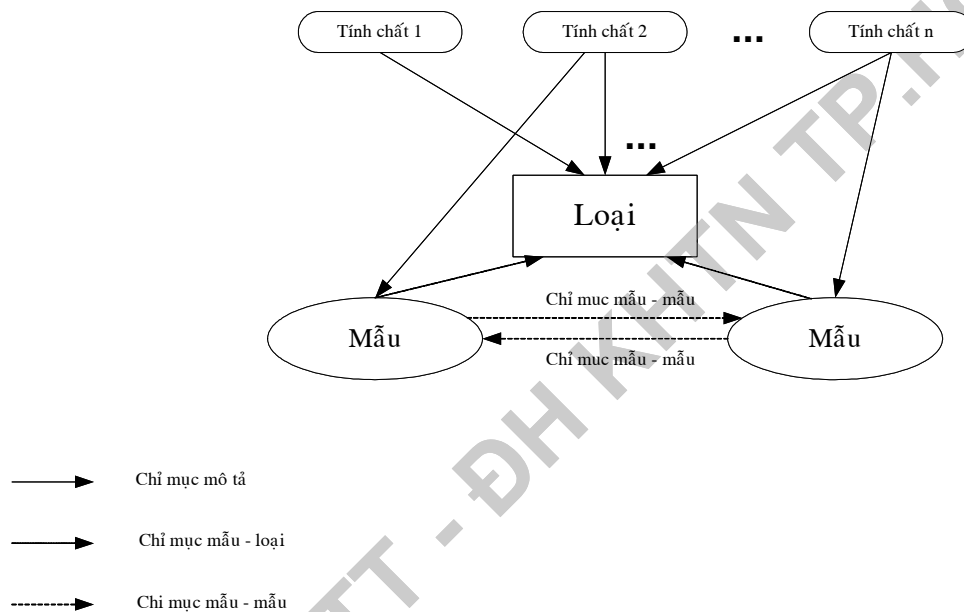
- + Chưa có cơ chế tích hợp rõ ràng nhiều kiểu kiến thức

2.6.2 Mô hình phân loại các mẫu :

Mô hình phân loại các mẫu được Bareiss và Porter đề xuất khi xây dựng hệ suy diễn dựa tình huống PROTOS. Trong mô hình này tình huống

được gọi là mẫu và cơ sở tình huống được tổ chức dựa trên cơ sở phân loại các mẫu này. Các mẫu tương tự nhau theo một tiêu chí nhất định sẽ được tổ chức trong cùng một loại và các loại này được tổ chức trong một cấu trúc mạng có quan hệ ngữ nghĩa.

Một phần của cơ sở tình huống tổ chức theo mô hình phân loại các mẫu được biểu diễn như hình vẽ sau :



Các chỉ mục mô tả là tập các tính chất mô tả mẫu và loại. Tập tính chất này ngoài việc mô tả loại còn được dùng phân biệt loại này với loại kia.

Chỉ mục mẫu loại là tập các tính chất của mẫu không phản ánh trên loại. Hai mẫu trong cùng một loại còn được phân biệt với nhau qua chỉ mục mẫu-mẫu. Đó là các tính chất mẫu này mà mẫu kia không có.

Các mẫu được tổ chức trong một loại dựa vào mức hành viên của nó trong loại và các mẫu được sắp xếp theo thứ tự mức thành viên này. Việc truy tìm trong mô hình này chính là việc phân loại tình huống bài toán mới.

Ta nhận thấy rằng các loại được phân chia và tổ chức cẩn thận trong mô hình. Các loại này sẽ không thay đổi trong suốt quá trình tiến hoá của việc suy diễn. Vì vậy mô hình chỉ phù hợp với các lĩnh vực ít biến đổi.

Việc tổ chức các loại trong cấu trúc mạng có quan hệ ngữ nghĩa cũng làm thuận lợi cho việc tích hợp nhiều kiểu tri thức. Tuy nhiên việc biểu diễn theo mô hình mạng này cũng là hạn chế với những lĩnh vực mà số loại là lớn bởi sự hạn chế của nó trong việc truy tìm.

2.6.3 Mô hình tình huống trừu tượng :

Mô hình này được xây dựng trên cơ sở nhận xét của Bergman và Wilke về lợi ích của tình huống trừu tượng :

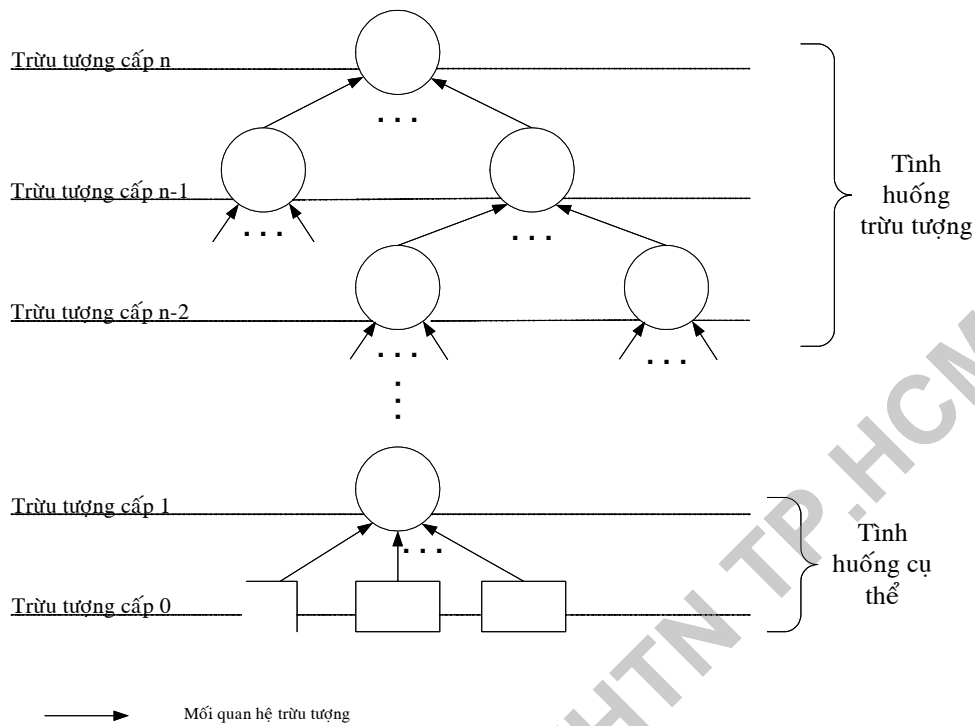
Tình huống trừu tượng có thể làm giảm độ phức tạp của các tình huống, và như vậy nó làm giảm công sức đánh giá tương tự và thích nghi

Tình huống trừu tượng được dùng để thay thế cho các tình huống cụ thể, điều này dẫn đến cơ sở tình huống được chia thành các phần nhỏ hơn làm thuận tiện cho việc truy tìm và cập nhật

Tình huống trừu tượng được dùng như là chỉ mục cho các tình huống ở mức trừu tượng thấp hơn, điều này có thể giúp cho việc truy tìm hiệu quả hơn

Sự trừu tượng làm tăng tính linh động của tiến trình tái sử dụng bởi vì các lời giải trừu tượng phủ một phần lớn hơn không gian bài toán

Một trong những hệ suy diễn dựa tình huống ứng dụng mô hình tình huống trừu tượng là hệ PARIS. Cấu trúc cơ bản của hệ này như hình vẽ sau :



Trong mô hình này tình huống ở mức trừu tượng càng cao thì mức chi tiết của nó càng giảm. Điều này có nghĩa là tình huống ở mức trừu tượng cao hơn sẽ ít tính chất hơn, ít quan hệ hơn, ít ràng buộc hơn, ... Tuy nhiên các tình huống trừu tượng vẫn nắm bắt các đặc trưng quan trọng.

Ta nhận thấy rằng tổ chức của mô hình này thể hiện 1 cây phân rã, và như vậy tiến trình truy tìm và cập nhật được thực hiện thông qua sự duyệt cây.

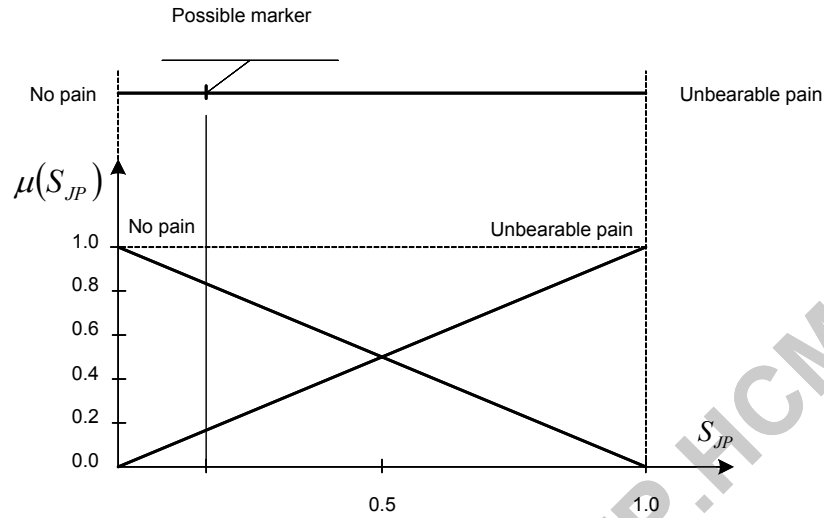
2.6.4 Tiếp cận dùng kĩ thuật mờ :

Các mô hình này dùng lý thuyết tập mờ và logic mờ trong việc biểu diễn tình huống và cơ sở tình huống để tích hợp tri thức không chắc chắn. Các hệ suy diễn nổi bật là ARC, MEPOS, PERCEPT, ...

ARC là hệ suy diễn dựa tình huống chính đoán viêm phổi và liệu pháp được Plaza và De Mántaras (Viện nghiên cứu trí tuệ nhân tạo Tây Ban Nha) phát triển năm 1980. Hệ này được xây dựng trên mô hình phân loại theo mẫu. Các mẫu được phân loại để chia nhỏ không gian cơ sở tình huống. Các loại được mô bởi những tính chất cần và đủ của các mẫu thuộc về nó. Như đã giới thiệu trong phần mô hình phân loại theo mẫu, hệ ARC phải có các chỉ mục mô tả, chỉ mục mẫu-loại chỉ mục mẫu-mẫu. Ngoài ra ARC còn đưa vào 1 khái niệm chỉ mục mới đó là chỉ mục near-miss. Chỉ mục này được dùng để trở đến các tình huống near-miss, đó là các tình huống được gán loại sai.

Điều đáng quan trọng hệ này là mức thành viên của các mẫu đối với loại của hệ ARC được biểu diễn bằng giá trị mờ. Cơ chế này cho phép chúng ta biểu diễn mức thành viên của tình huống trong tập lời giải bằng các thuật ngữ mờ hồ hoặc giá trị không chắc chắn.

Hệ Mepos là hệ suy diễn dựa tình huống mờ trong lĩnh vực gây tê được Petersen phát triển năm 1997. Điểm nổi bật trong hệ này là sử dụng cái gọi là visual analogous scale (tạm gọi là thực thể tương tự trực quan) để nắm bắt tri thức không chắc chắn và biểu diễn tình huống cũng như hỗ trợ đánh giá tương tự. Cơ chế hoạt động của visual analogous scale được minh họa như sau (ví dụ minh họa của chính Petersen) :



Phần trên của hình biểu diễn visual analogous scale của bệnh nhân và phần dưới mô tả tập mờ biểu diễn 2 tính chất nopain và unbearable pain .

Để đánh giá tương tự giữa hai tình huống trước hết xây dựng các luật mờ . Xem ví dụ sau với p là bệnh nhân :

$$[[(age(p) = old) \wedge (\max(dose(p)) = low)] \rightarrow [\max(selfJudgedpain(p)) = nopain]]$$

$$[[(age(p) = young) \wedge (\max(dose(p)) = normal)] \rightarrow [\max(selfJudgedpain(p)) = nopain]]$$

Sau đó đánh giá tương tự sẽ dựa vào mức thành viên của tình huống trên các luật này :

$$\sin(case_i, case_j) = \max_K (\min(\mu_{ruleK}(case_i), \mu_{ruleK}(case_j)) faktor_K)$$

Mô hình PERCEPT được Werner Dubitzky (Đại học Ulter , Vương quốc Anh) giới thiệu trong luận văn tiến sĩ của mình .Trong hệ này ông đề xuất ra hai khái niệm :tính chất nguyên đa hình và tính chất phức đa hình .Tính chất đa hình ở đây có nghĩa là có thể nhận nhiều dạng dữ liệu khác nhau bao gồm giá trị số ,giá trị thuật ngữ linguistic , giá trị dự báo mờ. Tính chất đa hình còn tạo cơ chế cho việc đánh giá tương tự giữa nhiều dạng giá trị khác

nhau . Để làm được điều này giá trị tính chất được mờ hoá thông qua kỹ thuật mờ . Ta sẽ xem kỹ hơn mô hình PERCEPT trong các phần sau .

2.6.5 Mô hình PERCEPT :

Mô hình PERCEPT dường như là mô hình tốt nhất trong các mô hình đã giới thiệu trên .Nó cho phép biểu diễn nhiều kiểu tri thức và đưa ra một framework cho việc xây dựng hệ CBR .Tổ chức của mô hình này cũng bao gồm sự kết hợp nhiều kỹ thuật khác nhau .

Trước hết PERCEPT tổ chức trên cơ sở phân loại các mẫu , các mẫu được biểu diễn trong một loại tùy thuộc vào mức thành viên của nó thuộc loại đó.Mức thành viên này được xác định bằng các đánh giá tương tự giữa nó và một tình huống mẫu,tình huống mẫu này là tình huống được xem là thuộc hoàn toàn vào lớp đang xét và được xây dựng một cách cẩn thận khi thiết kế

Trong PERCEPT cũng biểu diễn tổ chức phân cấp thông qua cơ chế trừu tượng hoá các lớp tình huống .Tri thức suy diễn được thể hiện thông qua cơ chế message của mô hình hướng đối tượng.Các lớp tình huống ở mức trừu tượng khác nhau được kết nối với nhau thông qua quan hệ thừa kế và quan hệ kết hợp.Các đối tượng tình huống chỉ được găng với các lớp tình huống ở mức trừu tượng thấp nhất thông qua quan hệ thể hiện và kết hợp

CHƯƠNG 3: XÂY DỰNG HỆ CBR CHẨN ĐOÁN BỆNH TIM

3.1 CBR trong y khoa

3.1.1 Các đặc trưng của y khoa

Y khoa là một lĩnh vực rất phức tạp và khác biệt so với các lĩnh vực khác. Điểm khác biệt cơ bản nhất là mối quan hệ tương hỗ giữa lý thuyết và thực hành. Đối tượng của y khoa là những bệnh nhân, những con người thực sự. Đó là những thực thể sống được tổ chức rất phức tạp về mặt sinh học kèm theo đó là hàng loạt những quá trình sống tác động qua lại, ảnh hưởng lẫn nhau. Những quá trình này luôn bị chi phối bởi điều kiện môi trường như: xuất hiện sự đối kháng mới, bệnh tật, mầm bệnh và nguồn bệnh...

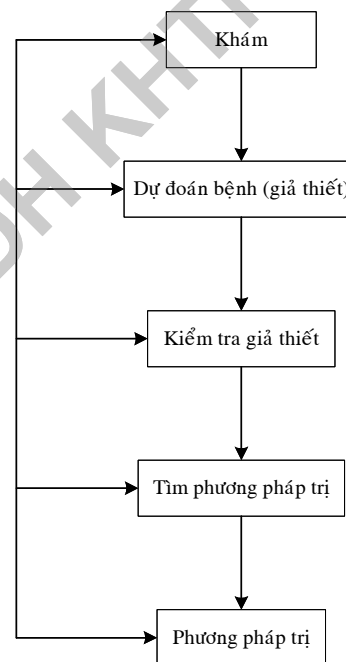
Những kiến thức về y khoa thường rất phức tạp. Để tìm ra những kiến thức mới, phương pháp truyền thống là dựa trên sự mô tả các ca bệnh, tập hợp những ca bệnh và các nghiên cứu tĩnh khác (thí nghiệm). Vì thế nó gần như tương phản với kiến thức của các lĩnh vực khác.

Kiến thức trong y khoa có tính tập trung cao. Tập hợp những ca bệnh tích lũy lại, kiến thức y khoa được sắp xếp trong những danh sách riêng và

những nguồn như hồ sơ bệnh án, nhật kí y khoa, cơ sở dữ liệu đề tài nghiên cứu, ...

Người hành nghề y khoa (bác sĩ) thường bị tràn ngập trong núi dữ liệu khổng lồ ngay cả khi chỉ với một bệnh nhân: hàng trăm thông số hoá dược, hàng tá hình X-quang và những tài liệu mẫu. Mặt khác những thông tin đó có thể trái ngược nhau vì thực nghiệm hoàn toàn khác với thí nghiệm. Đôi khi họ phải đau đầu khi gặp những triệu chứng mơ hồ của một bệnh nhân.

Họ luôn phải làm việc trong trạng thái căng thẳng. Họ phải chọn lượng thuốc thích hợp sao cho nó có được sự tác động tới người bệnh như mong muốn trong một thời gian ngắn. Thông thường ta làm việc theo sơ đồ sau:



Lý thuyết là như vậy nhưng thực tế các bác sĩ kinh nghiệm thường không làm như vậy. Họ không tuân theo sơ đồ trên. Khi gặp một ca bệnh với những triệu chứng quen thuộc, họ lập tức nhớ đến những ca bệnh tương tự về phương pháp trị liệu của căn bệnh đó. Từ đó họ có thể đưa ra những phương

pháp trị bệnh chắc chắn với xác suất đạt hiệu quả cao . Đây là cách giải quyết chiến lược và cũng tạo nên một hệ kiến thức cơ bản dựa trên những luận điểm sau :

Triệu chứng bệnh là bước khởi đầu để đưa ra các giả thiết

Phương pháp chữa trị để làm giảm sự nguy hiểm

Tính thường xuyên của bệnh có từ kinh nghiệm cá nhân và các học thuyết

Quyết định của bác sĩ đối với từng ca bệnh là hết sức quan trọng . Nó có thể ảnh hưởng đến tình trạng của bệnh nhân cũng như những hậu quả để lại sau này. Vì thế việc lựa chọn phương pháp điều trị là hết sức cần thiết .CBR (Case-Based Reasoning) là phương pháp thích hợp nhất để bác sĩ có thể chẩn đoán bệnh , rồi từ đó tìm ra cách điều trị

3.1.2 Thuận lợi của CBR trong y khoa

Suy diễn dựa tình huống phù hợp trong y khoa bởi những lý do sau:

Nhận thức: các hệ trên cơ sở tình huống tương ứng với việc tăng trí tuệ, sự trừu tượng và nhận thức của bác sĩ. Thường chúng dùng các kinh nghiệm từ chẩn đoán lâm sàng để đối xử với tình huống hiện tại một cách hiệu quả hơn, để chẩn đoán các ngoại lệ tương tự, ... Schank và các cộng sự đã nói rằng sự sáng tạo của con người là khả năng dựa trên các kinh nghiệm

Kinh nghiệm rõ ràng: các hệ CBR ít dùng tri thức cô đọng (ví dụ như các luật) mà mở rộng thay thế, thay đổi và quên tri thức bởi sự tích hợp các tình huống vào cơ sở tình huống. Như vậy mỗi hệ dựa tình huống có thể tự điều chỉnh để đặc tả các yêu cầu lâm sàng hay khám bệnh

Tính hai mặt chủ quan và khách quan của kiến thức : Trong y khoa , việc tìm ra kiến thức mới bằng cách ước lượng dữ liệu qua sự tích lũy các ca

bệnh lâm sàng là một sự qui ước . Kiến thức này là khách quan , nó không phụ thuộc vào người thăm dò . Kiến thức đó được chứa đựng trong sách giáo khoa , những qui tắc tiêu chuẩn và cách tiến hành nhưng không đủ đặc trưng rành mạch và thường thì nó không được chấp nhận bởi những quan điểm khác nhau về y học hoặc bởi những ý kiến khác nhau , những kinh nghiệm về chẩn đoán bệnh và điều trị bệnh .

Trong hệ chuyên gia , vấn đề thực hành lâm sàng thường khác nhau vì chúng sử dụng kiến thức chủ quan của một bác sĩ hoặc hơn . Kiến thức này là hạn chế trong không gian và theo thời gian . Tuy nhiên , kiến thức như thế vẫn có thể ứng dụng được bởi tính đặc trưng của nó và cũng bởi vì nó phần nào được các bác sĩ chấp nhận . Rõ ràng đây là một tình trạng "tiến thoái lưỡng nan" mà không thể giải quyết được trong hệ chuyên gia theo qui ước . Nhưng không phải như hệ chuyên gia , hệ cơ sở tình huống CB bao gồm những tình huống nghĩa là kiến thức chủ quan và nguyên nhân hình thức , kiến thức phổ biến (sinh lí học , vi trùng học , giải phẫu học ...) như kiến thức khách quan . Bằng sự phân tách của kiến thức khách quan và chủ quan , CBR đưa ra một tính khách quan và một nét đặc trưng riêng biệt của kiến thức cơ bản và giải quyết được tình trạng " tiến thoái lưỡng nan " như trên .

Thu thập tri thức tự động : Sự tích hợp liên tục những ca bệnh trong suốt quá trình sử dụng hệ CBR đưa đến một kiến thức phát triển có thể chấp nhận được . Hơn nữa , hệ thống đó còn có thể trừu tượng hóa kiến thức bằng cách tích hợp những nhóm ca bệnh thành những mẫu bệnh .

Tích hợp hệ thống : Việc sử dụng dữ liệu và hồ sơ bệnh án được sưu tầm trong các bệnh viện hay nhờ những người hành nghề y (được lưu trữ trong máy có khả năng truyền đạt) cho phép tích hợp trong hệ truyền đạt bệnh lâm sàng hiện hành . Kinh nghiệm cho thấy rằng CBR , kết hợp kiến thức nhận rồi qua quá trình thảo luận ta có thể giải quyết kiến thức đó .

Hướng thứ hai để ước định hệ CBR trong y khoa là xác định chúng bằng cách sử dụng dữ liệu lâm sàng thật .

Hầu hết các dấu hiệu được đưa ra bởi cơ sở lập luận dù hệ CBR có được dùng thường xuyên trong lâm sàng hay không . Và những chương trình y khoa hữu dụng khác thường được bổ sung bởi những điều kiện đặc trưng thiết yếu của những chỉnh thể đơn lẻ nhưng thống nhất . Có vài hệ được nguy biến hóa từ tất cả những quan điểm này .

Qua những lí do trên ta cũng thấy được phần nào rằng CBR là một kỹ thuật thích hợp cho hệ tri thức nền trong y khoa . Nó được áp dụng để chẩn đoán và chữa bệnh trong một chừng mực nào đó . Để hiểu rõ hơn chúng ta nên xem xét một số hệ thống chẩn đoán sử dụng kỹ thuật CBR :

3.1.3 Một số hệ CBR trong y khoa

Hệ chẩn đoán y khoa CASEY (được sử dụng trong chẩn đoán bệnh tim) hoạt động theo ba bước chính sau : tìm tình huống tương tự , tiến hành so sánh sự khác biệt giữa tình huống tương tự và tình huống đang xét , cuối cùng là chuyển kết quả chẩn đoán của tình huống tương tự cho tình huống đang xét nếu sự khác biệt không quá quan trọng ; và nó sẽ cố gắng giải thích và sửa đổi chẩn đoán . Nếu không có tình huống tương tự hoặc những sự sửa đổi chẩn đoán không thành công thì CASEY sẽ sử dụng những nguyên tắc giả định . Nói chung là nó cố gắng giải quyết bài toán thích nghi bằng những hoạt động thích nghi tổng quát . Mặc dù vậy không phải bất kì sự khác biệt nào giữa tình huống tương tự và tình huống đang xét cũng có thể giải quyết bởi những hoạt động đó .

HỆ FLORENCE : được sử dụng trong chương trình chăm sóc sức khỏe theo một hướng tổng quát hơn , trong kỹ thuật chăm sóc điều dưỡng , nó thi hành cả 3 nhiệm vụ cơ bản : chẩn đoán , dự đoán và kê đơn . Chẩn đoán bệnh bằng các dụng cụ đo đặc sức khỏe . Từ đó thu được những chỉ số nhằm xác định tình trạng sức khỏe để trả lời câu hỏi : " Tình trạng sức khỏe bệnh nhân hiện như thế nào ? " . Sau đó để dự đoán bệnh , nghĩa là dự đoán sự thay đổi của tình trạng bệnh nhân trong tương lai . Phương pháp tiếp cận cơ sở tình huống được sử dụng : người bệnh sẽ được so sánh với một " người bệnh tương tự " , người mà ta đã xác định rõ chuỗi tiến triển của tình trạng sức khỏe . Những "người bệnh tương tự " này đã được nghiên cứu về toàn bộ tình trạng đầu tiên và các đặc điểm được xác định ngay sau đó . Sự phát triển của một bệnh nhân không chỉ phụ thuộc vào tình trạng của anh ta (tình trạng sức khỏe , bệnh cũ và bệnh đang có) mà còn phụ thuộc vào những phương pháp trị liệu xa hơn , vài sự dự đoán cá nhân về những phương pháp trị liệu khác nhau sẽ được tổng hợp . Và tất nhiên là sau khi đưa ra những phương pháp trị liệu khác nhau thì mục đích cuối cùng vẫn là kê đơn thuốc : " Liệu tình trạng sức khỏe của bệnh nhân có thể cải thiện ra sao ? " . Ta có thể trả lời câu hỏi này bằng cách sử dụng kiến thức tổng quát về tác động của các phương pháp trị liệu và kết quả của việc trị liệu cho các bệnh nhân có tình huống tương tự . Đây là sự kết hợp giữa phương thức luật và phương thức tiếp cận cơ sở tình huống .

MEDIC : MEDIC lại quan tâm đến cách tổ chức cơ sở tình huống của nó . Đây là phương pháp chẩn đoán bằng đồ thị trong lĩnh vực bệnh phổi . Đồ thị trình bày kiến thức xử lý tình huống . Bộ nhớ không chỉ bao gồm đồ thị mà còn chứa các vùng tổ chức cơ sở dữ liệu về chẩn đoán của từng ca bệnh trong từng tình huống . Mỗi ngữ cảnh mô tả đồ thị của một tình huống riêng biệt

bằng một ví dụ cụ thể . Tổ chức bộ nhớ và hệ truy tìm cho phép một hệ suy diễn tìm ra qui trình giải quyết vấn đề một cách thích hợp nhất .

Trên đây là ba hệ sử dụng CBR để phục vụ trong lĩnh vực y khoa . Rõ ràng qua đó ta có thể phần nào nhận thấy rằng CBR là một kỹ thuật thích hợp cho hệ kiến thức nền trong y khoa . Tuy vậy ta lại gặp phải một số vấn đề nảy sinh khi ứng dụng CBR trong lĩnh vực y khoa : hình thức trình bày của một cơ sở tình huống phải được xác định rõ , phải lựa chọn một thuật toán đặc trưng và phải tránh tình trạng cơ sở tình huống tăng tiến đến vô cùng ; như nhóm các tình huống thành mẫu cơ sở , loại bỏ tình huống không cần thiết hoặc hạn chế số lượng cơ sở tình huống và cập nhật cơ sở tình huống trong suốt quá trình thảo luận chuyên môn , nhưng vấn đề chính của CBR vẫn là bài toán thích nghi . Sự thích nghi phụ thuộc vào từng lĩnh vực và công việc đặc thù . Nhiều lúc nó chỉ là những vấn đề hết sức đơn giản như trong FLORENCE hay là sự di chuyển đáp án hoặc hơn một tí (GS 52) nhưng cũng có khi phải xác định những điểm khác biệt giữa tình huống đang xét và các tình huống tương tự (CASEY) . Trong lĩnh vực y khoa các tình huống chứa đựng một số lượng vô cùng lớn các giá trị đặc trưng . CBR thì thường giải quyết các vấn đề dựa trên những giá trị đặc trưng có sẵn hoặc tích lũy từ kinh nghiệm hành nghề

3.2 CBR chẩn đoán bệnh tim

3.2.1 Giới thiệu

Bệnh tim là nguyên nhân chủ yếu ảnh hưởng đến tỷ lệ tử vong và tỷ lệ bệnh tật trên người sau 40 tuổi. Tử vong do bệnh tim chiếm khoảng 40% trên

tổng số tử vong .thống kê cho thấy có khoảng 800 000 trường hợp bị bệnh tim mới,450 000 trường hợp tái phát và 520 000 tử vong/năm

Hiện nay bệnh tim đứng thứ 3 về loại bệnh phải nằm viện ngắn ngày sau sinh đẻ và chấn thương.Việc chẩn đoán và điều trị bệnh tim rất tốn kém

Chẩn đoán bệnh tim là một chuyên ngành của chẩn đoán y khoa do đó những thuận lợi mà CBR cung cấp cho việc xây dựng hệ chuyên gia chẩn đoán y khoa cũng đúng với chẩn đoán bệnh tim.

3.2.2 Nguồn dữ liệu

Nguồn dữ liệu:www.comatech.com

Người tạo dữ liệu:

1. Bác sỹ Andras Janosi:Viện tim Budapest,Hungary
- 2.Bác sỹ William Steinbrunn:bệnh viện đại học Zurich,Thụy Điển
3. Bác sỹ Matthias Pfisterer:bệnh viện đại học Basel,Thụy Điển
4. Tiến sỹ Bác sỹ Robert Detrano:trung tâm y khoa V.A,Long Beach.

Người cung cấp: David W. Aha

3.2.3 Biểu diễn tình huống

Theo như nguồn dữ liệu cung cấp thì có khoảng 76 thuộc tính cho mỗi tình huống chẩn đoán bệnh tim tuy nhiên chỉ có 14 thuộc tính có ảnh hưởng tới chẩn đoán bệnh tim một cách rõ ràng nhất.

1. Age :biểu diễn tuổi cả bệnh nhân
2. Sex : biểu diễn giới tính của bệnh nhân
3. Chestpain : Biểu diễn các kiểu đau ngực của bệnh tim

4. Blood press : Biểu diễn huyết áp (mm Hg)
5. Cholesterol : Biểu diễn lượng cholesterol trong máu (mg/dl)
6. LowBloodsugar : Biểu diễn mức hạ đường trong máu
7. ECG : biểu diễn kết quả đo điện tim
8. Heartrate : số nhịp khi tim đập nhanh nhất
9. Angina : bị bệnh viêm họng
10. OldPeak : độ lõm của sóng điện tim
11. Slope : biểu diễn độ dốc của đỉnh cao nhất đoạn sóng ST
12. VesselCount : số mạch chính được định màu bởi flourosoby
13. Thal : độ nhấp nháy cơ tim
14. Sick:thuộc tính chẩn đoán bệnh

Ta nhận thấy rằng những tính chất này nhận giá trị trong miền giá trị số hay trong tập rời rạc các giá trị số

3.2.4 Đánh giá tương tự giữa các tình huống

Khoảng cách metric giữa hai tính chất càng nhỏ thì độ tương tự giữa chúng càng lớn. Hệ CBR truy tìm tình huống tương tự nhất, để thực hiện được nhiệm vụ này ta chỉ cần đi so sánh giá trị tương tự mà không cần biết độ tương tự là bao nhiêu. Như vậy việc đánh giá tương tự có thể thay thế bằng việc xác định khoảng cách. Khoảng cách metric giữa hai tính các tính chất được định nghĩa như sau

1. Male:

Miền giá trị: $D = \{\text{nam}, \text{nữ}\}$.

Định nghĩa khoảng cách :

$$d(x, y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

2. Age:

Miền giá trị : $D = [29,77]$

Định nghĩa khoảng cách :

$$d(x,y) = \frac{|x-y|}{48}$$

3. ChestPain :

Miền giá trị : $D = \{\text{typical angina, atypical angina, non-anginal pain, asymptomatic}\}$

Định nghĩa khoảng cách :

$$d(x,y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

4. BloodPress :

Miền giá trị : $D = [94,200]$

Định nghĩa khoảng cách :

$$d(x,y) = \frac{|x-y|}{106}$$

5. Cholesterol :

Miền giá trị : $D = [126,564]$

Định nghĩa khoảng cách :

$$d(x,y) = \frac{|x-y|}{106}$$

6. LowBloodSugar :

Miền giá trị : $D = \{\text{true, false}\}$

Định nghĩa khoảng cách :

$$d(x,y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

7. ECG :

Miền giá trị : $D = \{\text{normal, having ST-T wave abnormality, left ventricular hypertrophy}\}$

Định nghĩa khoảng cách :

$$d(x, y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

8. HeartRate :

Miền giá trị : $D = [71, 202]$

Định nghĩa khoảng cách :

$$d(x, y) = \frac{|x - y|}{131}$$

9. Angina :

Miền giá trị : $D = \{\text{yes, no}\}$

Định nghĩa khoảng cách :

$$d(x, y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

10. Oldpeak :

Miền giá trị : $D = [0, 6.2]$

Định nghĩa khoảng cách :

$$d(x, y) = \frac{|x - y|}{6.2}$$

11. Slope:

Miền giá trị : $D = \{\text{upsloping, flat, downsloping}\}$

Định nghĩa khoảng cách :

$$d(x, y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

12. VesselCount :

Miền giá trị : $D = \{0,1,2,3\}$

Định nghĩa khoảng cách :

$$d(x, y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

13. Thal :

Miền giá trị: $D = \{\text{normal, fixed defect, reversable defect}\}$

Định nghĩa khoảng cách :

$$d(x, y) = \begin{cases} 0 & x \equiv y \\ 1 & x \neq y \end{cases}$$

Ta nhận thấy rằng qua định nghĩa khoảng cách như trên sự khác biệt về mặt đơn vị không còn ý nghĩa nữa. Thực ra chúng ta đã ngầm quy miền giá trị của tất cả các tính chất về $[0,1]$.

Cũng tương tự như trên thay vì đi xét tương tự giữa các tình huống ta chỉ cần đi xét khoảng cách giữa chúng. Khoảng cách giữa hai tình huống sẽ được xác định theo khoảng cách Euclide

$$d(A, B) = \sqrt{\sum_{i=1}^n d_i^2}$$

d_i : khoảng cách giữa hai tính chất thứ i của hai tình huống A, B

3.2.5 Phân loại tình huống

ta sẽ chia tập dữ liệu thành 6 lớp như sau:

Lớp	1	2	3	4	5	6
ECG	0	0	1	1	2	2
Sick	0	1	0	1	0	1

Thủ tục huấn luyện: dùng để tạo một cơ sở tình huống ban đầu

x : tình huống

$f(x)$: lớp của tình huống x

Một cách hình thức hơn ta có thể định nghĩa $f(x)$ như sau

$$f : R^n \rightarrow V$$

với V là tập hữu hạn $\{v_1, v_2, \dots, v_s\}$

Thuật toán

Với mỗi tình huống huấn luyện $\langle x, f(x) \rangle$

thêm tình huống vào cơ sở tình huống

3.2.6 Truy tìm tình huống

Ý tưởng:

Cho tình huống truy vấn q , để truy tìm tình huống hay một số tình huống phù hợp ta làm hai bước như sau

Trước hết dùng giải thuật K láng giềng gần nhất để xác định lớp mà tình huống truy vấn q thuộc về

Sau đó ta chọn từ lớp đó một tình huống hay một số tình huống tương tự nhất với tình huống truy vấn q

Thuật toán**Đầu vào:** tình huống truy vấn q tập cơ sở tình huống có J lớp**Đầu ra :** tình huống được truy tìm p **bước 1: Tìm K tình huống láng giềng gần nhất**

```
for i=1 to k
     $t_i = x_i$ 
for i=k+1 to n do
     $i_{max} = 1$ 
    for j=2 to k do
        if  $d(q, x_j) > d(q, x_{i_{max}})$  then
             $i_{max} = j$ 
        if  $d(q, x_i) < d(q, t_{i_{max}})$  then
             $t_{i_{max}} = x_i$ 
```

bước 2: xác định lớp của tình huống truy vấn q

```
id=0
for i=1 to J do
     $id_i = 0$ 
for i=1 to J do
    for j=1 to k do
        if  $f(t_j) = i$  then
```

$id_i = id_i + 1$

for $i=1$ to J do

if $id_i > id$ then

$id = id + 1$

bước 3: Tìm tình huống p tương tự nhất

$p = t_1$

for $i=2$ to k do

if $d(p, q) > d(p, t_i)$ then

$p = t_i$

Giá trị k

Thực tế sau nhiều lần kiểm tra với nhiều giá trị k khác nhau nhận thấy rằng chương trình cho kết quả tốt với k nằm trong khoảng 20-30

3.2.7 Thích nghi tình huống

Xuất phát từ thực tế, để chẩn đoán bệnh bác sỹ phải cần một thời gian để theo dõi lại những triệu chứng

Mặt khác mặc dù có rất nhiều phương pháp thích nghi trong CBR được đề nghị nhưng chưa có phương pháp nào có thể khả dĩ thay thế được chuyên gia. Nhất là trong lĩnh vực liên quan tới mạng sống của con người thì việc để cho máy tính thích nghi hoàn toàn là điều không được tốt cho lắm. Như vậy tiến trình thích nghi trong bài toán này chỉ nên sao chép lại lời giải. Sau đó tình huống mới này sẽ được đưa vào hàng đợi để bác sỹ xác nhận lại

Tiến trình sửa đổi bây giờ hầu hết nằm bên ngoài hệ CBR

3.2.8 Tiến trình học

Cập nhật

Hàng đợi sẽ được cập nhật định kỳ vào cơ sở tình huống, dĩ nhiên là phải sau khi có ý kiến bác sỹ rằng tình huống đó đã hoàn thành

Xoá

Để tránh việc cơ sở tình huống trở nên quá lớn với những tình huống không bao giờ hay rất ít được truy cập đến ta phải có cơ chế xoá bớt những tình huống này ra khỏi tình huống.

Để thực hiện điều này ta đưa vào tình huống một thuộc tính để đếm sự truy cập đến nó. Mỗi lần tình huống được đưa vào danh sách k láng giềng gần nhất thì thuộc tính này giảm giá trị đi 1. Đồng thời các thuộc tính không được đưa vào danh sách k láng giềng gần nhất sẽ tăng giá trị lên 1. Khi tình huống nào mà giá trị tính chất này tăng lên đạt ngưỡng t_d đó thì n_l sẽ bị xoá khỏi cơ sở tình huống. Và như vậy cơ sở tình huống sẽ tốt hơn theo thời gian

CHƯƠNG 4: ĐÁNH GIÁ KẾT QUẢ VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỀ TÀI

4.1 Đánh giá kết quả

Chương trình có kết quả truy tìm khá tốt, tuy nhiên các tiến trình khác của CBR còn khá thô. Do đặc trưng của lĩnh vực nên hệ chỉ mang tính hỗ trợ, kết quả này là chấp nhận được

Ta xem một số kết quả khi thực hiện chương trình

Tình huống truy vấn

Giới tính	nam
Tuổi	67
đau ngực	0
Huyết áp	120
Lượng Cholesterol trong máu	229

Mức hạ đường trong máu	0
Kết quả điện tim	2
Nhịp tim	129
Viêm họng	1
Độ lồm của sóng điện tim	2.6
đường dốc của sóng điện tim	0
số mạch chính được định màu	2
độ nhấp nháy cơ tim	-1
Kết luận : ?	

Tình huống truy tìm

Giới tính	nam
Tuổi	56
đau ngực	2
Huyết áp	130
Lượng Cholesterol trong máu	256
Mức hạ đường trong máu	1
Kết quả điện tim	2
Nhịp tim	142
Viêm họng	1
Độ lồm của sóng điện tim	0.6
đường dốc của sóng điện tim	0
số mạch chính được định màu	1
độ nhấp nháy cơ tim	1
Kết luận: bị bệnh tim	

Tình huống truy vấn

Giới tính	nữ
Tuổi	41
đau ngực	1
Huyết áp	130
Lượng Cholesterol trong máu	204
Mức hạ đường trong máu	0
Kết quả điện tim	2
Nhịp tim	172
Viêm họng	0
Độ lồm của sóng điện tim	1.4
đường dốc của sóng điện tim	1
số mạch chính được định màu	0
độ nhấp nháy cơ tim	0
Kết luận : ?	

Tình huống truy tìm

Giới tính	nam
Tuổi	64
đau ngực	3
Huyết áp	110
Lượng Cholesterol trong máu	211
Mức hạ đường trong máu	0
Kết quả điện tim	2
Nhịp tim	144

Viêm họng	1
Độ lồi của sóng điện tim	1.8
đường dốc của sóng điện tim	0
số mạch chính được định màu	0
độ nhấp nháy cơ tim	0
Kết luận : không bị bệnh	

4.2 Các hướng phát triển

4.2.1 Đối với ứng dụng

+Nghiên cứu mức độ quan trọng của các tính chất đối với bệnh tim để truy tìm tình huống hiệu quả hơn

+Mở rộng thành hệ chẩn đoán y khoa tổng quát

4.2.2 Đối với CBR

+Nghiên cứu việc biểu diễn tri thức trong cơ sở tình huống để cung cấp cơ chế tích hợp nhiều kiểu tri thức hơn

+Nghiên cứu việc xây dựng một framework cho hệ CBR .Những người thiết kế hệ CBR sẽ dựa trên những framework này để xây dựng hệ CBR

+Phát triển các tool hỗ trợ một số bước hay toàn bộ chu trình CBR

PHỤ LỤC :DỮ LIỆU

dulieu															
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick	
1	1	63	3	145	233	1	2	150	0	2.3	-1	0	1	0	
2	1	67	0	160	286	0	2	108	1	1.5	0	3	0	1	
3	1	67	0	120	229	0	2	129	1	2.6	0	2	-1	1	
4	1	37	2	130	250	0	0	187	0	3.5	-1	0	0	0	
5	0	41	1	130	204	0	2	172	0	1.4	1	0	0	0	
6	1	56	1	120	236	0	0	178	0	0.8	1	0	0	0	
7	0	62	0	140	268	0	2	160	0	3.6	-1	2	0	1	
8	0	57	0	120	354	0	0	163	1	0.6	1	0	0	0	
9	1	63	0	130	254	0	2	147	0	1.4	0	1	-1	1	
10	1	53	0	140	203	1	2	155	1	3.1	-1	0	-1	1	
11	1	57	0	140	192	0	0	148	0	0.4	0	0	1	0	
12	0	56	1	140	294	0	2	153	0	1.3	0	0	0	0	
13	1	56	2	130	256	1	2	142	1	0.6	0	1	1	1	
14	1	44	1	120	263	0	0	173	0	0	1	0	-1	0	
15	1	49	1	130	266	0	0	171	0	0.6	1	0	0	0	
16	1	64	3	110	211	0	2	144	1	1.8	0	0	0	0	

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
17	0	58	3	150	283	1	2	162	0	1	1	0	0	0
18	1	58	1	120	284	0	2	160	0	1.8	0	0	0	1
19	1	58	2	132	224	0	2	173	0	3.2	1	2	-1	1
20	1	60	0	130	206	0	2	132	1	2.4	0	2	-1	1
21	0	50	2	120	219	0	0	158	0	1.6	0	0	0	0
22	0	58	2	120	340	0	0	172	0	0	1	0	0	0
23	0	66	3	150	226	0	0	114	0	2.6	-1	0	0	0
24	1	43	0	150	247	0	0	171	0	1.5	1	0	0	0
25	1	40	0	110	167	0	2	114	1	2	0	0	-1	1
26	0	69	3	140	239	0	0	151	0	1.8	1	2	0	0
27	1	60	0	117	230	1	0	160	1	1.4	1	2	-1	1
28	1	64	2	140	335	0	0	158	0	0	1	0	0	1
29	1	59	0	135	234	0	0	161	0	0.5	0	0	-1	0
30	1	44	2	130	233	0	0	179	1	0.4	1	0	0	0
31	1	42	0	140	226	0	0	178	0	0	1	0	0	0
32	1	43	0	120	177	0	2	120	1	2.5	0	0	-1	1
33	1	57	0	150	276	0	2	112	1	0.6	0	1	1	1
34	1	55	0	132	353	0	0	132	1	1.2	0	1	-1	1
35	1	61	2	150	243	1	0	137	1	1	0	0	0	0
36	0	65	0	150	225	0	2	114	0	1	0	3	-1	1
37	1	40	3	140	199	0	0	178	1	1.4	1	0	-1	0
38	0	71	1	160	302	0	0	162	0	0.4	1	2	0	0
39	1	59	2	150	212	1	0	157	0	1.6	1	0	0	0
40	0	61	0	130	330	0	2	169	0	0	1	0	0	1
41	1	58	2	112	230	0	2	165	0	2.5	0	1	-1	1
42	1	51	2	110	175	0	0	123	0	0.6	1	0	0	0
43	1	50	0	150	243	0	2	128	0	2.6	0	0	-1	1
44	0	65	2	140	417	1	2	157	0	0.8	1	1	0	0
45	1	53	2	130	197	1	2	152	0	1.2	-1	0	0	0
46	0	41	1	105	198	0	0	168	0	0	1	1	0	0
47	1	65	0	120	177	0	0	140	0	0.4	1	0	-1	0
48	1	44	0	112	290	0	2	153	0	0	1	1	0	1

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
49	1	44	1	130	219	0	2	188	0	0	1	0	0	0
50	1	60	0	130	253	0	0	144	1	14	1	1	-1	1
51	1	54	0	124	266	0	2	109	1	2.2	0	1	-1	1
52	1	50	2	140	233	0	0	163	0	0.6	0	1	-1	1
53	1	41	0	110	172	0	2	158	0	0	1	0	-1	1
54	1	54	2	125	273	0	2	152	0	0.5	-1	1	0	0
55	1	51	3	215	213	0	2	125	1	1.4	1	1	0	0
56	0	51	0	130	305	0	0	142	1	1.2	0	0	-1	1
57	0	46	2	142	177	0	2	160	1	1.4	-1	0	0	0
58	1	58	0	128	216	0	2	131	1	2.2	0	3	-1	1
59	0	54	2	135	304	1	0	170	0	1	0	0	0	0
60	1	54	0	120	188	0	0	113	0	1.4	0	1	-1	1
61	1	60	0	145	282	0	2	142	1	2.8	0	2	-1	1
62	1	60	2	140	185	0	2	155	0	3	0	0	0	1
63	1	54	2	150	232	0	2	165	0	1.6	1	0	-1	0
64	1	59	0	170	326	0	2	140	1	3.4	-1	0	-1	1
65	1	46	2	150	231	0	0	147	0	3.6	0	0	0	1
66	0	65	2	155	269	0	0	148	0	0.8	1	0	0	0
67	1	67	0	125	254	1	0	163	0	0.2	0	2	-1	1
68	1	62	0	120	267	0	0	99	1	1.8	0	2	-1	1
69	1	65	0	110	248	0	2	158	0	0.6	1	2	1	0
70	1	44	0	110	197	0	2	177	0	0	1	1	0	1
71	0	65	2	160	360	0	2	151	0	0.8	1	0	0	0
72	1	60	0	125	258	0	2	141	1	2.8	0	1	-1	1
73	0	51	2	140	308	0	2	142	0	1.5	1	1	0	0
74	1	48	1	130	245	0	2	180	0	0.2	0	0	0	0
75	1	58	0	150	270	0	2	111	1	0.8	1	0	-1	1
76	1	45	0	104	208	0	2	148	1	3	0	0	0	0
77	0	53	0	130	264	0	2	143	0	0.4	0	0	0	0
78	1	39	2	140	321	0	2	182	0	0	1	0	0	0
79	1	68	2	180	274	1	2	150	1	1.6	0	0	-1	1
80	1	52	1	120	325	0	0	172	0	0.2	1	0	0	0

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
81	1	44	2	140	235	0	2	180	0	0	1	0	0	0
82	1	47	2	138	257	0	2	156	0	0	1	0	0	0
83	0	53	2	128	216	0	2	115	0	0	1		0	0
84	0	53	0	138	234	0	2	160	0	0	1	0	0	0
85	0	51	2	130	256	0	2	149	0	0.5	1	0	0	0
86	1	66	0	120	302	0	2	151	0	0.4	0	0	0	0
87	0	62	0	160	164	0	2	145	0	6.2	-1	3	-1	1
88	1	62	2	130	231	0	0	146	0	1.8	0	3	-1	0
89	0	44	2	108	141	0	0	175	0	0.6	0	0	0	0
90	0	63	2	135	252	0	2	172	0	0	1	0	0	0
91	1	52	0	128	255	0	0	161	1	0	1	1	-1	1
92	1	59	0	110	239	0	2	142	1	1.2	0	1	-1	1
93	0	60	0	150	258	0	2	157	0	2.6	0	2	-1	1
94	1	52	1	134	201	0	0	158	0	0.8	1	1	0	0
95	1	48	0	122	222	0	2	186	0	0	1	0	0	0
96	1	45	0	115	260	0	2	185	0	0	1	0	0	0
97	1	61	0	120	260	0	0	140	1	3.6	0	1	-1	1
98	1	39	0	118	219	0	0	140	0	1.2	0	0	-1	1
99	0	61	0	145	307	0	2	146	1	1	0	0	-1	1
100	1	56	0	125	249	1	2	144	1	1.2	0	1	0	1
101	1	52	3	118	186	0	2	190	0	0	0	0	1	0
102	0	43	0	132	341	1	2	136	1	3	0	0	-1	1
103	0	62	2	130	263	0	0	97	0	1.2	0	1	-1	1
104	1	41	1	135	203	0	0	132	0	0	0	0	1	0
105	1	58	2	140	211	1	2	165	0	0	1	0	0	0
106	0	35	0	128	183	0	0	182	0	1.4	1	0	0	0
107	1	63	0	130	330	1	2	132	1	1.8	1	3	-1	1
108	1	65	0	135	254	0	2	127	0	2.8	0	1	-1	1
109	1	48	0	130	256	1	2	150	1	0	1	2	-1	1
110	0	63	0	150	407	0	2	154	0	4	0	3	-1	1
111	1	51	2	100	222	0	0	143	1	1.2	0	0	0	0
112	1	55	0	140	217	0	0	111	1	5.6	-1	0	-1	1

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
113	1	29	1	130	204	0	2	202	0	0	1	0	0	0
114	1	51	0	140	261	0	2	186	1	0	1	0	0	0
115	0	43	2	122	213	0	0	165	0	0.2	0	0	0	0
116	0	55	1	135	250	0	2	161	0	1.4	0	0	0	0
117	1	70	0	145	174	0	0	125	1	2.6	-1	0	-1	1
118	1	62	1	120	281	0	2	103	0	1.4	0	1	-1	1
119	1	35	0	120	198	0	0	130	1	1.6	0	0	-1	1
120	1	51	2	125	245	1	2	166	0	2.4	0	0	0	0
121	1	59	1	140	221	0	0	164	1	0	1	0	0	0
122	1	59	3	170	288	0	2	159	0	0.2	0	0	-1	1
123	1	52	1	128	205	1	0	184	0	0	1	0	0	0
124	1	64	2	125	309	0	0	131	1	1.8	0	0	-1	1
125	1	58	2	105	240	0	2	154	1	0.6	0	0	-1	0
126	1	47	2	108	243	0	0	152	0	0	1	0	0	1
127	1	57	0	165	289	1	2	124	0	1	0	3	-1	1
128	1	41	2	112	250	0	0	179	0	0	1	0	0	0
129	1	45	1	128	308	0	2	170	0	0	1	0	0	0
130	0	60	2	102	318	0	0	160	0	0	1	1	0	0
131	1	52	3	152	298	1	0	178	0	1.2	0	0	-1	0
132	0	42	0	102	265	0	2	122	0	0.6	0	0	0	0
133	0	67	2	115	564	0	2	160	0	1.6	0	0	-1	0
134	1	55	0	160	289	0	2	145	1	0.8	0	1	-1	1
135	1	64	0	120	246	0	2	96	1	2.2	-1	1	0	1
136	1	70	0	130	322	0	2	109	0	2.4	0	3	0	1
137	1	51	0	140	299	0	0	173	1	1.6	1	0	-1	1
138	1	58	0	125	300	0	2	171	0	0	1	2	-1	1
139	1	60	0	140	293	0	2	170	0	1.2	0	2	-1	1
140	0	54	1	132	288	1	2	159	1	0	1	1	0	0
141	1	35	0	126	282	0	2	156	1	0	1	0	-1	1
142	0	45	1	112	160	0	0	138	0	0	0	0	0	0
143	1	70	2	160	269	0	0	112	1	2.9	0	1	-1	1
144	1	52	2	138	223	0	0	169	0	0	1	0	0	0

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
145	1	53	0	142	226	0	2	111	1	0	1	0	-1	0
146	0	59	0	174	249	0	0	143	1	0	0	0	0	1
147	0	62	0	140	394	0	2	157	0	1.2	0	0	0	0
148	1	64	0	145	212	0	2	132	0	2	0	2	1	1
149	1	57	0	152	274	0	0	88	1	1.2	0	1	-1	1
150	1	52	0	108	233	1	0	147	0	0.1	1	3	-1	0
151	1	56	0	132	184	0	2	105	1	2.1	0	1	1	1
152	1	43	2	130	315	0	0	160	0	1.9	1	1	0	0
153	1	53	2	130	246	1	2	173	0	0	1	3	0	0
154	1	48	0	124	274	0	2	166	0	0.5	0	0	-1	1
155	0	56	0	134	409	0	2	150	1	1.9	0	2	-1	1
156	1	42	3	148	244	0	2	178	0	0.8	1	2	0	0
157	1	59	3	178	270	0	2	145	0	4.2	-1	0	-1	0
158	1	38	2	138	175	0	0	173	0	0	1	0	0	0
159	0	63	1	140	195	0	0	179	0	0	1	2	0	0
160	1	42	2	120	240	1	0	194	0	0.8	-1	0	-1	0
161	1	66	1	160	246	0	0	120	1	0	0	3	1	1
162	1	54	1	192	283	0	2	195	0	0	1	1	-1	1
163	1	69	2	140	254	0	2	146	0	2	0	3	-1	1
164	1	50	2	129	196	0	0	163	0	0	1	0	0	0
165	1	51	0	140	298	0	0	122	1	4.2	0	3	-1	1
166	1	43	0	132	247	1	2	143	1	0.1	0	0	-1	1
167	0	62	0	138	294	1	0	106	0	1.9	0	3	0	1
168	0	68	2	120	211	0	2	115	0	1.5	0	0	0	0
169	1	67	0	100	299	0	2	125	1	0.9	0	2	0	1
170	1	69	3	160	234	1	2	131	0	0.1	0	1	0	0
171	0	45	0	138	236	0	2	152	1	0.2	0	0	0	0
172	0	50	1	120	244	0	0	162	0	1.1	1	0	0	0
173	1	59	3	160	273	0	2	125	0	0	1	0	0	1
174	0	50	0	110	254	0	2	159	0	0	1	0	0	0
175	0	64	0	180	325	0	0	154	1	0	1	0	0	0
176	1	57	2	150	126	1	0	173	0	0.2	1	1	-1	0

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
177	0	64	2	140	313	0	0	133	0	0.2	1	0	-1	0
178	1	43	0	110	211	0	0	161	0	0	1	0	-1	0
179	1	45	0	142	309	0	2	147	1	0	0	3	-1	1
180	1	58	0	128	259	0	2	130	1	3	0	2	-1	1
181	1	50	0	144	200	0	2	126	1	0.9	0	0	-1	1
182	1	55	1	130	262	0	0	155	0	0	1	0	0	0
183	0	62	0	150	244	0	0	154	1	1.4	0	0	0	1
184	0	37	2	120	215	0	0	170	0	0	1	0	0	0
185	1	38	3	120	231	0	0	182	1	3.8	0	0	-1	1
186	1	41	2	130	214	0	2	168	0	2	0	0	0	0
187	0	66	0	178	228	1	0	165	1	1	0	2	-1	1
188	1	52	0	112	230	0	0	160	0	0	1	1	0	1
189	1	56	3	120	193	0	2	162	0	1.9	0	0	-1	0
190	0	46	1	105	204	0	0	172	0	0	1	1	0	0
191	0	46	0	138	243	0	2	152	1	0	0	0	0	0
192	0	64	0	130	303	0	0	122	0	2	0	2	0	0
193	1	59	0	138	271	0	2	182	0	0	1	0	0	0
194	0	41	2	112	268	0	2	172	1	0	1	0	0	0
195	0	54	2	108	267	0	2	167	0	0	1	0	0	0
196	0	39	2	94	199	0	0	179	0	0	1	0	0	0
197	1	53	0	123	282	0	0	95	1	2	0	2	-1	1
198	0	63	0	108	269	0	0	169	1	1.8	0	2	0	1
199	1	52	0	128	204	1	0	156	1	1	0	0	0	1
200	0	34	1	118	210	0	0	192	0	0.7	1	0	0	0
201	1	47	0	112	204	0	0	143	0	0.1	1	0	0	0
202	0	67	2	152	277	0	0	172	0	0	1	1	0	0
203	1	54	0	110	206	0	2	108	1	0	0	1	0	1
204	1	66	0	112	212	0	2	132	1	0.1	1	1	0	1
205	0	52	2	136	196	0	2	169	0	0.1	0	0	0	0
206	0	55	0	180	327	0	1	117	1	3.4	0	0	0	1
207	1	54	0	122	286	0	2	116	1	3.2	0	2	0	1
208	1	56	0	130	283	1	2	103	1	1.6	-1	0	-1	1

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
209	1	46	0	120	249	0	2	144	0	0.8	1	0	-1	1
210	0	49	1	134	271	0	0	162	0	0	0	0	0	0
211	1	42	1	120	295	0	0	162	0	0	1	0	0	0
212	1	41	1	110	235	0	0	153	0	0	1	0	0	0
213	0	41	1	126	306	0	0	163	0	0	1	0	0	0
214	0	49	0	130	269	0	0	163	0	0	1	0	0	0
215	1	61	3	134	234	0	0	145	0	2.6	0	2	0	1
216	0	60	2	120	178	1	0	96	0	0	1	0	0	0
217	1	67	0	120	237	0	0	71	0	1	0	0	0	1
218	1	58	0	100	234	0	0	156	0	0.1	1	1	-1	1
219	1	47	0	110	275	0	2	118	1	1	0	1	0	1
220	1	52	0	125	212	0	0	168	0	1	1	2	-1	1
221	1	62	1	128	208	1	2	140	0	0	1	0	0	0
222	1	57	0	110	201	0	0	126	1	1.5	0	0	1	0
223	1	58	0	146	218	0	0	105	0	2	0	1	-1	1
224	1	64	0	128	263	0	0	105	1	0.2	0	1	-1	0
225	0	51	2	120	295	0	2	157	0	0.6	1	0	0	0
226	1	43	0	115	303	0	0	181	0	1.2	0	0	0	0
227	0	42	2	120	209	0	0	173	0	0	0	0	0	0
228	0	67	0	106	223	0	0	142	0	0.3	1	2	0	0
229	0	76	2	140	197	0	1	116	0	1.1	0	0	0	0
230	1	70	1	156	245	0	2	143	0	0	1	0	0	0
231	1	57	1	124	216	0	0	141	0	0.3	1	0	-1	1
232	1	59	2	126	218	1	0	134	0	2.2	0	1	1	1
233	1	40	0	152	223	0	0	181	0	0	1	0	-1	1
234	1	42	2	130	180	0	0	150	0	0	1	0	0	0
235	1	61	0	140	207	0	2	138	1	1.9	1	1	-1	1
236	1	66	0	160	228	0	2	138	0	2.3	1	0	1	0
237	1	46	0	140	311	0	0	120	1	1.8	0	2	-1	1
238	0	71	0	112	149	0	0	125	0	1.6	0	0	0	0
239	1	59	3	134	204	0	0	162	0	0.8	1	2	0	1
240	1	64	3	170	227	0	2	155	0	0.6	0	0	-1	0

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
241	0	66	2	146	278	0	2	152	0	0	0	1	0	0
242	0	39	2	138	220	0	0	152	0	0	0	0	0	0
243	1	57	1	154	232	0	2	164	0	0	1	1	0	1
244	0	58	0	130	197	0	0	131	0	0.6	0	0	0	0
245	1	57	0	110	335	0	0	143	1	3	0	1	-1	1
246	1	47	2	130	253	0	0	179	0	0	1	0	0	0
247	0	55	0	128	205	0	1	130	1	2	0	1	-1	1
248	1	35	1	122	192	0	0	174	0	0	1	0	0	0
249	1	61	0	148	203	0	0	161	0	0	1	1	-1	1
250	1	58	0	114	318	0	1	140	0	4.4	-1	3	1	1
251	1	58	1	125	220	0	0	144	0	0.4	0	0	-1	0
252	0	55	1	132	342	0	0	166	0	1.2	1	0	0	0
253	1	44	0	120	169	0	0	144	1	2.8	-1	0	1	1
254	1	63	0	140	187	0	2	144	1	4	1	2	-1	1
255	0	63	0	124	197	0	0	136	1	0	0	0	0	1
256	1	41	1	120	157	0	0	182	0	0	1	0	0	0
257	1	59	0	164	176	0	2	90	0	1	0	2	1	1
258	0	57	0	140	241	0	0	123	1	0.2	0	0	-1	1
259	1	45	3	110	264	0	0	132	0	1.2	0	0	-1	1
260	1	68	0	144	133	1	0	141	0	3.4	0	2	-1	1
261	1	57	0	130	131	0	0	115	1	1.2	0	1	-1	1
262	0	57	1	130	236	0	2	174	0	0	0	1	0	1
263	1	38	2	138	175	0	0	173	0	0	1	0	0	0
264	0	58	0	170	225	1	2	146	1	2.8	0	2	1	1
265	1	36	1	130	221	0	2	163	0	0	1	0	-1	0
266	1	56	1	120	240	0	0	169	0	0	-1	0	0	0
267	1	67	2	152	212	0	2	150	0	0.8	0	0	-1	1
268	1	52	2	172	199	1	0	162	0	0.5	1	0	-1	0
269	1	57	2	150	168	0	0	174	0	1.6	1	0	0	0
270	1	48	1	110	229	0	0	168	0	1	-1	0	-1	1
271	1	54	0	140	239	0	0	160	0	1.2	1	0	0	0
272	0	48	2	130	275	0	0	139	0	0.2	1	0	0	0

dulieu														
Id	Male	Age	C	B	Ch	LB	ECG	HR	A	O	S	VC	Thal	Sick
273	1	34	3	118	182	0	2	174	0	0	1	0	0	0
274	0	57	0	128	303	0	2	159	0	0	1	1	0	0
275	0	71	2	110	265	1	2	130	0	0	1	1	0	0
276	1	49	2	120	188	0	0	139	0	2	0	3	-1	1
277	1	54	1	108	308	0	0	156	0	0	1	0	-1	0
278	1	59	0	140	177	0	0	162	1	0	1	1	-1	1
279	1	57	2	128	229	0	2	150	0	0.4	0	1	-1	1
280	1	65	3	138	282	1	2	174	0	1.4	0	1	0	1
281	0	45	1	130	234	0	2	175	0	0.6	0	0	0	0
282	0	56	0	200	288	1	2	133	1	4	-1	2	-1	1
283	1	54	0	110	239	0	0	126	1	2.8	0	1	-1	1
284	1	44	1	120	220	0	0	170	0	0	1	0	0	0
285	0	62	0	124	209	0	0	163	0	0	1	0	0	0
286	1	54	2	120	258	0	2	147	0	0.4	0	0	-1	0
287	1	51	2	94	227	0	0	154	1	0	1	1	-1	0
288	0	44	2	118	242	0	0	149	0	0.3	0	1	0	0
289	0	58	1	136	313	1	2	152	0	0	1	2	0	1
290	0	60	3	150	240	0	0	171	0	0.9	1	0	0	0
291	1	44	2	120	226	0	0	169	0	0	1	0	0	0
292	1	61	0	138	166	0	2	125	1	3.6	0	1	0	1
293	1	42	0	136	315	0	0	125	1	1.8	0	0	1	1
294	1	68	2	118	227	0	0	151	0	1	1	1	-1	0
295	1	46	1	101	197	1	0	156	0	0	1	0	-1	0
296	1	77	0	125	304	0	2	162	1	0	1	3	0	1
297	0	54	2	110	214	0	0	158	0	1.6	0	0	0	0
298	0	58	0	100	248	0	2	122	0	1	0	0	0	0
299	1	48	2	124	255	1	0	175	0	0	1	2	0	0
300	1	57	0	132	207	0	0	168	1	0	1	0	-1	0
301	1	49	2	118	149	0	2	126	0	0.8	1	3	0	1
302	0	74	1	120	269	0	2	121	1	0.2	1	1	0	0
303	0	54	2	160	201	2	2	163	0	0	1	1	0	0

TÀI LIỆU THAM KHẢO

- [1] Agnar Aamodt, Enric Plaza, Case-based reasoning: Foundational Issues, methodological variations, and system approaches
- [2] Ralph Bergmann, On the use of Taxonomies for representing case features and local similarity measures
- [3] Ralph Bergmann, introduce to case-based reasoning

- [4] David B. Leake, CBR in context : The present and future
- [5] Michael M Richter, CBR : past and future A personal view
- [6] Rainer Schmidt , Lothar Gierl. Case-based reasoning for medical knowledge-based systems
- [7] Changduk Jung, Ingo Han, Boomil Suh, Risk Analysis for electronic commerce using case-based reasoning
- [8] David B. Leake, Andrew Kinley, and David Wilson, Linking Adaptation and Similarity Learning
- [9] Rhodora L. Reyes and Raymund C. Sison, Representing and Indexing Cases in CBR-Tutor
- [10] Ramon López de Mántaras and Enric Plaza, Case-Based Reasoning : An Overview
- [11] Frank Heister, Wolfgang Wilke, An Architecture for maintaining case-based reasoning systems
- [12] Bergmann & Wilke, 1996
Bergmann R., W. Wilke, "On the Role of Abstraction in Case-Based Reasoning", in Proc. Advances in Case-Based Reasoning, 3rd European Workshop, EWCBR-96, pp28-43, Lausanne, Switzerland, November 1996.
- [13] Cheng, 1991
Cheng, Y., "Context-Dependent Similarity", in *Uncertainty in Artificial Intelligence 6*, 12, P.P. Bonissone, M. Henrion, L.N. Kanal, J.F. Lemmer (editors), pp41-47, North-Holland, NY, 1991.