Họ tên sinh viên: Nguyễn Quang Thắng

Mã số sinh viên: 20153524 lớp: KSTN-CNTT-K60

Granular Rules for Medical Diagnosis

Nguyễn Quang Thắng1

1-2 Viện công nghệ thông tin & Truyền thông, Trường Đại học Bách Khoa Hà Nội,

Giảng viên hướng dẫn: Phạm Văn Hải

**TÓM TẮT**— Bài báo này thảo luận về mô hình chi tiết cho quy tắc chuẩn đoán bệnh, nó là sự mở rộng của mô hình quy tắc tập thô. Việc chuẩn đoán bệnh được đặc trưng bởi ba yếu tố: cơ chế tập trung, chuẩn đoán riêng biệt và phát hiện các triệu chứng. Đầu tiên, cơ chế tập trung sử dụng một tập các triệu chứng cái mà luôn luôn được xem xét bởi hầu hết tất cả các trường hợp của một bệnh, nếu trường hợp bệnh đó không gồm bất kì một triệu chứng nào trong số đó, bệnh đó sẽ bị loại bỏ khỏi chuẩn đoán. Thứ hai, từ tập các bệnh đã được chọn, tập các triệu chứng được giám sát cao hơn trong trường hợp được sử dụng để xác nhận các chuẩn đoán khác nhau. Cuối cùng, phát hiện các triệu chứng là một tập các triệu chứng mà sự xuất hiện ở bệnh này rất thấp nhưng lại rất quan trọng để chuẩn đoán bệnh khác. Những mô hình quy tắc có thể dễ dàng được được mô tả bởi sự mở rộng của mô hình tập thô: các tập hỗ trợ của hai tập triệu chứng ban đầu đại diện cho xấp xỉ trên và xấp xỉ dưới của khái niệm mục tiêu. Cái cuối cùng được mô tả bởi sự liên quan giữa kết luận này với kết luận khác, cái sẽ là một loại mới của chi tiết thông tin.

# GIỚI THIỆU

Chuẩn đoán y khoa thông thường chỉ ra một bệnh được định nghĩa bởi một tập các triệu chứng. Triệu chứng là một quy tắc chẩn đoán chính trước khi kiểm tra, xét nghiệm. Mặc dù tác động của chẩn đoán phân biệt từ triệu chứng hiện nay thấp, nhưng sự thật là sự thay đổi của các triệu chứng là rất quan trọng để đánh giá tình trạng của bệnh. Ngay cả khi kiểm tra, xét nghiệm cũng không thể phát hiện sự thay đổi tình trạng bệnh nhân, khi đó tập hợp các triệu chứng có thể cung cấp thông tin quan trọng cho bác sĩ.

Việc chẩn bệnh được tiến hành như sau. Đầu tiên, các bác sĩ thực hiện kiểm tra thể chất cho bệnh nhân và thu thập các triệu chứng quan sát được. Nếu các triệu chứng được quan sát đủ, một tập hợp các triệu chứng sẽ giúp hệ thống tự tin chẩn đoán bệnh tương ứng. Do đó, sự tương ứng giữa một tập hợp các biểu hiện và một bệnh sẽ hữu ích cho chẩn đoán phân biệt. Hơn nữa, sự giống nhau của các bệnh sẽ được suy ra bằng các triệu chứng.

Tác giả đã được thảo luận mô hình hóa các lý luận chẩn đoán triệu chứng bằng cách sử dụng các ý tưởng cốt lõi của các tập thô vì [12]: lựa chọn (sàng lọc) và chẩn đoán phân biệt bệnh có liên quan chặt chẽ với các quy tắc chẩn đoán thu được bằng các xấp xỉ trên và dưới của một khái niệm được đưa ra. Do đó, bài viết này thảo luận về việc chính thức hóa các quy tắc chẩn đoán y tế có liên quan chặt chẽ với mô hình quy tắc tập thô. Điểm quan trọng là lý luận chẩn đoán y khoa được đặc trưng bởi cơ chế tập trung, bao gồm sàng lọc và chẩn đoán phân biệt, tương ứng với xấp xỉ trên và xấp xỉ dưới của một khái niệm mục tiêu. Hơn nữa, bài viết này tập trung vào việc phát hiện các triệu chứng, có thể được xem là mối quan hệ giữa các quy tắc của các bệnh khác nhau.

Bài viết được tổ chức như sau. Phần 2 cho thấy các đặc điểm của quá trình chẩn đoán y tế. Phần 3 giới thiệu các bộ thô và định nghĩa cơ bản của quy tắc xác suất. Phần 4 đưa ra hai cách chính thức hóa các quy tắc chẩn đoán y tế. Mô hình đầu tiên là một mô hình xác định, tương ứng với mô hình tập thô của Pawlak. Và một ý kiến ​​khác đưa ra một phần mở rộng của các ý tưởng trên trong miền xác suất, có thể được xem như là ứng dụng của mô hình tập thô có độ chính xác thay đổi [14]. Phần 5 đề xuất một mô hình quy tắc mới, bao gồm chính thức hóa các quy tắc để phát hiện các triệu chứng. Cuối cùng, phần 6 là kết luận.

# ĐẶC ĐIỂM QUÁ TRÌNH CHUẨN ĐOÁN Y TẾ

Phần này tập trung vào quá trình chẩn đoán y tế dựa trên quy tắc. Các thảo luận cơ bản về lý luận chẩn đoán y khoa liên quan đến các bộ thô được đưa ra trong [11]

**II.1 RHINOS**

RHINOS là một hệ thống chuyên gia chẩn đoán các trường hợp lâm sàng về đau đầu hoặc đau mặt dựa vào các biểu hiện. Trong hệ thống này, một mô hình chẩn đoán được đề xuất bởi Matsumura [1] được áp dụng cho lĩnh vực này, bao gồm ba loại quy trình lý luận sau: lý luận độc quyền, lý luận bao hàm và lý luận về các triệu chứng.

Đầu tiên, lý luận độc quyền loại trừ một bệnh từ tập các bệnh khi bệnh nhân không có triệu chứng cần thiết để chẩn đoán bệnh đó. Thứ hai, lý luận bao gồm nghi ngờ một bệnh trong đầu ra của quá trình độc quyền khi một bệnh nhân có các triệu chứng cụ thể đối với một bệnh. Cuối cùng, việc nghi ngờ các triệu chứng của các bệnh khác khi một số triệu chứng không thể giải thích được bằng kết luận chẩn đoán sẽ được phân tích gợi ý.

Mỗi lý do là dựa trên quy tắc và tất cả các quy tắc cần thiết cho quá trình chẩn đoán được lấy từ các chuyên gia y tế theo cách sau:

**Quy tắc độc quyền**

Những quy tắc này để sinh ra lý luận độc quyền. Nói cách khác, tiền đề của quy tắc này tương đương với điều kiện cần thiết để kết luận chẩn đoán. Từ cuộc thảo luận với các chuyên gia y tế, sáu thuộc tính cơ bản sau đây được chọn không thể thiếu để xác định tình trạng cần thiết:

1. Tuổi,

2. Vị trí đau,

3. Bản chất của cơn đau,

4. Mức độ nghiêm trọng của cơn đau,

5. Lịch sử kể từ khi khởi phát

6. Đau đầu có bị giật không.

Ví dụ: quy tắc độc quyền của chứng đau nửa đầu thông thường được định nghĩa là:

Để nghi ngờ chứng đau nửa đầu thông thường, cần có các triệu chứng sau:

1. Vị trí đau: không lan xuống mắt,

2. Bản chất: nhói hoặc dai dẳng hoặc tỏa ra,

3. Lịch sử kể tự khi khởi phát: kéo dài hoặc đột ngột

4 Đau đầu giật: có.

Một trong những lý do tại sao sáu thuộc tính được chọn là để giải quyết vấn đề giao diện của các hệ chuyên gia: nếu tất cả các thuộc tính được xem xét, tất cả các triệu chứng sẽ được đưa vào, bao gồm các triệu chứng không cần thiết cho chẩn đoán. Để làm cho lý luận độc quyền nhỏ gọn, chúng tôi chỉ chọn các yêu cầu tối thiểu. Đáng chú ý là loại lựa chọn này có thể được xem là thứ tự của các thuộc tính nhất định, dự kiến sẽ được tạo ra từ cơ sở dữ liệu. Vấn đề này sẽ được thảo luận sau trong chương 6.

**Quy tắc bao hàm:**

Các tiền đề của các quy tắc bao gồm bao gồm một tập hợp các biểu hiện cụ thể cho một bệnh. Nếu một bệnh nhân thỏa mãn một bộ, bệnh này nên được nghi ngờ với một số xác suất. Quy tắc này được bắt nguồn bằng cách hỏi các chuyên gia y tế về các mục sau đây cho mỗi bệnh:

1. Một tập hợp các biểu hiện mà chúng tôi nghi ngờ một bệnh.

2. Xác suất bệnh nhân mắc bệnh với tập hợp các biểu hiện này: SI (Chỉ số thỏa mãn)

3. Tỷ lệ bệnh nhân thỏa mãn tập hợp với tất cả các bệnh nhân mắc bệnh này: CI (Chỉ số bao phủ)

4. Đối với những bệnh nhân mắc bệnh này không thỏa mãn tất cả các biểu hiện thu thập được, quay lại 1.

Do đó, một quy tắc tích cực được mô tả bằng một tập hợp các biểu hiện, chỉ số thỏa đáng (SI), tương ứng với độ chính xác đo lường, và chỉ số bao phủ của nó (CI), tương ứng với tổng tỷ lệ positive.

Lưu ý rằng SI và CI được đưa ra theo kinh nghiệm bởi các chuyên gia y tế.

Ví dụ, một trong ba quy tắc tích cực đối với chứng đau nửa đầu thông thường được đưa ra như sau:

1. Lịch sử: đau kéo dài

2. Đau đầu: có

3. Bản chất: nhói hoặc dai dẳng

4. Triệu chứng không liên tục: không

5. Thời gian kéo dài: hơn 6 giờ

6. Vị trí: không lan xuống mắt

Từ đó ta sẽ suy ra độ chính xác 0,9 (SI = 0,9) và bao gồm 60 phần trăm của tổng số trường hợp (CI = 0,6).

**Phát hiện triệu chứng:**

Quy tắc này được sử dụng để phát hiện các triệu chứng của nhiều bệnh, có được bởi tất cả các biểu hiện có thể có của bệnh. Bằng cách sử dụng quy tắc này, các biểu hiện không thể giải thích bằng kết luận sẽ được kiểm tra, điều này cho thấy các triệu chứng của các bệnh khác. Ví dụ, triệu chứng bệnh của chứng đau nửa đầu phổ biến là:

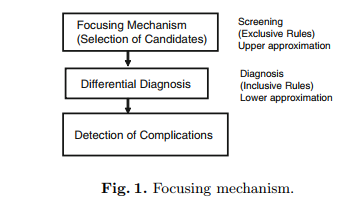
Các triệu chứng sau đây có thể được giải thích bởi

1. Bệnh đau nửa đầu thường gặp: vị trí đau: bất kỳ hoặc

2. Trầm cảm: không hoặc đau đầu giật: có hoặc ...

Do đó, khi một bệnh nhân mắc chứng đau nửa đầu bị trầm cảm, người ta nghi ngờ rằng mình cũng có thể mắc bệnh khác.

**II.2 CƠ CHẾ TẬP TRUNG**



Quá trình quan trọng nhất trong chẩn đoán y tế phân biệt bệnh được trình bày ở trên được gọi là cơ chế tập trung [7,13]. Ngay cả trong chẩn đoán phân biệt bệnh đau đầu, các chuyên gia y tế nên kiểm tra khả năng của hơn 100 ứng viên, mặc dù các bệnh thường gặp là 5 hoặc 6. Những ứng viên này sẽ được kiểm tra bằng lịch sử bệnh quá khứ và hiện tại, kiểm tra thể chất và kiểm tra trong phòng thí nghiệm. Trong các thủ tục chẩn đoán, một bệnh sẽ loại trừ từng người một nếu các triệu chứng cần thiết cho chẩn đoán không phát hiện ra.

Cơ chế tập trung bao gồm hai phong cách sau: lý luận độc quyền và lý luận bao gồm. Mối quan hệ của mô hình chẩn đoán này với một mô hình chẩn đoán khác được thảo luận trong [5,11], được tóm tắt trong hình. 1: Đầu tiên, lý luận độc quyền loại trừ bệnh từ các ứng cử viên khi bệnh nhân không có triệu chứng cần thiết để chẩn đoán bệnh đó. Tiếp theo, lý luận bao gồm nghi ngờ một bệnh trong đầu ra của quá trình độc quyền khi một bệnh nhân có các triệu chứng cụ thể của một bệnh. Dựa trên cuộc thảo luận với các chuyên gia y tế, các quy trình lý luận này được mô hình hóa thành hai loại quy tắc, quy tắc phủ định (hoặc quy tắc độc quyền) tương ứng với lý luận đọc quyền và quy tắc tích cực tương ứng với lý luận bao hàm [1].

# ĐỊNH NGHĨA CƠ BẢN CỦA QUY TẮC

**III.1 Tập thô**

Trong các phần sau, báo cáo sử dụng các ký hiệu sau được giới thiệu bởi GrzymalaBusse và Skowron [4], dựa trên lý thuyết tập thô [2]. Gọi U là tập hợp hữu hạn không trống được gọi là không gian mẫu và A biểu thị một tập hợp các thuộc tính không trống, hữu hạn. Ví dụ a: U → cho a ∈ A, trong đó được gọi là miền của a. Sau đó, một bảng quyết định được định nghĩa là một hệ thống thông tin, A = (U, A∪ {d}). Các công thức trên trường B ⊆ A ∪{d} và V là các biểu thức có dạng [a = v], được gọi là mô tả trên B, trong đó a ∈ B và v ∈ . Tập F (B, V) của các công thức trên B là tập hợp nhỏ nhất thỏa mãn tất cả các công thức trên B và đóng đối với tính phân biệt, kết hợp và phủ định.

Với mỗi f ∈ F (B, V), biểu thị ý nghĩa của f trong A, tức là, tập hợp tất cả các đối tượng trong U có thuộc tính f, được định nghĩa theo quy định như sau:

1. Nếu f có dạng [a = v], thì = {s U | a (s) = v}.

2. = ; = ∨ ; = U - .

**III.2 Phân loại độ chính xác và độ bao phủ**

**III.2.1 Định nghĩa về độ chính xác và độ bao phủ**

Độ chính xác phân loại và phạm vi bảo hiểm hoặc tỷ lệ dương thực sự được xác định như sau:

**Định nghĩa 1:** Gọi R và D là công thức F (B, V) và tập các đối tượng cái mà có được sự quyết đinh d. Khi đó độ chính xác và độ bao phủ (true positive) cho R→ d được định nghĩa như sau:

Đáng chú ý là đo lường mức độ đầy đủ của một đề xuất, R → D và đo lường mức độ cần thiết của nó. Ví dụ: nếu bằng 1.0, thì R → D là đúng. Mặt khác, nếu bằng 1.0, thì D → R là đúng. Do đó, nếu cả hai số đo là 1.0 thì R ↔ D.

**III.3 Quy tắc xác suất**

Bằng cách sử dụng của độ chính xác và độ bao phủ, một quy tắc xác xuất được định nghĩa như sau:

Trong đó D biểu thị một tập hợp các mẫu ra quyết định d. Nếu các ngưỡng cho độ chính xác và độ bao phủ được đạt thành giá trị cao, ý nghĩa của phần có điều kiện của quy tắc xác suất tương ứng với vùng bị chồng chéo cao. Quy tắc này là một loại đề xuất xác suất với hai biện pháp thống kê, đó là sự mở rộng của mô hình chính xác biến số Ziarko (VPRS) [14]

Điều đáng chú ý là cả quy tắc tích cực và quy tắc phủ định đều được xác định là trường hợp đặc biệt của điều này quy tắc, như thể hiện trong các phần tiếp theo.

# CHÍNH THỨC HÓA QUY TẮC CHUẨN ĐOÁN Y KHOA

**IV.1 Mô hình quyết đoán**

**IV.1.1 Quy tắc tích cực**

Một quy tắc tích cực được định nghĩa là một quy tắc chỉ được hỗ trợ bởi các ví dụ tích cực. Do đó, độ chính xác của phần có điều kiện của nó đối với một bệnh bằng 1. Mỗi bệnh có thể có nhiều quy tắc tích cực. Nếu chúng ta tập trung vào tập hợp hỗ trợ của một quy tắc, thì nó tương ứng với một tập hợp con của xấp xỉ thấp hơn của một khái niệm đích, được giới thiệu trong các tập thô [2]. Do đó, một quy tắc tích cực được định nghĩa là:

Trong đó D biểu thị một tập hợp các mẫu thuộc ra quyết định d.

Quy tắc tích cực này thường được gọi là quy tắc xác định. Quy tắc xác định chỉ được hỗ trợ bởi các ví dụ tiêu cực, được gọi là quy tắc phủ định, được giới thiệu giới đây

**IV.1.2 Quy tắc tiêu cực**

Điểm quan trọng là một quy tắc phủ định có thể được biểu diễn dưới dạng giao thoa của quy tắc độc quyền [13]. Một quy tắc độc quyền được định nghĩa là một quy tắc có bộ hỗ trợ bao gồm tất cả các ví dụ tích cực. Đó là, phạm vi bảo hiểm của quy tắc đối với một bệnh bằng 1. Đó là, một quy tắc độc quyền đại diện cho điều kiện cần thiết của một quyết định. Tập hợp hỗ trợ của một quy tắc độc quyền tương ứng với xấp xỉ trên của một khái niệm đích, được giới thiệu trong các tập thô [2]. Do đó, một quy tắc độc quyền được định nghĩa là:

trong đó D biểu thị một tập hợp các mẫu thuộc ra quyết định d.

Tiếp theo ta xem xét sự tương ứng của quy tắc tiêu cực như cách sau đây. Một quy tắc độc quyền được miêu tả như sau

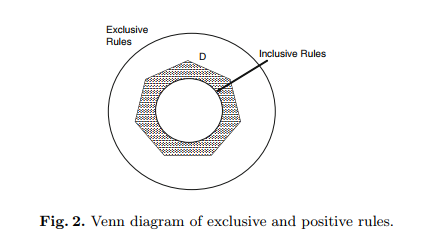
Bởi vì điều kiện của quy tắc độc quyền tương ứng với với các điều kiện cần để kết luận quyết định d. Vì quy tắc tiêu cực tương đương với tính tổng hợp của quy tắc độc quyền, ta được

có nghĩa là nếu một trường hợp không thỏa mãn bất kỳ cặp giá trị thuộc tính nào trong điều kiện của quy tắc phủ định, thì chúng ta có thể loại trừ quyết định d khỏi các ứng cử viên.

Do vậy, quy tắc tiêu cực được định nghĩa như sau :

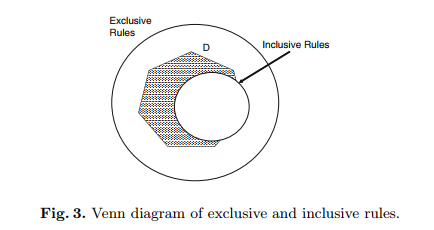
trong đó D biểu thị một tập hợp các mẫu thuộc ra quyết định d.

Các quy tắc tiêu cực cũng nên được đưa vào một danh mục các quy tắc xác định, bởi vì phạm vi bao phủ của chúng, thước đo các khái niệm phủ định, bằng 1,0. Đáng chú ý là tập hợp hỗ trợ quy tắc phủ định tương ứng với một tập hợp con của vùng phủ định, được giới thiệu trong các tập thô [2]. Tóm lại, quy tắc tích cực và tiêu cực tương ứng với các khu vực tích cực và tiêu cực được xác định trong các bộ thô. Hình 2 cho thấy sơ đồ Venn của các quy tắc đó.



**IV.2 Mô hình xác suất**

Mặc dù mô hình xác định ở trên hoàn toàn tương ứng với mô hình tập thô gốc của Pawlak, các quy tắc chẩn đoán phân biệt rất nghiêm ngặt đối với lâm sàng, bởi vì chẩn đoán lâm sàng có thể bao gồm các yếu tố không chắc chắn.Tsumoto [5] làm giảm tình trạng của các quy tắc tích cực và xác định quy tắc bao gồm, trong đó mô hình các quy tắc bao gồm của mô hình RHINOS. Định nghĩa gần giống như quy tắc xác suất được xác định trong chương 3, ngoại trừ các ràng buộc về độ chính xác: ngưỡng cho độ chính xác là đủ cao. Do đó, các định nghĩa của các quy tắc được tóm tắt như sau.



**IV.2.1 Quy tắc độc quyền**

**IV.2.2 Quy tắc bao hàm**

Tóm lại, quy tắc tích cực và tiêu cực tương ứng với các khu vực tích cực và tiêu cực được xác định trong mô hình tập thô thay đổi [14]. Hình 3 cho thấy sơ đồ Venn của các quy tắc đó.

Tsumoto giới thiệu một thuật toán để tạo ra các quy tắc độc quyền và toàn diện là PRIMEROSE-REX và tiến hành xác nhận thử nghiệm và so sánh các kết quả cảm ứng với các quy tắc được mua thủ công từ các chuyên gia y tế [5]. Kết quả cho thấy các quy tắc không bao gồm các thành phần của lý luận chẩn đoán phân cấp. Các chuyên gia y tế phân loại một tập hợp các bệnh thành các nhóm bệnh tương tự và lý do chẩn đoán của chúng là đa giai đoạn: đầu tiên, các nhóm bệnh khác nhau được kiểm tra, sau đó chẩn đoán phân biệt cuối cùng được thực hiện với nhóm bệnh được chọn. Để mở rộng phương pháp thành quy tắc chẩn đoán phân cấp, một trong các tác giả đề xuất một số cách tiếp cận để phân loại khai thác từ một bộ dữ liệu trong [6,9,10**].**

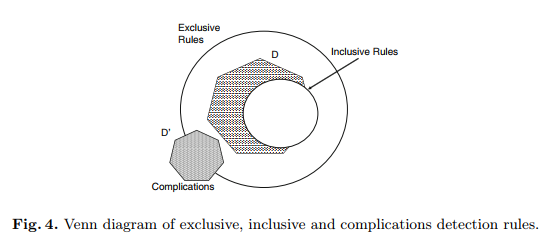
# PHÁT HIỆN TRIỆU CHỨNG

Các mô hình quy tắc trước đây không bao gồm lý do về việc phát hiện các triệu chứng, được giới thiệu là hình ảnh bệnh như trong chương 1. Ý tưởng cốt lõi là các chuyên gia y tế phát hiện các triệu chứng không xảy ra thường xuyên ở các ứng viên chẩn đoán cuối cùng. Ví dụ, chúng ta hãy giả sử rằng một bệnh nhân bị đau đầu do co cơ, người thường than phiền về cơn đau dai dẳng, cũng phàn nàn về cơn đau thần kinh tọa, nói rằng anh ấy cảm thấy đau dữ dội mỗi tháng một lần. Tình huống này rất bất thường và vì chứng đau nửa đầu thường xuyên được quan sát bởi chứng đau nửa đầu, các chuyên gia y tế nghi ngờ rằng anh ấy bị đau đầu do co cơ và đau nửa đầu thông thường. Do đó, một tập hợp các triệu chứng không hữu ích để chẩn đoán bệnh có thể quan trọng nếu chúng thuộc tập hợp các triệu chứng thường biểu hiện ở các bệnh khác. Nói cách khác, tập hợp các triệu chứng như vậy sẽ là yếu tố phát hiện các triệu chứng. Dựa trên những quan sát này, quy tắc phát hiện triệu chứng có thể được định nghĩa như sau:

**V.1 Quy tắc phát hiện triệu chứng**

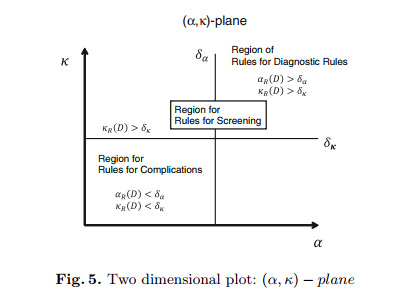
Quy tắc phát hiện triệu chứng của bệnh được định nghĩa là một bộ quy tắc mỗi quy tắc được đưa vào quy tắc bao gồm các bệnh khác:

Hình 4 đã miêu tả mỗi quan hệ giữa quy tắc độc quyền, bao hàm và phát hiện triệu chứng.



**V.2 Phân loại giám sát**

Mối quan hệ giữa ba loại quy tắc có thể được hình dung trong mặt phẳng hai chiều (α, như trong Hình 5. Trục tung và trục hoành biểu thị các giá trị của độ chính xác và độ che phủ. Sau đó, mỗi quy tắc có thể được vẽ trong mặt phẳng với các giá trị độ chính xác và độc che phủ của nó. Vùng dành cho các quy tắc bao hàm được hiển thị ở phía trên bên phải, trong khi khu vực dành cho phát hiện các triệu chứng nằm ở phía dưới bên trái. Khi một quy tắc của khu vực đó thuộc về một quy tắc bao hàm của bệnh khác, nó được đưa vào quy tắc phát hiện triệu chứng của các bệnh mục tiêu**.**



# Ví dụ minh họa

**VI.1 Bài toán**

Bài toán chuẩn đoán bệnh của bênh nhận:

* Input: tập các triệu chứng của bệnh nhân
* Output: Kết quả chuẩn đoán bệnh nhân mắc bênh gì với độ chính xác (%)

**VI.2 Cách giải quyết vấn đề**

* **Quy tắc độc quyền**

Bệnh nhân ban đầu đưa vào một tập triệu chứng tổng quá được gợi ý từ các chuyên gia y tế.

Ta có một bảng tri thức được thu thập từ các bệnh nhận bị bệnh từ trước như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Số thứ tự |  |  | …… |  | Bệnh |
| 1 |  |  | …… |  |  |
| 2 |  |  | …… |  |  |
| …… | …… | …… | …… | …… | …… |
| m |  |  | …… |  |  |

Trong đó m là số bệnh nhân bị bệnh trước đây, là giá trị xem người thứ i có xuất hiện triệu chứng hay không. là loại bệnh h, coi như là kết luận từ các triệu chứng .

Các thực hiện như sau:

1. Lặp trên từng bệnh trong bảng tri thức như miêu tả. Với mỗi bệnh ta có tập hợp các hàng D là tập các triệu chứng đưa ra bệnh
2. Từ tập các triệu chứng người dùng đưa ra, với mỗi triệu chứng , ta chiếu trên bảng tri thức trên, được tập hợp R là các hàng có bệnh và có triệu chứng . Nếu D R thì bệnh nhận có thể mắc bệnh , cho vào tập hợp G
3. Sau tất cả vòng lặp, ta sẽ có kết quả G= {} gồm các bệnh mà bệnh nhân có thể mắc phải.

* **Quy tắc bao hàm**

Từ tập các kết quả G= {} của quy tắc bao hàm trên, với mỗi bệnh người bệnh đưa ra các triệu chứng chi tiết , những thuộc tính trong tập các triệu chứng được đưa ra bởi các chuyên viên y tế.

Ta cũng có bảng tri thức như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Số thứ tự |  |  | …… |  | Bệnh |
| 1 |  |  | …… |  |  |
| 2 |  |  | …… |  |  |
| …… | …… | …… | …… | …… | …… |
| u |  |  | …… |  |  |

Trong đó u là số bệnh nhân bị bệnh trước đây, là giá trị xem người thứ i có xuất hiện triệu chứng hay không. là loại bệnh h, coi như là kết luận từ các triệu chứng .

Các bước thực hiện như sau:

1. Với mỗi bệnh thuộc G, ta thực hiện lặp từng bệnh đó. Với mỗi bệnh ta có tập hợp các hàng D là tập các triệu chứng đưa ra bệnh
2. Từ tập các triệu chứng người dùng đưa ra, với tập các triệu chứng , ta chiếu trên bảng tri thức trên, được tập hợp R là các hàng có bệnh và có các triệu chứng . Nếu DR/D>và DR/R>

thì ta kết luận được là bệnh nhân mặc bệnh đó ( cũng là các ngưỡng dược quyết định bởi chuyên gia y tế). Còn không thì ta tiếp tục thực hiện vòng lặp bước 1.

* **Phát hiện triệu chứng**

Nếu qua quy tắc bao hàm mà chưa kết luận được bệnh nhân mắc bệnh gì, thì ta sử dụng quy tắc phát hiện biến chứng, vì các triệu chứng có thể không quan trọng để kết luận ra bệnh này nhưng quan trọng để kết luận ra bệnh khác.

Các bước thực hiện như sau:

1. Với mỗi triệu chứng bệnh nhân đưa ra, ta chiếu trên bảng tri thức trong quy tắc bao hàm, rồi nhóm theo từng loại bệnh.
2. Với từng loại bệnh, ta tiếp tục có R là tập các hàng chứa bệnh đó trong nhóm ở bước 1 và D là tập các hàng chứa bệnh từ toàn bộ bảng tri thức.
3. Nếu DR/D>và DR/R> thì ta kết luận được là bệnh nhân mặc bệnh đó ( cũng là các ngưỡng dược quyết định bởi chuyên gia y tế). Còn không thì ta tiếp tục thực hiện vòng lặp bước 1.

# KẾT LUẬN

Bài báo đã thảo luận về việc chính thức hóa lý luận chẩn đoán y tế dựa trên triệu chứng. Lý luận bao gồm ba quá trình, lý luận độc quyền, lý luận bao hàm và phát hiện các triệu chứng, hai lý thuyết trước đây thuộc về một cơ chế tập trung. Trong lý luận độc quyền, một bệnh được loại trừ khỏi các ứng cử viên chẩn đoán khi bệnh nhân không có triệu chứng cần thiết để chẩn đoán. Quá trình tương ứng với sàng lọc. Thứ hai, trong lý luận bao gồm, một bệnh trong số các ứng cử viên được lựa chọn bị nghi ngờ khi một bệnh nhân có các triệu chứng cụ thể của một bệnh, tương ứng với chẩn đoán phân biệt. Cuối cùng, nếu các triệu chứng hiếm khi được quan sát ở bệnh cuối cùng, triệu chứng của các bệnh khác sẽ bị nghi ngờ.

Các nghiên cứu trước đây được khảo sát: một trong những tác giả tập trung vào cơ chế tập trung. Đầu tiên, trong một phiên bản xác định, hai bước được mô hình hóa thành hai loại quy tắc thu được từ các biểu diễn xấp xỉ trên và dưới của một bệnh đã cho. Sau đó, ông mở rộng nó thành quy tắc xác suất. Sau đó, các tác giả chính thức hóa các quy tắc phát hiện triệu chứng trong bài báo này. Ý tưởng cốt lõi là các quy tắc không chỉ đơn giản được chính thức hóa bởi các mối quan hệ giữa một tập hợp các triệu chứng và bệnh, mà bởi các quy tắc giữa một triệu chứng, bệnh mục tiêu và các bệnh khác. Bước tiếp theo sẽ là giới thiệu một thuật toán hiệu quả để tạo ra các quy tắc phát hiện từ dữ liệu phức tạp.

# TÀI LIỆU THAM KHẢO

1. Matsumura, Y., Matsunaga, T., Maeda, Y., Tsumoto, S., Matsumura, H., Kimura, M.: Consultation system for diagnosis of headache and facial pain: “rhinos”. In: Wada, E. (ed.) LP. Lecture Notes in Computer Science, vol. 221, pp. 287–298. Springer (1985)

2. Pawlak, Z.: Rough Sets. Kluwer Academic Publishers, Dordrecht (1991)

3. Pawlak, Z.: Rough modus ponens. In: Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems 98. Paris (1998)

4. Skowron, A., Grzymala-Busse, J.: From rough set theory to evidence theory. In: Yager, R., Fedrizzi, M., Kacprzyk, J. (eds.) Advances in the Dempster-Shafer Theory of Evidence, pp. 193–236. Wiley, New York (1994)

5. Tsumoto, S.: Automated induction of medical expert system rules from clinical databases based on rough set theory. Inf. Sci. 112, 67–84 (1998)

6. Tsumoto, S.: Extraction of experts’ decision rules from clinical databases using rough set model. Intell. Data Anal. 2(3), 215–227 (1998)

7. Tsumoto, S.: Modelling medical diagnostic rules based on rough sets. In: Polkowski, L., Skowron, A. (eds.) Rough Sets and Current Trends in Computing. Lecture Notes in Computer Science, vol. 1424, pp. 475–482. Springer, Heidelberg (1998)

8. Tsumoto, S.: Automated discovery of positive and negative knowledge in clinical databases based on rough set model (2000)

9. Tsumoto, S.: Extraction of hierarchical decision rules from clinical databases using rough sets. Inf. Sci. (2003)

10. Tsumoto, S.: Extraction of structure of medical diagnosis from clinical data. Fundam. Inform. 59(2–3), 271–285 (2004)

11. Tsumoto, S.: Rough sets and medical differential diagnosis. In: Skowron, A., Suraj, Z. (eds.) Rough Sets and Intelligent Systems. Intelligent Systems Reference Library, vol. 42, pp. 605–621. Springer (2013)

12. Tsumoto, S., Tanaka, H.: Induction of probabilistic rules based on rough set theory. In: Jantke, K.P., Kobayashi, S., Tomita, E., Yokomori, T. (eds.) Algorithmic Learning Theory, 4th International Workshop, ALT 1993, Tokyo, Japan, 8-10 November 1993, Proceedings. Lecture Notes in Computer Science, vol. 744, pp. 410–423. Springer (1993). https://doi.org/10.1007/3-540-57370-4

13. Tsumoto, S., Tanaka, H.: Automated discovery of medical expert system rules from clinical databases based on rough sets. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining 96, pp. 63–69. AAAI Press, Palo Alto (1996)

14. Ziarko, W.: Variable precision rough set model. J. Comput. Syst. Sci. 46, 39–59 (1993)