Contents

[I. Clinical Decision Support Systems - Theory and Practice 2](#_Toc337257012)

[A. Data Mining and Clinical Decision Support Systems 2](#_Toc337257013)

[1. Giới thiệu 2](#_Toc337257014)

[2. Khai phá dữ liệu và nhận dạng mô hình thống kê 2](#_Toc337257015)

[3. Khai phá dữ liệu và hệ hỗ trợ ra quyết định lâm sàn 3](#_Toc337257016)

[4. Học có giám sát và học không giám sát 4](#_Toc337257017)

# Clinical Decision Support Systems - Theory and Practice

## Data Mining and Clinical Decision Support Systems

Khai phá dữ liệu và hệ hỗ trợ ra quyết định lâm sàn

### Giới thiệu

Khai phá dữ liệu là một quá trình của việc tìm ra mô hình và mối quan hệ trong tập dữ liệu lớn. Nội dung bao gồm một số lĩnh vực như là nhận dạng mô hình, thống kê, khoa học máy tính, và quản lý cơ sở dữ liệu. Vì vậy, định nghĩa của khai phá dữ liệu chủ yếu phụ thuộc vào quan điểm của người đưa ra các định nghĩa đó. Ví dụ, từ việc nhận dạng mô hình, khai phá dữ liệu được định nghĩa như là quá trình xác nhận tính hợp lệ, mới lạ, và dễ hiểu của mô hình trong tập dữ liệu.

Trong ý nghĩa rộng lớn hơn, mục tiêu chính của khai phá dữ liệu là để chuyển đổi dữ liệu thành thông tin hữu ích hơn. Cụ thể hơn, một trong những mục tiêu chính của khai phá dữ liệu là để khám phá các mô hình mới cho người dùng. Việc phát hiện ra các mô hình mới có thể phục vụ hai mục đích: mô tả và dự báo. Trước đây chỉ tập trung vào việc tìm kiếm các mô hình và trình bày cho người sử dụng trong một hình thức dễ hiểu và giải thích được. Dự đoán liên quan đến việc xác định các biến hoặc các miền trong cơ sở dữ liệu và sử dụng chúng để dự đoán giá trị hoặc hành vi trong tương lai của một số thực thể.

Khai phá dữ liệu thích hợp để hỗ trợ quyết định trong chăm sóc y tế. Các tổ chức y tế phải đối mặt với áp lực ngày càng tăng để cải thiện chất lượng chăm sóc trong khi giảm chi phí. Bởi vì khối lượng lớn dữ liệu được tạo ra trong các cơ sở y tế, không phải ngạc nhiên khi tổ chức y tế có quan tâm đến khai phá dữ liệu để tăng cường thực hành, quản lý bệnh tật, và sử dụng tài nguyên.

Ví dụ:

Một ứng dụng khai phá dữ liệu để chăm sóc sức khỏe đã được thực hiện vào đầu những năm 1990 của công ty United HealthCare. United HealthCare là một công ty quản lý chăm sóc sức khoẻ và đã phát triển hệ thống khai phá dữ liệu đầu tiên, Quality Screening and Management (QSM), để phân tích hồ sơ điều trị từ các thành viên của công ty. QSM kiểm tra 15 phương pháp để nghiên cứu những bệnh nhân có bệnh mãn tính và so sánh sự chăm sóc nhận được từ các thành viên với các tiêu chuẩn và hướng dẫn của quốc gia. Sau đó, kết quả phân tích được sử dụng để xác định các chiến lược nâng cao chất lượng quản lý, và giám sát hiệu quả của các hoạt động đó.Mặc dù không cung cấp hỗ trợ trực tiếp cho việc ra quyết định tại thời điểm đó, những dữ liệu này có thể được sử dụng để cải thiện hướng dẫn lâm sàng.

### Khai phá dữ liệu và nhận dạng mô hình thống kê

Nhận dạng mô hình là một phần trong lĩnh vực khai phá dữ liệu. Nó là khoa học tìm kiếm các mô hình phân tích với khả năng mô tả hoặc phân loại dữ liệu. Mục tiêu là để suy ra từ một tập hợp các dữ liệu làm thuận lợi hơn quá trình ra quyết định. Qua thời gian, phương pháp nhận dạng mô hình đã phát triển thành một lĩnh vực rộng lớn bao gồm nhiều lĩnh vực con, bao gồm cả số liệu thống kê, kỹ thuật, khoa học máy tính, và trí tuệ nhân tạo. Bởi vì sự quan tâm và tham gia nhiều ngành, không ngạc nhiên khi nhận dạng mô hình bao gồm một loạt các phương pháp tiếp cận. Một cách tiếp cận để nhận dạng mô hình được gọi là nhận dạng mô hình thống kê.

Nhận dạng mô hình thống kê chỉ việc sử dụng các phương pháp thống kê để mô hình hoá các dữ liệu.Tóm lại, mỗi mô hình được đại diện bởi một tập hợp các tính năng hoặc các biến liên quan đến một đối tượng. Mục đích là để chọn các tính năng cho phép các đối tượng được phân loại vào một hoặc nhiều nhóm hoặc các lớp.

### Khai phá dữ liệu và hệ hỗ trợ ra quyết định lâm sàn

Với sự phát triển của sức mạnh điện toán và công nghệ y học , các bộ dữ liệu lớn cũng như các phương pháp phân loại dữ liệu đa dạng và phức tạp đã được phát triển và nghiên cứu. Kết quả là, khai phá dữ liệu đã thu hút được sự chú ý đáng kể trong vài thập kỷ vừa qua, và đã có một số lượng lớn các ứng dụng bao gồm cả khai phá dữ liệu và các hệ  hỗ trợ quyết định lâm sàng . Hệ hỗ trợ quyết định được đề cập như là một lớp học của hệ dựa trên máy tính nhằm hỗ trợ quá trình ra quyết định. Bảng 3.1 liệt kê một số ví dụ về các hệ hỗ trợ quyết định sử dụng các công cụ khai phá dữ liệu trong cơ sở y tế.

|  |  |
| --- | --- |
| **Hệ thống (tham khảo)** | **Giải thích** |
| **Hệ thống nhận diện và diễn giải hình ảnh y khoa** | |
| Máy tính hỗ trợ chẩn đoán ung thư vú | Sự khác biệt giữa các nốt vú lành tính và ác tính, dựa trên nhiều tính năng siêu âm |
| Chẩn đoán của chứng rối loạn thần kinh cơ | Phân loại tín hiệu điện đồ (EMG), dựa trên các hình dạng và tỷ lệ phát của đơn vị động cơ hành động tiềm năng(MUAPs) |
| **Hệ thống giáo dục** | |
| Khai thác tài liệu sinh y học | Hệ thống tự động để khai phá MEDLINE cho các tham khảo về gen và protein và đánh giá sự phù hợp của mỗi tham khảo được phân công |

Bảng 3.1 – Ví dụ về hệ hỗ trỡ quyết định lâm sàn và công cụ khai phá dữ liệu dùng nhận dạng mô hình thống kê.

Một hệ hỗ trợ quyết định điển hình bao gồm năm thành phần: quản lý dữ liệu, quản lý mô hình, động cơ tri ​​thức, giao diện người dùn, và người dùng. Một trong những sự khác biệt lớn giữa các hệ hỗ trợ quyết định sử dụng các công cụ khai phá dữ liệu và những hệ thống chuyên gia dựa trên luật là các động cơ tri thức.Trong các hệ thống hỗ trợ quyết định sử dụng hệ thống chuyên gia dựa trên nguyên tắc, các công cụ suy luận phải được cung cấp các sự kiện và các quy tắc liên quan đến chúng, thường được thể hiện trong luật "nếu-thì".Trong ý nghĩa này, hệ thống hỗ trợ quyết định đòi hỏi một lượng lớn các kiến thức tiên nghiệm của người ra quyết định để cung cấp câu trả lời đúng cho các câu hỏi .Ngược lại, hệ thống hỗ trợ quyết định sử dụng công cụ khai thác dữ liệu không đòi hỏi phải có kiến thức tiên nghiệm trên người ra quyết định.Thay vào đó, hệ thống được thiết kế để tìm mô hình mới và các mối quan hệ trong một tập hợp các dữ liệu, hệ thống sau đó áp dụng tri thức mới được phát hiện với một tập dữ liệu mới. Điều này là hữu ích khi một kiến ​​thức tiên nghiệm là hạn chế hoặc không tồn tại.

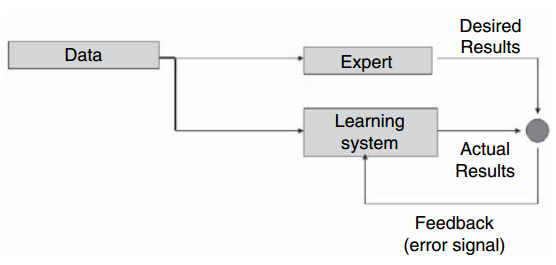
Nhiều hệ thống hỗ trợ quyết định lâm sàng đã thành công bằng cách sử dụng các hệ thống chuyên gia dựa trên nguyên tắc đã được phát triển cho các khu vực chuyên ngành trong lĩnh vực chăm sóc sức khoẻ.Một thí dụ ban đầu của một hệ thống chuyên gia dựa trên nguyên tắc là MYCIN, sử dụng các quy tắc của nó để xác định các vi sinh vật gây ra nhiễm khuẩn huyết và viêm màng não.Tuy nhiên, các hệ thống như vậy có thể được thử thách khả năng duy trì do thực tế các hệ thống thường chứa hàng ngàn các quy tắc hay. Ngoài ra, các hệ thống quy tắc "nếu-thì" có khó khăn trong việc đối phó với sự không chắc chắn. Hệ thống Bayesian là một cách để giải quyết vấn đề không chắc chắn.

### Học có giám sát và học không giám sát

Khai phá dữ liệu và mô hình dự báo có thể được xem như việc học từ dữ liệu. Trong bối cảnh này, khai phá dữ liệu có hai hướng: học có giám sát và học không giám sát.

#### Học có giám sát

Học có giám sát, còn gọi là phá dữ liệu định hướng, giả định rằng người sử dụng biết trước các lớp và có mẫu của các lớp hiện có.(Hình 4.1) Tri thức này được chuyển tới hệ thống thông qua một quá trình gọi là huấn luyện. Các tập dữ liệu được sử dụng trong quá trình này được gọi là mẫu huấn luyện. Các mẫu huấn luyện bao gồm các biến phụ thuộc hoặc biến mục tiêu, và các biến độc lập hoặc đầu vào. Hệ thống này được điều chỉnh dựa trên các mẫu huấn luyện và tín hiệu lỗi (sự khác biệt giữa các phản hồi mong đợi và phản hồi thực tế của hệ thống).Nói cách khác, một hệ thống học có giám sát có thể được xem như là một hoạt động làm giảm sự khác biệt giữa các giá trị mong đợi và giá trị quan sát đoực như là sự tiến bộ quá trình đào tạo.Với đủ các mẫu trong dữ liệu huấn luyện, sự khác biệt này sẽ được giảm thiểu và sự nhận dạng mô hình sẽ được chính xác hơn.



Bảng 4.1 – Học có giám sát.

Mục tiêu của phương pháp này là để thiết lập một mối quan hệ hoặc dự đoán mô hình giữa các biến phụ thuộc và độc lập. Mô hình dự đoán thuộc về học có giám sát vì một biến được qui định là mục tiêu sẽ được giải thích như là một chức năng của biến khác.Mô hình dự báo thường được xây dựng để dự đoán giá trị trong tương lai hoặc hành vi của một đối tượng hoặc thực thể. Bản chất của biến mục tiêu/phụ thuộc xác định loại mô hình: một mô hình được gọi là một mô hình phân loại nếu biến mục tiêu là rời rạc, và một mô hình hồi quy nếu biến mục tiêu là liên tục.

Ví dụ:

Goldman mô tả việc xây dựng một hệ hỗ trợ quyết định lâm sàng để dự đoán sự hiện diện của nhồi máu cơ tim trong 4.770 bệnh nhân có biểu hiện đau ngực cấp tính tại hai bệnh viện trường đại học và bốn bệnh viện cộng đồng. Dựa trên các triệu chứng và dấu hiệu của bệnh nhân, các hệ hỗ trợ quyết định lâm sàng có độ nhạy tương tự nhau (88,0% so với 87,8%) nhưng đặc biệt cao hơn đáng kể (74% so với 71%) ở dự đoán các trường hợp không có nhồi máu cơ tim khi so với quyết định của bác sĩ nếu bệnh nhân đã được yêu cầu vào khoa tim mạch. Nếu quyết định thừa nhận được chỉ dựa trên hệ hỗ trợ quyết định, việc tiếp nhận bệnh nhân không có nhồi máu cơ tim đến các đơn vị chăm sóc mạch vành đã được giảm 11,5% mà không gây ảnh hưởng xấu đến kết quả của bệnh nhân hoặc chất lượng chăm sóc.

#### Học không giám sát

Trong học không giám sát hoặc học vô hướng, hệ thống được trình bày với một tập các dữ liệu, nhưng không có thông tin như thế nào để nhóm các dữ liệu vào các lớp học có ý nghĩa hơn (Hình 4.2). Dựa trên nhận thức sự tương đồng mà hệ thống học phát hiện trong tập dữ liệu, hệ thống phát triển các lớp hoặc cụm cho đến khi một tập các mô hình có thể định nghĩa được bắt đầu xuất hiện. Không có biến mục tiêu, tất cả các biến được đối xử theo cùng một cách mà không có sự phân biệt giữa biến phụ thuộc và biến độc lập.

C:\Users\Binh\Pictures\Capture2.PNG

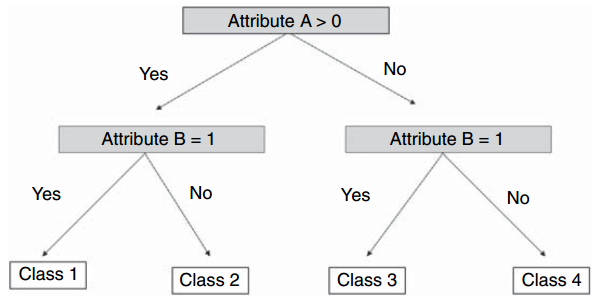
Bảng 4.2 – Học không giám sát

Ví dụ:

Avanzolini phân tích 13 yếu tố sinh lý được theo dõi thường xuyên trong một nhóm 200 bệnh nhân trong khoảng thời gian sáu giờ ngay sau phẫu thuật tim trong một nỗ lực để xác định bệnh nhân có nguy cơ phát triển các biến chứng sau phẫu thuật. Sử dụng một phương pháp học không giám sát (clus-tering), các nhà điều tra cho thấy sự tồn tại của hai loại nguy cơ được xác định rõ bệnh nhân: những người có nguy cơ phát triển các biến chứng sau mổ và những người có nguy cơ thấp.

### Phân lớp với học có giám sát

Trong học có giám sát, phân lớp đề cập đến việc bản đồ hoá của các mục dữ liệu vào một trong những lớp đã được xác định trước. Trong sự phát triển của các công cụ khai phá dữ liệu và các hệ hỗ trợ quyết định lâm sàng sử dụng phương pháp thống kê như mô tả sau, một trong những nhiệm vụ quan trọng là tạo ra một mô hình phân lớp, mô hình sẽ dự đoán các lớp của một số thực thể hoặc các mẫu dựa trên các giá trị của các thuộc tính đầu vào. Chọn đúng phương pháp phân lớp là một bước quan trọng trong quá trình nhận dạng mô hình. Một loạt các kỹ thuật đã được sử dụng để có được sự phân lớp tốt. Một số các kỹ thuật được sử dụng rộng rãi và được nhiều người biết đến được sử dụng trong khai phá dữ liệu bao gồm các cây quyết định, hồi qui logic, mạng thần kinh, và cách tiếp cận gần nhất.



Bảng 5.1 – Cây quyết định đơn giản.

#### Cây quyết định

Việc sử dụng cây quyết định có lẽ là phương pháp dễ hiểu nhất và được sử dụng rộng rãi nhất của học có giám sát.Hình 5.1 là một cây quyết định đơn giản sử dụng hai thuộc tính. Một hệ thống cây quyết định điển hình thông qua một chiến lược từ trên xuống trong việc tìm kiếm một giải pháp.Nó bao gồm các nút nơi mà thuộc tính được dự đoán được thử nghiệm. Tại mỗi nút, thuật toán kiểm tra tất cả các thuộc tính và tất cả các giá trị của mỗi thuộc tính liên quan để xác định thuộc tính và giá trị của thuộc tính sẽ tách dữ liệu "tốt nhất" vào đồng nhất hơn nhóm phụ có liên quan đến các biến mục tiêu.Nói cách khác, mỗi nút là một câu hỏi phân lớp và các nhánh của cây được phân vùng dữ liệu vào các lớp khác nhau. Quá trình này lặp đi lặp lại cho đến khi không thể tách dữ liệu nữa. Do vậy, các nút cuối ở phần cuối của các cây đại diện cho các lớp khác nhau.