ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

LÊ THÀNH CÔNG - 1612842

ĐỒ ÁN (seminar) MÔN HỌC
TRÍ TUỆ NHÂN TẠO

ĐỀ TÀI : MẠNG BAYES – GIỚI THIỆU VÀ TÍNH CHẤT ĐỘC LẬP
DỰA TRÊN TÀI LIỆU : BAYESIAN NETWORK REPRESENTATION AND INDEPENDENCE – UC BERKELEY CS188 INTRO TO AI

TP.HCM – /2018

Mục lục

TÓM TẮT ĐỒ ÁN

Đồ án trình bày những vấn đề cơ bản về mạng Bayes chủ yếu gồm giới thiệu các khái niệm cơ bản cấu thành mạng Bayes, tập trung khai thác tính chất độc lập có điều kiện và áp dụng tính chất đó để xác định quan hệ giữa các biến trong đồ thị của mạng Bayes. Nội dung đồ án gồm 4 chương tất cả. Chương 1 giới thiệu sơ lược về các mô hình xác suất. Chương 2 đi vào khái niệm độc lập, độc lập có điều kiện và quy tắc mắc xích kèm theo các ví dụ minh họa. Chương 3 tập trung làm rõ mạng Bayes gồm các khái niệm, kí hiệu, đặc điểm, mối quan hệ, công thức giữa các phần tử trong mạng. Chương 4 chủ yếu trình bày thuật toán D-seperation giúp xác định các biến nào độc lập hay độc lập có điều kiện với nhau trong mạng Bayes để từ đó suy ra các phân phối phù hợp với điều kiện. Phần tổng kết tóm gọn lại những gì đã trình bày cũng như nội dung cốt lõi của đồ án. Phụ lục gồm danh mục các tài liệu tham khảo.

NỘI DUNG

Trong thực tế, có rất nhiều vấn đề đặt ra yêu cầu thao tác trên một lượng lớn dữ liệu sưu tập từ hiện tượng, đối tượng xảy ra nào đó, chẳng hạn: triệu chứng bệnh nhân, trang web, thông tin tín dụng,…Khi đó, việc cần làm đó là rút trích các thông tin hữu ích từ nguồn thông tin khổng lồ đó để có thể đưa ra quyết định tối ưu đáp ứng sự kiện nào đó hoặc dự đoán tương lai. Đây là một dạng bài toán điển hình trong lĩnh vực Trí Tuệ Nhân Tạo, từ một bộ dữ liệu cho trước ta dùng một phân lớp phù hợp với nó, dùng một số dữ liệu để đánh giá bộ phân lớp và sau đó dùng bộ phân lớp này để phân lớp các dữ liệu mới. Có nhiều hướng tiếp cận để giải quyết bài toán này như cây quyết định, mạng neural,…trong đó mạng Bayes là cách tiếp cận đơn giản và có nền tảng toán học chặt chẽ. Đây là mô hình đồ thị xác suất được áp dụng trong các bài toán liên quan đến một lượng lớn các biến ngẫu nhiên và có thể biểu diễn một cách tự nhiên các mối quan hệ độc lập của các biến bằng đồ thị. Với ưu điểm trong sáng về mặt ngữ nghĩa, dễ kết hợp với các phương pháp quyết định tối ưu, phương pháp này thường được dùng chủ yếu biểu diễn tri thức không chắc chắn và ngày càng đóng vai trò quan trọng trong các hệ chuyên gia, hệ hỗ trợ quyết định.

Mạng Bayes phát triển từ những năm cuối thập kỷ 70 của thế kỉ XX, kết hợp từ nhu cầu mô hình hóa, biến cố quan sát và tri thức có trước. Kết hợp suy luận 2 chiều với nền tảng chặt chẽ của lý thuyết xác suất làm cho mạng Bayes trở thành một lựa chọn ưu tiên trong việc xây dựng hệ chuyên gia.

Chương 1. Các mô hình xác suất - Probabilistic models

1.1. Mô hình xác suất

Các mô hình xuất suất về cơ bản chỉ có thể xấp xỉ so với những gì diễn ra trong thực tế, chỉ có thể mô phỏng một phần nào đó cách thế giới làm việc.

Ví dụ: Kết quả tung đồng xu về mặt lý thuyết là 50% cho mỗi mặt đồng xu, nhưng trên thực tế xác suất này chỉ xấp xỉ và tiệm cận 50% chứ không thể hoàn toàn chính xác. Số lần tung đồng xu càng nhiều, xác suất càng tiệm cận với giá trị trên lý thuyết.

Các mô hình này luôn luôn đơn giản hóa nhiều nhất có thể, nó không thể đại diện được mọi biến cố cũng như mọi sự tương tác qua lại giữa các biến cố trong thế giới thực. Điều quan trọng chúng ta cần lưu ý: "All models are wrong; but some are useful." – George E. P. Box. Tất cả các mô hình đều sai, không có cái nào đúng cả nhưng chúng sẽ có tác dụng để chúng ta xử lý một phần vấn đề nào đó. Như vậy, trong quá trình thu thập dữ liệu, người thu thập không nên bỏ sót bất kỳ biến cố nào có lý bởi nhất định chúng sẽ có ích sau này.

Thông thường, chúng ta làm việc với các mô hình luôn có những xác suất của các biến cố không được cho trực tiếp mà phải thông qua tính toán. Cách đơn giản nhất là áp dụng luật Bayes (Bayes rule). Ta có điều kiện nhưng điều kiện đối mới là cái chúng ta cần tìm và luật Bayes cho phép chúng ta giải quyết vấn đề đó và đạt được xác suất cần tìm. Khi chúng ta xác định một mô hình, chúng ta không có một mô hình cụ thể về khả năng trạng thái ở thời điểm T và cho tất cả thông tin của thời gian T sau khi quan sát được thì chúng ta có một mô hình mịn hơn chỉ để cho tương tác địa phương và bằng cách nào đó từ việc tính toán điều kiện chúng ta quan tâm.

Lý luận chuẩn đoán (diagnostic reasoning): trên các tập lớn các biến cố ngẫu nhiên. Lý luận nhân quả (causal reasoning) khai phá các mẫu có thể xảy ra dựa trên tiền đề cho trước, và điều chúng ta có thể làm tính toán giá trị thông tin (value of information). Có một khái niệm khá hấp dẫn. Ý tưởng ở đây là chúng ta có thông tin không đầy đủ về các biến cố ngẫu nhiên thì bạn muốn trả bao nhiêu cho ai đó để cho bạn biết giá trị của biến cố ngẫu nhiên đó. Đó là câu hỏi mà chúng ta sẽ có thể trả lời ngay khi chúng ta biết được giá trị thông tin.

1.2. Công thức Bayes

$$P(h|D) = \frac{P(h)P(D|h)}{P(D)}$$

Công thức Bayes là tâm điểm và là nền tảng của phương pháp học Bayes, nó cho phép tính giả thiết có khả năng cao nhất P(h|D) từ xác suất tiên nghiệm P(h) của giả thiết, xác suất quan sát dữ liệu P(D) và xác suất dữ liệu được quan sát khi biết trước giả thiết P(D|h).

Ý nghĩa:

- Đây là luật được sử dụng hầu hết trong các hệ thống AI hiện đại sử dụng lập luận xác suất, đặc biệt để đánh giá xác suất chuẩn đoán từ xác suất nguyên n hân
- Rất hữu dụng vì trong thực tế người ta thường biết phần bên phải công thức và cần ước lượng phần bên trái.

Ví dụ: m là bệnh viêm màng não và s là bệnh cứng cổ

P(s|m) = 0.5

P(m) = 1/50000

P(s) = 1/20

P(m|s) = $\frac{P(S|m)P(m)}{P(s)} = 0.5 * \frac{20}{50000} = 0.0002$

Chương 2. Độc lập giữa các biến ngẫu nhiên

2.1. Độc lập tuyệt đối - Independence

Tóm tắt nhanh một cặp tính chất mà chúng ta cần xem xét trước và rất quan trọng trong đề tài này chính là tính độc lập (independence). Hai biến cố A và B gọi là độc lập nhau nếu việc xảy ra hay không xảy ra biến cố này không làm thay đổi xác suất xảy ra của biến cố kia và ngược lại.

Hai biến gọi là độc lập với nhau khi và chỉ khi: $\forall x, y : P(x, y) = P(x)P(y)$

Xác suất hợp (Joint probability – là xác suất của hai biến cố cùng xảy ra) bằng tích hai xác suất biên (Marginal probability – xác suất của một biến cố mà không quan tâm đến biến cố khác)

Ta có thể viết dưới dạng khác: $\forall x, y : P(x|y) = P(x)$

$$\forall x, y : P(y|x) = P(y)$$

Ký hiệu giản lược: $X \perp\!\!\!\perp Y$

Tính độc lập tuyệt đối này giúp giảm độ phức tạp của mô hình giả định. Nhưng trong thực tế, hai biến cố hiếm khi độc lập tuyệt đối với nhau bởi thì thường các biến cố tác động qua lại lẫn nhau bằng cách nào đó nhưng có lẽ trong vài trường hợp tập hợp con của các biến này độc lập với tập hợp con biến khác.

Ví dụ: Nếu chúng ta xem xét các biến cố sau: thời tiết, giao thông tắc nghẽn, lỗ hỏng và đau răng. Có lẽ lỗ hỏng và đau răng độc lập với thời tiết và giao thông tắc nghẽn. Không có sự tác động trực tiếp ở đây, đây là mô hình giả định nên bạn sẽ không bao giờ biết thực sự nó đúng hay sai nhưng đó là cách ước đoán xấp xỉ tốt với điều xảy ra trong thực tế.

Ví dụ: $P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Bảng 1. Phân bố xác suất của P_1(T, W)

Ta được cho trước bảng phân phối chung ở trên. Làm cách nào để kiểm tra tính độc lập các biến cố. Ta xây dựng bảng xác suất không điều kiện cho mỗi biến trong bảng phân phối chung trên. Trong trường hợp này là 2 biến T, W nên ta được hai bảng xác suất không điều kiện sau:

P(T)

T	P
hot	0.5
cold	0.5

P(W)

W	P
sun	0.6
rain	0.4

Bảng 2-3. Xác suất P(T) và P(W)

Sau đó ta nhân lại các xác suất biên để xây dựng bảng phân phối mới theo công thức $\forall x, y : P(x, y) = P(x)P(y)$

$P_2(T, W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3

cold	rain	0.2

Bảng 4. Phân bố xác suất của P_2(T, W)

Nếu giá trị được xây dựng ở bảng mới trùng với bảng cũ, ta kết luật hai biến này độc lập với nhau nhưng nếu có bất kỳ giá trị nào tương ứng khác nhau thì hai biến không độc lập. Trong trường hợp này, hai biến T và W không độc lập tuyệt đối.

Ví dụ: Tung đồng xu N đồng xu cùng lúc. Giả sử 8 người cùng tung đồng xu cùng lúc. Bạn xem xét phân phối chung dựa trên kết quả tung đồng xu, đó là phân phối chung trên 8 biến ngẫu nhiên, cái mà có 2 – 8 mục, 1 mục cho mỗi khởi tạo chung có thể. Nhưng kết quả của mỗi lần tung đồng xu là độc lập nên chúng ta có thể tạo được cùng thông tin bằng việc giả định chúng độc lập và chỉ cần 8 bảng như sau:

P(X_1)

H	0.5
T	0.5

P(X_2)

H	0.5
T	0.5

P(X_n)

H	0.5
T	0.5

Bảng 5-6-7. Xác suất của P(X_1), P(X_2) và P(X_3)

Mỗi bảng chỉ cần 2 mục nếu có n đồng xu thì sẽ có 2n mục tại bảng phân phối chung đầy đủ. Đây là cách khá vắn tắt khi trình bày trong trường hợp này. Chúng ta giả định chúng độc lập và lập lại tương tự cho các biến còn lại, sau đó tiếp tục nhân các biến tương ứng lại với nhau để được bảng phân phối xác suất chung.

Đây thực sự là giả định có hạn chế để giả định độc lập hoàn toàn, và điểm hạn chế bởi vì A thường ra giá trị true và B nếu nó độc lập, nếu ta đo một biến thì không có nghĩa ta biết được mọi thứ về các biến khác. Có một số biến ta không thể đo lường được nhưng một số biến lại có thể đo và chúng ta muốn suy luận về các biến chúng ta không thể đo nhờ vào việc suy luận từ các biến chúng ta có thể đo và ta có thể thực hiện sự suy luận đó. Có sự kết nối giữa các biến chúng không thể độc lập bởi nếu chúng độc lập một sự đo đạc một biến không cho ta biết bất kì về những cái khác.

2.2. Độc lập có điều kiện - Conditional independence

Đây là sự giả định ước đoán yếu hơn độc lập tuyệt đối. Chúng ta cần ít nhất 3 biến để có thể áp dụng. Về mặt trực quan, hai biến ngẫu nhiên X và Y được gọi là độc lập có điều kiện với Z cho trước, nếu một khi Z được cho trước thì giá trị của Y không cung cấp thêm được thông tin gì về X.

Xét ngữ cảnh: một người có thể bị đau răng (toothache) có thể có lỗ sâu (cavity) và nếu robot nha sĩ thăm dò có thể tìm được lỗ sâu (catch).

$$P(Toothache, Cavity, Catch)$$

Nếu ta có lỗ sâu, xác suất để tìm được lỗ sâu đó không phụ thuộc vào việc ta bị đau răng.

$$P(+catch \mid +toothache, +cavity) = P(+catch \mid +cavity)$$

Đây là một giả định có lý do, việc dò được lỗ sâu dựa vào hình dạng của lỗ sâu chứ không phải lỗ sâu đó nó gây đau, hay bản chất có hoặc không có lỗ sâu. Khi ta biết cavity = true, ta sẽ có một phân phối để xem là máy dò tìm sẽ thành công hay không. Nếu ai đó nói với bạn có bị đau răng hay không điều đó không cho bạn thêm bất cứ thông tin gì về việc tìm được lỗ sâu hay không.

Trường hợp nếu không có lỗ sâu ta có công thức tương tự:

$$P(+catch \mid +toothache, -cavity) = P(+catch \mid -cavity)$$

Như vậy, Catch độc lập có điều kiện với Toothache khi biết Cavity

$$P(Catch \mid Toothache, Cavity) = P(Catch \mid Cavity)$$

Ta suy ra được các phát biểu tương đương sau:

P(Toothache | Catch, Cavity) = P(Toothache | Cavity)

P(Toothache, Catch | Cavity) = P(Toothache | Cavity).P(Catch | Cavity) (Do Toothache và Catch độc lập với nhau dựa trên biến điều kiện Cavity nên ta có thể áp dụng công thức suy ra từ độc lập 2 biến kèm theo điều kiện Cavity)

Tóm lại:

- Độc lập không điều kiện hay độc lập tuyệt đối hiếm xảy ra trong thực tế và hầu như không thú vị bởi chúng ta không thể làm được gì nhiều với những phân phối này
- Độc lập có điều kiện là một khuôn mẫu cơ bản và khá chắc chắn về kiến thức để có thể mã hóa hay chuyển thể vào mạng Bayes (Bayesian network)
- X độc lập có điều kiện với Y với Z được cho trước khi và chỉ khi:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

$$\text{Hoặc } \forall x, y, z : P(x|z, y) = P(x|z)$$

$$\text{Hoặc } \forall x, y, z : P(y|z, x) = P(y|z)$$

Chứng minh:

$$P(x|z, y) = \frac{P(x,y,z)}{P(y,z)} = P(z).\frac{P(x, y|z)}{P(y,z)} = P(z).P(x|z).\frac{P(y|z)}{P(z).P(y|z)} = P(x|z)$$

Ký hiệu rút gọn: $X \perp\!\!\!\perp Y | Z$

Ví dụ 1:

Ta xét đến 3 biến sau: giao thông ùn tắc (Traffic), mang dù (Umbrella) và trời mưa (Raining).

Ta chọn Raining là điều kiện cho trước thì 2 biến còn lại sẽ độc lập với nhau:

Traffic $\perp\!\!\!\perp$ Umbrella | Raining

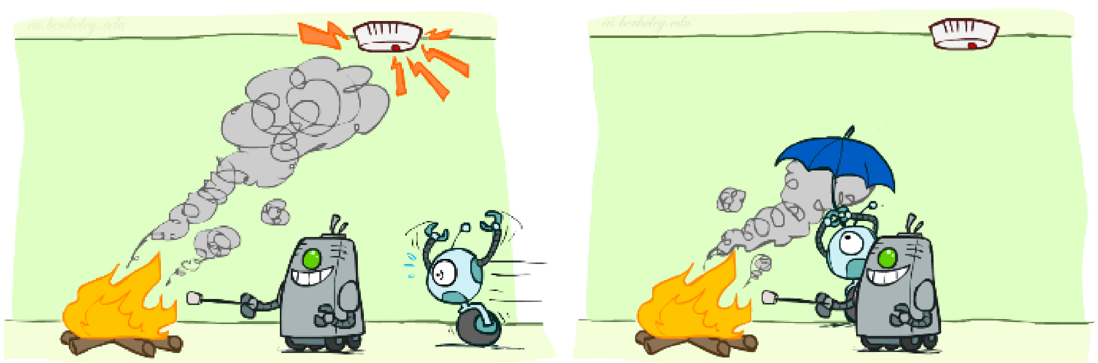
Traffic độc lập với Umbrella với Raining cho trước. Trời mưa hay không mưa sẽ quyết định người ta có mang dù hay không và một cách độc lập đó là mưa hay không thì sẽ quyết định giao thông có ùn tắc hay không, nhưng không có sự tác động ở đây: khi người mang dù làm nhiều người gây ùn tắc. Nhưng nếu chúng ta nghĩ đó là sự tác động qua lại, nếu mọi người lái xe và bắt đầu nhìn vào những người mang dù bởi vì chúng khá ngầu thì khi đó lại gây ùn tắc dẫn đến chúng không độc lập nữa.

Ví dụ 2:

Ta xét: Lửa (Fire), Khói (Smoke), Báo động (Alarm)

$$\text{Fire} \perp\!\!\!\perp \text{Alarm} \mid \text{Smoke}$$

Fire độc lập với Alarm, với Smoke được cho trước.



Hình 1. Hình minh họa cho ví dụ

Nếu có khói, còi báo động sẽ thông báo dựa trên khói, nó không báo động dựa trên việc có nóng hay không mà dựa trên việc có khói vào trong còi báo động vì thế ngay lập tức chúng ta biết có khói. Còn việc có lửa hay không lại không tác động đến việc làm còi báo động hay không. Nếu ta biết có khói nhờ người khác thông báo rằng còi báo động đang reo thì điều này cũng không thể cho ta biết rằng thực sự có lửa hay không. Trong ngữ cảnh này, việc có khói hay không đã cung cấp đầy đủ thông tin về việc có lửa hay không chứ không dựa vào lửa.

2.3. Độc lập có điều kiện và quy tắc mắc xích – Conditional independence and chain rule

Độc lập có điều kiện là chìa khóa để giải quyết tất cả những thứ liên quan sau này mà chúng ta làm về mạng Bayes. Quy tắc mắc xích (chain rule):

$$P(X_1, X_2, \dots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots$$

Điều này luôn đúng đối với bất cứ phân bố chung nào, chúng ta có viết xác suất chung đầy đủ dựa trên tích của các xác suất có điều kiện, cần chọn một thứ tự biến và theo thứ tự đó để biểu diễn công thức trên. Ta có n cách chọn thứ tự nếu có n biến nên có n cách để áp dụng quy tắc mắc xích.

Phân rã tầm thường:

P(Traffic, Rain, Umbrella) = P(Rain).P(Traffic|Rain).P(Umbrella|Rain, Traffic)

Kết hợp quy tắc mắc xích và độc lập có điều kiện:

P(Traffic, Rain, Umbrella) = P(Rain).P(Traffic|Rain).P(Umbrella|Rain) (Do Traffic $\perp\!\!\!\perp$ Umbrella | Rain như đề cập ở ví dụ trên).

Kích thước của bảng phân phối sẽ rất khác biệt, nếu P(Y| X_1 ~ X_k), từ X_1 ~ X_k sẽ tạo nên tổ hợp khá lớn nếu k lớn, mỗi biến ở điều kiện cho trước có 2 lựa chọn true/false thì với k điều kiện thì tổ hợp sẽ có 2^k trường hợp phân phối. Như vậy trong bảng phân phối sẽ có từ 2 ~ 2k entry. Nếu giảm thiểu tối đa được lượng biến cho trước sẽ rất tốt trong việc xử lý dữ liệu sau này, tránh lãng phí được lượng dữ liệu không cần thiết cũng như thời gian xử lý.

Thứ tự của quy tắc mắc xích cần được lưu ý sao cho hợp lý. Do có nhiều cách áp dụng quy tắc mắc xích, chúng ta cần áp dụng nó theo một thứ tự mà độc lập có điều kiện tương ứng phù hợp với biểu thức chúng ta có và ngoài ra có thể loại bỏ đơn giản phần điều kiện cho trước. Một cách hay được áp dụng đó là xét mọi thứ này như một quá trình nhân quả nếu các biến của chúng ta bằng cách nào đó theo trình tự nối tiếp nhau mà không nhất thiết phải chính xác nếu chúng ta nghĩ về nó bằng cách này và thứ tự khi áp dụng quy tắc mắc xích theo quá trình nhân quả đó sẽ cho chúng ta gợi ý để loại bỏ bớt những điều kiện không cần thiết liên quan đến độc lập có điều kiện một cách hợp lý.

Mạng Bayes cũng như các mô hình đồ họa giúp thể hiện các quan hệ độc lập có điều kiện. Dựa trên việc lược bớt những điều kiện cho trước không cần thiết giúp chúng ta xây dựng mạng Bayes tiện lợi hơn kết hợp với độc lập có điều kiện và các ký hiệu trên mô hình giúp việc giải quyết các vấn đề sáng sủa hơn. Ý tưởng của mạng Bayes sẽ khai thác việc áp dụng quy tắc mắc xích cũng như áp dụng độc lập có điều kiện và đơn giản hóa các phân phối có điều kiện giúp tối thiểu hóa các biểu thức để dễ lưu trữ, dễ xác định, dễ làm việc.

Ví dụ: Quy tắc mắc xích mô phỏng quá trình bắt ma

Chúng ta có một mô hình có gắn cảm biến có thể cho biết điều gì đang xảy ra khi nó hoạt động đo đạc. Trường hợp đang xét chỉ có 2 vị trí trên (top) và dưới (bottom). Ta đo đạc trên mỗi hình vuông, mỗi hình vuông có thể trả về màu đỏ hoặc không phải màu đỏ, như vậy có 2 biến ngẫu nhiên. Ngoài ra, trạng thái con ma ở trên top là một biến ngẫu nhiên khác với mục đích xác nhận con ma có ở trên top hay ở dưới bottom. Ta xác định sự tương tác giữa các biến này, nói chung, ta cần xác định phân phối chung đầy đủ (full joint distribution). Ta bắt đầu xác định tương tác địa phương, chúng ta có trước về nơi con ma ở, chúng ta giả định những cảm biến (T, B) độc lập với điều kiện cho trước là nơi con ma ở (G).

T: Cảm biến trên cao là màu đỏ

B: Cảm biến ở dưới là màu đỏ

G: Con ma ở trên (top)

Hai cảm biến T và B có mục đích nhận biết có ma hay không nên chúng hoạt động độc lập có điều kiện với nhau trong trường hợp này dựa trên vị trí G của con ma cho trước, chúng ta xác định các phân phối: 4 trường hợp của T với G, 4 trường hợp của B với G.

Áp dụng quy tắc mắc xích:

$$P(G, B, T) = P(G).\ P(T|G).\ P(B|G, T)$$

Hai cảm biến B và T độc lập có điều kiện với vị trí G con ma cho trước nên chúng ta có thể giản lược:

$$P(G, B, T) = P(G).\ P(T|G).\ P(B|G)\ (*)$$

Cho trước:

P(+g) = 0.5

P(-g) = 0.5

P(+t | +g) = 0.8

P(+t | -g) = 0.4

P(+b | +g) = 0.4

P(+b | -g) = 0.8

T	B	G	P(T,B,G)
+t	+b	+g	0.16
+t	+b	-g	0.16
+t	-b	+g	0.24
+t	-b	-g	0.04
-t	+b	+g	0.04
-t	+b	-g	0.24
-t	-b	+g	0.06
-t	-b	-g	0.06

Bảng 8. Bảng phân bố xác suất của ví dụ tính theo (*)

Sau khi áp dụng công thức (*) ta suy ra được bảng phân phối chung đầy đủ (full joint distribution) của P(T, B, G). Điểm tuyệt vời của quy tắc mắc xích đó là chúng ta không cần phải tính toán toàn bộ mới biết được xác suất tại một vị trí bất kỳ mà chỉ cần có thông tin cho trước và công thức mắc xích ta có thể tính được giá trị xác suất tại vị trí bất kỳ trên bảng phân phối.

Chương 3. Mạng Bayes – Bayesian network

3.1. Bức tranh toàn cảnh

Bayes network là một phương pháp mới để biểu diễn phân phối chung, và rất gần với quy tắc mắc xích. Đây là một cách nghĩ mới, một cách nghĩ trực quan hơn, chúng ta sẽ làm việc liên quan đến đồ thị, và đồ thị này thể hiện phân bố xác suất. Khi chúng ta xét đồ thị này, chúng ta có thể tìm ra được phân phối mà không cần nhìn vào số liệu trong bảng phân phối chung đầy đủ.

Thông thường nếu là biến ngẫu nhiên rời rạc, ta có thể trình bày dưới dạng bảng biểu, nhưng với biến ngẫu nhiên liên tục ta sử dụng hàm mật độ xác suất. Nếu là biến ngẫu nhiên liên tục, ta vẫn có thể dùng Bayes network, những phân phối có điều kiện mà chúng ta làm việc sẽ là hàm mật độ chứ không còn là biến rời rạc nữa. Các bước tiến hành vẫn tương tự như trường hợp biến ngẫu nhiên rời rạc.

3.2. Vấn đề khi sử dụng bảng phân phối chung đầy đủ như một mô hình xác suất

Nếu là trường hợp ít biến, việc sử dụng bảng là hợp lý và chấp nhận được. Nhưng trong trường hợp có nhiều biến, việc sử dụng bảng không còn hiệu quả vì dữ liệu được tổ hợp lại quá lớn gây khó khăn để có thể trình bày một cách rõ ràng. Rất khó để nghiên cứu hay ước tính một cách thực tiễn đối với nhiều biến ở một thời điểm. Thật vậy, chúng ta không thể nào học quá nhiều dữ liệu mà chỉ đạt hiệu quả khi học một tập hợp nhỏ các tham số. Nếu như phải trình bày toàn bộ đặc điểm bảng phân phối dữ liệu bằng tay, thủ công thì thực sự không khả thi. Rõ ràng, chúng ta cần suy nghĩ khác đi, sẽ dễ dàng hơn nếu như tìm được mối liên hệ giữa các biến và sự tương tác lẫn nhau, đó là ý tưởng xây dựng mạng Bayes sau này.

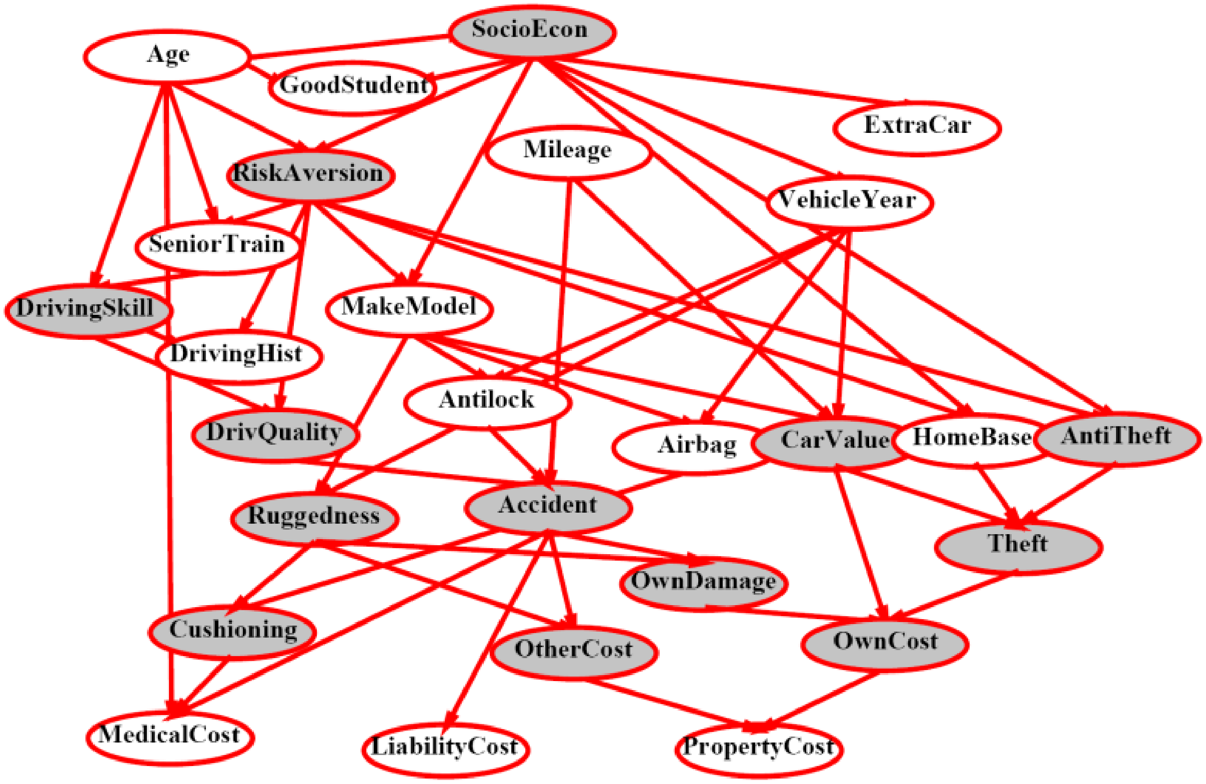
3.3. Đặc điểm mạng Bayes

Định nghĩa: Mạng Bayes là một kỹ thuật cho việc mô tả các phân phối chung phức tạp bằng việc sử dụng phân phối đơn giản dựa vào xác suất có điều kiện.

- Thường có tên gọi là mô hình đồ thị (graphical models). Về kỹ thuật, không chỉ là áp dụng luật Bayes trong mạng Bayes mà là được áp dụng một cách hiệu quả khi có nhiều hơn 2 biến trở nên phức tạp hơn
- Mô tả các biến tương tác nội bộ với nhau
- Tương tác nội bộ mắc xích với nhau tạo thành tương tác toàn cục, tương tác gián tiếp.

Ví dụ 1:

Mạng Bayes được cho như sau:



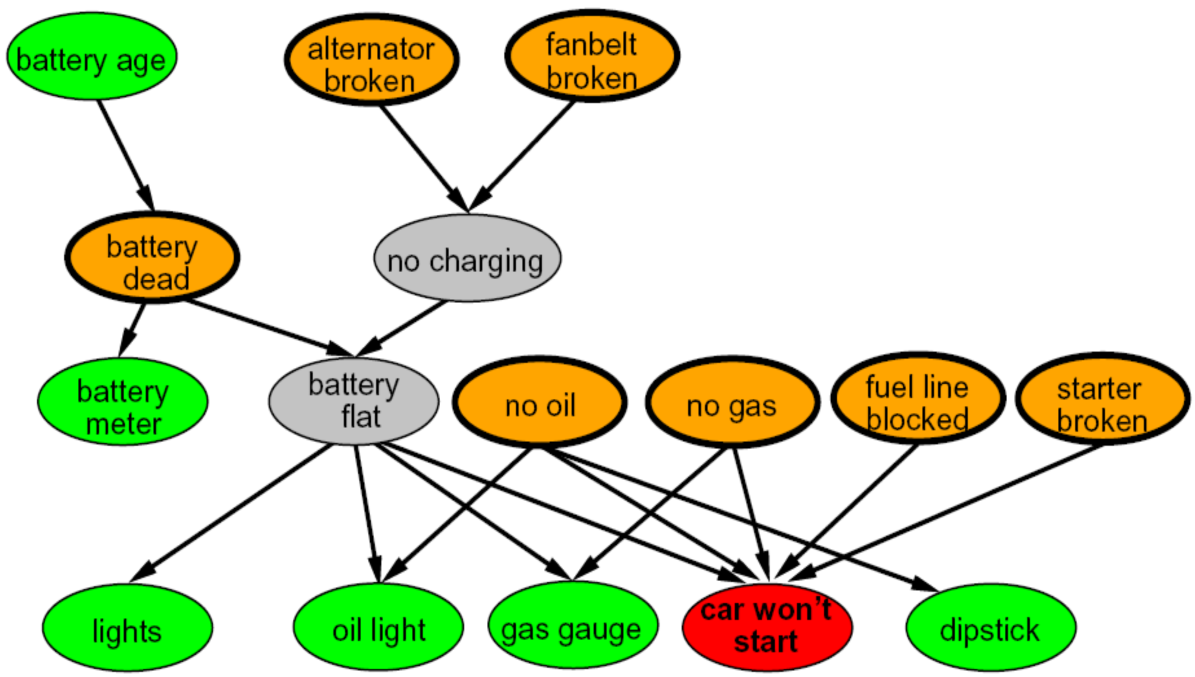
Hình 2. Mạng Bayes mô phỏng các thông tin phục vụ phân tích bảo hiểm

Chúng ta muốn xác định một phân phối dựa trên các mối hệ của các biến như hình trên giải quyết bài toán bảo hiểm. Ở đây có tất cả 27 biến, mỗi biến có giá trị true/false như vậy sẽ có khoảng $2^{27} \approx 134$ triệu entry của bảng phân phối. Nếu sử dụng biện pháp thông thường một cách máy móc thì sẽ phải xây dựng một bảng phân phối với 134 triệu dòng, điều này thực sự bất khả thi. Hầu như không ai có thể dành thời gian để làm chuyện này.

Chúng ta cần mô hình hóa sự tương tác nội bộ giữa các biến. Ví dụ, SocioEcon có một mũi tên hướng tới ExtraCar. Liệu có một chiếc xe hơi (ExtraCar) tại gia hay không thì phụ thuộc vào tình trạng kinh tế (SocioEcon), mối liên hệ này hợp lý và dựa trên 2 biến nội bộ. Một học sinh giỏi (GoodStudent) được giả định phụ thuộc vào tình trạng kinh tế (SocioEcon) và độ tuổi (Age) trong mô hình này. Dựa vào mô hình ta có thể thấy các biến ảnh hưởng trực tiếp với nhau, AntiTheft là thiết bị chống trộm được gắn vào xe có vẻ như phụ thuộc vào điều kiện tài chính (SocioEcon) và lo ngại rủi ro (Risk Aversion) của người đó. Nhìn chung, có rất nhiều biến ở đây và thử đặt ra câu hỏi tai nạn có thể xảy ra hoặc không thì đối với một công ty bảo hiểm bạn muốn kết nối để xây dựng nên mô hình quan hệ tương tác dựa trên dữ liệu bạn có thể quan sát về một người và tính toán xác suất xảy ra tai nạn hoặc có thể xác suất của tai nạn đó gây ra nhiều thiệt hại và v.v Dựa trên tính toán mà những rủi ro kỳ vọng mà bạn đảm bảo cho cá nhân nào đó sẽ trả lời được số tiền bảo hiểm hợp lý để trả cho cá nhân đó.

Ví dụ 2:

Cho mạng Bayes về xe hơi:



Hình 3. Mạng Bayes mô phỏng quá trình kiểm tra xử lý sự cố xe hơi

Xét ngữ cảnh: Quá trình xử lý sự cố cho xe hơi

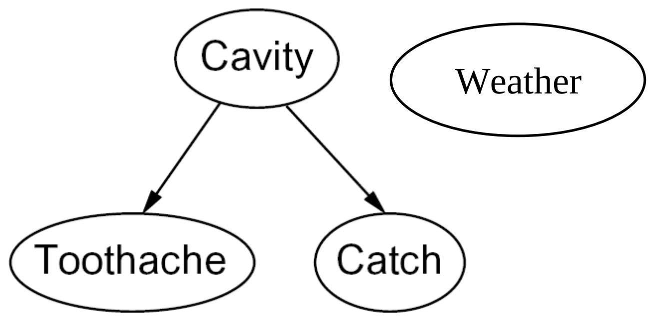
Có những nguyên nhân có thể đi sai hướng và chúng là nguyên nhân gây nên các biến khác cũng đi sai hướng. Trong trường hợp này, có thể máy phát điện bị hư (alternator broken), có thể vành quạt hư (fanbelt broken), chúng lần lượt có thể dẫn đến ắc quy không được sạc (no charging) dẫn đến ắc quy hết năng lượng (battery flat). Cùng lúc đó, ắc quy nếu bị chết (battery dead) cũng có thể dẫn đến ắc quy hết năng lượng (battery flat), và việc này có thể phụ thuộc vào 2 biết trên battery dead và no charging. Chúng ta có nhiều những thông tin chuẩn đoán ở dưới, đó là những thứ có thể đo đạc, chúng là kết quả đo đạc từ các cảm biến biến và máy móc. Trên thực tế, ta có thể đo đạc tất cả biến này nếu có công cụ hợp lý để đo đạc, đặc biệt trong trường hợp này 16 biến nên có khoảng $2^{16} \approx 65000$ trường hợp. Để xác định rõ 65000 trường hợp trong phân phối chung này thực sự rất tốn thời gian cũng như khó để tóm lược những xác suất cho việc khởi tạo mô hình đầy đủ. Tất cả những gì ta cần làm là trình bày cách chúng tương tác và tác động lẫn nhau sau đó xác định tương tác địa phương nội bộ giữa từng biến với nhau. Điều tệ nhất cần xác định trong trường hợp này đó là xe không khởi động được (car won’t start), chúng ta biết kích thước của phân phối có điều kiện là 2^k nếu mỗi biến xác định true/false. k quyết định kích thước phân phối. Biến car won’t start có nhiều biến cha nhất nên phân phối có điều kiện của biến này sẽ là tên lớn nhất trong trường hợp mạng Bayes này.

3.4. Ký hiệu của mô hình đồ thị - Graphical model notation

Nodes: là các biến ngẫu nhiên, có thể được quan sát rồi và có kết quả (assigned) hoặc chưa quan sát (unassigned) tùy thuộc vào các tình huống. Biến chưa quan sát có nền màu trắng, đã quan sát rồi có số liệu rồi có nền màu xám.

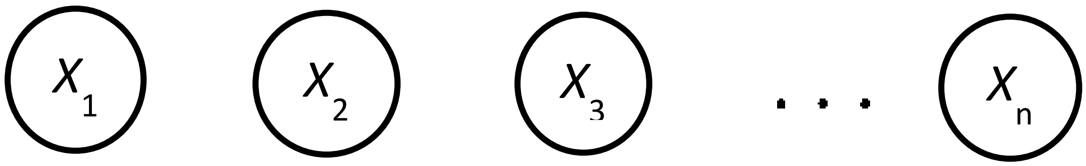
Arcs: thể hiện mối liên hệ, tương tác trực tiếp giữa các biến. Đối với tương tác gián tiếp giữa các biến, đó là đường đi đi qua nhiều nodes gồm nhiều arcs. Nếu

giữa 2 biến không có cung nối, chúng độc lập có điều kiện. Để xác định phân phối của các biến này, chỉ cần xét điều kiện dựa trên node cha (parent) chứ không cần lấy điều kiện trên tất cả các biến khác, đó là cách chính thống. Nhưng bằng trực giác, khi xác định mối liên hệ trực tiếp qua các biến bằng việc thêm cung (arcs).
Ví dụ: Toothache và Catch độc lập có điều kiện theo Cavity và Weather độc lập so với các biến khác



Hình 4. Mạng Bayes minh họa ví dụ

Nếu có lỗ sâu (cavity), điều này dẫn đến đau răng (toothache) hoặc cũng có thể dẫn đến việc tìm được lỗ sâu (catch). Giả thuyết các mũi tên có nghĩa là nhân quả trực tiếp, điều này thường đúng nhưng không phải lúc nào cũng đúng. Về mặt toán học, nếu là quan hệ nhân quả trực tiếp chúng ta sẽ thấy chúng không đúng trong thực tế, nó đúng vì chúng là các loại mô hình dễ làm việc.
Ví dụ:
Tung N đồng xu độc lập thì mạng Bayes lúc này như thế nào?
Đồ thị lúc này không có cạnh vì không có tương tác trực tiếp vì thế nó sẽ như thế này:



Chúng hoàn toàn độc lập với nhau.

Ví dụ: Traffic

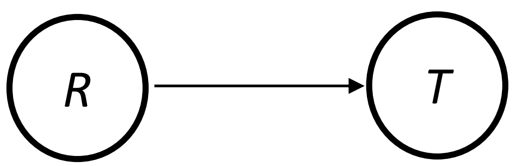
Các biến gồm:

R: Trời mưa

T: Kẹt xe

Model 1: Giả định 2 biến này độc lập với nhau

Model 2: Giả định trời mưa gây nên kẹt xe



Agent nếu sử dụng model 2 sẽ tốt hơn. Nếu sử dụng model 1 nếu ta biết được thông tin về T thì model này không cung cấp được bất cứ thông tin gì thêm về R. Nhưng nếu sử dụng model 2, hiểu được sự tương tác giữa chúng có thể từ một biến với một biến cho trước mà suy luận thêm được thông tin nào đó.

Ví dụ: Traffic II

Xét về giao thông nhưng với ngữ cảnh rộng hơn

Có một con rồng, nó có thể bị đau răng. Có một tình huống kẹt xe. Có một sân vận động bóng đá. Trời có thể mưa hoặc không. Có một ngôi nhà với mái có thể nhỏ giọt nước hoặc không.

Các biến:

T: Kẹt xe

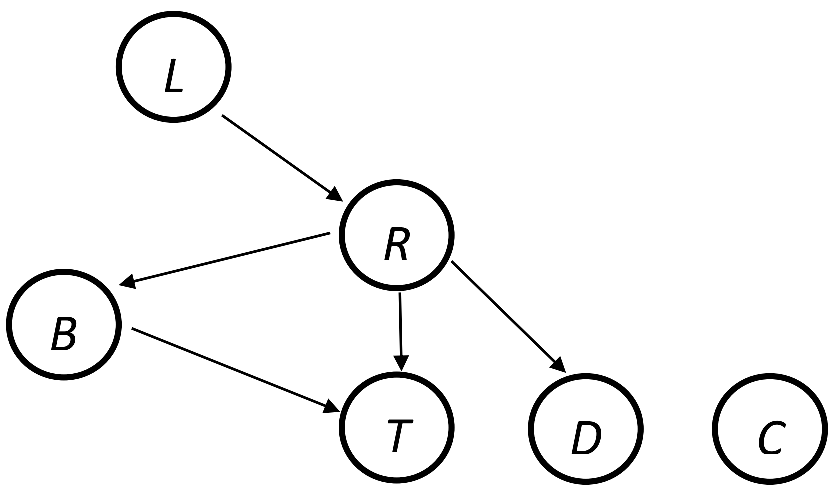
R: Trời mưa

L: Áp suất thấp

D: Mái nhà có nước rỉ

B: Trận bóng

C: Lỗ sâu răng



Hình 5. Mạng Bayes minh họa

Ví dụ: Mạng báo động

Các biến:

B: Tên trộm

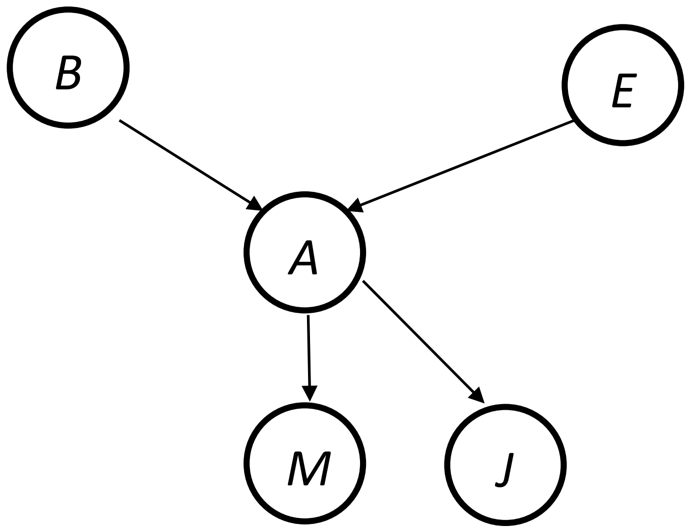
A: Chuông báo động reo

M: Mary gọi

J: John gọi

E: Động đất

Ngữ cảnh: Chuông báo động khi nhận biết có trộm vào nhà sẽ reo, hoặc khi động đất cũng có thể ảnh hưởng làm chuông reo. Khi chuông reo sẽ ảnh hưởng đến nhà hàng xóm là Mary và John, điều này dẫn đến có thể là John hoặc Mary sẽ gọi điện thông báo cho người chủ nhà biết.



Hình 6. Mạng Bayes minh họa

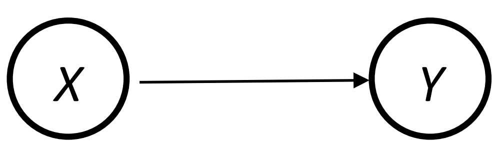
3.5. Ngữ nghĩa mạng Bayes - Bayesian network semantics

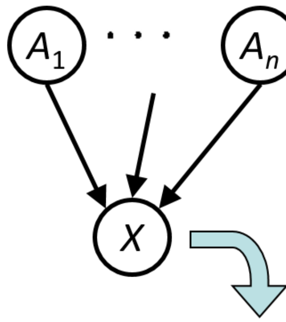
Mạng Bayes: đồ thị có hướng không chu trình mà mỗi nốt được dẫn giải bằng một bảng phân bố xác suất có điều kiện (conditional probability table - CPT)

Tổ chức:

- Tập biến ngẫu nhiên, mỗi biến đại diện một nốt
- Tập các cung kết nối hai nốt. Nếu X nối với Y, ta nói rằng X là cha-parent của Y
- Mỗi nốt có một phân bố xác suất có điều kiện xác định từ các nốt cha:

$$P(X_i \mid \text{Parents}(X_i))$$





$$P(X|A_1 \ldots A_n)$$

Hình 7. Từ đồ thị chuyển thành công thức

A Bayes net = Topology (graph) + Local Conditional Probabilities

Mạng Bayes là sự kết hợp giữa đồ thị và xác suất có điều kiện giữa các biến nội bộ với nhau.

3.6. Kích thước mạng Bayes

Một phân phối chung với N biến kiểu logic: 2^N

Một mạng Bayes gồm N nốt, mỗi nốt có tới k nốt cha: $O(N*2^{k+1})$

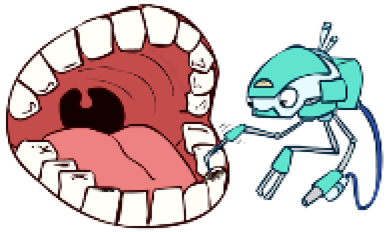
3.7. Xác suất trong mạng Bayes – Probabilities in Bayesian network

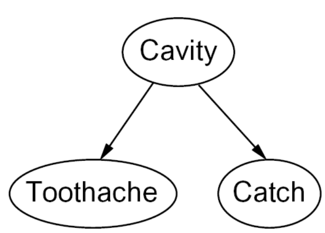
Đây là cách chúng ta kết hợp xác suất vào mạng Bayes. Cách để tính toán một xác suất nào đó từ mạng Bayes đó là nhân lại tất cả xác suất có điều kiện lại với nhau.

Nếu cho trước một mạng Bayes, đó là một đồ thị có hướng không chu trình và mỗi biến có điều kiện cho trước là nốt cha của nó. Chúng ta có thể khôi phục lại phân phối chung đầy đủ dựa trên tập biến bằng việc nhân tất cả phân phối có điều kiện của nốt hiện tại cần tính. Công thức như sau:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\mathit{parents}(X_i))$$

Ví dụ:





Hình 8. Mạng Bayes minh họa

Quay trở lại ví dụ bàn về Toothache, Catch và Cavity. Giả sử ta muốn tính toán:

P(+cavity, +catch, -toothache) =

P(+cavity).P(+catch | +cavity).P(-toothache | +cavity)

Công thức trên được suy ra tương ứng từ mạng Bayes đã cho. Nốt Cavity không có nốt cha nên chỉ cần dùng xác suất không điều kiện P(+cavity). Nốt Catch và Toothache có nốt cha là Cavity nên áp dụng công thức xác suất có điều kiện cho mỗi nốt và tiến hành nhân lại với nhau. Như vậy, khi cho trước một mạng bayes và các thông số xác suất cần thiết, ta có thể tính được một trường hợp xác suất bất kỳ trong bảng phân phối xác suất chung đầy đủ.

Nhưng làm thế nào để đảm bảo rằng $P(x_1, x_2, \dots, x_n) = \prod_1^n P(x_i | parents(x_i))$ trả về kết quả của phân phối chung thích hợp bởi vì không chắc đảm bảo công thức này tính ra số mà khi cộng các xác suất trong bảng phân phối lại ta không được kết quả bằng 1.

Ta xét lại quy tắc mắc xích $P(x_1, x_2, \dots, x_n) = \prod_1^n P(x_i | x_{1 \dots x_{i-1}})$

Quy tắc mắc xích như đã đề cập là luôn đúng với mọi phân phối, có thể có n cách sắp xếp thứ tự các biến khác nhau.

Mặt khác, ta giả định độc lập có điều kiện có công thức như sau:

$$P(x_i | x_1, \dots, x_{i-1}) = P(x_i | parents(x_i))$$

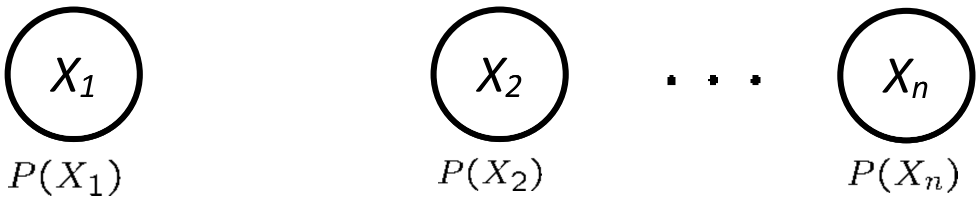
Giả định rằng thứ tự các nốt trước i là nốt cha của I nên $x_1 \sim x_{i-1}$được xem như cha của x_i theo $X_i \perp\!\!\!\perp \{X_1, \dots, X_{i-1}\} \backslash Parents(X_i) | Parents(X_i)$. Nếu độc lập với một số biến, ta hoàn toàn có thể loại bỏ điều ở biến độc lập đó để giảm độ phức tạp. Lưu ý nên sắp xếp thứ tự các biến trong công thức tương ứng với thứ tự các nốt trong đồ thị mạng Bayes.

Không phải mạng Bayes nào cũng có thể đại diện được mọi phân phối chung, mà chỉ có thể đại diện được cho các phân phối chung thỏa giả sử về độc lập có điều kiện.

Ví dụ:

Quay trở lại ví dụ về tung N đồng xu độc lập, đây là mạng Bayes đơn giản nhất với N đồng xu được tung độc lập với nhau, không có mối liên kết giữa các nốt.

Làm thế nào để tính xác suất chung của trường hợp này. Với h là ngửa (head) và t là xấp (tail), xét trường hợp:



h	0.5
t	0.5

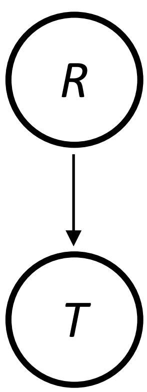
h	0.5
t	0.5

h	0.5
t	0.5

Bảng 9-10-11. Bảng xác suất của các biến $P(X_1)$, $P(X_2)$ và $P(X_n)$

P(h, h, t, h) = 0.5 x 0.5 x 0.5 x 0.5 = 0.0625

Ví dụ: Traffic



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

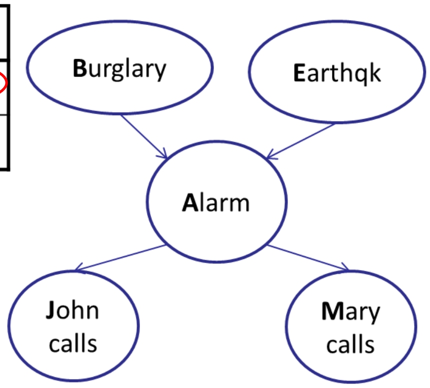
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

Bảng 12-13. Bảng xác suất của các biến P(R) và P(T|R)

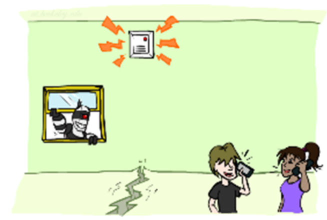
P(+r, -t) = P(+r).P(-t|+r) = 1/4 x 1/4 =1/16

Ví dụ: Mạng báo động

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Bảng 14. Các bảng phân phối của từng biến

Quay trở lại với câu chuyện chuông báo động, giả sử ta cần tính:

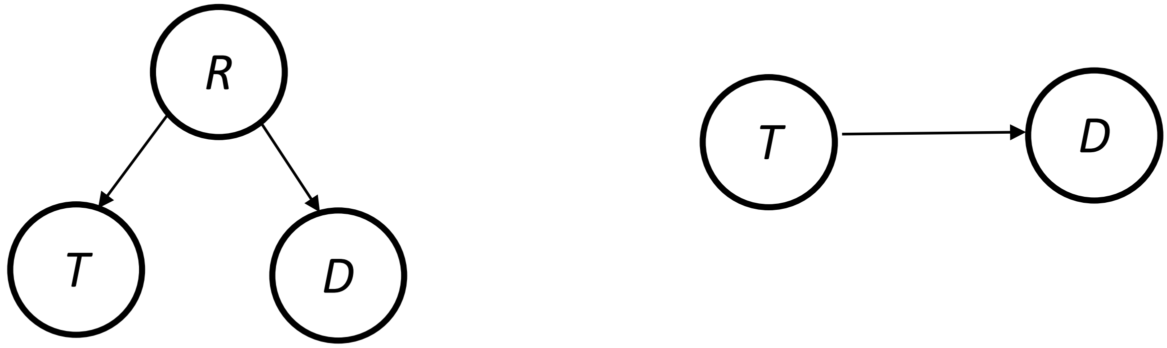
P(+b, -e, +a, +j, +m) = 0.001 x 0.998 x 0.94 x 0.9 x 0.7 = 0.0005910156

3.8. Quan hệ nhân quả? – Causality?

Khi mạng Bayes phản ánh mô hình nhân quả thực sự thường sẽ đơn giản hơn, có ít nốt cha hơn. Hơn nữa, mô hình này sẽ dễ dàng suy nghĩ hơn và dễ hiểu hơn khi người đọc nhìn vào mạng Bayes này. Ngoài ra, chúng sẽ dễ dàng hơn khi suy luận từ các chuyên gia, dễ nói chuyện với họ và đặt các câu hỏi phù hợp với mô hình. Ví dụ hỏi xác suất của việc kẹt xe xảy ra nếu biết đội tuyển Việt Nam vừa thắng đội tuyển Malaysia ở trận chung kết bóng đá vô địch Đông Nam Á? Khi đó chuyên gia học đã có mô hình sẵn và dựa vào mô hình kết quả trận đấu có ảnh hưởng đến giao thông nên họ có thể trả lời cho người hỏi một cách dễ dàng về xác suất xảy ra có kẹt xe hay không.

Nhưng thực tế thì mạng Bayes không cần phải có quan hệ nhân quả. Thỉnh thoảng quan hệ nhân quả không tồn tại trên miền.

Ví dụ:



Hình 9. Mạng Bayes minh họa

R: rain – trời mưa

T: traffic – kẹt xe

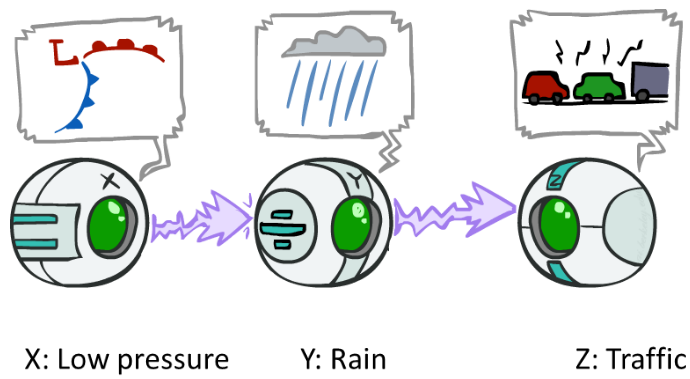
D: drips – mái nhà bị rỉ nước

Ban đầu xét 3 biến ta được mối quan hệ nhân quả nếu có mưa thì gây kẹt xe và nhà bị rỉ nước. Giả sử vì lí do nào đó, biến R – rain không thể sử dụng, mạng Bayes lúc này chỉ còn 2 biến là T – traffic và D – drips, nhưng người dùng lại không muốn 2 biến này độc lập tuyệt đối, phải biểu diễn mối liên hệ tương quan nào đó, lúc này mũi tên vẫn được sử dụng nhưng không phải thể hiện mối quan hệ nhân quả mà là thể hiện sự tương quan giữa các biến.

Chương 4. D-separation

4.1. Nghiên cứu tính chất độc lập với bộ ba

4.1.1. Bộ ba đầu tiên: Chuỗi nhân quả - Causal chains



Hình 10. Bộ ba chuỗi nhân quả - causal chain

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

Chuỗi này được hình thành từ 3 biến, biến X là nguyên nhân gây ra Y, Y là nguyên nhân gây ra Z.

Lúc này X không được đảm bảo sẽ độc lập tuyệt đối với Z bởi vì ta có thể luôn chọn được tập đầu vào (entries) của phân phối có điều kiện như là Z tất yếu sẽ bằng X nên không có sự độc lập giữa X và Z. Chúng ta có thể kết luận được gì từ 2 biến nằm ngoài X và Z nếu cho trước điều kiện của biến ở trong Y.

Ví dụ áp suất thấp gây mưa gây kẹt xe, áp suất cao gây không mưa gây không kẹt xe.

P(+y|+x) =1, P(-y|-x) = 1, P(+z|+y) = 1, P(-z|-y) = 1

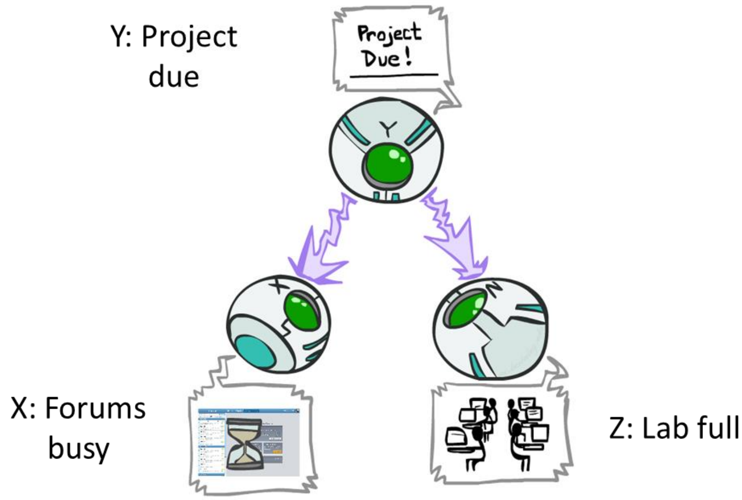
Liệu X có độc lập có điều kiện với Z cho trước Y:

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} = P(z|y)$$

Vậy $X \perp\!\!\!\perp Z \mid Y$

Các biến đã quan sát dọc theo mắc xích ngăn chặn các ảnh hưởng, có một con đường ảnh hưởng từ X đến Z nhưng nếu có điều kiện Y chen giữa làm ngăn chặn ảnh hưởng trực tiếp từ X đến Z.

4.1.2. Bộ ba thứ hai: nguyên nhân chung – common cause



$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

Hình 11. Bộ ba chung nguyên nhân – common cause

Trường hợp này, nốt X và Z đều có nốt cha là Y. X và Z không độc lập hoàn toàn với nhau bởi vì nếu điền vào bảng phân phối có điều kiện đặc biệt, X cùng giá trị với Y, Z cùng giá trị với Y nên X cùng giá trị với Z.

Ví dụ: Đồ án đến hạn làm cho các diễn đàn đông nghẹt và phòng lab đầy.

P(+x|+y) = 1, P(-x|-y) = 1, P(+z|+y) = 1, P(-z|-y) = 1

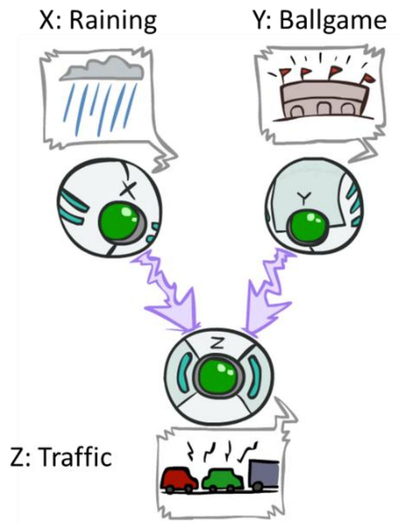
Liệu X có độc lập có điều kiện với Z cho trước Y:

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} = P(z|y)$$

Vậy X $\perp\!\!\!\perp$ Z | Y

Việc quan sát nguyên nhân ngăn chặn ảnh hưởng giữa 2 kết quả bởi vì những lỗi sai biểu thị ảnh hưởng trực tiếp nhưng có biến ở giữa khi được quan sát sẽ ngăn chặn ảnh hưởng của các lỗi này.

4.1.3. Bộ ba thức ba: Kết quả chung – Common effect



Hình 12. Bộ ba chung kết quả - common effect

Hai nguyên nhân, một kết quả.

X và Y liệu có độc lập? P(X,Y,Z) =P(X)P(Y)P(Z|X,Y)

$$\sum_{Z} P(X,Y,Z) = \sum_{Z} P(X)P(Y)P(Z|X,Y)$$

$$\sum_{Z} P(X,Y,Z) = P(X)P(Y)\sum_{Z} P(Z|X,Y)$$

$$\sum_{Z} P(X,Y,Z) = P(X)P(Y)$$

X và Y có độc lập ví dụ như: trận bóng (ball game) và mưa (raining) gây kẹt xe (traffic) nhưng chúng không liên quan với nhau.

X và Y không độc lập có điều kiện với nhau khi Z cho trước bởi vì X và Y đã độc lập tuyệt đối với nhau.

Trường hợp này ngược lại so với các trường hợp khác

Việc quan sát kết quả ảnh hưởng làm kích hoạt ảnh hưởng giữa các nguyên nhân có thể. Nếu nguyên nhân đầu tiên đã đúng thì nguyên nhân thứ 2 không nhất thiết phải đúng nữa.

4.2. Đánh giá các trường hợp phức tạp liên quan đến bộ ba phần tử

4.2.1. Trường hợp phổ quát – The general case

Câu hỏi chung: trong một mạng Bayes cho trước, liệu có 2 biến độc lập hoàn toàn với nhau?

Để giải quyết câu hỏi này cần phân tích đồ thị và cách chúng ta thực hiện nó. Phân rã đồ thị thành các bộ ba kiểm tra xem liệu có ảnh hưởng chạy dọc theo các bộ ba này và nhờ đó chúng ta biết rõ ràng đồ thị rằng liệu có ảnh hưởng nào đang thực hiện hay có sự ngăn cách. Kiểm tra xem giữa 2 biến ta có thể tìm được đường đi mà ảnh hưởng có thể chạy giữa các biến.

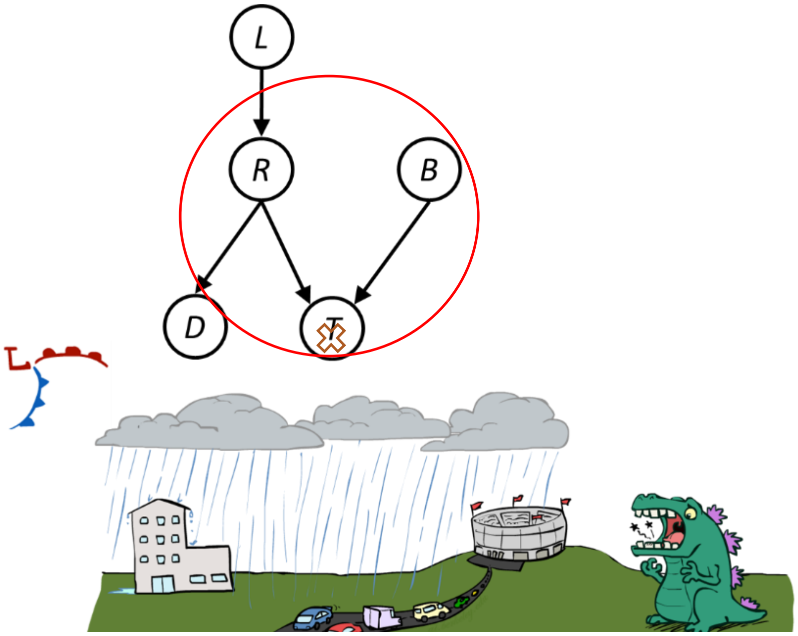
4.2.2. Khả năng tiếp cận - Reachability

Công thức: các nốt quan sát có màu xám, tìm kiếm các con đường trong đồ thị kết quả.

Nếu 2 nốt kết nối gián tiếp nhau không bị chặn bởi nốt xám quan sát thì chúng độc lập có điều kiện

Kiểm tra xem liệu có đường đi nào giữa 2 biến sau đó kiểm tra liệu đường đó có bị ngăn chặn bởi biến đang được quan sát nằm trên đường đi đó, nếu không bị chặn thì có ảnh hưởng dọc theo đường đi. Chúng ta không thể chỉ nhìn vào các con đường và xem chúng bị chặn hay không bởi các biến của chúng có được quan sát hay không. Chúng ta biết rằng chúng ta hầu như ở đó trước các loại cấu trúc này. Quan sát các biến thực sự làm kích hoạt các ảnh hưởng và không quan sát thì ngăn cách con đường.

Ví dụ:

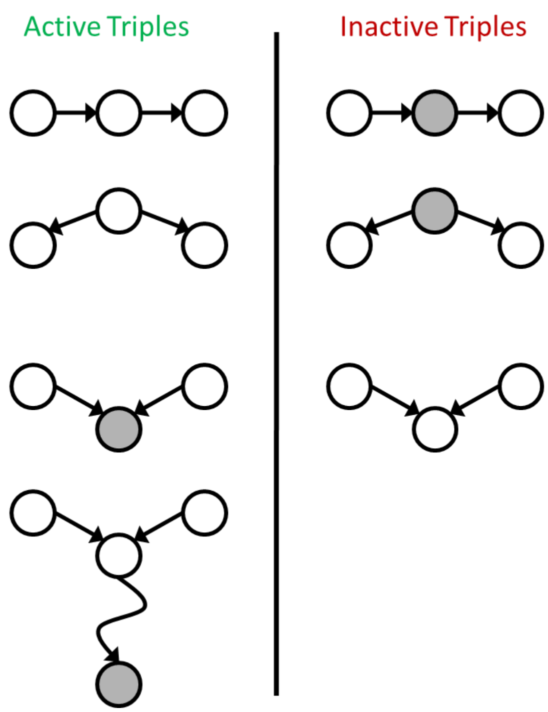


Hình 13. Hình minh họa

Ở tình huống ví dụ trên, cấu trúc v tại T không được tính như một liên kết trên đường đi trừ phi nó được kích hoạt tức là nó đã được quan sát.

4.2.3. Đường đi có hoạt động và không hoạt động – Active/inactive paths

Có 2 loại bộ ba (triple): bộ ba có hoạt động và không hoạt động



Hình 14. Các loại bộ ba hoạt động và không hoạt động

Có thể thấy trường hợp đầu tiên là causal chain với nốt ở giữa chưa được quan sát, đây là bộ ba hoạt động. Nhưng nếu nốt ở giữa được quan sát thì 2 nốt ngoài không còn bị ảnh hưởng nữa và bộ ba này trở về trạng thái không hoạt động. Trường hợp hai là common cause tương tự so với causal chain. Trường hợp thứ 3 là common effect, nếu nốt ở giữa được quan sát thì bộ ba này hoạt động và không hoạt động nếu nốt giữa chưa quan sát. Nhưng có thể có trường hợp nốt giữa chưa quan sát nhưng nốt con của nó có quan sát thì đây cũng gọi là bộ ba hoạt động (active triple).

Một con đường gọi là hoạt động nếu như mỗi bộ ba của nó hoạt động, không hoạt động nếu có một bộ ba không hoạt động.

4.3. D-separation: thuật toán trả lời các truy vấn

$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

Mục đích: Kiểm tra giữa 2 biến bất kỳ có độc lập có điều kiện với nhau hay không?

Kiểm tra tất cả các đường đi giữa X_i và X_j

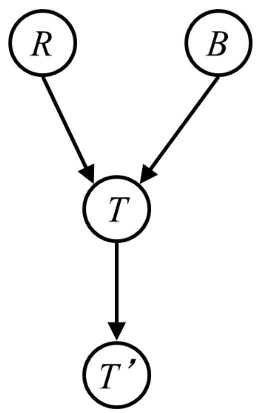
Nếu có 1 hay nhiều đường hoạt động (active), sự độc lập không được đảm bảo:

$$X_i \not\perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

Ngược lại, nếu tất cả đường đều không hoạt động (inactive), độc lập được đảm bảo:

$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

Ví dụ 1:

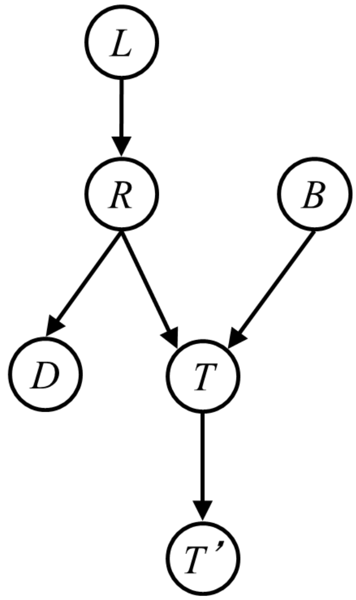


R $\perp\!\!\!\perp$ B ? Đúng

R $\perp\!\!\!\perp$ B | T ? Không đảm bảo đúng

R $\perp\!\!\!\perp$ B | T' ? Không đảm bảo đúng

Ví dụ 2:



L $\perp\!\!\!\perp$ T’ | T ? Đúng

L $\perp\!\!\!\perp$ B ? Đúng

L $\perp\!\!\!\perp$ B | T ? Không đảm bảo đúng

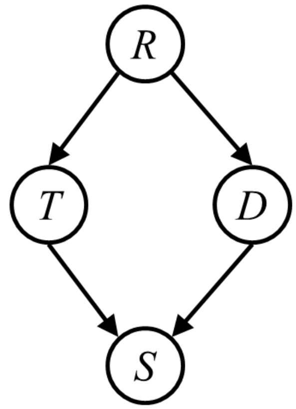
L $\perp\!\!\!\perp$ B | T’ ? Không đảm bảo đúng

L $\perp\!\!\!\perp$ B | T, R ? Đúng

Giải thích mẫu: L $\perp\!\!\!\perp$ B | T, R ? Đúng

Ta xét nốt bắt đầu là L và kết thúc là B, nốt được quan sát là T và R, đường L đến P chỉ có một đường duy nhất, trên đường đó có 2 bộ ba, bộ ba thứ nhất là causal chain L – R – T và bộ ba thứ hai common effect R – T – B. Xét bộ ba thứ nhất causal chain L – R – T, nốt R được quan sát nên bộ ba này inactive, dẫn đến con đường duy nhất này inactive. Vậy L $\perp\!\!\!\perp$ B | T, R là đúng.

Ví dụ 3:



Các biến: R: Raining, T: Traffic, D: Roof drips, S: I'm sad

Các câu hỏi: T $\perp\!\!\!\perp$ D ? Không đảm bảo đúng

T $\perp\!\!\!\perp$ D | R ? Đúng

T $\perp\!\!\!\perp$ D | R, S ? Không đảm bảo đúng

Lý luận tương tự ví dụ trên.

Ý nghĩa cấu trúc – Structure implications

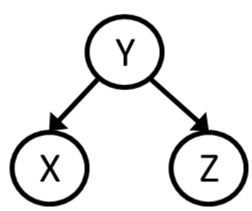
Một khi đã hiểu và áp dụng được D-seperation tức là chúng ta đã có thể đọc hiểu được rõ ràng đồ thị của mạng Bayes. Cho bất cứ tập biến nào, ta vẫn có thể quyết định xem chúng có đảm bảo độc lập có điều kiện với nhau hay không chỉ bằng cách nhìn vào đồ thị.

Cho trước một mạng Bayes, ta có thể xây dựng một danh sách hoàn chỉnh các độc lập có điều kiện theo định dạng:

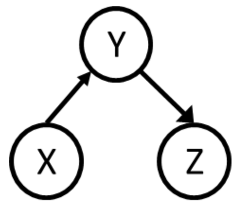
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

Danh sách này quyết định tập hợp các phân bố xác suất có thể được thể hiện.

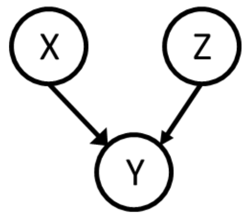
Ví dụ 4: Tìm tất cả các độc lập trong các mạng Bayes sau:



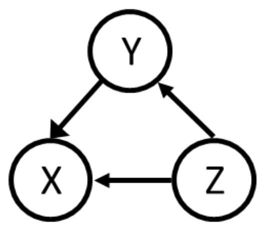
Trường hợp này là common cause: $\{X \perp\!\!\!\perp Z \mid Y\}$



Trường hợp này là causal chain: $\{X \perp\!\!\!\perp Z \mid Y\}$



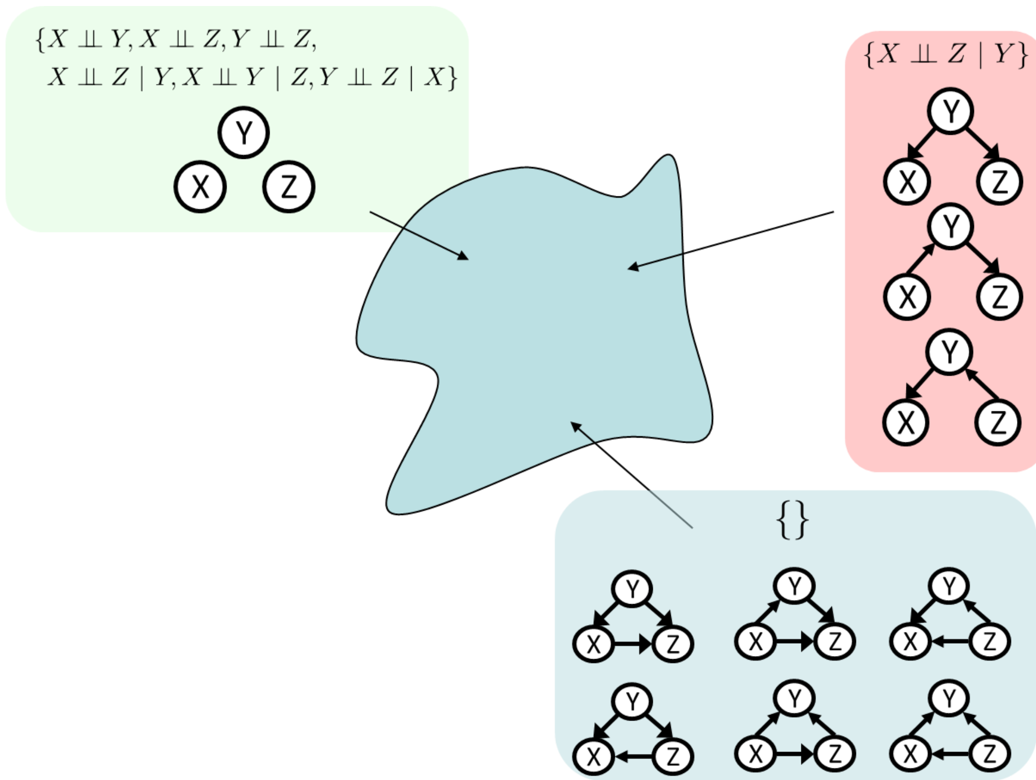
Trường hợp này là common effect: $\{X \perp\!\!\!\perp Z\}$



Trường hợp này tất cả các nốt đều có nốt cha nên: $\{\emptyset\}$

Ý nghĩa, với tập độc lập là rỗng thì có thể biểu diễn phân phối bất kỳ với 3 biến X, Y, Z. Nếu tập độc lập khác rỗng, ta chỉ có thể biểu diễn phân phối với X, Y, Z sao cho thỏa mãn độc lập có điều kiện nằm trong danh sách độc lập đã liệt kê.

4.4. Phân phối giới hạn liên kết – Topology limits distribution



Hình 15. Sơ đồ Venn của các tập phân phối trên bộ ba X, Y, Z

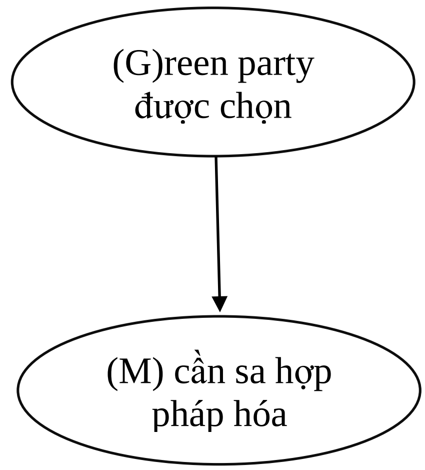
Tập màu xanh dương đại diện cho tất cả phân phối trên 3 biến X, Y, Z, tập này không có phân biệt các phần tử có độc lập hay không nên tập độc lập là rỗng. Một tập con của tất cả phân phối thỏa mãn hoàn toàn tất cả các độc lập và độc lập có điều kiện có thể từ 3 biến X, Y, Z là tập màu xanh lá. Tập màu đỏ còn lại là tập có chung 1 độc lập có điều kiện. Tất cả hoàn thành sơ đồ Venn như trên.

Nếu mạng Bayes có cấu trúc như vậy, và giờ ta muốn đại diện cho cùng một phân phối với một mạng Bayes có cấu trúc khác liệu có khả thi không? Vẫn có thể ví dụ như trường hợp ở tập màu đỏ, tuy cấu trúc khác nhau nhưng có cùng độc lập có điều kiện, nên phân phối thể hiện ra cũng như nhau.

CHƯƠNG 5. BÀI TẬP

Đến mùa bầu cử, tổng thống được chọn có thể hoặc không thể là ứng cử viên của Green Party. Các học giả tin rằng các chủ tịch của Green Party có khả năng hợp pháp hóa marijuana (cần sa) hơn các ứng cử viên từ đảng khác, nhưng hợp pháp hóa có thể xảy ra ở bất kỳ chính quyền nào. Được sự hỗ trợ của xác suất, các nhà phân tích mô hình hóa tình hướng với mạng Bayes bên dưới

	+g	-g
P(G)	0.1	0.9



	P(+m|G)	P(-m|G)
+g	2/3	1/3
-g	0.25	0.75

Câu hỏi 1. Tìm P(+m), xác suất biên của cần sa được hợp pháp hóa?

P(+m) = P(+m, +g) + P(+m, -g) = P(+m | +g)P(+g) + P(+m | -g)P(-g)
= 2/3 *1/10 + 1/4 *9/10 = 7/24

Câu hỏi 2. Các cơ quan 24/7 đưa tin về cần sa được hợp pháp hóa (+m), nhưng mà chưa thể tìm ra người trúng cử. Tìm xác suất có điều kiện P(+g | +m) mà chủ tịch Green Party được chọn?

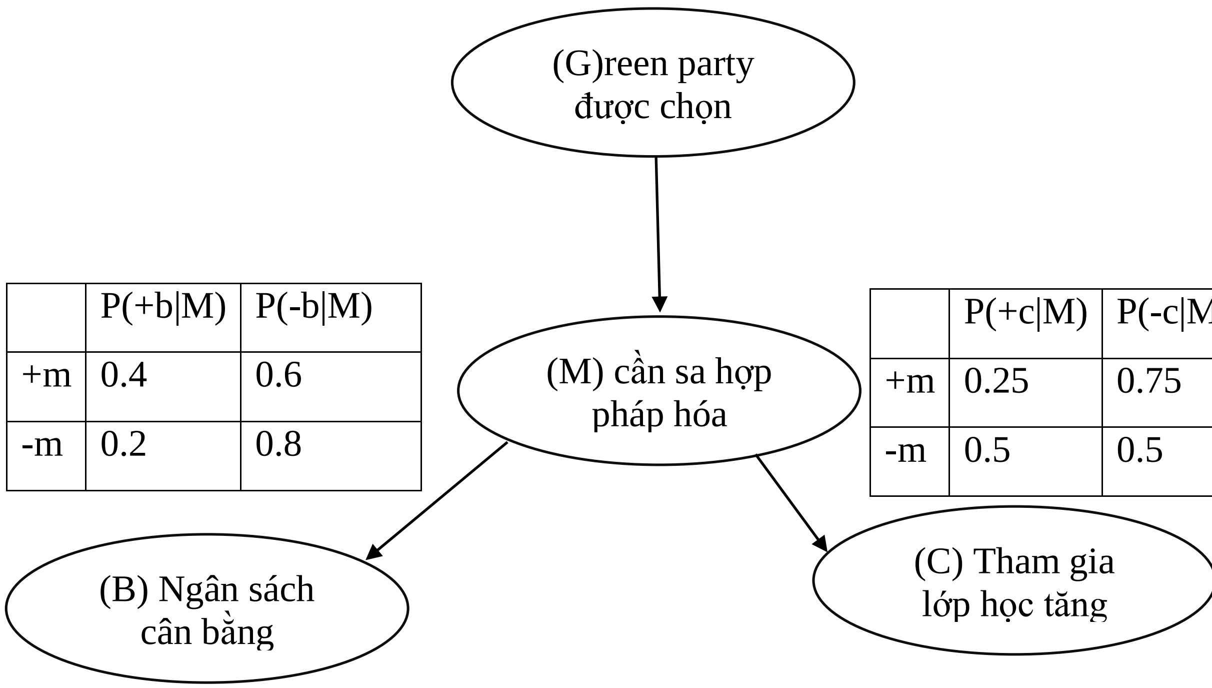
$$P(+g, +m) = \frac{P(+m|+g)P(+g)}{P(+m)} = \frac{2}{3} * \frac{1}{10} * \frac{24}{7} = \frac{8}{35}$$

Câu hỏi 3. Điền vào bảng xác suất chung dựa trên G và M

G	M	P(G, M)
+g	+m	1/15
+g	-m	1/30

-g	+m	9/40
-g	-m	27/40

Bây giờ, ta mở thêm 2 biến ngẫu nhiên để tăng cường suy luận trên mạng Bayes: (B) ngân sách cân bằng hay không và C (tham gia lớp học có tăng không)



	P(+b|M)	P(-b|M)
+m	0.4	0.6
-m	0.2	0.8

	P(+c|M)	P(-c|M)
+m	0.25	0.75
-m	0.5	0.5

Câu hỏi 4.

Phân phối chung đầy đủ sau khi thêm 2 biến. Điền đầy đủ giá trị:

G	M	B	C	P(G, M, B, C)
+	+	+	+	1/150
+	+	+	-	1/50
+	+	-	+	1/100
+	+	-	-	3/100
+	-	+	+	1/300

+	-	+	-	1/300
+	-	-	+	1/75
+	-	-	-	1/75

Câu hỏi 5.

Tính toán các giá trị sau:

P(+b | +m) = 4/10 (có được từ bảng xác suất có điều kiện của B)

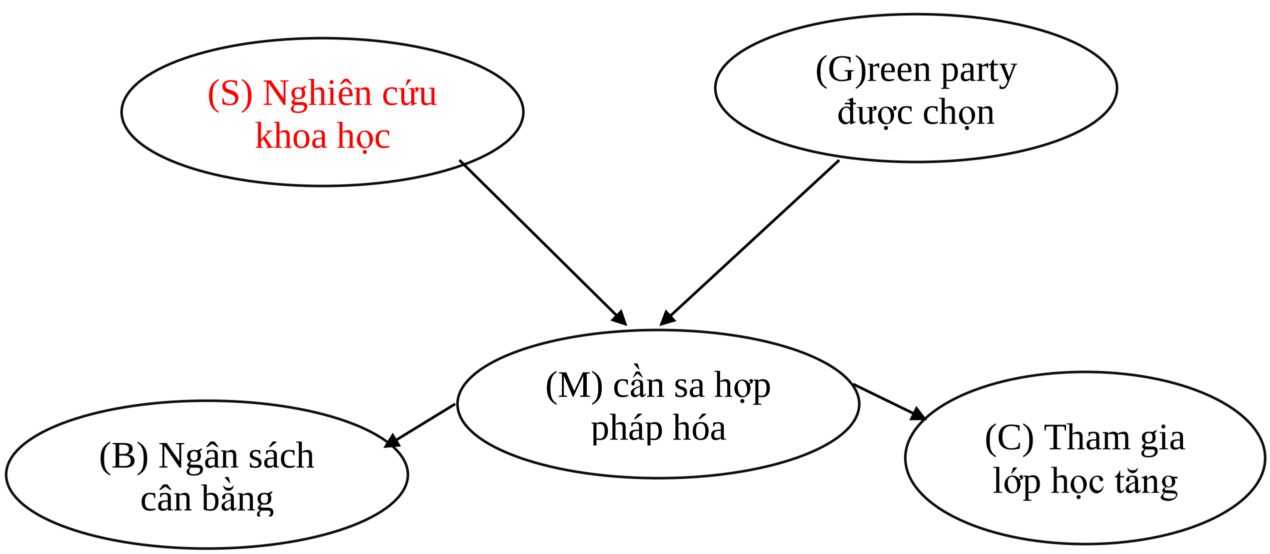
P(+b | +m, +g) = 4/10 (do B $\perp\!\!\!\perp$ G | M)

P(+b) = $\sum_{G,M,C} P(G,M,+b,C) = \frac{8}{1200} + \frac{24}{1200} + \frac{4}{1200} + \frac{4}{1200} + \frac{27}{1200} + \frac{81}{1200} + \frac{81}{1200} + \frac{81}{1200} = \frac{31}{120}$

P(+c|+b) = $\frac{P(+b,+c)}{P(+b)} = \sum_{G,M} P(G,M,+b,+c)\frac{120}{31} = \frac{\left(\frac{8}{1200}+\frac{4}{1200}+\frac{27}{1200}+\frac{81}{1200}\right)120}{31} = \frac{12}{31}$

Câu hỏi 6.

Thêm một node S vào mạng Bayes phản ánh xác suất mà nghiên cứu khoa học có thể ảnh hưởng đến xác suất cần sa hợp pháp hóa. Giả định rằng viêc nghiên cứu không ảnh hưởng trực tiếp B hoặc C. Vẽ sơ đồ mạng Bayes cho phù hợp. CPT nào cần được điều chỉnh?



Ta cần điều chỉnh P(M|G) thành P(M|G,S) và gồm 8 dòng entry thay vì 4

Câu hỏi 7. Xác định đúng sai?

B $\perp\!\!\!\perp$ G ? Không đảm bảo

C $\perp\!\!\!\perp$ G | M ? Đúng

G $\perp\!\!\!\perp$ S ? Đúng

B $\perp\!\!\!\perp$ C ? Không đảm bảo

B $\perp\!\!\!\perp$ C | G ? Không đảm bảo

G $\perp\!\!\!\perp$ S | M ? Không đảm bảo

G $\perp\!\!\!\perp$ S | B ? Không đảm bảo

TỔNG KẾT

Mạng Bayes mã hóa một cách khá gọn các phân phối chung xây dựng thành đồ thị trong sáng về mặt ngữ nghĩa, phản ảnh tự nhiên cấu trúc tri thức trong bộ óc con người, dễ thu thập và tích hợp tri thức tiên nghiệm. Tuy nhiên cần phải có tri thức khởi tạo chính xác, cần phải tính toán trên mọi giả thiết trong không gian tham số của nó để hình thành mô hình tối ưu.

Đảm bảo được các độc lập của phân phối có thể suy được từ cấu trúc đồ thị của mạng Bayes.

D-seperation giúp xác định độc lập cũng như độc lập có điều kiện từ đồ thị của mạng Bayes.

Phân phối chung của mạng Bayes có thể có nhiều hơn các độc lập có điều kiện mà chưa được phát hiện cho tới khi xem xét được phân phối đặc biệt của nó.

PHỤ LỤC

Danh mục cách tài liệu tham khảo:

1. Stuart Russel, Peter Norvig, Artificial Intelligence A Modern Approach, Third Edition, pp. 510 – 520.
2. Slide lý thuyết Bayesian Networks I & II
3. UC Berkeley CS188, Slide bài giảng Bayes Net I & II và Bài tập