

HỆ THỐNG GỢI Ý ÁP DỤNG TRONG QUÁ TRÌNH KÊ ĐƠN THUỐC

Trần Nguyễn Minh Thu¹, Đào Minh Trung Tuấn², Lưu Tiến Đạo² và Trần Quốc Lĩnh³

¹ Khoa Công nghệ Thông tin & Truyền thông, Trường Đại học Cần Thơ

² Trung tâm Công nghệ Phần mềm Đại học Cần Thơ

³ Bệnh viện Đa khoa tỉnh Hậu Giang

Thông tin chung:

Ngày nhận: 19/09/2015

Ngày chấp nhận: 10/10/2015

Title:

Recommender system for creating a prescription

Từ khóa:

Hệ thống gợi ý, luật kết hợp, kê đơn thuốc

Keywords:

Recommender systems, MedRES, association rules, drug prescribing

ABSTRACT

This subject uses the association rule-based recommender method to build the recommender system (MedRES – Medicine Recommender System) for the process of creating a prescription. The MedRES was built for the purpose of helping physicians choose the appropriate medicinal products to prescribe in the shortest period of time. The MedRES helps the young doctors with less experience can learn the prescribing method from other doctors. From the data set of prescription transactions, this research applies the Apriori mining algorithm to build a model on each disease. This model is a set of association rules, which represents the relationship between the drug's active ingredients. The MedRES system has good suggestions in most cases even for new doctors. The MedRES was evaluated at 2 phases: the model construction phase (the models were evaluated by experts) as well as the phase of evaluating the system after construction. In the phase after construction, the system was evaluated by using the Given-N method with the indices recall, precision and F1.

TÓM TẮT

Nghiên cứu này ứng dụng phương pháp gợi ý dựa trên luật kết hợp để xây dựng hệ thống gợi ý (MedRES – Medicine Recommender System) cho quá trình tạo đơn thuốc của bác sĩ. MedRES được xây dựng với mục đích giúp bác sĩ chọn được các sản phẩm thuốc phù hợp để kê đơn trong khoảng thời gian ngắn nhất. MedRES cũng giúp các bác sĩ trẻ, ít kinh nghiệm có thể học hỏi phương pháp kê đơn của các bác sĩ khác. Từ tập dữ liệu giao dịch về đơn thuốc, nghiên cứu này áp dụng giải thuật khai mỏ Apriori để xây dựng mô hình cho từng bệnh. Mô hình này chính là tập các luật kết hợp, nó biểu diễn mối liên quan giữa các hoạt chất của thuốc. Hệ thống MedRES gợi ý tốt trong đa số trường hợp kể cả đối với bác sĩ mới. MedRES được đánh giá ở 2 giai đoạn: giai đoạn xây dựng mô hình (các mô hình được đánh giá bởi các chuyên gia) cũng như giai đoạn đánh giá hệ thống sau khi xây dựng. Ở giai đoạn sau khi xây dựng, hệ thống được đánh giá bằng phương pháp Given-N với các chỉ số F1.

1 GIỚI THIỆU

Hiện nay, hầu hết các bệnh viện đều được tin học hóa trong khâu quản lý khám chữa bệnh. Các

bác sĩ, sau khi khám và chẩn đoán bệnh, thường sử dụng phần mềm máy tính để kê đơn thuốc cho bệnh nhân. Thay vì phải viết toàn bộ tên thuốc thì hiện tại bác sĩ chỉ cần nhớ mã hoặc một phần của

tên thuốc để nhập vào đơn thuốc, chương trình sẽ giúp bác sĩ tìm ra tên thuốc đầy đủ. Ngoài ra, các phần mềm quản lý bệnh viện hiện nay còn cung cấp nhiều tiện ích khác hỗ trợ cho bác sĩ trong việc kê đơn như: soạn thảo cách dùng thuốc, sao chép đơn thuốc cũ, soạn đơn thuốc mẫu cho một số bệnh thường gặp,... Chính vì thế, việc ứng dụng phần mềm quản lý bệnh viện đã làm giảm đáng kể thời gian kê đơn so với việc kê đơn thủ công như trước kia, góp phần làm giảm bớt sự quá tải của các bệnh viện.

Tuy được hỗ trợ nhiều tiện ích, nhưng do sản phẩm thuốc ngày càng đa dạng, danh mục thuốc của các bệnh viện ngày càng có nhiều sản phẩm, Danh mục thuốc ban hành kèm theo Thông tư số 31/2011/TT-BYT ngày 11/7/2011 của Bộ Y tế (Bộ Y tế, 2011) có trên 1000 sản phẩm thuốc, các thuốc có tên biệt dược tương tự nhau cũng ngày càng nhiều nên các bác sĩ gặp không ít khó khăn trong việc kê đơn, đặc biệt đối với bác sĩ trẻ. Các bác sĩ cần phải nhớ rất nhiều mã thuốc, do đó việc tìm kiếm một sản phẩm thuốc phù hợp để kê đơn vẫn mất không ít thời gian. Câu hỏi đặt ra là làm thế nào để hỗ trợ tốt hơn cho bác sĩ trong việc kê đơn thuốc, giúp bác sĩ kê đơn một cách dễ dàng và nhanh chóng.

Vào giữa những năm 90, hệ thống gợi ý (recommender system) đã nổi lên như là một lĩnh vực nghiên cứu độc lập khi mà phương pháp lọc cộng tác được sử dụng để xây dựng hệ thống gợi ý. Các hệ thống gợi ý được xây dựng để giúp người dùng tìm thấy thông tin có liên quan một cách nhanh chóng và dễ dàng trên một cơ sở dữ liệu lớn. Các hệ thống gợi ý được ứng dụng trong nhiều lĩnh vực như thương mại điện tử, giải trí, khoa học, tin tức... Trong lĩnh vực y khoa, hệ thống gợi ý đã đạt được một số kết quả khả quan ban đầu. Các hệ thống đều dựa vào tiền sử bệnh và hồ sơ bệnh án để chẩn đoán và phòng ngừa bệnh tật (Davis *et al.*, 2010; Folino, 2010). Vừa mới đây, người Thái Lan đã ứng dụng hệ thống gợi ý kết hợp cùng với các giải thuật thông minh để xây dựng hệ thống chẩn đoán từ xa nhằm nâng cao hiệu quả trong công tác khám chữa bệnh của các cơ sở y tế ở vùng sâu, vùng xa (Sunantha Sodseea and Maytayanin Komkhaio, 2013).

Xuất phát từ vấn đề thực tiễn nêu trên, theo cách tiếp cận “hướng người dùng”, chúng tôi xây dựng hệ thống gợi ý áp dụng cho quá trình kê đơn thuốc của bác sĩ (gọi tắt là *MedRES*), dựa trên cơ sở dữ liệu các đơn thuốc đã được hình thành của

các bác sĩ, sử dụng giải thuật khai mở luật kết hợp *Apriori*.

2 HỆ THỐNG GỢI Ý

2.1 Hệ thống gợi ý

Theo Adomavicius và Tuzhilin (2005), hệ thống gợi ý có ứng dụng rất đa dạng trong thực tế, giúp người dùng giải quyết vấn đề quá tải thông tin và cung cấp cho họ những gợi ý được cá nhân hóa về các sản phẩm và dịch vụ. Do đó, việc ứng dụng hệ thống gợi ý trong các hệ thống thông tin, đặc biệt là hệ thống thông tin với cơ sở dữ liệu lớn, mang lại nhiều lợi ích cho cả người dùng lẫn hệ thống.

Đối với người dùng, hệ thống cung cấp các gợi ý để giúp họ chọn lựa các sản phẩm hay dịch vụ phù hợp với nhu cầu và sở thích một cách nhanh chóng và dễ dàng. Hệ thống, còn có thể cung cấp cho người dùng những gợi ý về các sản phẩm hay dịch vụ mới mà trước đây họ chưa từng biết đến, cái mà được cho là người dùng có thể có nhu cầu.

Đối với hệ thống thông tin, hệ thống gợi ý có khả năng học được nhu cầu và sở thích của người dùng thông qua thông tin phản hồi của họ khi sử dụng hệ thống. Nhờ đó, các gợi ý mà hệ thống cung cấp cho người dùng ngày càng phù hợp hơn, hệ thống ngày càng thân thiện với người dùng hơn, hiệu quả hoạt động của hệ thống cũng tăng lên.

Theo Balabanovic và Shoham (1997), hệ thống gợi ý thường được chia ra thành ba loại dựa trên cách mà các gợi ý được tạo ra: (1) gợi ý dựa trên nội dung: người dùng được gợi ý những mục dữ liệu tương tự với những mục dữ liệu mà chính họ thích trong quá khứ; (2) gợi ý dựa trên lọc cộng tác: người dùng được gợi ý những mục dữ liệu mà những người dùng khác (có cùng thị hiếu và sở thích) thích trong quá khứ; (3) gợi ý hỗn hợp: kết hợp hai phương pháp trên.

Ngoài ba phương pháp gợi ý trên, chúng ta cũng có thể sử dụng luật kết hợp tạo ra các gợi ý dựa trên một mô hình phụ thuộc của các mục dữ liệu.

2.2 Phương pháp gợi ý dựa trên luật kết hợp

Khai mở dữ liệu dựa trên luật kết hợp là một phương pháp được nghiên cứu phổ biến trong việc khám phá mối liên quan giữa các biến trong cơ sở dữ liệu lớn. Mối liên quan này mang tính quy luật, có nghĩa là các biến đồng thời xuất hiện một cách thường xuyên. Ví dụ, khách hàng của siêu thị

thường xuyên mua chung tã giấy và sữa cho em bé. Các giải thuật luật kết hợp cho phép tìm kiếm các mối liên quan giữa các biến và biểu diễn những mối liên quan này thành tập các luật kết hợp. Giải thuật *Apriori* là một giải thuật luật kết hợp được đề xuất đầu tiên bởi Agrawal *et al.* (1993) dựa trên quan điểm tìm các luật mạnh.

Phương pháp gợi ý dựa trên luật kết hợp là đưa ra danh sách gợi ý cho người dùng hiện hành dựa trên tập các mục dữ liệu được đánh giá cao bởi người dùng và tập các luật kết hợp (hay mô hình phụ thuộc) (Hahsler, 2011).

Để tạo ra một danh sách gợi ý cho người dùng hiện hành u_a dựa trên tập các mục dữ liệu I_a được đánh giá cao bởi người dùng và tập các luật kết hợp R , cần thực hiện các bước sau:

- Tìm trong R tất cả các luật $X \Rightarrow Y$ với $X \subseteq I_a$.

Bảng 1: Cơ sở dữ liệu đơn thuốc với 5 giao dịch

Transaction ID (số hiệu đơn thuốc)	Items (mã hoạt chất)
1307000005	Nabum01, Thioc01, Amlod01
1307000011	Amiod02, Nitro01, Esome01
1307000017	Amlod01, Ibers01, Nabum01, Thioc01
1307000019	Telmi01, Clopi01, Melox01, Metho01
1307000072	Telmi01, Nitro01, Clopi01, Melox01, Metho01

2.3 Một số vấn đề của hệ thống gợi ý dựa trên luật kết hợp

Tính cá nhân là đặc điểm thiết yếu của hệ thống gợi ý. Nghĩa là, hệ thống phải phù hợp với nhu cầu và sở thích của từng người dùng (hoặc từng nhóm người dùng). Tuy nhiên, hệ thống gợi ý dựa trên luật kết hợp thường không đảm bảo được tính cá nhân hóa. Hệ thống gợi ý không quan tâm người dùng hiện hành là ai mà chỉ quan tâm đến phiên giao dịch của người dùng. Nghĩa là, nếu hai khách hàng khác nhau chọn các sản phẩm giống nhau cho vào giỏ hàng thì hệ thống bán hàng sẽ đưa ra cùng các mục sản phẩm gợi ý cho cả hai mà không quan tâm đến lịch sử mua hàng của họ. Hơn nữa, luật kết hợp thường hoạt động với một ngưỡng *support* và *confidence* tối thiểu. Điều đó chỉ đúng trên quan điểm của hệ thống mà không đúng đối với quan điểm của người dùng. Do đó, để làm tăng tính cá nhân hóa của hệ thống dựa trên luật kết hợp, cần phải có một qui tắc đặc biệt cho mỗi người dùng. Hệ thống ASARM (Adaptative Support Association Rules Mining) được phát triển bởi Lin *et al.* (2002), tạo ra danh sách các luật phù hợp nhất dựa trên giá trị *support* tương ứng với từng cá nhân người dùng.

- Gợi ý N mục dữ liệu duy nhất ở về phải (Y) của các luật tìm được với giá trị *confidence* tốt nhất.

Lấy ví dụ minh họa với tập giao dịch (5 đơn thuốc) như trong Bảng 1 được tạo ra từ các đơn thuốc có mã hoạt chất chứa trong tập $I = \{Amiod02, Amlod01, Clopi01, Esome01, Ibers01, Melox01, Metho01, Nabum01, Nitro01, Telmi01, Thioc01\}$.

Từ tập dữ liệu minh họa ta có thể sinh ra luật $\{Nabum01, Thioc01\} \Rightarrow \{Amlod01\}$. Luật này có độ hỗ trợ là: $\text{sup}(R) = \text{sup}(\{Nabum01, Thioc01, Amlod01\}) = 2/5 = 0.4$ và có độ tin cậy là: $\text{conf}(R) = \text{sup}(R) / \text{sup}(\{Nabum01, Thioc01\}) = 0.4 / 0.4 = 1$. Với luật được sinh ra như trên, nếu bác sĩ đã chọn mã hoạt chất *Nabum01* và *Thioc01* trong đơn thuốc thì hệ thống sẽ gợi ý thuốc *Amlod01* cho quá trình kê đơn của bác sĩ vì xác suất bác sĩ sẽ cho thuốc *Amlod01* là rất cao.

Phác đồ điều trị hướng dẫn sử dụng thuốc dựa trên hoạt chất của thuốc. Trong thực tế, mỗi hoạt chất có chứa trong nhiều sản phẩm thuốc với tên biệt dược khác nhau (do nhiều công ty sản xuất). Ví dụ, hoạt chất *Amlodipine* (thường được dùng trong điều trị bệnh tăng huyết áp) có chứa trong các sản phẩm thuốc có tên biệt dược như: *Amlodac, AMLODIPIN, Apitim,...* Do vậy, cùng một hoạt chất, các bác sĩ khác nhau có thể lựa chọn các sản phẩm thuốc khác nhau phù hợp nhất với thói quen/kinh nghiệm cá nhân của họ. Vì vậy, trong nghiên cứu của chúng tôi, luật kết hợp được xây dựng dựa trên mã hoạt chất. Ứng với mỗi bác sĩ, chúng tôi sẽ gợi ý những thuốc có biệt dược khác nhau dựa trên mã hoạt chất được gợi ý và dựa trên thói quen dùng thuốc của bác sĩ.

3 GIẢI THUẬT XÂY DỰNG MEDRES

Để hệ thống *MedRES* tạo ra các gợi ý phù hợp với từng bác sĩ, cần tạo ra tập tin lưu thông tin tổng hợp về thói quen sử dụng thuốc biệt dược của từng bác sĩ dựa trên lịch sử kê đơn của họ. Đối với mỗi bác sĩ, tập tin này cho biết mức độ ưu tiên/tỷ lệ sử dụng từng sản phẩm thuốc trong cùng hoạt chất được dùng điều trị một bệnh cụ thể. Các thông tin

lưu trong tập tin này bao gồm: *bác sĩ kê đơn, mã bệnh, mã hoạt chất, mã thuốc và tỷ lệ*. Bảng 2 biểu

diễn một phần dữ liệu về thói quen sử dụng thuốc của bác sĩ.

Bảng 2: Một phần dữ liệu về thói quen sử dụng thuốc của bác sĩ

Kdth ID	Username	Mabenh	Mahoatchat	Mathuoc	Tyle
1	bhtri	A06.2	Alver01	ALV01	100
2	bhtri	A06.2	Amyla01	ENT01	100
3	bhtri	A06.2	Lacto01	YBI01	100
4	bhtri	A06.2	Para01	HAP03	100
5	bhtri	A37	Acety03	ACE08	100
6	bhtri	A37	Cefix01	BIC01	100
7	bhtri	A37	Predn01	PRE03	100
8	bhtri	C77.3	Amoxi01	PMS06	100
9	bhtri	C77.3	Chlor01	CHL01	100
10	bhtri	C77.3	HOASTEX	THU02	100
11	bhtri	C77.3	Para02	AGI03	100
12	bhtri	D21.0	Cefet01	cet02	100
13	bhtri	D21.0	Para01	PAR03	100
14	bhtri	D21.0	Rabep01	MAX03	100
15	bhtri	E11	Amiod02	aml02	100
16	bhtri	E11	Amoxi01	KLA01	100
17	bhtri	E11	Capto01	CAP02	100
18	bhtri	E11	CDL_CBQ	HOA06	100
19	bhtri	E11	Felod01	FEL01	100
20	bhtri	E11	Glicl01	DOR06	100
21	bhtri	E11	Hyosc01	BIS01	100
22	bhtri	E11	Metfo01	GLU08	72.73
23	bhtri	E11	Metfo01	met09	27.27

Bảng 2 cho thấy, các dòng có giá trị *Kdth_ID* từ 15 đến 23 biểu diễn thói quen dùng thuốc của bác sĩ *bhtri* đối với mã bệnh *E11*. Trong đó, đối với mã hoạt chất *Amiod02*, bác sĩ *bhtri* từng sử dụng chỉ một sản phẩm thuốc duy nhất có mã là *aml02*, do đó tỷ lệ sử dụng của thuốc này là 100%. Đối với mã hoạt chất *Metfo01*, bác sĩ *bhtri* từng sử dụng hai sản phẩm thuốc có mã là *GLU08* và *met09* với tỷ lệ sử dụng tương ứng là 72.73% và 27.27%, do đó thuốc có mã *GLU08* được ưu tiên cao hơn thuốc có mã *met09*.

Từ tập dữ liệu giao dịch về đơn thuốc, áp dụng giải thuật *Apriori* (sẵn dùng trong gói thư viện *arules* của ngôn ngữ *R*, được tải về từ địa chỉ <http://cran.r-project.org>) để xây dựng mô hình phụ thuộc của các hoạt chất, nghĩa là tìm ra các luật hay mối liên hệ giữa các hoạt chất.

3.1 Xây dựng mô hình

Để có được mô hình phù hợp, phản ánh đúng đắn thói quen/quy luật kê đơn của các bác sĩ, cần xây dựng mô hình phụ thuộc cho từng mã bệnh bởi vì các bác sĩ có thói quen/cách thức kê đơn khác nhau đối với từng mã bệnh. Việc xây dựng mô

hình phụ thuộc đối với từng mã bệnh còn giúp tránh được tình trạng gợi ý nhầm giữa bệnh này với bệnh khác. Quá trình xây dựng mô hình được chia làm ba giai đoạn:

- Chia tập dữ liệu giao dịch về đơn thuốc ra thành các tập con theo từng mã bệnh.
- Xây dựng mô hình cho từng mã bệnh.
- Đánh giá mô hình.

3.1.1 Chia tập dữ liệu giao dịch về đơn thuốc thành các tập con theo từng mã bệnh

Tập dữ liệu giao dịch về đơn thuốc tương ứng với từng mã bệnh chỉ chứa thông tin: *mã giao dịch* (số hiệu đơn thuốc) và *mục dữ liệu* (mã hoạt chất). Bảng 3 biểu diễn một phần tập dữ liệu giao dịch tương ứng với mã bệnh *I10 - Cao huyết áp vô căn (nguyên phát)*. Mỗi giao dịch được đại diện bởi một mã giao dịch (hay số hiệu đơn thuốc) và được biểu diễn bởi một số dòng dữ liệu. Các mã hoạt chất thuộc các dòng của cùng một giao dịch chính là các mục dữ liệu của giao dịch đó. Bảng 3 cho thấy, giao dịch có mã “1307000005” chứa 03 mục dữ liệu là: *Nabum01*, *Thioc01* và *Amlod01*.

Bảng 3: Một phần tập dữ liệu giao dịch tương ứng với mã bệnh I10 - Cao huyết áp vô

Transaction ID (Số hiệu đơn thuốc)	Items (Mã hoạt chất)
1307000005	Nabum01
1307000005	Thioc01
1307000005	Amlod01
1307000011	Amiod02
1307000011	Nitro01
1307000011	Esome01
1307000017	Amlod01
1307000017	Ibers01
1307000017	Nabum01
1307000017	Thioc01
...	...

3.1.2 Xây dựng mô hình cho từng mã bệnh

Giai đoạn này được thực hiện qua 03 bước:

Bước 1: Chuyển đổi tập dữ liệu giao dịch sang dạng ma trận nhị phân

Nghiên cứu này sử dụng cách biểu diễn tập dữ liệu giao dịch dưới dạng ma trận chỉ mục nhị phân theo nghiên cứu của Hahsler *et al.* (2005). Theo đó, 05 giao dịch đầu tiên trong tập dữ liệu giao dịch ở Bảng 3 được chuyển đổi sang dạng ma trận nhị phân như Hình 1.

Bước 2: Ước tính các tham số

Do mô hình cho một mã bệnh được xây dựng dựa trên tập dữ liệu giao dịch tương ứng với mã

bệnh đó nên việc chọn các tham số phụ thuộc vào đặc điểm của tập dữ liệu này. Chính vì thế, các tham số cần được ước tính riêng đối với mỗi tập dữ liệu giao dịch.

Theo nghiên cứu này, các tham số được ước tính như sau:

- *Độ hỗ trợ tối thiểu (minsup)*: được tính bằng giá trị thấp nhất của độ hỗ trợ của tất cả tập mục dữ liệu thường xuyên (được khai mở bởi giải thuật *ecolat* sẵn dùng trong gói thư viện *arules* của ngôn ngữ *R*, tải về từ địa chỉ <http://cran.r-project.org>).

- *Độ tin cậy tối thiểu (minconf)*: được chọn bằng 0.6.

- *Độ dài tối đa của luật (maxlen)*: được chọn bằng 4 (bằng độ dài của một đơn thuốc tiêu chuẩn, theo ý kiến của chuyên gia).

Việc chọn tham số như trên có thể đảm bảo được yêu cầu: hầu hết các mã bệnh đều có mô hình tương ứng với tổng số luật lớn hơn hoặc bằng 1. Nghĩa là hệ thống có thể gợi ý thuốc cho bác sĩ đối với hầu hết các mã bệnh được chẩn đoán.

Với *maxlen* = 4, tập luật kết hợp chỉ bao gồm các luật có độ dài nhỏ hơn hoặc bằng 4. Theo Hahsler, (2011) mô hình chỉ chứa những luật có độ dài về phải (RHS) bằng 1, nghĩa là về trái của luật (LHS) có độ dài nhỏ hơn hoặc bằng 3.

Bước 3: Sinh tập luật kết hợp bằng cách áp dụng giải thuật *Apriori*

Transaction (Số hiệu đơn thuốc)	Items (Mã hoạt chất)													
	Amiod02	Amlod01	Cefdi01	Clopi01	Esome01	Euglc01	Ibers01	Melox01	Metho01	Nabum01	Nitro01	Telmi01	Terpi01	Thioc01
1307000005	0	1	0	0	0	0	0	0	0	1	0	0	0	1
1307000011	1	0	0	0	1	0	0	0	0	0	1	0	0	0
1307000017	0	1	0	0	0	0	1	0	0	1	0	0	0	1
1307000019	0	0	0	1	0	0	0	1	1	0	0	1	0	0
1307000027	0	0	1	0	0	1	0	0	0	0	0	1	1	0

Hình 1: Biểu diễn ma trận nhị phân của 05 giao dịch đầu tiên tương ứng với mã bệnh I10

Sau khi ước tính được các tham số cần thiết, áp dụng giải thuật *Apriori* với các tham số này để xây dựng mô hình luật kết hợp giữa các hoạt chất đối

với từng mã bệnh. Các tập luật kết hợp được lưu trữ chung trong một tập tin *Rules.csv* có dạng như Bảng 4.

Bảng 4: Một phần dữ liệu của tập tin Rules.csv (lưu trữ tập luật kết hợp)

RulesID	Mabenh	LHS	RHS	Support
1	A02.1	NA	Para01	1
2	A02.1	Amoxi01	Para01	0.5
3	A02.1	Cefur01	Levof02	0.5
4	A02.1	Levof02	Cefur01	0.5
5	A02.1	Cefur01	VitamC	0.5
...

Trên Bảng 4, cột *RulesID* lưu định danh của luật (phân biệt các luật với nhau), cột *Mabenh* lưu mã bệnh, cột *LHS* lưu danh sách các mã hoạt chất bên trái của luật, cột *RHS* lưu mã hoạt chất bên phải của luật. Giá trị NA trong cột *LHS* có nghĩa là phần bên trái của luật kết hợp tương ứng là rỗng (không chứa mã hoạt chất nào).

Đối với mỗi mã bệnh, các luật kết hợp được sắp xếp theo thứ tự giảm dần của độ hỗ trợ (*support*), số lượng luật kết hợp được sinh ra nhiều hay ít tùy thuộc vào các tham số được chọn (*minsup*, *minconf* và *maxlen*) và đặc điểm của tập dữ liệu giao dịch.

3.1.3 Đánh giá mô hình

Các mô hình cần phải được đánh giá bởi chuyên gia trong lĩnh vực khám chữa bệnh và kê đơn thuốc. Việc đánh giá này nhằm xác định tính đúng đắn của mô hình trước khi ứng dụng nó vào hệ thống *MedRES*.

Do có quá nhiều mã bệnh nên chuyên gia chỉ đánh giá mô hình của một vài mã bệnh tiêu biểu. Đối với nghiên cứu này, chuyên gia đánh giá 05 mô hình tương ứng với 05 mã bệnh được chia làm 03 nhóm: *bệnh rất phổ biến*, *bệnh phổ biến ở mức trung bình* và *bệnh rất hiếm gặp*. Chuyên gia chỉ đánh giá độ phù hợp của luật ở một trong hai trạng thái: *đúng* hoặc *sai*.

3.2 Tạo danh sách thuốc gợi ý dựa trên mô hình luật kết hợp

Để tạo ra một danh sách thuốc gợi ý cho bác sĩ *U* dựa trên tập các hoạt chất *I* đã được chọn và tập các luật kết hợp *R* (tương ứng với mã bệnh được chọn làm chẩn đoán), cần thực hiện 03 bước sau đây:

- Tìm trong *R* tất cả các luật có dạng $X \Rightarrow Y$ với $X \subseteq I$.
- Chọn ra *N* hoạt chất đầu tiên (duy nhất) ở vế phải (*Y*) của các luật tìm được.
- Dựa trên thông tin thói quen sử dụng thuốc của bác sĩ *U*, với mỗi hoạt chất trong danh sách *N*

hoạt chất được chọn, chọn ra *M* thuốc (biệt dược) có giá trị *tỷ lệ* (mức ưu tiên) cao nhất.

Qua 03 bước nêu trên, hệ thống sẽ tạo ra một danh sách thuốc gợi ý phù hợp với bác sĩ *U*. Thứ tự các thuốc trong danh sách gợi ý được sắp xếp theo 02 mức:

- Các hoạt chất được sắp xếp giảm dần theo độ hỗ trợ của luật kết hợp.
- Các thuốc trong cùng hoạt chất được sắp xếp theo giá trị *tỷ lệ* (mức ưu tiên) của nó.

Để gợi ý thuốc cho một bác sĩ mới (bác sĩ chưa từng kê đơn đối với một bệnh cụ thể), thay vì sử dụng thông tin thói quen của bác sĩ mới này, hệ thống sẽ dựa trên thông tin thói quen sử dụng thuốc của các bác sĩ khác. Khi đó, ở bước 03 của quá trình tạo danh sách gợi ý, với mỗi hoạt chất trong danh sách *N* hoạt chất được chọn, hệ thống sẽ chọn ra *M* thuốc (biệt dược) có giá trị *tỷ lệ* cao nhất đối với tất cả bác sĩ khác.

4 THỰC NGHIỆM VÀ ĐÁNH GIÁ HỆ THỐNG

4.1 Giới thiệu về cơ sở dữ liệu sử dụng cho thực nghiệm

Để triển khai thử nghiệm hệ thống *MedRES*, chúng tôi trích lọc dữ liệu từ cơ sở dữ liệu thực tế đang triển khai tại Bệnh viện Đa khoa tỉnh Hậu Giang trong khoảng thời gian từ ngày 01/7/2013 đến ngày 14/3/2014 thông qua phần mềm CUSC-HIS Hậu Giang (Hệ thống thông tin tổng thể tại Bệnh viện Đa khoa tỉnh Hậu Giang). Đối với hệ thống *MedRES*, đối tượng gợi ý là *thuốc*, đối tượng được gợi ý là *bác sĩ* và *đơn thuốc* thể hiện mối quan hệ giữa *bác sĩ* và *thuốc*. Do đó, chúng tôi thu thập dữ liệu về *thuốc*, *bác sĩ* và *đơn thuốc* cũng như dữ liệu về *mã bệnh* (theo ICD - International Classification of Diseases) được bác sĩ dùng làm chẩn đoán trong quá trình kê đơn thuốc của phần mềm CUSC-HIS Hậu Giang. Tập dữ liệu được thu thập bao gồm: 65.838 đơn thuốc, 69 bác sĩ kê đơn, 553 sản phẩm thuốc, 1108 mã bệnh.

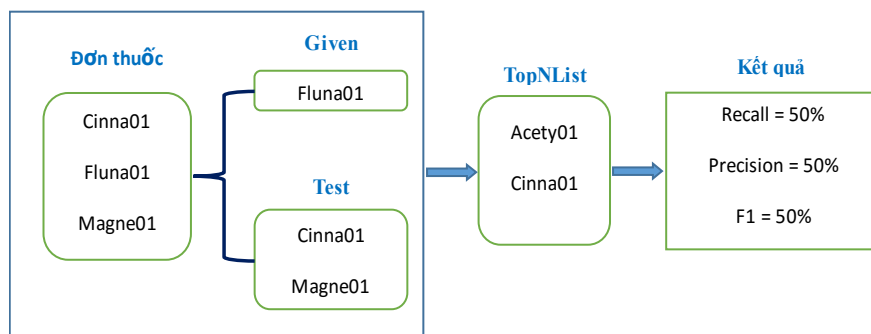
Để xây dựng hệ thống theo giải thuật đề nghị ở phần 2, chúng tôi cũng tiến hành tiền xử lý dữ liệu bằng các xóa bỏ những đơn thuốc thiếu chẩn đoán, loại bỏ các đơn thuốc có chứa hơn 4 sản phẩm thuốc nhằm loại bỏ những thuốc “dư thừa” (theo ý kiến của chuyên gia trong lĩnh vực khám chữa bệnh và kê đơn thuốc). Mặt khác, do có nhiều sản phẩm thuốc thiếu thông tin về hoạt chất (chiếm tỷ lệ ~ 31% tổng số sản phẩm thuốc) nên thông tin về hoạt chất của thuốc cần phải được cập nhật đầy đủ. Việc cập nhật thông tin về hoạt chất được thực hiện

dưới sự hỗ trợ của một chuyên gia về Dược của Bệnh viện Đa khoa tỉnh Hậu Giang.

4.2 Phương pháp đánh giá hệ thống MedRES

Hệ thống gợi ý tạo đơn thuốc *MedRES* là bước tiền khả thi để tích hợp chức năng gợi ý vào CUSC-HIS Hậu Giang. Hệ thống *MedRES* được đánh giá ở 2 giai đoạn: giai đoạn xây dựng mô hình (các tập luật sau khi xây dựng được đánh giá bởi các chuyên gia); giai đoạn đánh giá hệ thống sau khi xây dựng.

Đánh giá trực tiếp (online) được thực hiện với sự tham gia của 14 bác sĩ, thực hiện 73 đơn thuốc trong khoảng thời gian từ 09/4/2015 đến 16/4/2015.



Hình 2: Ví dụ phương pháp đánh giá Given-1

4.3 Đánh giá kết quả của hệ thống MedRES

Đánh giá mô hình

Do có quá nhiều mã bệnh nên chuyên gia chỉ đánh giá mô hình của một vài mã bệnh tiêu biểu. Đối với đề tài này, chuyên gia đánh giá 05 mô hình

Nghi thức đánh giá được sử dụng trong thực nghiệm này là nghi thức *Given-N* với $N=1,2,3$ và nghi thức *AllButOne* (Herlocker J.L *et al.*, 2004), với danh sách gợi ý có độ dài lần lượt từ 1 đến 4 sản phẩm thuốc. Nghi thức *Given-N* được tiến hành bằng cách chia một “giao dịch” thành hai phần: một tập được gọi là “*Given*” và một tập được gọi là “*Test*”. Tập “*Given*” dùng để tính các gợi ý. Sau đó, so sánh các sản phẩm thực tế (*Test*) và các sản phẩm gợi ý đề nghị, mỗi gợi ý đúng độ chính xác của hệ thống tăng lên 1. Đặc biệt, khi sử dụng phương pháp *Given-N* bắt buộc giao dịch sử dụng để đánh giá phải có ít nhất $N+1$ mục dữ liệu. Hình 2 là một ví dụ về phương pháp đánh giá *Given-1*.

Bảng 5: Thống kê tỷ lệ luật đúng đối với 05 mô hình được chuyên gia đánh giá

Stt	Mã bệnh	Tên bệnh	Tổng số đơn thuốc	Mức độ phổ biến	Tỷ lệ luật đúng
1	I10	Cao huyết áp vô căn (nguyên phát)	3966	rất phổ biến	7/8 ~ 87,5%
2	K29	Viêm dạ dày và tá tràng	3198		49/68 ~ 72%
3	J11	Cảm cúm không định danh virus	2486	phổ biến ở mức trung bình	20/25 ~ 80%
4	J35.1	Phi đại amyđan	7		44/44 ~ 100%
5	H65	Viêm tai giữa không nung mủ	1	rất hiếm gặp	18/32 ~ 56,3%
Trung bình					79,2%

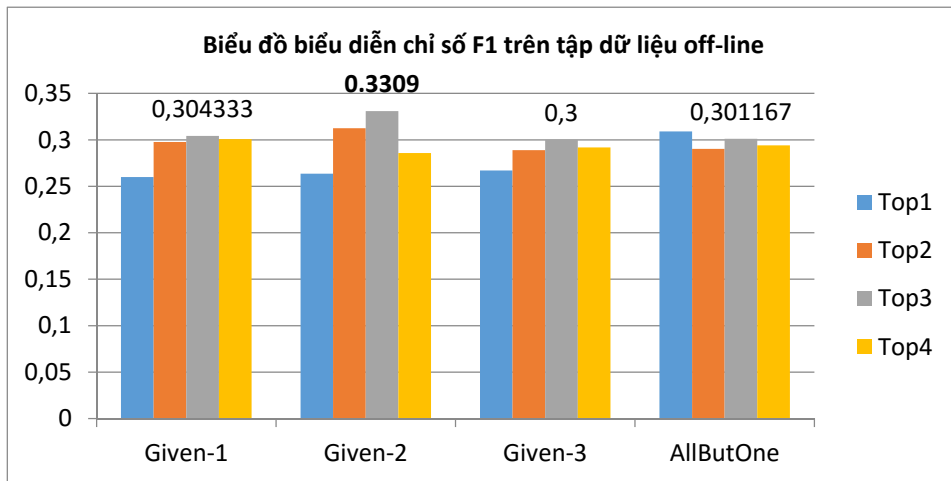
Dựa trên kết quả đánh giá, tỷ lệ luật đúng được thống kê trong Bảng 5. Từ kết quả thống kê, cả 05 mô hình được chuyên gia đánh giá đều có tỷ lệ luật đúng trên 50% (trung bình của 05 mô hình là 79.2%).

Đánh giá hệ thống

Kết quả đánh giá theo chỉ số *F1* trên tập dữ liệu off-line được biểu diễn bởi đồ thị ở Hình 3.

tương ứng với 05 mã bệnh được chia làm 03 nhóm dựa trên mức độ phổ biến như trong Bảng 5. Sự phân chia mức độ phổ biến của các mã bệnh chỉ mang tính tương đối và dựa trên tổng số đơn thuốc các bác sĩ đã kê trong tập dữ liệu giao dịch.

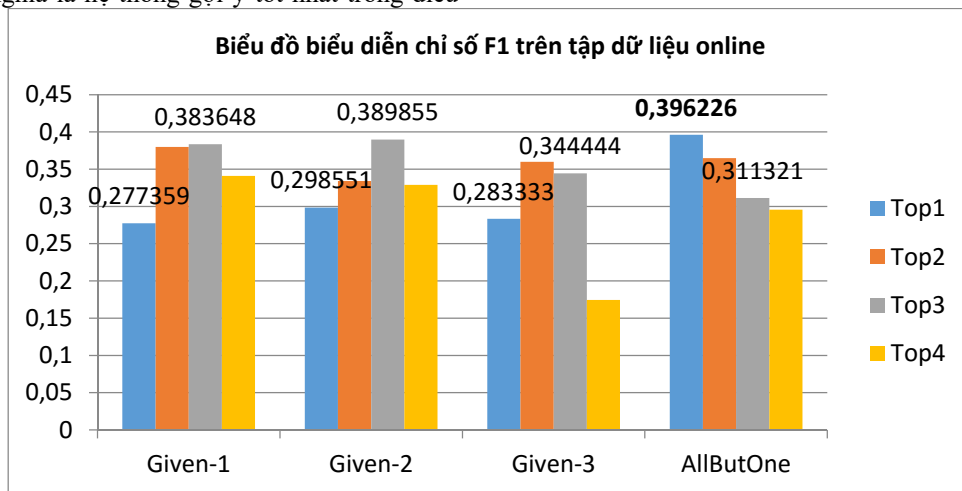
Đồ thị cho thấy hệ thống đạt hiệu suất cao nhất tại *Top3* (hay *TopN = 3*) và *Given-2* với chỉ số *F1* đạt giá trị lớn nhất bằng 33.09%. Chỉ số *recall*, *precision* tương ứng lần lượt là: 32.55% và 36.65%. Điều này có nghĩa là hệ thống gợi ý tốt nhất trong điều kiện: số hoạt chất tối đa trong danh sách gợi ý bằng 3 và số hoạt chất mà bác sĩ đã chọn (trước khi gợi ý) bằng 2.



Hình 3: Biểu đồ biểu diễn chỉ số F1 trên tập dữ liệu off-line

Kết quả đánh giá theo chỉ số *F1* trên tập dữ liệu online được biểu diễn bởi đồ thị ở Hình 4. Đồ thị cho thấy hệ thống đạt hiệu suất cao nhất tại *Top1* (hay *TopN* = 1) và *AllButOne* với chỉ số *F1* đạt giá trị lớn nhất bằng 39.62%. Chỉ số *recall*, *precision* tương ứng lần lượt là: 39.62% và 39.62%. Điều này có nghĩa là hệ thống gợi ý tốt nhất trong điều

kiện: số hoạt chất tối đa trong danh sách gợi ý bằng 1 và số hoạt chất mà bác sĩ đã chọn (trước khi gợi ý) nằm trong khoảng từ 1 đến 3. Tuy nhiên, trong điều kiện *Top3* và *Given-2*, hệ thống cũng đạt hiệu suất cao với các chỉ số: *recall* = 36.95%, *precision* = 45.28% và *F1* = 38.98%.



Hình 4: Biểu đồ biểu diễn chỉ số F1 trên tập dữ liệu online

Kết quả đánh giá trong hai trường hợp cho thấy hệ thống đạt hiệu suất tốt hơn trong trường hợp dữ liệu online. Điều này cũng chứng minh được tính khả thi của hệ thống *MedRES*. Kết quả đánh giá cũng là cơ sở cho việc chọn lựa tham số phù hợp nhất cho hệ thống *MedRES* như: độ dài tối đa của danh sách các hoạt chất được gợi ý (hay *TopN*).

5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Nghiên cứu của chúng tôi đã áp dụng thành công luật kết hợp để xây dựng hệ thống gợi ý

MedRES phục vụ quá trình kê đơn thuốc của bác sĩ. *MedRES* giúp bác sĩ chọn được các sản phẩm thuốc phù hợp đề kê đơn trong khoảng thời gian ngắn nhất cũng như giúp các bác sĩ trẻ, ít kinh nghiệm có thể học hỏi phương pháp kê đơn của các bác sĩ khác.

Hệ thống *MedRES* là bước tiền khả thi để tích hợp chức năng gợi ý vào CUSC-HIS Hậu Giang. Qua quá trình thử nghiệm và đánh giá hệ thống, kết quả cho thấy hệ thống có thể gợi ý tốt trong đa số trường hợp, ngay cả khi bác sĩ chưa chọn thuốc nào

và kể cả đối với các bác sĩ mới. Nhờ sử dụng thông tin về thói quen sử dụng thuốc (được tổng hợp từ lịch sử kê đơn của bác sĩ), *MedRES* có thể gợi ý thuốc phù hợp với từng bác sĩ. *MedRES* đảm bảo tính đa dạng cũng như tính mới của các gợi ý. Tính đa dạng thể hiện ở chỗ *MedRES* có thể gợi ý nhiều thuốc trong cùng một hoạt chất. Tính mới thể hiện ở chỗ *MedRES* có thể gợi ý thuốc thuộc hoạt chất mà bác sĩ chưa từng dùng.

Tuy nhiên, tập dữ liệu về đơn thuốc mà chúng tôi sử dụng để xây dựng luật kết hợp được thu thập từ CUSC-HIS Hậu Giang bao gồm các đơn thuốc của tất cả các bác sĩ (chưa quan tâm đến kinh nghiệm của bác sĩ). Do đó, các luật trong mô hình được xây dựng có độ tin cậy chưa cao, dẫn đến hệ thống gợi ý chưa tốt. Trong tương lai, để tăng độ tin cậy của các luật chúng tôi dự kiến thu thập dữ liệu về đơn thuốc của các bác sĩ có nhiều kinh nghiệm hoặc là gán cho mỗi đơn thuốc một trọng số riêng (tương ứng với khả năng kinh nghiệm của bác sĩ).

Hệ thống *MedRES* có thể cung cấp gợi ý cho các bác sĩ mới dựa trên qui luật kê đơn của các bác sĩ khác. Tuy nhiên, đối với thuốc mới (thuốc chưa từng được kê đơn) và mã bệnh mới, *MedRES* chưa gợi ý được do chưa có đơn thuốc liên quan. Để gợi ý được thuốc mới, chúng tôi dự kiến định nghĩa thêm một thuộc tính cho biết thuốc là thuốc mới. Khi đó, đối với mỗi hoạt chất trong danh sách gợi ý, một vài thuốc mới thuộc hoạt chất này sẽ có thể được thêm vào danh sách.

Bên cạnh đó, danh sách thuốc gợi ý hiện tại được sắp xếp theo từng hoạt chất (dựa trên độ hỗ trợ của luật kết hợp) nhưng thứ tự này chưa phù hợp với thứ tự kê đơn trong thực tế của các bác sĩ tại Bệnh viện Đa khoa tỉnh Hậu Giang (ví dụ: thuốc Kháng Sinh -> Kháng Viêm -> Giảm Đau -> Vitamin). Để cải tiến vấn đề này, trong tương lai, chúng tôi sẽ định nghĩa thêm thứ tự ưu tiên hiển thị của danh sách gợi ý theo nhu cầu của bác sĩ.

Nghiên cứu này chỉ mới tập trung vào chẩn đoán chính trong đơn thuốc, chưa quan tâm đến các yếu tố khác như: chẩn đoán phụ; độ tuổi, giới tính, chỉ số sinh hiệu và các yếu tố khác liên quan đến bệnh nhân. Do đó, trong một số trường hợp, hệ thống gợi ý chưa chính xác. Để có được mô hình tổng quát hơn, phù hợp hơn, cần đưa thêm các tiêu chí này vào quá trình xây dựng hệ thống.

Thói quen kê đơn của bác sĩ sẽ thay đổi theo thời gian do nhiều yếu tố tác động như: thay đổi *phác đồ điều trị*, thay đổi *danh mục thuốc của bệnh viện*,... Do đó, cần phải cập nhật mô hình sau một

thời gian nhất định. Hiện tại, do dữ liệu về đơn thuốc được tiền xử lý dưới sự hỗ trợ của chuyên gia về Dược nên mô hình được cập nhật một cách thủ công. Chúng tôi dự kiến định kỳ 3 tháng huấn luyện lại toàn bộ mô hình.

LỜI CẢM ƠN

Nhóm tác giả xin chân thành cảm ơn Trung tâm Công nghệ Phần mềm Đại học Cần Thơ đã hỗ trợ trong việc cung cấp thông tin về hệ thống quản lý bệnh viện đang triển khai tại Bệnh viện Đa khoa tỉnh Hậu Giang (CUSC-HIS). Xin cảm ơn quý lãnh đạo Bệnh viện Đa khoa tỉnh Hậu Giang đã tạo điều kiện thuận lợi cho quá trình nghiên cứu và thực nghiệm. Chân thành cảm ơn sự hỗ trợ của các bác sĩ và dược sĩ (Bệnh viện Đa khoa tỉnh Hậu Giang) trong việc xử lý dữ liệu, kiểm thử và đánh giá kết quả thực nghiệm.

TÀI LIỆU THAM KHẢO

1. Adomavicius G and Tuzhilin A, 2005. Towards the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17(6): 734-749.
2. Agrawal R, Imielinski T, and Swami A, 1993. Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 207-216, Washington D.C.
3. Balabanovic, M. and Y. Shoham, 1997. Fab: Content-based, collaborative recommendation. *Communications of the ACM* 40(3): 66-72.
4. Bộ Y tế, 2011. Thông tư số 31/2011/TT-BYT ngày 11/7/2011 của Bộ Y tế. <http://www.moh.gov.vn>, truy cập ngày 01/10/2013.
5. Breese JS, Heckerman D, and Kadie C, 1998. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *Proc. 14th Conf. Uncertainty in Artificial Intelligence*.
6. D.A. Davis, N.V. Chawla, N. Christakis and A.-L. Barabási, 2010. Time to CARE: A Collaborative Engine for Practical Disease Prediction. *Data Mining and Knowledge Discovery*, vol.20, pp.388-415.
7. F. Folino and C. Pizzuti, 2010. A Comorbidity-Based Recommendation

- Engine for Disease Prediction. In Proceedings of 23rd IEEE International Symposium on Computer-Based Medical Systems, pp. 6-12.
8. Hahsler M, Grün B, and Hornik K, 2005. arules -- A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software* 14(15): 1-25.
 9. Hahsler M. 2011 recommenderlab: A Framework for Developing and Testing Recommendation Algorithms. [Online]. Available: <http://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf>
 10. Lin W, Alvarez SA and Ruiz C, 2002. Efficient Adaptive-Support Association Rule Mining for Recommender Systems. *Data Mining and Knowledge Discovery* 6(1): 83-105.
 11. Sunantha Sodsee and Maytiyanin Komkhao, 2013. Evidence-based Medical Recommender Systems: A Review. *International Journal of Information Processing and Management (IJIPM)*, Volume4, Number6, pp. 114- 120.
 12. Herlocker J.L et al, 2004. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53.