

Hệ cơ sở tri thức

HỆ TƯ VẤN THỜI TRANG TÍCH HỢP DATA MINING

GV: PHẠM VĂN HẢI

NHÓM SINH VIÊN THỰC HIỆN:

LŨ MẠNH HÙNG

PHAN HỒNG LĨNH

NGUYỄN VĂN TÚC

VŨ CÔNG LUẬT

NỘI DUNG THUYẾT TRÌNH

- Giới thiệu về ý tưởng.
- Chi tiết các giải pháp được áp dụng để thực hiện ý tưởng.

GIỚI THIỆU VỀ Ý TƯỞNG START UP

- *Xây dựng website bán hàng thời trang tích hợp hệ tư vấn về trang phục và khai phá dữ liệu.*
- *Input:* các lựa chọn của người sử dụng trên các form có sẵn
- *Output:* các mô tả về trang phục phù hợp với yêu cầu của người dùng, có kèm theo hình ảnh được lấy từ cơ sở dữ liệu nhóm xây dựng.

MÔ TẢ BÀI TOÁN THỰC TẾ

- *Mục đích:* Để nâng cao hiệu quả bán hàng của Website, nhóm sử dụng:
- *Hệ tư vấn:* Người sử dụng sẽ nhập thông tin về trang phục bằng các form gợi ý trên Website. Dựa vào đó ứng dụng sẽ trả về kết quả là mô tả chi tiết của trang phục phù hợp nhất.

MÔ TẢ BÀI TOÁN THỰC TẾ (tiếp)

- *Khai phá dữ liệu*: Hệ thống có khả năng phân tích các đơn hàng để tìm ra các quy luật mua sắm phổ biến của người dùng.
- Tri thức đó được dùng để đưa ra gợi ý mua sắm cho người dùng và được chuyển thành luật bổ sung vào cơ sở tri thức của hệ tư vấn.

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

- Thuật toán suy diễn tiến.
- Thuật toán tìm bao đóng.
- Thuật toán tìm luật dư thừa.
- Giải thuật Apriori trong khai phá dữ liệu.

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

- *Thuật toán suy diễn tiến*: suy diễn từ một tập giả thiết ban đầu sẽ đưa ra các kết luận về trạng phục phù hợp.
- Cơ sở tri thức thuật toán:
 - Tập các mệnh đề giả thiết.
 - Tập các luật R.
 - Tập các kết luận.

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

Web Hust

Quản lý Người dùng

Quản lý Danh Mục

Quản Lý Sản Phẩm

Quản Lý Xuất nhập

ADMIN: Linh Phan

Tài khoản

Người dùng

Danh mục

Sản Phẩm

Luật

Hóa đơn

Quản lý luật

Danh sách luật:

Tự động cập nhật luật

Xóa luật dư thừa

Thêm Luật

ID	Vế Trái	Vế Phải	Độ tin cậy	Hành động
1	[25]	29	8	<div>Sửa</div> <div>Xóa</div>
2	[25, 28]	27	4	<div>Sửa</div> <div>Xóa</div>
3	[11, 22]	32	9	<div>Sửa</div> <div>Xóa</div>
4	[11, 24]	33	12	<div>Sửa</div> <div>Xóa</div>
5	[36]	31	31	<div>Sửa</div> <div>Xóa</div>
6	[17]	36	10	<div>Sửa</div> <div>Xóa</div>
7	[18]	36	23	<div>Sửa</div> <div>Xóa</div>
8	[15]	35	25	<div>Sửa</div> <div>Xóa</div>
9	[12, 34]	36	5	<div>Sửa</div> <div>Xóa</div>
10	[16]	34	7	<div>Sửa</div> <div>Xóa</div>
11	[25]	34	11	<div>Sửa</div> <div>Xóa</div>
12	[26]	37	7	<div>Sửa</div> <div>Xóa</div>
13	[37]	40	8	<div>Sửa</div> <div>Xóa</div>
14	[15, 28]	38	3	<div>Sửa</div> <div>Xóa</div>
15	[23]	28	11	<div>Sửa</div> <div>Xóa</div>
16	[36]	38	32	<div>Sửa</div> <div>Xóa</div>
17	[12]	26	11	<div>Sửa</div> <div>Xóa</div>

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

- *Thuật toán tìm bao đóng* của tập giả thiết X trên một tập luật cho ta tập tất cả các sự kiện có thể suy ra được từ X dựa trên tập luật đó.
- *Thuật toán tìm luật dư thừa*: Với mỗi luật nằm trong tập luật ta sẽ loại nó khỏi tập luật, rồi tiến hành tìm bao đóng của vế trái luật đó dựa trên tập luật mới. Nếu bao đóng tìm được có chứa vế phải của luật đó thì luật đó là luật dư thừa (trả về true) còn ngược lại trả về false.

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

➤ *Giải thuật Apriori:*

- *Input:* Tập dữ liệu đơn hàng.
 - giá trị minsup.
 - giá trị minconf.
- *Output:* Tập tất cả các luật kết hợp $X \rightarrow Y$ sao cho:
 - $\text{sup}(X \rightarrow Y) \geq \text{minsup}$
 - $\text{conf}(X \rightarrow Y) \geq \text{minconf}$

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

➤ *Giải thuật Apriori*: gồm hai bước chính:

- *Bước 1*: tìm tất cả tập các món hàng có độ hỗ trợ lớn hơn bằng *minsup*.
- *Bước 2*: dùng tập trên để sinh ra các luật kết hợp có độ tin cậy lớn hơn bằng *minconf*.

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

➤ Giải thuật Apriori:

- Quét T $\Rightarrow C_1 = \{\{A\}: 2, \{B\}: 3, \{C\}: 3, \{D\}: 1, \{E\}: 3\}$
 $F_1 = \{\{A\}: 2, \{B\}: 3, \{C\}: 3, \{E\}: 3\}$
 $\Rightarrow C_2 = \{\{AB\}, \{AC\}, \{AE\}, \{BC\}, \{BE\}, \{CE\}\}$
- Quét T $\Rightarrow C_2 = \{\{AB\}: 1, \{AC\}: 2, \{AE\}: 1, \{BC\}: 2, \{BE\}: 3, \{CE\}: 2\}$
 $F_2 = \{\{AC\}: 2, \{BC\}: 2, \{BE\}: 3, \{CE\}: 2\}$
 $C_3 = \{BCE\}$
- Quét T $\Rightarrow C_3 = \{\{BCE\}: 2\} \Rightarrow F_3 = \{BCE\}$

➤ Bước 1: tìm tất cả tập các món hàng có độ hỗ trợ lớn hơn bằng *minsup*.

➤ Dữ liệu đơn hàng T với *minsup* = 50%

TID	Món hàng
T1	A, C, D
T2	B, C, E
T3	A, B, C, E
T4	B, E

CHI TIẾT CÁC PHƯƠNG PHÁP ĐƯỢC SỬ DỤNG

➤ *Giải thuật Apriori:*

- *Bước 2:* dùng tập trên để sinh ra các luật kết hợp có độ tin cậy lớn hơn bằng *minconf*.
- Các tập được xét: {A}: 50%, {B}: 75%, {C}: 75%, {E}: 75%, {AC}: 50%, {BC}: 50%, {BE}: 75%, {CE}: 50%, {BCE}: 50%.
- Cho $minconf \geq 50\%$
- Xét tập thường xuyên {A, C}
- $A \rightarrow C \text{ conf} = 100\% > minconf$.
- $C \rightarrow A \text{ conf} = 67\% > minconf$.
- *Tính tương tự với các trường hợp còn lại.*

KẾT QUẢ ĐÁNH GIÁ

- Nhóm xây dựng tập luật gồm 31 luật đã bao phủ toàn bộ tập Sản phẩm nên mọi truy vấn thì hệ chuyên gia đều trả về kết quả chính xác
- Độ bao phủ: 100% mọi truy vấn đều trả về kết quả
- Độ chính xác: 100% mọi kết quả đều chính xác
- Thêm luật tự động với độ chính xác 100%

KẾT QUẢ SO SÁNH

- Nhóm xây dựng thuật toán: Loại bỏ luật dư thừa ra khỏi tập luật giúp tối ưu tập luật, giảm chi phí khi chạy thuật toán (thời gian, bộ nhớ).
- Sau đây là thời gian thực thi truy vấn với tập luật ban đầu (chưa loại bỏ luật dư thừa) và tập luật tối ưu (đã loại bỏ luật dư thừa):

Thời gian thực thi (s)	Tập luật dư thừa	Tập luật tối ưu
1	0.056	0.036
2	0.064	0.025
3	0.050	0.022
4	0.057	0.031
5	0.055	0.025
6	0.056	0.027
7	0.059	0.023
8	0.062	0.033
9	0.055	0.024
10	0.051	0.026