**KHAI THÁC DỮ LIỆU VÀ ỨNG DỤNG**

**Association Rules Report**

C:\Users\tdqua_000\Dropbox\SS-Slides\DeCuong-CDIO\Template CDIO v4.2\Templates\Hinh anh\LogoTruong.png

Bộ môn Khoa học Máy tính

Khoa Công nghệ thông tin

Đại học Khoa học tự nhiên TP HCM

## Mục lục

[1 Thông tin nhóm 3](#_Toc509925002)

[2 Apriori 4](#_Toc509925003)

[3 FP-Growth 6](#_Toc509925004)

[4 Tập phổ biến đóng và tập phổ biến tối đại 7](#_Toc509925005)

[5 Các độ đo lý thú 8](#_Toc509925006)

# Thông tin nhóm

|  |  |  |
| --- | --- | --- |
| **MSSV** | **Họ Tên** | **Email** |
| 1512203 | Nguyễn Quốc Huy | 1512203@student.hcmus.edu.vn |
| 1512262 | Võ Anh Khoa | 1512262@student.hcmus.edu.vn |

# Apriori

Các bước tìm các frequent itemsets với minsup = 60% (min.count >= 4):

***Bước 1:***

|  |  |
| --- | --- |
| C1 | Support |
| Bia | 3 |
| Bánh mì | 4 |
| Hành | 4 |
| Sữa | 4 |
| Khoai tây | 5 |

|  |  |
| --- | --- |
| L1 | Support |
| Sữa | 4 |
| Hành | 4 |
| Bánh mì | 4 |
| Khoai tây | 5 |

***Bước 2***

|  |  |
| --- | --- |
| C2 | Support |
| Sữa, Hành | 2 |
| Sữa, Bánh mì | 2 |
| Sữa, Khoai tây | 3 |
| Hành, Bánh mì | 3 |
| Khoai tây, Hành | 4 |
| Khoai tây, Bánh mì | 4 |

|  |  |
| --- | --- |
| L2 | Support |
| Khoai tây, Hành | 4 |
| Khoai tây, Bánh mì | 4 |

***Bước 3***

|  |  |
| --- | --- |
| C3 | Support |
| Khoai tây, Hành, Bánh mì | 3 |

: Các itemsets có support < minsup

**Các frequent itemsets:   
{Sữa}, {Khoai tây}, {Hành}, {Bánh mì}, {Khoai tây, Hành}, {Khoai tây, Bành mì}**

***Bước 4:*** Tính confident, min conf = 80%

|  |  |  |  |
| --- | --- | --- | --- |
| Associcatịon Rules | Support(A,B) | Support(A) | Confidence |
| Bánh mì => Khoai tây | 4 | 4 | 100% |
| Hành => Khoai tây | 4 | 4 | 100% |
| Khoai tây => Bánh mì | 4 | 5 | 80% |
| Khoai tây => Hành | 4 | 5 | 80% |

Vậy ta tìm được 4 luật kết hợp thỏa mãn minsup >= 60% và minconf >= 80%:

R1: Khoai tây => Hành (support = 67%, confidence = 80%)

R1: Hành => Khoai tây (support = 67%, confidence = 100%)

R1: Khoai tây => Bánh mì (support = 67%, confidence = 80%)

R1: Bánh mì => Khoai tây (support = 67%, confidence = 100%)

# FP-Growth

### Mô tả source code

Cấu trúc của một transaction:

### So sánh thuật toán Apriori và FP-Growth

# Tập phổ biến đóng và tập phổ biến tối đại

### Tập phổ biến đóng

Tập phổ biến đóng là tập phổ biến mà không có tập nào bao nó có cùng độ phổ biến.

- Với F là tập hợp gồm tất cả tập phổ biến.

- Gọi C là tập hợp gồm tất cả tập phổ biến đóng.

### Tập phổ biến tối đại

Tập phổ biến tối đại là tập phổ biến mà không có tập nào bao nó là phổ biến.

### Tập phổ biến đóng và tối đại cho bài 1

Tập phổ biến đóng:

{Sữa}, {Khoai tây}, {Khoai tây, Hành}, {Khoai tây, Bánh mì}.

Tập phổ biến tối đại:

{Sữa}, {Khoai tây, Hành}, {Khoai tây, Bánh mì}.

# Các độ đo lý thú

### Công thức tính các độ đo confidence, lift, conviction, leverage

Độ đo Confidence trình bày mức độ thường xuyên của tính đúng đắn của luật

Độ đo của luật , ứng theo tập hợp giao tác T là tỉ lệ giao tác có chứa X và Y so với giao tác chỉ chứaa X.

Độ đo Confidence được định nghĩa như sau:

Độ đo Lift là tỷ lệ của sự hỗ trợ quan sát được đối với X và Y là độc lập

Độ đo Conviction được định nghĩa:

Độ đo Leverage được định nghĩa

### Tính các độ đo cho các luật kết hợp ở bài 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lift | Conviction | Leverage |
| Khoai tây Hành | 1.2 | 1.67 | 0.11 |
| Hành Khoai tây | 1.2 |  | 0.11 |
| Khoai tây Bánh mì | 1.2 | 1.67 | 0.11 |
| Bánh mì Khoai tây | 1.2 |  | 0.11 |

* Đối với Luật :   
  Trước hết ta ta tính   
  Ta có :
* Đối với Luật :

Trước hết ta ta tính   
Ta có :

* Đối với Luật :   
  Trước hết ta ta tính   
  Ta có :
* Đối với Luật :

Trước hết ta ta tính   
Ta có :

### Nhận xét

Các độ đo này có sự khác biệt, vì ý nghĩa của chúng có sự khác biệt.

**Confidence**: Confidence cho biết tỉ lệ phần trăm một luật tìm được là đúng.

**Lif**t: Lift cho biết độ độc lập giữa 2 transaction. Nếu lift xấp xỉ với 1 thì A và B độc lập. Nếu lift > 1, giá trị của nó cho ta biết mức độ phụ thuộc giữa 2 transaction, và dẫn đến tăng tính hữu ích. Thường hiệu quả với những item có support nhỏ.

**Conviction**: Tương tự như Lift, nhưng khắc phục được nhược điểm của lift là không phân biệt được chiều của luật.

**Leverage**: Leverage cho biết mức độ độc lập giữa 2 items. Thường hiệu quả với những item có support lớn.