

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH**

Khoa Công nghệ thông tin

Môn Khai thác dữ liệu và ứng dụng

**Báo cáo bài tập thực hành 2:
TẬP PHỔ BIẾN VÀ LUẬT KẾT HỢP**

GV: Bùi Thị Danh
Lê Hoài Bắc
Nguyễn Hoàng Khai
Nguyễn Ngọc Thảo

Sinh viên thực hiện:

Vũ Hồng Anh	1412019
Trần Thiên Hoàng	1412185
Võ Phương Hòa	1412192

Tháng 10/2016, Tp. Hồ Chí Minh

Mục lục

I. Nội dung thực hành	1
II. Phân chia nhiệm vụ và kết quả.....	1
III. Test.....	1
1. Yêu cầu 1	1
2. Yêu cầu 2	2
3. Yêu cầu 3	4
IV. Tài liệu tham khảo	5

I. Nội dung thực hành

Sinh viên cài đặt thuật toán Apriori để phát sinh các tập phổ biến và luật kết hợp từ tập dữ liệu đã cho (reetail.dat).

1. Cài đặt hàm phát sinh ứng viên.
2. Cài đặt thuật toán Apriori sử dụng hàm phát sinh ứng viên ở yêu cầu 1.
3. Cài đặt thuật toán phát sinh luật kết hợp thỏa giá trị minconf cho trước.

II. Phân chia nhiệm vụ và kết quả

STT	MSSV	Thành viên	Nhiệm vụ	Kết quả
1	1412019	Vũ Hồng Anh	Yêu cầu 1	Hoàn thành
2	1412192	Võ Phương Hòa	Yêu cầu 2	Hoàn thành
3	1412185	Trần Thiên Hoàng	Yêu cầu 3	Hoàn thành

III. Test

1. Yêu cầu 1

Input 1:

1
2
3

Output 1:

1 2
1 3
2 3

Input 2:

1 2
1 3
2 3

Output 2:

1 2 3

Input 3:

1 2 3
1 2 4
1 3 4
2 3 4
1 3 5
1 2 5
2 3 5

Output 3:

1 2 3 4
1 2 3 5

Input 4:

1 2 3
1 2 4
1 3 5
2 3 5
3 4 5

Output 4: rỗng**2. Yêu cầu 2**

(sử dụng thư viện numpy trong file myApriori.py)

Input: retail.dat

Output:

- mins = 0.03 : (time = 396.01 s)

0.17 32
0.03 36
0.18 38
0.57 39
0.17 41
0.48 48
0.05 65
0.04 89
0.03 110
0.04 170
0.04 225
0.03 237

0.03 32 38
0.1 32 39
0.04 32 41
0.09 32 48
0.03 36 38
0.12 38 39
0.04 38 41
0.09 38 48
0.03 38 110
0.03 38 170
0.13 39 41
0.33 39 48
0.03 39 65
0.03 39 89
0.1 41 48
0.03 48 89
0.06 32 39 48
0.03 38 39 41
0.07 38 39 48
0.08 39 41 48

- mins = 0.05: (time = 375.12 s)

0.17 32
0.18 38
0.57 39
0.17 41
0.48 48
0.05 65
0.1 32 39
0.09 32 48
0.12 38 39
0.09 38 48
0.13 39 41
0.33 39 48
0.1 41 48
0.06 32 39 48
0.07 38 39 48
0.08 39 41 48

- mins = 0.1 : (time = 338.23 s)

0.17 32
0.18 38
0.57 39
0.17 41
0.48 48
0.12 38 39
0.13 39 41
0.33 39 48
0.1 41 48

- mins = 0.2 : (time = 370.90 s)

0.57 39
0.48 48
0.33 39 48

3. Yêu cầu 3

Input: file output mins = 0.05 (yêu cầu 2)

Output:

- mincons = 0.5 , k = 1:

0.59 32 -> 32 39
0.53 32 -> 32 48
0.67 38 -> 38 39
0.5 38 -> 48 38
0.76 41 -> 41 39
0.58 39 -> 48 39
0.69 48 -> 48 39
0.59 41 -> 48 41

- mincons = 0.5, k = 3:

0.59 32 -> 32 39

0.53 32 -> 32 48

0.67 38 -> 38 39

0.5 38 -> 48 38

0.76 41 -> 41 39

0.58 39 -> 48 39

0.69 48 -> 48 39

0.59 41 -> 48 41

- mincons = 0.7, k = 1:

0.76 41 -> 41 39

- mincons = 0.7, k = 3:

0.76 41 -> 41 39

IV. Tài liệu tham khảo

1. tutorialspoint.com/python/python_dictionary.htm
2. docs.python.org/2.4/lib/types-set.html
3. stackoverflow.com/questions/11200137/installing-numpy-on-64bit-windows-7-with-python-2-7-3