Kiểm tra chất lượng dữ liệu click

Sử dụng các trường dữ liệu sub\_category, lognum, function, số lượng bản ghi NULL của tất cả các trường là 0. Do số lượng bản ghi NULL là 0, dữ liệu đã đáp ứng đủ chất lượng để làm đầu vào cho mô hình. Thống kê số lượng bản ghi NULL được ghi lại trong bảng dưới đây.

Đề xuất giải pháp mô hình tương quan các sản phẩm dựa trên luật kết hợp

Để xây dựng mô hình tương quan sản phẩm, với mỗi chủ đề (sub\_category), chúng tôi chọn ra top 2 icon (function) có lượt click cao nhất thu thập được từ dữ liệu 3 tháng gần nhất. Sau đó, sử dụng luật kết hợp để tìm ra các sub\_category khách hàng thường click cùng nhau. Như vậy, với mỗi thuê bao định danh (isdn) của khách hàng, output gồm các sub\_category được đề xuất và các top 2 function tương ứng được đề xuất cho mỗi sub\_category được gợi ý.

Ví dụ về đầu ra của mô hình gợi ý

|  |  |  |
| --- | --- | --- |
| isdn | Sub\_category | Function |
|  | [‘Mua sắm - Giải trí’, ‘Du lịch – Đi lại] | [] |
|  | [‘Tài chính’, ‘Ngân hàng’] | [] |

Các tham số để finetune mô hình:

Min\_support – Độ hỗ trợ tối thiểu

Chỉ số đánh giá: Lift – Chỉ số lift

Xây dựng mô hình tương quan các sản phẩm dựa trên luật kết hợp

Input:

Output

Bối cảnh dữ liệu: một isdn có thể xuất hiện nhiều lần, mỗi lần thể hiện tổng số lần click một function và sub\_category tương ứng. Ví dụ về dữ liệu

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Các bước tạo ra luật kết hợp

Bước 1: nhóm theo isdn, gom các sub\_category thành một list tổng hợp tất cả các chủ đề mà người dùng MyViettel đã click

Bước 2: Tính support và lift cho từng cặp antecedents, consequents

Bước 3: Finetune chỉ số lift và ngưỡng support cho các cặp antecedents, consequents của mô hình sao cho tất cả các sub\_category xuất hiện ít nhất một lần trong antecedents hoặc consequents.

Thực hiện fine tune mô hình tương quan các sản phẩm dựa trên luật kết hợp theo tham số cao

Các trường dữ liệu ở kết quả thống kê cần quan tâm trong mô hình này là…

Antecedents: (các) chủ đề mà khách hàng click vào trước

Consequents: (các) chủ đề mà khách hàng click vào sau

Lift: Chỉ số lift, chỉ số lift lớn hơn 1 và càng cao, cho thấy xác suất mà khách hàng click theo thứ tự được dự đoán là càng lớn.

Finetune lần 1 với min\_support = , lift =

Với các tham số min\_support =, lift = , kết quả mô hình được thống kê trong bảng dưới đây.

Với kết quả mô hình trên, antecedents và consequents còn thiếu chủ đề …, nên với tham số này, mô hình sẽ không thể dự đoán liệu với những khách hàng truy cập chủ đề…, thì đề xuất thêm chủ đề nào. Vì vậy, lần finetune thứ 2 sẽ được thực hiện để cho ra kết quả có chủ đề... xuất hiện ít nhất 1 lần trong antecedents hoặc consequents

Thực hiện fine tune mô hình tương quan các sản phẩm dựa trên luật kết hợp theo tham số trung bình

Finetune lần 2 với min\_support = , lift =

Với các tham số min\_support =, lift = , kết quả mô hình được thống kê trong bảng dưới đây.

Với kết quả mô hình trên, antecedents và consequents còn thiếu chủ đề …, nên với tham số này, mô hình sẽ không thể dự đoán liệu với những khách hàng truy cập chủ đề…, thì đề xuất thêm chủ đề nào. Vì vậy, lần finetune thứ 3 sẽ được thực hiện để cho ra kết quả có chủ đề... xuất hiện ít nhất 1 lần trong antecedents hoặc consequents

Thực hiện fine tune mô hình tương quan các sản phẩm dựa trên luật kết hợp theo tham số thấp

Finetune lần 3 với min\_support = , lift =

Với các tham số min\_support =, lift = , kết quả mô hình được thống kê trong bảng dưới đây.

Với kết quả mô hình trên, antecedents và consequents đã đầy đủ các chủ đề, tuy nhiên, để đảm bảo cân bằng giữa tính đầy đủ các chủ đề và tham số min\_support, tham số trung bình ở lần finetune thứ 2 sẽ được sử dụng.

Trực quan hoá và trinh bày kết quả dự đoán của mô hình tương quan các sản phẩm dựa trên luật kết hợp

Phân tích kết quả dự đoán của mô hình tương quan các sản phẩm dựa trên luật kết hợp

Đề xuất mô hình gợi ý sản phẩm/khách hàng dựa trên tương quan sản phẩm

Xây dựng mô hình mô hình gợi ý sản phẩm/khách hàng dựa trên tương quan sản phẩm

Input

Đầu vào của mô hình bao gồm nội dung và các trường dữ liệu tương ứng sau.

Thuê bao định danh khách hàng (isdn)

Số lần truy cập icon (lognum)

Chủ đề cố định trên app MyViettel (sub\_category)

Icon trên app MyViettel (function)

Output:

Đầu ra của mô hình bao gồm nội dung và các trường dữ liệu tương ứng sau.

Thuê bao định danh khách hàng (isdn)

List các chủ đề được mô hình đề xuất trên app MyViettel (sub\_category)

List các function được mô hình đề xuất trên app MyViettel (function)

Tên mô hình/thuật toán: Apriori

Các bước xây dựng mô hình

Bước 1: tổng hợp các lần truy cập vào các sub\_category của từng khách hàng MyViettel theo từng isdn.

Bước 2: lấy top 2 function của mỗi sub\_category có lượt click cao nhất

Bước 3: chạy mô hình apriori

Bước 4: từ bảng tổng hợp kết quả các luật, finetune tham số min\_support và lift >1 sao cho sao cho từng sub\_category xuất hiện ít nhất 1 lần trong antecedents hoặc consequents

Bước 5: lấy kết quả finetune phù hợp nhất làm tham số cuối của mô hình.

Thư viện sử dụng: mlxtend, numpy, pandas