Interestingness measure

Bài 3. Độ đo lý thú.

3.1

Độ tin cậy (confidence)

Độ tin cậy của luật X 🡪 Y được định nghĩa:

Conf(X 🡺 Y) = supp(X U Y) / supp(X)

Trong đó supp(X U Y) là độ phổ biến của X và Y trong cùng một giao dịch.

Supp(X) là độ phổ biến của X.

Hiểu một cách khác, độ tin cậy của X 🡪 Y chính là xác xuất các bộ chứa Y trên điều kiện các bộ giá trị chứa X. Độ tin cậy có giá trị trong khoảng [0,1], nếu X và Y độc lập nhau thì độtin cậy của X 🡪 Y bằng 0, ngược lại nếu Y xuất hiện trong mọi dòng dữ liệu chứa X thì độ tin cậy của luật bằng 1.

Độ đo Lift

Độ đo này dùng để đánh giá mối quan hệ giữa X và Ytrong luật X 🡪 Y

Lift(X 🡺 Y) =

Trong đó supp(X U Y) là độ phổ biến của X và Y trong cùng một giao dịch.

Supp(X) là độ phổ biến của X, và supp(Y) là độ phổ biến của Y.

Giá trị của Lift thuộc khoảng [0,∞) , các luật kết hợp với Lift>1 được xem là hữu ích vì khi đó, Conf(X 🡪 Y) > Supp(Y) có nghĩa xác xuất của Y thỏa điều kiện X lớn hơn xác xuất của Y không cần thỏa điều kiện nào, nói cách khác sự tồn tại của Y phụ thuộc vào sự tồn tại của X. Nếu Lift =1 thì X và Y là độc lập.

Độ đo chắc chắn (Conviction)

Độ chắc chắn của luật X 🡪 Y được định nghĩa:

conv(X 🡺 Y) =

Supp(Y) là độ phổ biến của Y

Conf(X 🡺 Y) là độ tinh cậy của tập X, Y

Conviction được xem như là sự thay thế cho độ tin cậy trong trường hợp không thu được kết quả thỏa đáng từ độ tin cậy, công thức Conviction khá giống Lift nhưng không như Lift, Conviction phụ thuộc vào hướng của luật (Conviction(X 🡪 Y) # Conviction(Y 🡪 X)). Giá tr ịcủa Conviction thuộc khoảng [0, ∞), các luật kết hợp có Conviction càng cao (>1) thì càng hữu ích, X và Y là độc lập nếu Conviction bằng 1.

Độ đo Leverage

Độ độ Leverage được xem như là độ mạnh của luật và được định nghĩa:

Lever(X 🡪 Y) = P(X U Y) – P(X)P(Y) = Supp(X 🡪 Y) – Supp(X)\*Supp(Y)

Trong đó supp(X U Y) là độ phổ biến của X và Y trong cùng một giao dịch.

Supp(X) là độ phổ biến của X, và supp(Y) là độ phổ biến của Y.

Leverge dùng để đo khoảng cách xác xuất giữa X, Y xuất hiện cùng nhau và xác xuất mà X và Y thỏa điều kiện phụ thuộc. Giá trị của Leverage thuộc khoảng [-0.25, 0.25], nếu Levearge = 0 thì X độc lập với Y.

Ghi chú: Supp(X) = Count(X)/|D|. Trong đó Count(X) là số lượng các giá trị chứa X trong tổng số bộ giá trị |D|.

3.2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Luật | Supp(X U Y) | Supp(X) | Supp(Y) | Confidence | Lift | Conviction | Leverage |
| Bread 🡪 Jam | 0.5 | 0.5 | 0.625 | 1 | 1.6 | Chia cho 0 | 0.1875 |
| Chips 🡪 Soda | 0.5 | 0.5 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.125 |
| Peanuts 🡪 Fruit | 0.5 | 0.5 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.125 |
| Milk 🡪 Fruit | 0.625 | 0.75 | 0.75 | 0.83 | 1.11 | 1.4705882 | 0.0625 |
| Fruit 🡪 Milk | 0.625 | 0.75 | 0.75 | 0.83 | 1.11 | 1.4705882 | 0.0625 |
| Soda 🡪 Milk | 0.625 | 0.75 | 0.75 | 0.83 | 1.11 | 1.4705882 | 0.0625 |
| Milk 🡪 Soda | 0.625 | 0.75 | 0.75 | 0.83 | 1.11 | 1.4705882 | 0.0625 |
| Jam 🡪 Bread | 0.5 | 0.625 | 0.5 | 0.8 | 1.6 | 2.5 | 0.1875 |
| Jam 🡪 Milk | 0.5 | 0.625 | 0.75 | 0.8 | 1.06 | 1.25 | 0.03125 |
| Jam 🡪 Soda | 0.5 | 0.625 | 0.75 | 0.8 | 1.06 | 1.25 | 0.03125 |
| Fruit, Soda 🡪 Milk | 0.5 | 0.5 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.125 |
| Chips, Jam 🡪 Bread | 0.375 | 0.375 | 0.5 | 1 | 2 | Chia cho 0 | 0.1875 |
| Bread, Chips 🡪 Jam | 0.375 | 0.375 | 0.625 | 1 | 1.6 | Chia cho 0 | 0.140625 |
| Bread, Soda 🡪 Chips | 0.375 | 0.375 | 0.5 | 1 | 2 | Chia cho 0 | 0.1875 |
| Bread, Chips 🡪 Soda | 0.375 | 0.375 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.09375 |
| Bread, Milk 🡪 Jam | 0.375 | 0.375 | 0.625 | 1 | 1.6 | Chia cho 0 | 0.140625 |
| Bread, Soda 🡪 Jam | 0.375 | 0.375 | 0.625 | 1 | 1.6 | Chia cho 0 | 0.140625 |
| Chips, Jam 🡪 Soda | 0.375 | 0.375 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.09375 |
| Chips, Milk 🡪 Soda | 0.375 | 0.375 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.09375 |
| Fruit, Jam 🡪 Milk | 0.375 | 0.375 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.09375 |
| Milk, Peanuts 🡪 Fruit | 0.375 | 0.375 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.09375 |
| Chips, Jam 🡪 Bread, Soda | 0.375 | 0.375 | 0.375 | 1 | 2.67 | Chia cho 0 | 0.234375 |
| Bread, Soda 🡪 Chips, Jam | 0.375 | 0.375 | 0.375 | 1 | 2.67 | Chia cho 0 | 0.234375 |
| Bread, Chips 🡪 Jam, Soda | 0.375 | 0.375 | 0.5 | 1 | 2 | Chia cho 0 | 0.1875 |
| Milk, Soda 🡪 Fruit | 0.5 | 0.625 | 0.75 | 0.8 | 1.06 | 1.25 | 0.03125 |
| Fruit, Milk 🡪 Soda | 0.5 | 0.625 | 0.75 | 0.8 | 1.06 | 1.25 | 0.03125 |
| Chips, Jam, Soda 🡪 Bread | 0.375 | 0.375 | 0.5 | 1 | 2 | Chia cho 0 | 0.1875 |
| Bread, Jam, Soda 🡪 Chips | 0.375 | 0.375 | 0.5 | 1 | 2 | Chia cho 0 | 0.1875 |
| Bread, Chips, Soda 🡪 Jam | 0.375 | 0.375 | 0.625 | 1 | 1.6 | Chia cho 0 | 0.140625 |
| Bread, Chips, Jam 🡪 Soda | 0.375 | 0.375 | 0.75 | 1 | 1.33 | Chia cho 0 | 0.09375 |

3.3

So sánh với weka

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Luật | Confidence | Lift | Conviction | Leverage |
| Bread 🡪 Jam | 1 | 1.6 | 1.5 | 0.19 |
| Chips 🡪 Soda | 1 | 1.33 | 1 | 0.13 |
| Peanuts 🡪 Fruit | 1 | 1.33 | 1 | 0.13 |
| Milk 🡪 Fruit | 0.83 | 1.11 | 0.75 | 0.06 |
| Fruit 🡪 Milk | 0.83 | 1.11 | 0.75 | 0.06 |
| Soda 🡪 Milk | 0.83 | 1.11 | 0.75 | 0.06 |
| Milk 🡪 Soda | 0.83 | 1.11 | 0.75 | 0.06 |
| Jam 🡪 Bread | 0.8 | 1.6 | 1.25 | 0.19 |
| Jam 🡪 Milk | 0.8 | 1.07 | 0.63 | 0.03 |
| Jam 🡪 Soda | 0.8 | 1.07 | 0.63 | 0.03 |
| Fruit, Soda 🡪 Milk | 1 | 1.33 | 1 | 0.13 |
| Chips, Jam 🡪 Bread | 1 | 2 | 1.5 | 0.19 |
| Bread, Chips 🡪 Jam | 1 | 1.6 | 1.5 | 0.19 |
| Bread, Soda 🡪 Chips | 1 | 2 | 1.5 | 0.19 |
| Bread, Chips 🡪 Soda | 1 | 1.33 | 0.75 | 0.09 |
| Bread, Milk 🡪 Jam | 1 | 1.6 | 1.13 | 0.14 |
| Bread, Soda 🡪 Jam | 1 | 1.6 | 1.13 | 0.14 |
| Chips, Jam 🡪 Soda | 1 | 1.33 | 0.75 | 0.09 |
| Chips, Milk 🡪 Soda | 1 | 1.33 | 0.75 | 0.09 |
| Fruit, Jam 🡪 Milk | 1 | 1.33 | 0.75 | 0.09 |
| Milk, Peanuts 🡪 Fruit | 1 | 1.33 | 0.75 | 0.09 |
| Chips, Jam 🡪 Bread, Soda | 1 | 2.67 | 1.88 | 0.23 |
| Bread, Soda 🡪 Chips, Jam | 1 | 2.67 | 1.88 | 0.23 |
| Bread, Chips 🡪 Jam, Soda | 1 | 2 | 1.5 | 0.19 |
| Milk, Soda 🡪 Fruit | 0.8 | 1.07 | 0.63 | 0.03 |
| Fruit, Milk 🡪 Soda | 0.8 | 1.07 | 0.63 | 0.03 |
| Chips, Jam, Soda 🡪 Bread | 1 | 2 | 1.5 | 0.19 |
| Bread, Jam, Soda 🡪 Chips | 1 | 2 | 1.5 | 0.19 |
| Bread, Chips, Soda 🡪 Jam | 1 | 1.6 | 1.13 | 0.14 |
| Bread, Chips, Jam 🡪 Soda | 1 | 1.33 | 0.75 | 0.09 |

3.4

Nhận xét: Qua đối chiếu kết quả giữa tính tay và bằng weka là tương đương nhau, riêng độ đo leverage weka làm tròn 2 chữ số, riêng độ đo conviction tính tay thì lại mẫu bằng 0 🡪 không chia được, weka lại cho kết quả.