# Decision rules

Iván Andrés Trujillo Abella

June 6, 2022

**Abstract**

The aim of this document is introduce to recommendation systems, and outline the application of to social analysis and another fields.

## 1 Introduction

brush up: repasar, refrescar. wrap up: Envolver. blunted: Embotado. runway: planes taking off: landings: shoot them down: cushion: glitch: shifting: take out: lining up: hacer fila queue; fila, hacer cola.

## 2 Basic concepts

a set of items of one transaction is termed **itemset** the number of elements could be expresed as **k-itemset**.

### 2.1 stack

is a data structure that con only remove and add over the top of the structure, think in a list with only two methods, pop and append. also denominated lifo (Last In, First Out).

### 2.2 queue

### 2.3 Heaps

### 2.4 Market Based Analysis

### 2.5 Association Rule

Here is important the concept of Co occurrence.

itemset $a, b, c$, a k-itemset is a set of k elements. $x \longrightarrow y$

we need define two metrics, support (s) and confidence(c). we need to define, $|s|$ the size or cardinality, and $\alpha(s)$ is the time that appear a sequence, for instance $\{x, y\}$

| id | itemsets |
|---|---|
| 1 | {a,b,c} |
| 2 | {x,y} |
| 3 | {x,y,z} |

One association rule could be $\{x\} \longrightarrow \{y\}$ or $\{x, y\} \longrightarrow \{z\}$. Note that we could be defined basket of products or disease and mobility's.

According to the last table $\alpha(\{x, y\}) = 2$.

we define that $D$ is dataset, teherefore the support is defined as.

$$s(x \longrightarrow y) = \frac{\alpha(x \cup y)}{|D|} \tag{1}$$

The confidence is therefore:

$$c(x \longrightarrow y) = \frac{\alpha(x \cup y)}{\alpha(x)} \tag{2}$$

Note here that is as a conditional probability, means, the probability of occur $x$ or $y$ given the number of times that occur $x$.

lift

$$l(x \longrightarrow y) = \frac{s(x \longrightarrow y)}{s(x)s(y)} \tag{3}$$

the independent occurrence, note that lift it is measure of actual confidence regard to the expected confidence ( the random occurrence of $x$ and $y$). if lift is equal to one then is not correlated.

The rule is good if improve the decision regarding a random decision. This measure allow us have a insigth of the relevance of the rule.

In other words lift is the increase in the empirical probability of see the consequent given that known the antecedent

$$l(x \longrightarrow y) = \frac{P(y \mid x)}{p(y)} \tag{4}$$

The data mining purpose is find all possible combination of rules of association that satisfy $c(x_i \longrightarrow y_i > limit$ and also to support, the limits could be differ.

If we have $\{a, b\} \longrightarrow \{c\}$. we search in transactions $a, b, c$ not matter until now the order. now we call **frequent itemset** a the collection of rules that the limits.

## 2.6 A priori algorithm

key idea: *if s is a itemset of $|s| = k$ is infrecuent all possible super set that could contain s also will be infrequent.*

### 2.6.1 Example A prior algorithm

| id | itemset | Count |
|---|---|---|
| 1 | A | 4 |
| 2 | B | 3 |
| 3 | C | 4 |
| 4 | D | 2 |
| 5 | E | 4 |
| 6 | F | 3 |

Assume that we establish a limit of count, then process if $limit \geq 3$ then the possible candidates are; Is the form of combine 5 in 2 is $\frac{5!}{(5-2)!2!}$. Note here that we need make combinations of elements:

| id | itemset | Count |
|---|---|---|
| 1 | A | 4 |
| 2 | B | 3 |
| 3 | C | 4 |
| 5 | E | 4 |
| 6 | F | 3 |

We need compare, each minimum support. and the candidates will be,

in the order ideas, the following

Note here the following problem, generate the tuple of three items or combinations, that differ. $(A, B, C), (A, B, F), (A, C, F), (B, C, E)$. the way of combine three different letters in three ways, and the way of order two letters in one. Note that here the letters are ordered and only differ in the last letter.

| id | itemset | Count |
|---|---|---|
| 1 | A,B | 3 |
| 2 | A,C | 3 |
| 3 | A,E | 2 |
| 4 | A,F | 3 |
| 5 | B,C | 3 |
| 6 | B,E | 3 |
| 7 | B,F | 2 |
| 8 | C,E | 2 |
| 9 | C,F | 2 |
| 10 | E,F | 1 |

| id | itemset | Count |
|---|---|---|
| 1 | A,B | 3 |
| 2 | A,C | 3 |
| 3 | A,F | 4 |
| 4 | B,C | 3 |
| 5 | B,E | 3 |

```
 for each k item in dataset of |k|=1:
    let be a candidate C1 if s(k) > limit:
for each combination (k) of candidates of |k|=2 :
    let be a candidate C2 if s(k) > limit:
repeat the process until reach the |k| to the |j| (maximun number of item in a transaction).
```

Is more efficient initialize the algorithm with greater $k$ due we can check in database all possible major sets.

then for instance $[A][][]$ and we can combine three different letters in two spaces and for the other only two different letters by two spaces. The algotrithm finish with the length of maximun items in dataset.

### 2.7 Principle of monotonicidad

Python to extract data from twitter:

Consumer Key Consumer secret Access Token Access Token Secret

### 2.8 Tweepy package

## 3 OAuth security protocol

copy your acces token in a file with extension .k,

A recommender system,

take in mind that the number of is $2^n - 1$ the number of combinations of a itemset of len n.

## 4 Understanding lexicographical order

We need understand a little about some set definitions and concepts,

### 4.1 Cartesian product

### 4.2 Power set

### 4.3 Binary relation

#### 4.3.1 order pairs

as defined as $(a, b) = (c, d)$ if $a = c$ and $c = d$.

in this order Cartesian product ($AxB$) is defined as all pair of order elements.

It is important note that under this definition cartesian product not is conmmutative.

**The Cartesian product of two sets are all possible order pairs of elements what could be make with both**

#### 4.3.2 Power set

is the number of posible subset that could be formed in a set, for instance $\{a, b, c\}$ the following are subsets: $\emptyset$, $\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}$ the possible number of combinations are $2^n, n = 3$.

# 5 Binomial theorem proof by combinatorics

# 6 Jacobian

Remember that we have a defition of matrix as a linear transformation;

https://www.youtube.com/watch?v=VmfTXVG9S0U https://www.youtube.com/watch?v=wCZ1VEmVjVolist=RDC

## 6.1 Integrals

## 6.2 Polar coordinates

## 6.3 Linear locality and Multivariate Calculus

sweat over: speed up: drag: fleet: unroll: straight: