Recommender System II

Start at 9:05

Agenda

- Apriori
- Market basket Analysis
- Association rule mining

⇒ Customers who bought this also bought something else

⇒

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01/12/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	01/12/10 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01/12/10 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01/12/10 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/10 8:26	3.39	17850.0	United Kingdom

⇒ Based on your current cart item ⟶ recommend what else you should by

↳ apriori

⇒ (A) You buy an iphone ⟶ after 1 month you by a Cover

Not going to do this

(B) You buy an iphone ⟶ In the same order what else did you buy

Long term memory

Idea Conditional probability using Past data

grocery store milk, egg, bread.

ItemSet	Freq ← Amount of dat
{milk, eggs}	375
{milk, bread}	600
{milk, Jam}	200

↳ add milk to Cart

↳ bread recommend as highest frequency.

$$\text{Support} = \frac{\#\text{ occurance of Aset}}{\#\text{ Total baskets}} = \frac{600}{600+200+375}$$

⇒

Invoice no.	SKU	QTY
1	milk	3
1	bread	4
2	milk	2
2	bread	2
2	Egg	1
3	milk	3

grows

	Freq
$\{M\}$	3
$\{B\}$	2
$\{E\}$	1
$\{M, B\}$	2
$\{M, E\}$	1
$\{E, B\}$	1
$\{M, B, E\}$	1
$\{B, M, E\}$	

7 sets

⇒ Customer who bought milk → Bought 3 kgs of bread

⇒ milk → bread

[M, C, B]

⇒ Total 100 unique items

Max sets possible ⇒ $100C_1 + 100C_2 + \dots + 100C_{100}$

$2^{100} - 1$

⇒ milk 600

bread 250

→ egg 20 → remove values which have support less than a threshold

→ Pen 2 →

Hyper Parameter = min support →

⇒ If any set A have low Support

↓

all other set comprising Set A will also have low support

Support (Itemset)= $\frac{\text{Frequency of Itemset (Support Count)}}{\text{Total Number of Transactions}}$

TID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}

Itemset	Support	%
[Beer]	3/5	0.6
[Bread]	4/5	0.8
[Cola]	2/5	0.4
[Diapers]	4/5	0.8
[Milk]	4/5	0.8
[Eggs]	1/5	0.2

⇒ Support for {Bread, Milk} 3/5

{Diaper, Beer} 3/5

⇒ Milk is very popular → 1000 order → 900 milks, 100 → Support (creates a problem)

40 time Pen → Books

30 with milk, 10 without milk

⇒ Confidence →

$$P(X \mid Milk) = \frac{\{X, Milk\}}{\{Milk\}} =$$

{X, Milk}: # of times the combination occurs

{Milk}: of the existing Cart basket

{Egg}	30	30/120
{Milk}	100	100/120
{Bread}	50	50/120
{Milk, B}	25	25/120
{Milk, E}	25	25/120
{B, E}	30	(30/120) 70/120
{M, B, E}	20	

Total Carts ⇒ 120

Bread in basket

{B, E} / {B}

$\left(\frac{30}{50}\right)$

⇒ 90% of Cases based on your existing basket → recommend in the order of Confidence

⇒ Confidence is Symmetric

Milk → Bread

Bread → Milk No

⇐ Not Symmetric

$$P(A|B) \Rightarrow \frac{\{A, B\}}{\{B\}}$$

$$P(B|A) \Rightarrow \frac{\{A, B\}}{\{A\}}$$

⇒ iphone → charger

Charger ↛ iphone

---

① $P(A) = \frac{\#\{A\}}{\#}$

Support $P(A \cap B) = \frac{\#\{A, B\}}{\#}$

✓ Confidence $P(A|B) = \frac{\#\{A, B\}}{\#\{B\}}$ ✓ This eqⁿ is used to predict

$$= \frac{\text{Support}\{A, B\}}{\text{Support}\{B\}}$$

⇒ we Calc. support for all the itemsets (min threshold) → store in memory

Algo

Calculate Confidence online / on the go

⇒ • Support
• distances } → 10/15 day during night time → Confidence, Similarity

MLOPS

⇒ Eggs, Milk 30 | 30.01 | ⇒ (A) Milk to the Cart →
bread, milk 30 | 29.99
Jam, milk 30 | 28.87
→ Pen, milk 30 | 32.01
Pen, Pencil 10

Popular product to cart ⇒ we will not know what to recommend

Support ⇓

(B) ⇒ Pen to Cart →

Popularity Bins

⇒ Job Portal → 1000 Jobs

900 → 4+ years of experience in Data Scientist →

If someone actual comes Finance → 5+ years of Final
→ 4+ years Data

→ Spotify →

Lift $\Rightarrow P(M,T) = \dfrac{P(M \wedge T)}{P(T)\ P(M)} = \dfrac{\text{Support}(M,T)}{\text{Support}(T)\ \text{Support}(M)}$

$0 - \infty$

Lift $(x,y) = 1 \Rightarrow x, y$ are independent

Lift $(x,y) < 1 \Rightarrow$ unlikely to be bought

Lift $(x,y) > 1 \Rightarrow$ likely to be bought

rare item to sell $\Rightarrow$ milk $\Rightarrow \dfrac{P(\text{milk} \wedge \text{Pen})}{P(\text{Pen})\ P(\text{milk})}$

Lift is symmetric.

$L(A,B) = L(B,A)$

$\Rightarrow$ (Egg, milk) (milk, egg) $\Rightarrow$

$\Rightarrow$ By default $\rightarrow$ we always use Confidence

$\downarrow$

use lift

Cons of Apriori →

(1) Very Expensive

(2) Cold Start Problem →

⇒ Apriori → Confidence → Lift → 4 items → Predict the next most common item using Support.

5 items