

7 Tuesday

127-238 / Week 19

May 2019

$$P(A/\text{Spam}) = 0.2$$

$$P(B/\text{Spam}) = 0.3$$

$$P(C/\text{Spam}) = 0.1$$

$$P(ABC/\text{Spam}) = 0.1$$

$$P(\text{Spam}/ABC) = \frac{P(ABC/\text{Spam}) \times P(\text{Spam})}{P(ABC)}$$

If A, B, C are independent events

$$\approx P(A/\text{Spam}) \times P(B/\text{Spam})$$

$$P(C/\text{Spam}) \times P(\text{Spam})$$

$$\approx 0.2 \times 0.3 \times 0.1 \times P(\text{Spam})$$

$$P(\text{Spam}/ABC) \approx 0.6 P(\text{Spam})$$

$$\text{Entropy} = - \sum_{i=1}^2 p_i \log p_i$$

$$= - [0.01 \log 0.01 + 0.1 \log 0.1]$$

May 2019

ABC

Friday

3

123-242 / Week 18

$S = \{A, B, C, D, AB, AC, BC, AD, ABC\}$

Positive border

- Items in the border that are also frequent in S

Negative border

- Items in the border are not frequent in S

Let $\text{min-sup} = 2$

Item 1

Item 2

A

$|A| = 5$

$|A| = 2$

B

$|B| = 4$

~~$|AB| = 2$~~ ✓

C

$|C| = 4$

~~$|AC| = 1 < \text{min-sup}$~~

D

$|D| = 2$

$|BC| = 2$

AB ✓

All $\geq \text{min-sup}$

~~$|BD| = 0 < \text{min-sup}$~~

AC

A, B, C, D

~~$|CD| = 0 < \text{min-sup}$~~

BC

AD

Item 3

ABC

$ABC \rightarrow 1 < \text{min-sup}$

A, B, BC, AC $\rightarrow \text{min-sup satisfied}$

ABC is in negative border

Saturday

41 / Week 18

New Moon

TF-IDF of name 0.176

TF-IDF of privacy 0.477

cosine = ?

Term frequency \Rightarrow No of documents
 $N = 2$

Inverse doc freq \nearrow
 \downarrow
IDF = $1 \times \log \left(\frac{N+1}{df} \right)$ \nearrow $df = 2$ for name
 \searrow $df = 1$ for privacy
 $\log \left(\frac{3}{1} \right)$

$$= \log \left(\frac{2+1}{2} \right) = 1 \times \log 3$$

let $u = \log(3/2)$ let $v = \log 3$

unday ●

cosine similarity - $\frac{0.176}{\sqrt{0.176^2 + 0.477^2}} = \frac{\log(3/2)}{\log 3}$

4

Saturday

124-241 Week 18

● New Moon

TF-IDF

May 2019

TF-IDF of name 0.176TF-IDF of privacy 0.477

Cosine = ?

Term frequency \Rightarrow No of document
 $N = 2$

Inverse doc freq

IDF = $1 \times \log \left(\frac{N+1}{df} \right)$ $\log \left(\frac{3}{1} \right)$

$$= \log \left(\frac{2+1}{2} \right)$$

$$= 1 \times \log 3$$

let

$$u = \log \left(\frac{3}{2} \right)$$

let

$$v = \log 3$$

Sunday ●

$$\text{Cosine similarity} = \frac{u \cdot v}{\|u\| \|v\|} = \frac{\log \left(\frac{3}{2} \right) \log 3}{\log \left(\frac{3}{2} \right) \log 3}$$

$$= 1$$

May 2019

Friday 3

123-242 / Week 18

$$S = \{A, B, C, D, AB, AC, BC, AD, ABC\}$$

Positive border

- Items in the border that are also frequent in S

Negative border

- Items in the border are not frequent in S

Let $\text{min-supp} = 2$

	$ S $	Itemset 2
A	$ A = 5$	$ A = 2$
B	$ B = 4$	$ AB = 2$ ✓
C	$ C = 4$	$ AD = 1 < \text{min}$
D	$ D = 2$	$ BC = 2$
AB ✓	All $\geq \text{min-supp}$	$ BD = 0 < \text{min}$
AC		$ CD = 0 < \text{min}$
BC		

AD
ABC

Itemset 3

ABC $\rightarrow 1 < \text{min supp}$
 A B, BC, AC $\rightarrow \text{min-supp satisfied}$

ABC is in negative border

May 2019

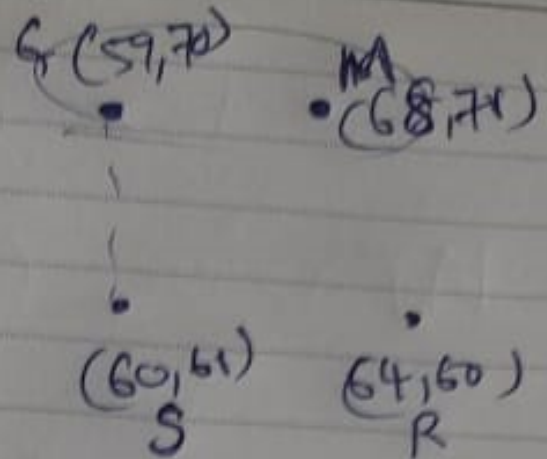
Wednesday

8

128-237 / Week 19

$$\begin{aligned} \text{Single Link (GS)} \\ &= \sqrt{(60-59)^2 + (70-61)^2} \\ &= \sqrt{1^2 + 9^2} = \sqrt{1+81} \\ &= \sqrt{82} \end{aligned}$$

$$\text{Avg link} = \left(\begin{array}{cc} \frac{59+68}{2} & \frac{70+71}{2} \\ \frac{60+64}{2} & \frac{61+60}{2} \end{array} \right)$$



$S = \{A, B, C, AB, AC, ABC\}$, Let $\text{minsep} = 2$

A
B
C

$A \Rightarrow 4$
 $B \Rightarrow 4$
 $C \Rightarrow 4$

$AB \rightarrow 2$ ✓
 $AC \rightarrow 2$ ✓
 $BC \rightarrow 2$ ✓

$\{AB, AC, BC\}$
 ABC

$ABC \rightarrow 1 < \text{minsep}$
 \therefore pruned