Tony Ho

CPSC 4310
Assignment 3

1/ Cosine similarity:

• cosine ($\vec{apricot}$, $\vec{pineapple}$)

$$= \frac{(0\times0)+(0\times0)+(0\times0)+(1\times1)+(0\times0)+(1\times1)}{\sqrt{0^2+0^2+0^2+1^2+0^2+1^2}\cdot\sqrt{0^2+0^2+0^2+1^2+0^2+1^2}}$$

$$= \frac{0+0+0+1+0+1}{\sqrt{2}\cdot\sqrt{2}}$$

$$= \frac{2}{2}$$

$$= 1$$

• cosine ($\vec{apricot}$, $\vec{digital}$)

$$= \frac{0+0+0+0+0+0}{\sqrt{1^2+1^2}\quad\sqrt{2^2+1^2+1^2}}$$

$$= \frac{0}{\sqrt{2}\cdot\sqrt{6}}$$

$$= 0$$

• cosine ($\vec{apricot}$, $\vec{information}$)

$$= \frac{0+0+0+0+0+0}{\sqrt{2}\cdot\sqrt{1^2+6^2+4^2}}$$

$$= 0$$

• cosine ($\vec{pineapple}$, $\vec{digital}$)

$$= \frac{0+0+0+0+0+0}{\sqrt{2}\cdot\sqrt{6}}$$

$$= 0$$

• cosine ($\vec{pineapple}$, $\vec{information}$)

$$= \frac{0+0+0+0+0+0}{\sqrt{2}\cdot\sqrt{53}}$$

$$= 0$$

- cosine ($\vec{digital}$, $\vec{information}$)

$$= \frac{0 + 2 + 6 + 0 + 4 + 0}{\sqrt{6} \cdot \sqrt{53}}$$

$$= \frac{12}{\sqrt{6} \cdot \sqrt{53}}$$

$$= 0.672927$$

2/

	Doc 1	Doc 2	Doc 3
food	1	1	0
restaurant	2	0	0
customer	1	1	1
waitress	1	0	0
store	0	1	2
cashier	0	1	1
appliance	0	0	1

$$w_{t,d} = tf_{t,d} \times idf_t$$

- food :

- $w_{food, Doc1}$

$$= tf_{food, Doc1} \times idf_{food}$$

$$= (1 + \log 1) \times \log (3/2)$$

$$= 0.176091$$

- $w_{food, Doc2} = tf_{food, Doc2} \times idf_{food}$

$$= (1 + \log 1) \times \log (3/2)$$

$$= 0.176091$$

- $w_{food, Doc3}$

$$= 0 \times \log (3/2)$$

$$= 0$$

- restaurant :
  - $W_{restaurant, Doc\,1}$
  $= (1 + \log 2) \times \log(3/1)$
  $= 0.820749$
  - $W_{restaurant, Doc\,2}$
  $= 0 \times \log(3/1)$
  $= 0$
  - $W_{restaurant, Doc\,3}$
  $= 0 \times \log(3/1)$
  $= 0$

- customer
  - $W_{customer, Doc\,1}$
  $= (1 + \log 1) \times \log(3/3)$
  $= 0$
  - $W_{customer, Doc\,2} = 0$
  - $W_{customer, Doc\,3} = 0$

- Waitress
  - $W_{waitress, Doc\,1} = (1 + \log 1) \times \log(3/1) = 0.477121$
  - $W_{waitress, Doc\,2} = 0$
  - $W_{waitress, Doc\,3} = 0$

- store
  - $W_{store, Doc\,1} = 0$
  - $W_{store, Doc\,2} = (1 + \log 1) \times \log(3/2) = 0.176091$
  - $W_{store, Doc\,3} = (1 + \log 2) \times \log(3/2) = 0.229100$

- Cashier :
  - $W_{cashier, Doc\,1} = 0$
  - $W_{cashier, Doc\,2} = (1 + \log 1) \times (\log 3/2) > 0.176091$
  - $W_{cashier, Doc\,3} = 0.176091$

- appliance :
  - $w_{appliance, Doc1} = 0$
  - $w_{appliance, Doc2} = 0$
  - $w_{appliance, Doc3} = 0.176091$

⇒ TF-IDF of each word for the 3 documents:

	Doc 1	Doc 2	Doc 3
food	0.176091	0.176091	0
restaurant	0.620749	0	0
customer	0	0	0
waitress	0.477121	0	0
store	0	0.176091	0.229100
cashier	0	0.176091	0.176091
appliance	0	0	0.176091

⇒ Comparing pair of documents

- cosine (Doc 1, Doc 2):

$$= \frac{(0.176091 \times 0.176091) + 0 + 0 + 0 + 0 + 0 + 0}{\sqrt{0.176091^2 + 0.620749^2 + 0.477121^2} \cdot \sqrt{0.176091^2 + 0.176091^2 + 0.176091^2}}$$

$$= \frac{0.031008 04}{0.802484772 \times 0.304998558}$$

$$= 0.126689$$

- cosine (Doc 1, Doc 3)

$$= 0$$

- cosine (Doc 2, Doc 3)

$$= \frac{0 + 0 + 0 + 0 + (0.176091 \times 0.229100) + (0.176091 \times 0.176091) + 0}{\sqrt{0.176091^2 + 0.176091^2 + 0.176091^2} \times \sqrt{0.229100^2 + 0.176091^2 + 0.176091^2}}$$

$$= \frac{0.071350488}{0.304998558 \times 0.338382757}$$

$$= 0.691339$$

So Doc 2 is more similar to Doc 3 than Doc 2 to Doc 1 because cosine (Doc 2, Doc 3) has the biggest value

Q3:

a)

I/PRP booked/VB a/DT flight/NNP from/IN Lethbridge/NNP to/TO Calgary/NNP ./.

Error:

- booked should be VBD because past tense verb
- flight should be NN because it is a noun, not a proper noun

Correction:

I/PRP booked/VBD a/DT flight/NN from/IN Lethbridge/NNP to/TO Calgary/NNP ./.

b) Error:

- flight should be NN
- serve should be VBP

Correction:

Does/VBZ this/DT flight/NN serve/VBP complementary/JJ drinks/NNS ?/?

c) Error:

- living should be VBG

Correction:

I/PRP have/VBP a/DT friend/NN living/VBG in/IN Calgary/NNP./.

d) Error:

- flights should be NNS
- afternoon should be JJ

Correction:
Can/MD you/PRP list/VB the/DT afternoon/JJ flights/NNS ?/?

Q4:

a)
It/PRP is/VBZ a/DT sweet/JJ dream/NN ./.

b)
The/DT new/JJ store/NN is/VBZ close/JJ to/TO the/DT restaurant/NN on/IN the/DT 13th/JJ street/NN ./.

c)
Give/VB it/PRP a/DT quick/JJ thought/NN when/WRB you/PRP have/VBP a/DT spare/JJ time/NN to/TO kill/VB ./.