
Problem 1: Thematic Role

Question	Noun Phrase	Thematic Role
a	John	agent
a	the sidewalk	theme
a	snow shovel	instrument
b	A raffle	theme
b	the local charity	beneficiary
c	Joe	agent
c	Susan	co-agent
d	Tom	recipient
d	a bike	theme
e	The dress	theme
e	Lisa	agent
e	matching shoes	co-theme
f	Julie	agent
f	grandmother	recipient
f	a letter	theme
g	George	experiencer
g	car	theme
h	The image	theme
h	Adobe Photoshop	instrument
h	cartoonist	agent

Problem 2: Similarities Between Vectors

(a)

$$\text{ManhattanDistance}(x, y) = \sum_{i=1}^N |x_i - y_i| = 3 + 5 + 4 + 5 = 17$$

Manhattan Distance of 17

(b)

$$\text{ManhattanDistance}(x, y) = \sum_{i=1}^N |x_i - y_i| = 5 + 4 + 5 + 3 = 17$$

Manhattan Distance of 17

(c)

$$Jaccard(x, y) = \frac{\sum_{i=1}^N \min(x_i, y_i)}{\sum_{i=1}^N \max(x_i, y_i)} = \frac{6 + 2 + 4 + 0}{9 + 7 + 8 + 5} = \frac{12}{29}$$

Jaccard Similarity of $\frac{12}{29}$

(d)

$$Jaccard(x, y) = \frac{\sum_{i=1}^N \min(x_i, y_i)}{\sum_{i=1}^N \max(x_i, y_i)} = \frac{1 + 3 + 4 + 2}{6 + 7 + 9 + 5} = \frac{10}{27}$$

Jaccard Similarity of $\frac{10}{27}$

(e)

$$Cosine(x, y) = \frac{\sum_{i=1}^N (x_i * y_i)}{\sqrt{\sum_{i=1}^N x_i^2} \sqrt{\sum_{i=1}^N y_i^2}} = \frac{54 + 14 + 32 + 0}{\sqrt{81 + 4 + 64 + 0} * \sqrt{36 + 49 + 16 + 25}} = 0.73$$

Cosine Similarity of 0.73

(f)

$$Cosine(x, y) = \frac{\sum_{i=1}^N (x_i * y_i)}{\sqrt{\sum_{i=1}^N x_i^2} \sqrt{\sum_{i=1}^N y_i^2}} = \frac{6 + 21 + 36 + 10}{\sqrt{36 + 49 + 16 + 25} * \sqrt{1 + 9 + 81 + 4}} = 0.67$$

Cosine Similarity of 0.67

Problem 3: Collins & Singer

(a)

Rule	Probability
If Contains(apple) → PRODUCT	1/3
If Contains(apple) → COMPANY	2/3
If Contains(tablet) → PRODUCT	3/4
If Contains(tablet) → COMPANY	1/4
If Contains(microsoft) → PRODUCT	1/3
If Contains(microsoft) → COMPANY	2/3
If Contains(british) → PRODUCT	1/2
If Contains(british) → COMPANY	1/2
If Contains(corporation) → PRODUCT	0/3
If Contains(corporation) → COMPANY	3/3

(b)

Rule	Probability
If Contains(mobile) \rightarrow PRODUCT	2/3
If Contains(mobile) \rightarrow COMPANY	1/3
If Contains(computer) \rightarrow PRODUCT	3/4
If Contains(computer) \rightarrow COMPANY	1/4
If Contains(tech) \rightarrow PRODUCT	1/4
If Contains(tech) \rightarrow COMPANY	3/4
If Contains(giant) \rightarrow PRODUCT	0/3
If Contains(giant) \rightarrow COMPANY	3/3
If Contains(leader) \rightarrow PRODUCT	0/2
If Contains(leader) \rightarrow COMPANY	2/2

Problem 4: Salience Values

There are four sentences with a FISH context: **S1**, **S5**, **S6**, and **S7**.

There are five sentences with a MUSIC context: **S2**, **S3**, **S4**, **S5**, and **S6**.

Note that **S5** and **S6** have both a FISH and MUSIC context.

(a) *Utah* appears in three sentences with a FISH context: **S1**, **S6**, and **S7**. *Utah* appears 4 times in the corpus.

$$\text{salience}(\text{Utah}, \text{FISH}) = \frac{P(\text{Utah}|\text{FISH})}{P(\text{Utah})} = \frac{3/4}{4/60} = 11.25$$

Salience value of 11.25.

(b) *Electric* appears in two sentences with a FISH context: **S5** and **S6**. *Electric* appears 3 times in the corpus.

$$\text{salience}(\text{electric}, \text{FISH}) = \frac{P(\text{electric}|\text{FISH})}{P(\text{electric})} = \frac{2/4}{3/60} = 10$$

Salience value of 10.

(c) *Bass* appears in two sentences with a FISH context: **S5** and **S6**. *Bass* appears 5 times in the corpus.

$$\text{salience}(\text{bass}, \text{FISH}) = \frac{P(\text{bass}|\text{FISH})}{P(\text{bass})} = \frac{2/4}{5/60} = 6$$

Salience value of 6.

(d) *Utah* appears in two sentences with a MUSIC context: **S2** and **S6**. *Utah* appears 4 times in the corpus.

$$\text{salience}(\text{Utah}, \text{MUSIC}) = \frac{P(\text{Utah}|\text{MUSIC})}{P(\text{Utah})} = \frac{2/5}{4/60} = 6$$

Salience value of 6.

(e) *Electric* appears in three sentences with a MUSIC context: **S3**, **S5** and **S6**. *Electric* appears 3 times in the corpus.

$$\text{salience}(\text{electric}, \text{MUSIC}) = \frac{P(\text{electric}|\text{MUSIC})}{P(\text{electric})} = \frac{3/5}{3/60} = 12$$

Salience value of 12.

(f) *Bass* appears in five sentences with a MUSIC context: **S2**, **S3**, **S4**, **S5** and **S6**. *Bass* appears 5 times in the corpus.

$$\text{salience}(\text{bass}, \text{MUSIC}) = \frac{P(\text{bass}|\text{MUSIC})}{P(\text{bass})} = \frac{5/5}{5/60} = 12$$

Salience value of 12.

Problem 5: Antecedents and Pronouns

(a)

- John Smith
- John
- his
- him

(b)

- John Smith
- John
- Mary
- his
- She

- him
- He
- he

(c)

- 10 oranges
- too many groceries

(d)

- himself

(e)

- his
- his
- their
- her

(f)

- It
- it

(g)

- his neighbor
- George