Name :- Sumit Kumar Yadav

Roll No. :- 18CS30042

Page No.

Date

Sol^n:- 4:- (a) "the dog likes that monkey"

for model M1:

~~P(M1)~~ $P(\text{string} \mid M1) = (0.2)(0.005)(0.02)(0.04)(0.001)$

$= 2 \times 5 \times 2 \times 4 \times 10^{-11}$

$= 8 \times 10^{-10}$

$P(\text{string} \mid M2) = (0.15)(0.01)(0.04)(0.04)(0.002)$

$= 15 \times 4 \times 4 \times 2 \times 10^{-11}$

$= 3 \times 4 \times 4 \times 10^{-10}$

$= 48 \times 10^{-10}$

$\therefore$ $P(\text{string} \mid M2) > P(\text{string} \mid M1)$

$\therefore$ ~~a~~ language model M2 is more relevant to the given string than model M1.

& likelihood ratio $= \frac{48}{8} = 6$

(b) $\because$ document does not contain the word 'STOP'

$\therefore$ It does not change the likelihood ratio

(c) $\because \lambda = 1/4$.

D1 → 9

D2 → 5

Q = "Semester exams"

$\therefore P(Q \mid d_1) = \left(\frac{1}{4} \times \frac{1}{9} + \frac{3}{4} \times \frac{2}{14}\right) \cdot \left(\frac{1}{4} \times \frac{0}{9} + \frac{3}{4} \times \frac{1}{14}\right)$

$= \left(\frac{1}{36} + \frac{\cancel{6}3}{14 \times \cancel{4}_2}\right)\left(0 + \frac{3}{14 \times 4}\right)$

Page No.
Date

$= (0.0277 + 0.107) \times 0.053$

$= 0.0071$

$= 7.1 \times 10^{-3}$

now, $P(Q \mid d_2) = \left(\frac{1}{4} \times \frac{1}{5} + \frac{3}{4} \times \frac{2}{14}\right) \cdot \left(\frac{1}{4} \times \frac{1}{5} + \frac{3}{4} \times \frac{1}{14}\right)$

$= \left(\frac{1}{20} + \frac{6}{56}\right) \cdot \left(\frac{1}{20} + \frac{3}{56}\right)$

$= (0.05 + 0.107)(0.05 + 0.053)$

$= 0.157 \times 0.103$

$= 0.0161$

$= 16 \times 10^{-3}$

$\therefore$ ranking : $d_2 > d_1$

as $P(Q \mid d_2) > P(Q \mid d_1)$.

Date / /

Sol^n:- 3:- (a) Main stages in summarization:-

→ content selection

Choose sentences to extract from document

→ Information ordering

Choose an order to place them in summary

→ Sentence realization

Simplify the sentences

→ Removing redundacy

Increase diversification by removing redundant sentences

(b) Underlying hypothesis of luhn :-

→ they choose sentences that have silent or informative words.

they uses tf-idf, then calculated

$$w_i = tf_{i,j} \times idf.$$

then finally weight(s) = $\frac{1}{|S|} \sum_{w \in S} weight(w)$

Underlying hypothesis of lex Rank:-

→ Sentences that convey the theme of the document are more similar to each other.

(c) Rouge-1 = $\frac{9}{9} = 1$

Rouge-2 = $\frac{5}{8}$

∵ Rouge-1 also ~~give~~ considered if the words are jumbled while in Rouge-2 considerd bigram so it managed the jumbled word in the sentence.

∴ Rouge-1 is very less correct compare to Rouge-2.

Solⁿ:- 1:(a) Ergodic markov chain :-

a discrete time stochastic process following markov property (next ~~state~~ state depends only on current state)

A markov chain is said to be ergodic if there exists a positive integer such that for all pairs of states in markov chain.

Ergodicity imp. for convergence of page as it consider $N \times N$ transition prob. matrix. it taken less iteration untill convergence.

~~to~~ ∵ transition probability is irreducible.

(b) steps :-

→ If a row of A has no 1's replace each element by $1/N$ (to ensure the sum row of prob = 1)

→ for all other row: divide each 1 by number of 1's in the row.

→ multiply the resultant matrix by a

→ add (1-a)/N to every entry of the resulting matrix.

(c) Adjacency matrix

$$\begin{array}{c|cccccc} & A & B & C & D & E & F \\ \hline A & 0 & 1 & 1 & 0 & 0 & 0 \\ B & 0 & 0 & 0 & 0 & 0 & 0 \\ C & 1 & 1 & 0 & 0 & 0 & 0 \\ D & 0 & 0 & 0 & 0 & 1 & 1 \\ E & 0 & 0 & 0 & 1 & 0 & 1 \\ F & 0 & 0 & 0 & 1 & 0 & 0 \end{array}$$

⇓ after step 1

$$\begin{array}{c|cccccc} & A & B & C & D & E & F \\ \hline A & 0 & 1 & 1 & 0 & 0 & 0 \\ B & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ C & 1 & 1 & 0 & 0 & 0 & 0 \\ D & 0 & 0 & 0 & 0 & 1 & 1 \\ E & 0 & 0 & 0 & 1 & 0 & 1 \\ F & 0 & 0 & 0 & 1 & 0 & 0 \end{array}$$

⇓ step 2

$$\begin{array}{c|cccccc} & A & B & C & D & E & F \\ \hline A & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ B & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ C & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ D & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ E & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ F & 0 & 0 & 0 & 1 & 0 & 0 \end{array}$$

⇒ Step 3 : $\alpha = 0.8$

$$
\begin{array}{c|cccccc}
 & A & B & C & D & E & F \\
\hline
A & 0 & \frac{0.8}{2} & \frac{0.8}{2} & 0 & 0 & 0 \\
B & \frac{0.8}{6} & \frac{0.8}{6} & \frac{0.8}{6} & \frac{0.8}{6} & \frac{0.8}{6} & \frac{0.8}{6} \\
C & \frac{0.8}{3} & \frac{0.8}{3} & 0 & 0 & \frac{0.8}{3} & 0 \\
D & 0 & 0 & 0 & 0 & \frac{0.8}{2} & \frac{0.8}{2} \\
E & 0 & 0 & 0 & \frac{0.8}{2} & 0 & \frac{0.8}{2} \\
F & 0 & 0 & 0 & 0.8 & 0 & 0
\end{array}
$$

II step 4 : add $(1-0.8)/N = 0.2/6 = \frac{0.1}{3}$

we get

$$
\begin{array}{c|cccccc}
 & A & B & C & D & E & F \\
\hline
A & \frac{0.1}{3} & 0.4+\frac{0.1}{3} & 0.4+\frac{0.1}{3} & 0 & 0 & 0 \\
B & \frac{0.8}{6}+\frac{0.1}{3} & \frac{0.8}{6}+\frac{0.1}{3} & \frac{0.8}{6}+\frac{0.1}{3} & \frac{0.8}{6}+\frac{0.1}{3} & " & " \\
C & 0.4+\frac{0.1}{3} & 0.4+\frac{0.1}{3} & 0 & 0 & 0 & 0 \\
D & 0 & 0 & 0 & & & \\
E & & & & & & \\
F & & & & & &
\end{array}
$$

resultant

we get,

$$
\begin{array}{c|cccccc}
 & A & B & C & D & E & F \\
\hline
A & 0.03 & 0.43 & 0.43 & 0.03 & 0.03 & 0.03 \\
B & 0.16 & 0.16 & 0.16 & 0.16 & 0.16 & 0.16 \\
C & 0.29 & 0.29 & 0.03 & 0.03 & 0.29 & 0.03 \\
D & 0.03 & 0.03 & 0.03 & 0.03 & 0.43 & 0.43 \\
E & 0.03 & 0.03 & 0.03 & 0.43 & 0.03 & 0.43 \\
F & 0.03 & 0.03 & 0.03 & 0.83 & 0.03 & 0.03
\end{array}
$$