V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1

National University of Computer & Emerging Sciences
FAST-Karachi Campus
CS4051- Information Retrieval
Quiz#3

Dated: April 22, 2024 Marks: 20
Time: 20 min.
Std-ID: _____Sol_____

Question No. 1

Outline three difference between Probabilistic Model for IR and Language Model for IR. [5]

Probabilistic Information Retrieval Model	Language Model for Information Retrieval
- In this model we try to present document in the decreasing value of P(R=1/ q,d). - All documents are evaluated as per degree of relevance. - Very relax assumptions.	- In language model for IR we estimate the probability of a query generating from the same document model that is P(q/Md). - Every document is treated as different model and the query generation process for each is estimated. - Very restricted.

Question No. 2

Consider making a language model from the following training text:

Arre Re Arre Ye Kya Hua Maine Na Ye Jaana. Arre Re Arre Ban Jaaye Na Kahin Koi Afsaana

a. Under a MLE-estimated unigram probability model, what are P(Arre) and P(Afsaana)? [2.5]

- P(Arre) = 4/19

- P(Afsaana) = 1/19

b. Under a MLE-estimated bigram model, what are P(Arre/ Re) and P(Re/Arre)? [2.5]

- P(Arre/Re) = P(Re,Arre) / P(Re) = 2/2 = 1

- P(Re /Arre) = P(Arre,Re) / P(Arre) = 2/4= 0.5

V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1V1

Question No. 3

Suppose we have a collection that consists of the 3 documents given in the below table.

DocID	Doc-content
1	w3 w2 w3 w1
2	w2 w3 w2 w4
3	w2 w1 w3

Build a query likelihood language model for this document collection. Assume a mixture model between the documents and the collection, with λ =0.5. Give the document ranking against the query =”w3 w2”. You can apply Laplace smoothing. [10]

Language Model

Doc-Model	w1	w2	w3	w4
D1	1/4	1/4	1/2	0
D2	0	1/2	1/4	1/4
D3	1/3	1/3	1/3	0
Collection Model	2/11	4/11	4/11	1/11

Model Probabilities for query =” w3 w2”

P(MD1/q) = [1/2*1/4+1/2*4/11] + [1/2*1/2+ 1/2 *4/11] = 2*[1/2*1/4 +1/2*4/11] = 2*[1/8+4/22] =19/44

P(MD2/q) = [1/2*1/4+1/2*4/11] + [1/2*1/2+ 1/2 * 4/11] = 2*[1/8+4/22] =19/44

P(MD3/q) = [1/2*1/3+1/2 *4/11] + [1/2*1/3+1/2*4/11] = 2*[1/6+4/22] =23/66

Ranking will be: D1,D2 D3 or D2,D1, D3