BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
HYDERABAD CAMPUS
FIRST SEMESTER 2020 – 2021
INFORMATION RETREIVAL (CS F469) – TEST-3 SOLUTION

Date: 19.11.2020 Weightage: 12% [24 Marks] Duration: 30mins. Type: Open Book

Q1. CUR Decomposition

A. Which of the CUR decomposition matrices uses SVD? **[1M]**
   A. C
   B. U
   C. R
   D. C and U

B. The advantage of CUR decomposition is that if the original matrix is sparse then the C and R will be **Sparse** **[1M]**

C. If a matrix has K singular values, then CUR selects (**k log k)** number of singular vectors. **[1M]**

D. The matrix U in the CUR deposition is sparse and small. **[T/F]** **False** **[1 M]**

E. CUR with no duplicates perform better than the one with duplicates in terms of space and time complexity. **[T/F]** **True** **[1 M]**

F. Give one-line justification as to why CUR decomposition uses SVD for computing the matrix U. **[1M]**

Since it preserves the richness of the data.

G. The rank of the matrices from CUR decomposition will be less than the number of row and column choices made when _____ . **[1M]**
duplicates are found in the matrix and we scale the rows/columns by the square root of the number of times it occurs.

I. Which of the following metric was used to measure the performance of the algorithm in NETFLIX recommender systems competition? **[1 M]**
   A. MAE
   B. RMSE
   C. Precision
   D. F-Score

Q2. Language Model

A. What type of language model P(E) should be used to generate better meaningful sentences in English?
   A. Unigram **[1 M]**
   B. bi-gram
   C. trigram
   D. n-gram

B. How do you express a Trigram language model mathematically? [**2 M]**
$P(W_1,W_2,\ldots,W_n)=P(W_1)XP(W_2|W_1)XP(W_3|W_1W_2)\ldots. P(W_n|W_{n-2}W_{n-1})$

$P(W_i) = P(W_i| W_{i-1}, _{i-2})$

C. What is the problem when a Unigram model considered for Language Modelling? **[2 M]**
When the Unigram Model is used to generate language the sentence formation may not be syntactically and semantically correct.For example the following two sentences will have same probability.

Dog barked at Ramu
Ramu at barked dog

D. What is the Markov assumption and how it helps in the simplifying the Language Model? **[3 M]**

The Markov assumption states that only a few number of previous words affect the probability of the current word. This is used to avoid computing the joint probability distribution.

Q3. Statistical Machine Translation

A. In which phase of IBM Models is the parallel corpus used and why? **[2 M]**

It is used in the training phase to estimate the word translation parameters in the Expectation Maximization step..

B. In IBM model-1 if the following three parallel sentences are used as training data, how many translation parameters need to be estimated? **[2 M]**

Note: English (e) vocabulary size is 15 and Hindi (h) vocabulary size is 18 also you may ignore Null word.

आकाश बैंक जाने के रस्ते पर चला	Akash walked on the road to the bank
श्याम नदी तट पर चला	Shyam walked on the river bank
आकाश द्वारा नदी तट से रेट की चोरी हो रही है	Sand on the banks of the river is being stolen by Akash

15X18 = 270

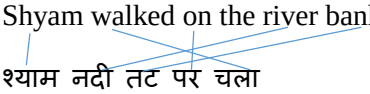
We need to estimate 270 translation parameters.

C. State the basic assumption used in the IBM model-1. **[2 M]**

The naïve assumption of IBM model-1 is that all alignments are equiprobable. That mean all the alignments have same probability which is $1/(1+l_t)^{l_s}$

D. Given the following sentence pair write the alignment variable a. **[2 M]**

Shyam walked on the river bank



श्याम नदी तट पर चला

A=[1,5,6,3,2] or A=[1->1,2->5,3->6,4->3,5->2]