

CS 5316 – Natural Language Processing

Quiz 5 Solution

(Time limit: 12 minutes)

Instructions: (1) Please write legibly. Unreadable answers will NOT be graded; (2) Write in the spaces provided for the questions only.

1. (3 points) Suppose a logistic regression model is trained for a POS tagging task by considering the previous two words and two labels as input, i.e., $P(q_n | q_{n-2}, q_{n-1}, w_n, w_{n-1}, w_{n-2})$. Describe how you would determine the best label (Noun, Verb, Preposition) for the word LUMS in the sentence “I study at LUMS” using (a) greedy inference, and (b) beam inference with beam length of 2. Show with example numbers.

Greedy inference

Suppose the most probable labels for study and at is Verb and Preposition. Then, the best label for LUMS is given by:

$\text{Argmax}_y \{ \text{Noun, Verb, Preposition} \} P(q_n = y | \text{Verb, Preposition, LUMS, at, study})$

Beam inference

Suppose top two labels for ‘at’ are Preposition and Verb, and top 2 labels for ‘study’ are Verb and Noun.

LUMS	study	At
Noun (0.6), Verb (0.3)	Verb	Preposition
Noun (0.4), Preposition (0.4)	Verb	Verb
Verb (0.5), Noun (0.3)	Noun	Preposition
Preposition (0.3), Verb (0.3)	Noun	Verb

Given the above example, the top two labels for LUMS are Noun (first row) and Verb (3rd row).

(3 points) Construct the word-by-word co-occurrence matrix for the sentence: the dog jumped over the lazy dog. Use context size of +1 and -1 words.

Vocabulary: the dog jumped over lazy

	the	dog	jumped	Over	Lazy	
The	2	1	0	1	1	
Dog	1	2	1	0	1	
Jumped	0	1	1	1	0	
Over	1	0	1	1	0	
Lazy	1	1	0	0	1	

2. (3 points) For the sentence given in Q 2 above, construct the training dataset for learning word2vec based on skip gram (context is + 1 and -1 words).

Input word, output word

The, dog; the, over; the, lazy

Dog, the; dog, jumped; dog, lazy

Jumped, dog; jumped, over

Over, jumped; over, the

Lazy, the; lazy, dog

3. (1 point) Distinguish between word2vec and Glove for learning word embeddings. Give at least 2 points.
- Word2vec uses a predictive model for learning embeddings, while Glove uses a co-occurrence based model
 - Word2vec is given by a vector in the learned weight matrix of a neural network, while Glove is given by a vector constructed from discriminating words in context.