Natural Language Processing Applications

Mid Semester Exam II

March 8, 2019

MM: 30 Time: 1.5 hrs

Note: Marks are mentioned next to the questions

1. Describe the various word based MT Models? What are the failures of Word Based MT? [5]

2. What is alignment? How is the phrase table (along with phrase translation probability) created for Phrase based MT? [5] 1 + 4

3. What are the advantages of LSTMs over RNNs? Clearly explain the architecture of LSTMs and GRUs [5] 2 + 3

4. Most Indo Aryan and Dravidian (Indian) Languages exhibit rich morphology. This means, the words are attached with some grammatical markers. This results in a large number of word forms for words belonging to certain categories (noun, verbs (think of "eat" and all its forms in your native language). Typically, vector embedding methods take words as input. Can you design a vector embedding mechanism which can learn embeddings such that different morphological variations (for eg walking, walks, walked, walk) of the same root (for eg walk) are closer in the vector space? [5]

5. Sequence to sequence models have been employed for Machine translation with some success. But, Attention based models have overtaken these models. What value do Attention models bring in in order to outperform bi-directional LSTM based seq-seq models? [5]

6. Will Attention based models as applied to English MT work as well for Dravidian Languages as they do for English, given all other aspects (training data etc) remain the same? What changes are required, if any? [5]