CSE 576 Quiz 14, April 16, 2019

NAME:
ID:

Q 1: What is the difference between Simplified Lesk Algorithm and Corpus Lesk Algorithm for Word Sense Disambiguation?

Simplified lesk –

given a word with multiple senses, disambiguate it based on overlap of the gloss and context using dictionary / thesaurus (eg. WordNet)

ex.

bank
- financial institue – surrounding text has words like deposit, mortgage etc.
- shore of a river – surrounding text has words like water, river, currents etc.

(unnecessary words are ignored)

Corpus lesk

use the huge labelled data which has [gloss + examples of each sense of a word] – jointly called signature

then choose the sense with most overlap of context and signature

IDF (inverse document frequency) weighing is done to weight the sense of a word

(which sense of a word is more frequent.)

Q 2: What is the difference between "in-vitro" and "in-vivo" evaluations of a Word Sense Disambiguation system?

In-vitro

- corpus based WSD evaluation (direct evaluation)
- create a corpus which has all senses of a word labelled differently
- split this into train & test
- check how many / how much percentage of word senses are correctly disambiguated

In-vivo

- incorporate WSD as a part of larger application system / NLP task like machine translation, information retrieval and question answering (indirect evaluation)
- measure the performance of the overall system using different WSD mechanisms and see which performs better