NLP HW#1

20180368 An Haechan

1. **Specify your own problem.**

Consider the following pairs of expressions.

1. Extol, praise highly (2) Destitute, very poor.

Let’s call such adverbs like *highly* and *very* as ‘intensity modifying adverbs’.

Also, call pairs like *(extol, praise)* and *(destitute, poor)* ‘intensity pairs’.

First, with the aid of a method ‘similar()’ in nltk.Text, manually determine good enough ‘intensity modifying adverbs’.

Secondly, using those modifying adverbs, find ‘intensity pairs’ from arbitrary nltk text corpus, in the form of list of triples, (modified word, intensity modifying adverb, word).

**(C) Use your own argument in English to show why your output is of reasonable quality, and how can you improve its quality further**

**C-1) Why your output is of reasonable quality?**

Much of my algorithm is explained in code, so I’ll briefly talk about it here.

First, and most importantly, I filtered my results using Wordnet. Suppose machine found that ‘very hot’ and ‘attractive’ has the same context, like “I am very hot guy” and “I am attractive guy”(context- am\_guy). Then it checks path similarity of ‘hot’ and ‘attractive’.

Additionally, I omitted be-verbs intentionally. Since be-verb has no intensities but frequently used.

**C-2) How can you improve its quality?**

Followings are the ideas I came out with.

(1) Cheap algorithm that can distinguish part-of-speech

(2) Removal of extremely frequent words

(3) Data set of words of which intensity cannot change