CS372 Homework #2

20160632 Janghyun Cho

1. Problem Statement

Given homework was to find lexical items D that increase the intensity of lexical items E that they modify. I set up the problem as “Find set of ADVERB + ADJECTIVE, which has intensifier as adverb”. With this problems, I made several assumptions based on example sentence data that has adverb-adjective set defined in problem. Those assumptions are like following: 1. Most of adverb-adjective set is in object space of sentence, which means it comes after a verb. 2. Most of intensifiers are used in objective sentence, not subjective. 3. By observing many intensifiers, many of them ends with phrase ‘ly’. I implemented my codes based on these assumptions.

1. Implementation

I used tagged sentence corpus (tagged\_sents()) from brown corpus which has tagged sentences with a tag to each word) and sentiwordnet corpus for finding how objective the word is. So the first part was to find adverb-adjective set which follows verb. Here I had to be aware of words that can come between verb and adverb-adjective set, which are prepositions and articles. Propositions and articles can be used both or one of them or none. With a found set, we get objective score of found adverb by using sentiwordnet. Then add it to word list if there are no duplicates, objective score in not 0, and adverb ends with ‘ly’. Last, sort it by objective score and return it as csv file.

1. Result Analysis

As we look at the result from the top, we can find words that looks obviously like an intensifier and which does not. From the top 20 adverbs in adverb-adjective set that my code found as an intensifier, are basically, dangerously, comically, basically, fantastically, obviously, plainly, fundamentally, duly, extraordinarily, typically, startlingly, previously, empirically, predominantly, openly, essentially, strangely, naively, invariably. Among these 20 adverbs, there are only 3 adverbs that we can definitely say it is not an intensifier even with its adjective, and these are comically, previously, and naively. By the result, this code got 17/20, which is a decent score. As in their rank, these words are least objective, which is most subjective, with much exaggeration. My first two sets are “basically anti-democratic” and “dangerously erroneous”, which are restrictive set.

1. Improvements

I think improvements can be made at three parts. First improvement is getting words score. Currently, only restriction in getting synset of a word is word itself and its position, whether it is verb, adjective, or adverb. And add all objective points for multiple synsets of the word. We need to get one specific synset as used in sentence or perhaps use average of objective score. Second, the sorting is done only by adding objective score of adverb and adjective currently. Maybe objective score can be widened to subject or verb of the sentence. Or even get objective score of whole sentence. Third, I intentionally got words that end with ‘ly’ as result. This filtered many unwanted words, but should have missed intensifiers. Using different decent filter will produce richer output.