Natural Language Processing

HW – 2

0816175 高靖

3 methods:

First methods I use is take each word in sentence, check its dep_ using **spacy**, and check S,V,O value by 'nsubj', 'ROOT', 'dobj', and only if both 3 dep_ matches the S,V,O in sentence, the id's label is 1, otherwise, set label to 0, and consequence just pass the baseline, I think the main reason is now I consider only a 'word', instead of real possible of S O V, which may be a phrase or consist of more than 1 word. So if change the way of determining flags, to check if the word 'in' S O V, instead of equal to S O V. Obviously, the label 1 is now more easy to appear. But accuracy was lower, because the V is actually, for example, 'seemed', but if we decide label using way mentioned above, the V such as 'and seemed', 'to be seemed', will also count for label 1. Thus, to cause error when labeling, so accuracy is lower than first label configuration setting.

Accuracy : 74.8%

The second methods I use the same structure as method 1. And I use NLTK, to do the stem first, rest of the part is much the same as method 1. I use **PorterStemmer() and also WordNetLemmatizer(),** to do stem of the 'V' part of sentence, and sometimes the stem may result in wrong or unappropriated condition, like 'playing' after stem is sometimes also 'playing', for 'playing' is actually been seen as Noun. So I also add pos='v', to specify pos. Combine with stem and lemmatize, I can barely get precise V & S & O of each sentence, and also config label by if the SVO after lemmatize is in the provided SVO.

Accuracy : 72.3%

Method 3 I use parsing tree to determine SVO in different way, I first catch the Verb by dep_ = 'ROOT', and catch the O by its head, if its head is point to ROOT, then it is viewed as O, and S is also check by V == 'ROOT's children, if the noun is in S and also in ROOT's children, I set the S's flag = True.

However, this method's accuracy is instead lower than 2

methods above, and I think the main reason is, sometimes, dep_ = ROOT is not really Verb of a sentence (although I'm not sure about answer via manually observe the sentence that which combination of SVO is correct answer) I found that sometime, ROOT is picked even the clause is a noun clause been placed in beginning of the sentence, so the ROOT is wrong, thus to cause S & O's error.

Leader board:

Rank 1st!!!!

(actually because at the time, there is little person who have made submission, so I get highest score XD)