

0716026 蕭楚濤

HW2 report

簡介：

這次作業我的判斷方法大體上都是以找關鍵字為主，也就是分別檢查S,V,O當中是否有包含我希望能包含的幾個關鍵字，如果都有包含，就算正確，反之錯誤。缺點是如果有冗字無法判斷出來。

方法一：

我的第一個嘗試是利用詞性做判斷，先利用spacy的pos_挑出所有動詞(verb)和助動詞(aux)，再去尋找這些token當中的children有沒有dep_恰好是nsubj, nsubjpass...這種主詞或者是dobj, pobj...這種受詞，如果存在就將subject或object的subtree當作關鍵字來檢查，因為考量部分主詞在其他動詞後也會擔任受詞，所以關鍵字只要符合其中一個組合即可，不用全部相同。

(f1 score = 0.730)

方法二：

方法二的想法是因為有時候動詞不會只有一個單字，而是類似片語有一整串，比如說：eased up on如果只算verb只會取道eased，而the game這個受詞因為是on的child，所以部會被算到，因此在這個方法中我會去檢查動詞後面幾個單字是否是advmod或prep，如果是，那他們的child也要一併檢查。但是結果反而比較差，所以可能是我的做法還不夠周密。

(f1 score = 0.715)

方法三：

方法三新增了兩個重點，第一個是只取dep_的root而不在取全部的verb和aux了，因為發現如果全部都取的話很容易會被判定成正確，但是其實這句話根本不是主旨。第二個是增加刪冗句的方法，因為sub tree中常常會把整個關係子句放進來，但是我們在檢查SVO的時候是不需要的，所以我這邊多了一個步驟把關係子句刪掉。

(f1 score = 0.755