

# NLP Hw2 報告

資工系 0816004 吳原博

## Overview

為了完成這次作業，我採取多種不同的做法，主要都是透過 spacy 的 dependency parse tree 完成，並且都是以謂語為中心檢查 SVO。不過，與其說是不同作法，我認為這更像是一種過程，以下介紹其中我認為比較具有代表性的三個部分：

### I. 找左找右

- 規則

遍歷 V 裡面的動詞，如果任一個動詞的 lefts 包含主詞且 rights 包含受詞，我就會認為那是合法的 SVO。

- 說明

這是一開始在熟悉 spacy 時所採用的方法，當初想說丟上去試試看沒想到就過 baseline 了。當然，這種作法一點也不合理，因為他假設主詞和受詞會皆在同一個動詞上，但是至少能正確判斷部分語句。

- 結論

在這裡我得到最大的收穫是發現如果任一個 SVO 不是原語句裡連續的詞彙就必須排除掉，光是做這件事就讓分數有顯卓的提升。

### II. 找頭找尾

- 規則

以 V 為中心，從 V 的第一個動詞 token 的 lefts 往前找到主詞，並且以 V 的最後一個介係詞或動詞 token 的 rights 找受詞

- 說明

當初採取這樣的規則，是因為在觀察句子裡的 SVO 結構以後，發現 V 可能不會是一個單詞，他可能是片語，含介係詞的不及物動詞，甚至是其他不完整的片段。

我認為，要找到主詞必須要從動詞下手，但是受詞可能會接在對應的介係詞後，也可能謂語裡包含多個動詞 (例如：plan to continue providing)，這種情況下就必須以最後面的動詞或介係詞去找受詞。

我是以 dep tag 來檢查找到的名詞是否為 S 和 O，我沒有用太多的 tag 去判斷 (在下面附上表格) 就能拿到不錯的分數了。

SVO	tag
S	(dep) nsubj, nsubjpass
V	(pos) AUX, VERB
O	(dep) dobj, pobj, attr

- **結論**

這樣的做法拿到的成績還不差，但是根據不同語法詞性必須要一一處理，例如當 V 裡面有助動詞時，而接下來的方法可以解決這樣的問題。

### III. Subtree

- **規則**

首先，從 V 裡面找到一個動詞，其 dependence parse tree 的 subtree 必須包含 V 裡面全部的詞，並且以這個動詞做 V 的代表。再來檢查主詞和受詞是否屬於上述動詞的 subtree，並且檢查其中是否包含一個詞具有我規定的 dep tag，並且指向 V 的一部分。

- **說明**

從 V 裡面找一個代表有兩個原因，一是確定這是一個合理的 V，另外是為了解決主受詞要從不同詞彙去找的問題。subtree 的想法來自 spacy 裡面的 subtree 功能，對於 SVO 都不是一個單詞的情況特別適用。

- **結論**

這個方法拿到我目前為止最好的分數，而且使用的 tag 跟第二個方法並無二致，相信花更多時間調整可以有更好的表現。

### 額外規則

- **說明**

除了上述基本規則，我還額外做了幾個判斷以針對那些較特殊的文法，之所以額外列出來是因為這些想法主要是之後微調時加上去的，跟前三者比較無關。我其實還有試一些其他方法，不過效果未必有提昇，因此在這裡我僅提最後版本所使用的策略。

- **不在句子裡面**

這是一個看似理所當然，但是花了我好一陣子才發現的問題。我發現光是過濾 SVO 不屬於原語句的情況就足以讓任何方法提昇大量的分數。

- **conj 動詞**

當一句話有多個謂語(這裡假設都是動詞)，並且用 and 這樣的連接詞串在一起，即使用 subtree 也無法正確找到其主詞，因此我以一個迴圈找到這串動詞的源頭並且以其找到主詞。

## Conclusion

---

這些規則看起來不難但其實並不好想，首先對於一句話裡面“怎樣才算是 SVO”這件事情就夠令人匪夷所思了，在這樣的前提要生出一個規則能適用所有語句更是困難，我能做的就是不斷猜測各種可能然後提交看分數怎麼樣。

無論如何，這次作業幫助我了解 spacy 這項工具的用法，並且讓我有機會仔細思考英文句子的詞性結構，算是讓我獲益良多。