

What can and cannot be modeled with Sequence Labeling

Adam Meyers
New York University



Summary

- What is sequence labeling?
- Modeling constituents with BIO tags
- Breaking up Relations into multiple tasks.



What is Sequence Labeling

- Find a sequence of labels that correspond to a sequence of tokens, each token being assigned one label
- Example: POS tagging
 - The dog walked into the park. → The/**DT**
dog/**NN** walked/**VBD** into/**IN** the/**DT** park/**NN** ./
PU



BIO Tags

- We can mark one type of constituent at a time with BIO tags.
 - B → beginning of constituent
 - I → inside constituent
 - O → outside constituent
- Suppose we are marking names:
 - Adam Meyers teaches at New York University in the United States.
 - Adam/**B-Name** Meyers/**I-Name** teaches/**O-Name** at/**O-Name** New/**B-Name** York/**I-Name** University/**I-Name** in/**O-Name** the/**O-Name** United/**B-Name** States/**I-Name** ./**O-Name**
- We can use methods similar to the HMM model we just discussed to handle some types of constituents (more detail in a later class).
- The transition probability is the same, but the likelihood probability is replaced with sets of features.



Relations cannot be represented as BIO sequences

- The price of chocolate rose 5%.
 - **PART**(%, price)
 - **Support**(rose, %)
- Relations are not constituents
- They are essentially pairs of constituents
- Finding relations is a sequence of 2 (or more) tasks
 - Find predicate of type X, e.g., %
 - Find arguments of type Y, given predicate of type X
 - Each of these can be a sequence labeling task (e.g., BIO)



The % Relation

- An instance of semantic role labeling
 - Discussed later in the semester
- It is part of many such relationships annotated during the NomBank project
 - <https://nlp.cs.nyu.edu/meyers/NomBank.html>
- There is also some NomBank data available for final projects
 - See Partitive-Files.zip and nombank_tasks.tgz in the shared NYU Drive directory



Relation Extraction Tasks

- Find argument, **given predicate**
- Find argument, **given support**
- Find argument, given **predicate and support**
- Realistic systems
 - Given → generated by automatic system(s)
 - First run systems recognizing all instances of % that take arguments and all related support items, e.g., **rose**
 - **output of such systems = input to argument finding system**
- Idealized system (using gold data)
 - Given taken from answer key
 - Use same system for finding arguments as above
 - Naturally, gold data should yield better results



More Semantic Role Labeling Examples

- The clam that the fish ate was complaining
 - Agent(**eat**, **the clam**) or ARG0
 - Patient(**eat**, **the fish**) or ARG1 or theme
 - Agent(**complain**, **the clam**) or ARG0
- In each case, a relationship between 2 words or phrases is being marked.
- Identifying single string spans are ill-suited to model these relationships.

