Dependency Parsing



Outline

Two formalisms for syntactic structure: Phrase structure and dependencies

Two algorithms for dependency parsing

- Transition based dependency parsing
- Graph based dependency parsing

Evaluating dependencies

Evaluating dependency parsers

The goal: Given a predicted tree and a reference tree

- How different is the structure of the predicted tree from that of the reference?
- How different are the labels of the predicted tree edges from the labels on the reference?

Correspondingly, there are two scores:

- Labeled attachment score
- Unlabeled attachment score

Accuracy= $\frac{\text{Number of correct dependencies}}{\text{Number of dependencies}}$

1 2 3 4 5 6

The tabby cat scratched the couch

Word	Dependent	Head	Label
The	1	3	det
tabby	2	3	amod
cat	3	4	subj
scratched	4	0	root
the	5	6	det
couch	6	4	obj

Accuracy= $\frac{\text{Number of correct dependencies}}{\text{Number of dependencies}}$

1 2 3 4 5 6

The tabby cat scratched the couch

	Word	Dependent	Head	Label
	The	1	3	det
	tabby	2	3	amod
~	cat	3	4	subj
	-scratched	4	0	root
	the	5	6	det
	≤couch	6	4	obj

Accuracy= $\frac{\text{Number of correct dependencies}}{\text{Number of dependencies}}$

1 2 3 4 5 6

The tabby cat scratched the couch

Word	Dependent	Head	Label
→The	1	3	det
⇒tabby	2	3	amod
≻ cat	3	4	subj
scratched	4	0	root
⇒the	5	6	det
∝ couch	6	4	obj

Accuracy= $\frac{\text{Number of correct dependencies}}{\text{Number of dependencies}}$

1 2 3 4 5 6

The tabby cat scratched the couch

Word	Dependent	Head	Label
⇒The	1	3	det
∍tabby	2	3	amod
cat	3	4	subj
-scratched	4	0	root
≥the	5	6	det
⊾couch	6	4	obj

Dependent Word Head Label The 1 2 det tabby 2 obj 4 subj cat 3 4 scratched 4 0 root 5 det the 6 couch 6 4 subj

Predicted dependencies

 $Accuracy = \frac{Number of correct dependencies}{Number of dependencies}$

1 2 3 4 5 6

The tabby cat scratched the couch

Word	Dependent	Head	Label
⇒The	1	3	det
⇒tabby	2	3	amod
≻ cat	3	4	subj
-scratched	4	0	root
⇒the	5	6	det
scouch	6	4	obj

Dependent Word Head Label *≫*The 1 2 det 2 →tabby obj 4 subj 3 >cat 4 scratched 4 0 root 5 det *⇒*the 6 **→**couch 6 4 subj

Predicted dependencies

 $Accuracy = \frac{Number of correct dependencies}{Number of dependencies}$

1 2 3 4 5 6

The tabby cat scratched the couch

Two edges are structurally incorrect

Word	Dependent	Head	Label
→The	1	3	det
	2	3	amod
>> cat	3	4	subj
scratched	4	0	root
≥the	5	6	det
couch	6	4	obj

Gold standard dependencies

 $Accuracy = \frac{Number of correct dependencies}{Number of dependencies}$

1 2 3 4 5 6

The tabby cat scratched the couch

Two edges are structurally incorrect



Gold standard dependencies

1 2 3

Label

det

amod

subj

root

The tabby cat scratched the couch

4

Two edges are structurally incorrect

Head

3

3

4

0

A worked example

Word

The

→tabby

cat

-scratched

Unlabeled attachment score = 4/6 = 66.6%

Accuracy=

6

5

the	5	6	det
couch	6	4	obj
	Gold standard de	ependenc	ies

Dependent

2

3

4



Number of correct dependencies

Number of dependencies

 $Accuracy = \frac{Number of correct dependencies}{Number of dependencies}$

1 2 3 4 5 6

The tabby cat scratched the couch

Three labeled edges are incorrect



Gold standard dependencies

 $Accuracy = \frac{Number of correct dependencies}{Number of dependencies}$

1 2 3 4 5 6

The tabby cat scratched the couch

Three labeled edges are incorrect

	Word	De	pendent	Head	Label
	>The		1	3	det
	⇒tabby		2	3	amod
	cat		3	4	subj
	-scratched		4	0	root
	⇒the		5	6	det
	∍couch		6	4	obj

Labeled attachment score = 3/6 = 50%

Word	Dependent	Head	Label
∕>The	1	2	det
tabby	2	4	obj
≥cat	3	4	subj
scratched	4	0	root
>the	5	6	det
couch	6	4	subj

Predicted dependencies