

Dependency parsing using neural networks.

Michał Zapotoczny
promotor: dr Jan Chorowski

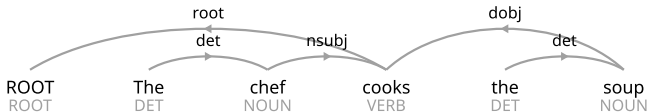
Instytut Informatyki
Uniwersytet Wrocławski

19.09.2017

What is dependency parsing and why it matters?

Definition

In dependency parsing we create a tree having words as nodes, representing structure of the sentence

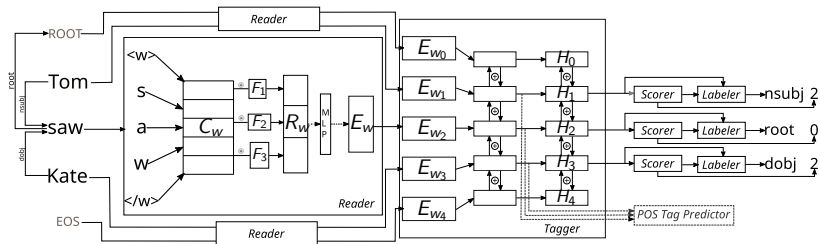


- Each word is connected to another word (or special ROOT token) via a labelled arc, forming a directed tree.
- Head-Dependent relationship is good approximation to semantic relationship between words
- Is used as part of NLP pipeline for many tasks (named entity recognition, relation extraction etc.)

- We train our network with supervised learning, so we need large amount of parsing examples
- Universal Dependencies is a large dataset of dependency parse trees for many languages (version 2.0 contains 70 treebanks for 50 languages). All treebanks in UD use unified format

- UAS - How many words have right parent?
- LAS - How many words have right parent and arc label?

Network architecture



Results

language	#sentences	Ours		SyntaxNet[2016]		ParseySaurus[2017]	
		UAS	LAS	UAS	LAS	UAS	LAS
Czech	87 913	91.41	88.18	89.47	85.93	89.09	84.99
Polish	8 227	90.26	85.32	88.30	82.71	91.86	87.49
Russian	5 030	83.29	79.22	81.75	77.71	84.27	80.65
German	15 892	82.67	76.51	79.73	74.07	84.12	79.05
English	16 622	87.44	83.94	84.79	80.38	87.86	84.45
French	16 448	87.25	83.50	84.68	81.05	86.61	83.1
Ancient Greek	25 251	78.96	72.36	68.98	62.07	73.85	68.1

Table: Baseline results of models trained on single languages from UD v1.3. Our models use only the orthographic representation of tokenized words during inference and works without a separate POS tagger.

Multilingual training

- Polish language has small number of dependency parsing examples (8 227 sentences, 83 571 words)
- Czech language has much better dataset (77 765 sentences, 1 332 566 words)
- Polish and Czech are similar.
- Use multitask learning to learn both languages at the same time
- How? Have 2 copies of the network and share part of the parameters between them

Multilingual results

Shared parts	Main lang	Aux lang	UAS	LAS
-	Polish	-	90.31	85.21
<i>Parser</i>	Polish	Czech	90.72	85.57
<i>Tagger, Parser</i>	Polish	Czech	91.19	86.37
<i>Tagger, POS Predictor, Parser</i>	Polish	Czech	91.65	86.88
<i>Reader, Tagger, POS Predictor, Parser</i>	Polish	Czech	91.91	87.77
<i>Parser</i>	Polish	Russian	90.31	85.07
<i>Tagger, POS Predictor, Parser</i>	Polish	Russian	91.34	86.36
<i>Reader, Tagger, POS Predictor, Parser</i>	Polish	Russian	89.16	82.94
-	Russian	-	83.43	79.24
<i>Parser</i>	Russian	Czech	83.15	78.69
<i>Tagger, POS Predictor, Parser</i>	Russian	Czech	83.91	79.79
<i>Reader, Tagger, POS Predictor, Parser</i>	Russian	Czech	84.78	80.35

Table: Impact of parameter sharing strategies on main language parsing accuracy when multilingual training is used for additional supervision.

Thank you

Questions?