# Project II: Cross-Domain Dependency Parsing

Weiwei Sun

Institute of Computer Science and Technology
Peking University

April 15, 2019

# Project II & III

**Project II**: Cross-domain dependency parsing

- Use 1% data for training, development and test
- 2019/05/15: test data release (official evaluation). **Submit your results to TAs before 2019/05/01.** We will provide further computational resources for high-performance systems.

**Project III**: Open Domain Semantic Parsing

- We will provide a *subtask* using a small data set.

# Cross-domain dependency parsing

**Practical suggestions**

- Data, data and data.
- Tools, tools and tools.

# Neural dependency parsing

- Deep Learning: A method of representation learning.
  Induce features from data automatically.
- Neural network models are powerful by approximating non-linear functions

Recurrent Neural Networks

- An initial state vector $h_0$
- In step $t$, it take input $x$ and calculate a new state $h_t$:

$$h_t = f(h_{t-1}, x)$$

- $h_t$ not only contains information from $x$, but also contextual information
- Given a sequence $x_1, x_2, \cdots, x_n$, iteratively, we can get an output sequence $h_1, h_2, \cdots, h_n$.
- Powerful to encode the context.

# Linguistic resources

Two types

- Labeled data. Here, treebanks!
- Unlabeled data. Useful for cross-domain evaluations.

Heterogeneous annotations

- Different but comparable annotation specifications.
- Chinese TreeBank (CTB)
  - Annotated by UPenn since 1998
  - Including word segmentation, POS tagging and phrase structure
  - The latest version is CTB 9.0
  - Including news, magazine, radio and blog articles

# Open-source tools

Stanford CoreNLP

- Phrase Structure Parsing / Dependency Parsing
- `https://stanfordnlp.github.io/CoreNLP/download.html`

Attentive Span Parser

- `https://github.com/nikitakit/self-attentive-parser/`

Biaffine-attention-based Dependency Parser

- `https://github.com/tdozat/Parser-v2`

| Data | UAS | LAS |
|---|---|---|
| PTB-SD | 95.8 | 94.6 |
| CTB-UD | 92.1 | 90.5 |