上海科技大学
ShanghaiTech University

A Hybrid Dependency Parsing Model Enhanced by Shared Feature

——Encoded in Shared Bidirectional LSTM @CS 280

Yunzhe Yuan   Zhiqiang Xie
SIST@ShanghaiTech University

Dependency Parsing & Grammar

Example From Kubler, McDonald, and Nivre, 2009

$$G = h(S, \Gamma, \boldsymbol{\lambda})$$

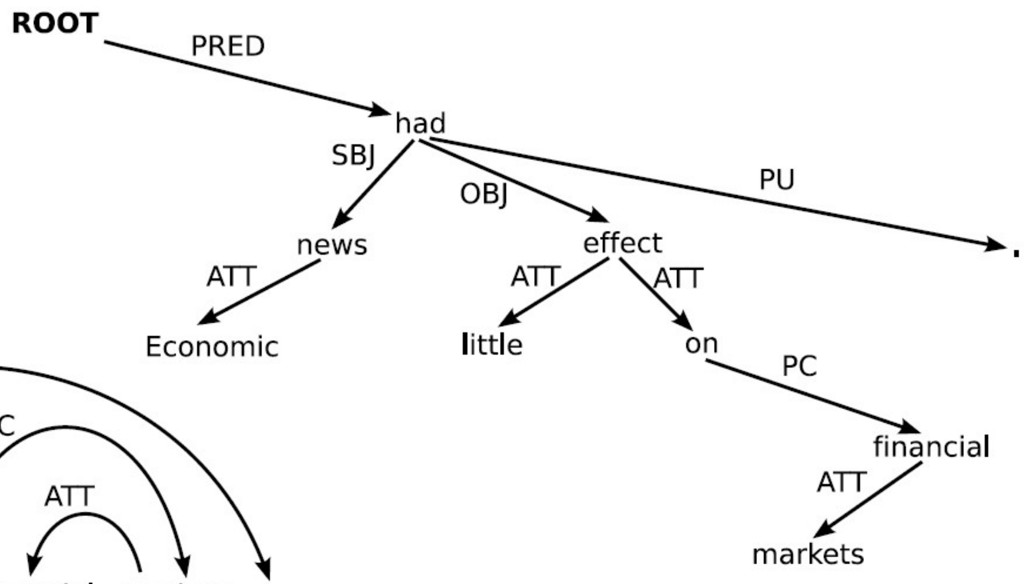
G: Dependency Tree

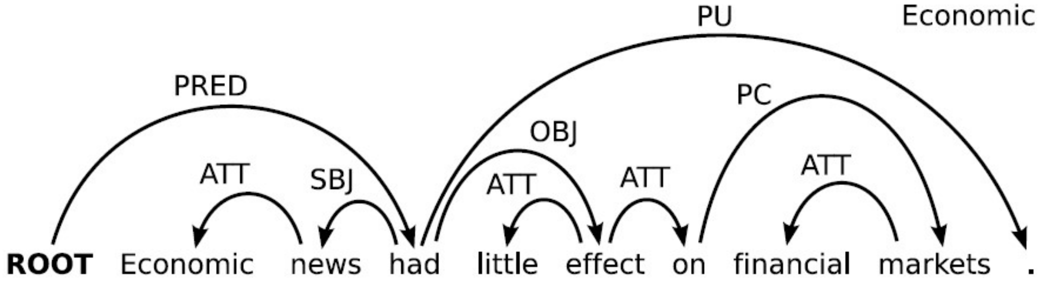
h: A search over the set of G.

S: Sentence

Γ: Rules

λ: Parameter set





Dependency Parsing is the task of recognizing a sentence and assigning a syntactic structure to it. The most widely used syntactic structure is the parse tree which can be generated using some parsing algorithms.

Graph-based & Transition-based Parser

Graph-based parser from graph theory:

$$h(S, \Gamma, \boldsymbol{\lambda}) = \underset{G=(V,A)\in\mathcal{G}_S}{\operatorname{argmax}} \prod_{(w_i,r,w_j)\in A} \lambda_{(w_i,r,w_j)}$$

Impoverished feature representation
Better search space

Transition-based parser from abstract machine:

t: Transition
c: Configuration

$\mathbf{h}(S, \Gamma, \boldsymbol{\lambda})$

$c \leftarrow c_0(S)$
while c is not terminal
  $t \leftarrow \lambda_c$
  $c \leftarrow t(c)$
return G_c

Rich feature representations
Error propagation

Motivation & Baseline

Table 7.1: Labeled parsing accuracy for top scoring graph and transition-based systems at CoNLL 2006.

Language	Ar	Bu	Ch	Cz	Da	Du	Ge	Ja
Transition-based	66.71	87.41	86.92	78.42	84.77	78.59	85.82	91.65
Graph-based	66.91	87.57	85.90	80.18	84.79	79.19	87.34	90.71

Language	Po	Sl	Sp	Sw	Tu	Average
Transition-based	87.60	70.30	81.29	84.58	65.68	80.75
Graph-based	86.82	73.44	82.25	82.55	63.19	80.83

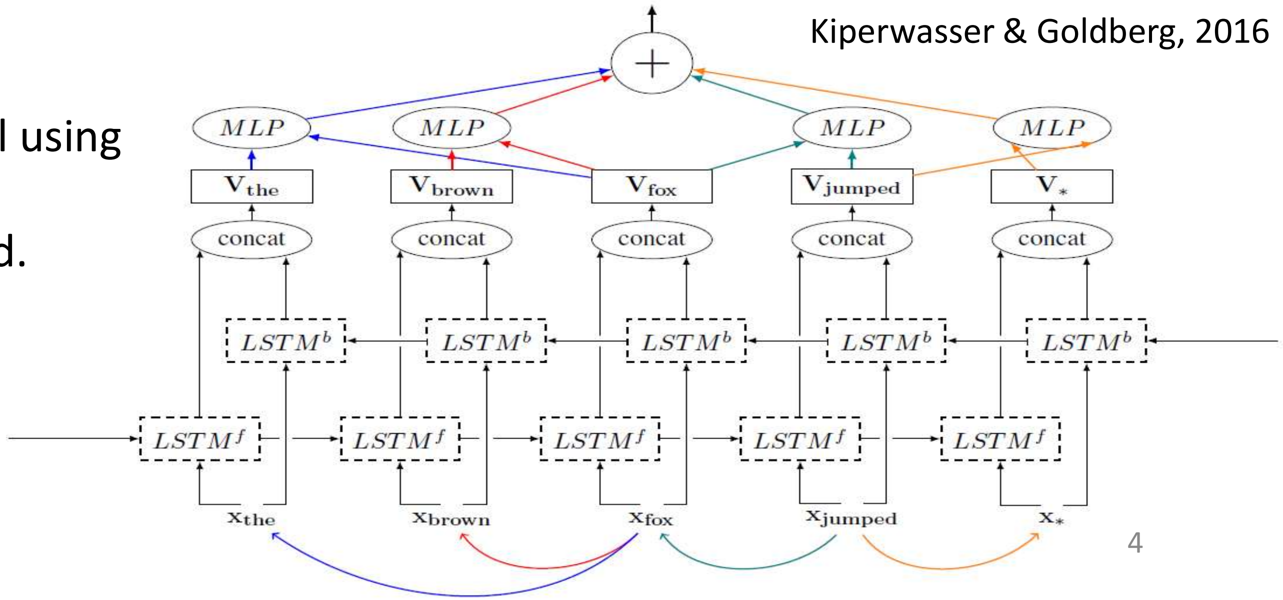
Kubler, McDonald, and Nivre, 2009

Similar performance are achieved by these two parsing methods, but difference can differ much from the different languages.

Baseline:

A simple and accurate model using Bidirectional-LSTM feature representations is introduced.

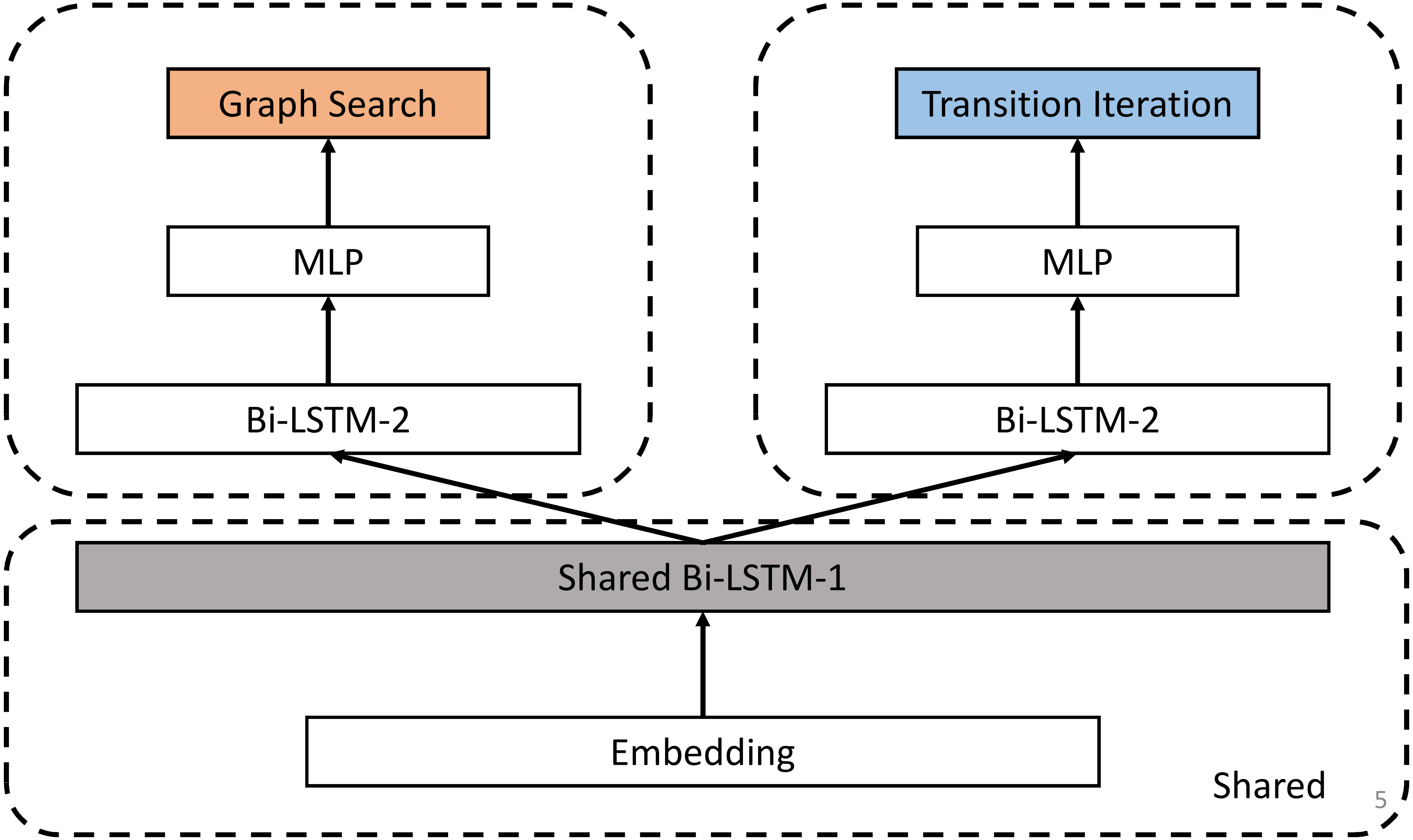
Kiperwasser & Goldberg, 2016



Hybrid Model

Insight:

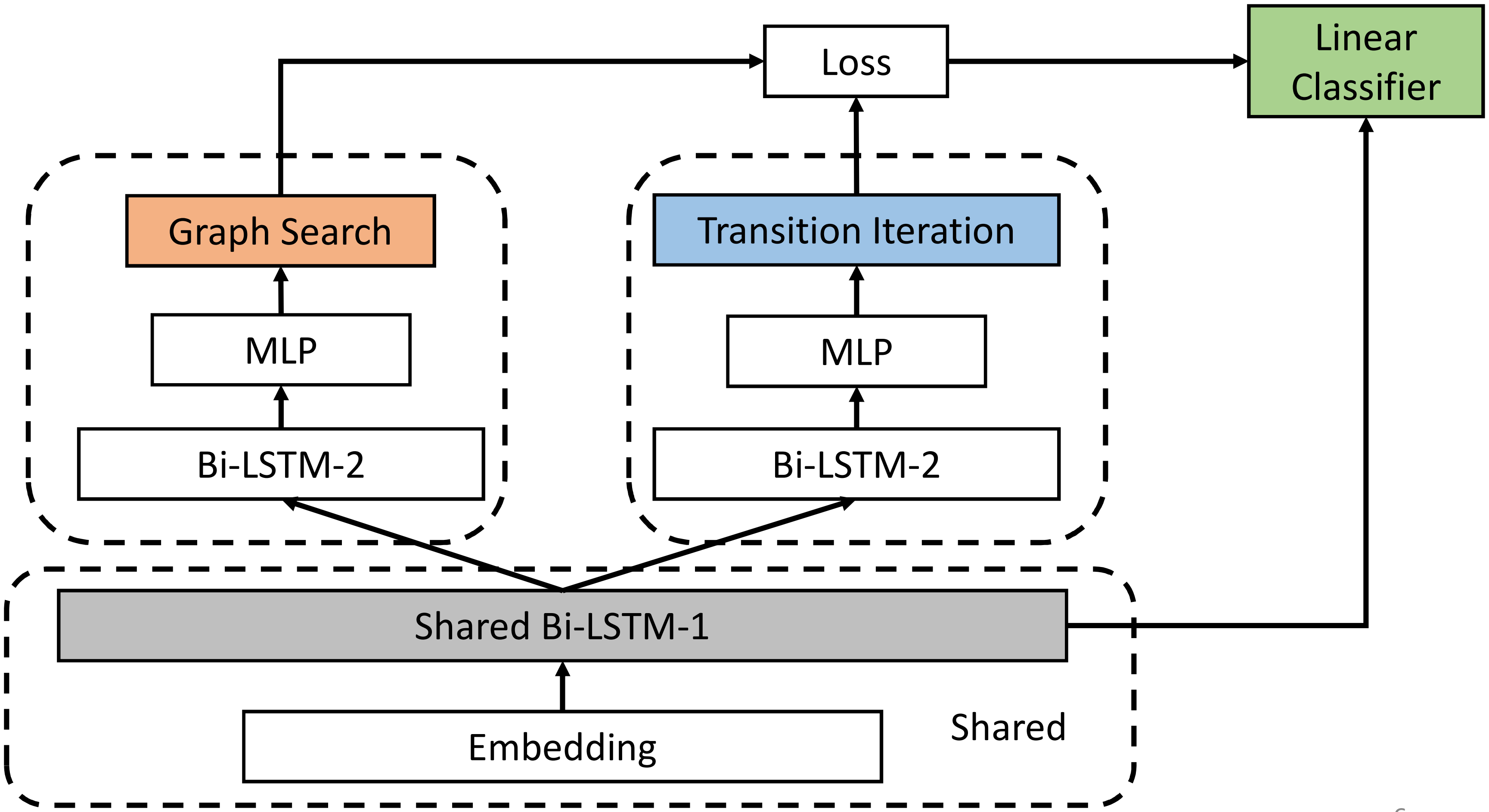
There are some intrinsic features of natural language are model-invariant.



Hybrid Model

Insight:

Integrate a classifier (online re-ranking) to choose the fitted model.



Result & Conclusion

Compared to the baseline, our hybrid model achieves a few improvement.

Model	Dataset	UAS	LAS
Graph	UD-en	88.38	86.32
Transition	UD-en	87.36	85.25
Hybrid + Classifier	UD-en	**88.85**	86.66
Hybrid + Random	UD-en	88.73	86.55
Hybrid + Graph	UD-en	88.65	86.26
Hybrid + Transition	UD-en	88.49	86.29
Graph	UD-zh	82.85	79.88
Transition	UD-zh	**83.28**	80.55
Hybrid + Classifier	UD-zh	82.52	79.71
Hybrid + Random	UD-zh	82.56	79.73
Hybrid + Graph	UD-zh	82.27	89.18
Hybrid + Transition	UD-zh	83.2	80.19

Model	Dataset	Data 1 Accuracy	Data 2 Accuracy
Graph	taobao	86.30%	87.82%
Graph + Embedding	taobao	87.64%	89.18%
Graph + Embedding + Ontology	taobao	87.47%	89.12%
Hybrid + Classifier	taobao	**88.74%**	88.58%
Hybrid + Classifier + Ontology	taobao	88.68%	88.92%
Hybrid + Random + Ontology	taobao	88.73%	88.87%
Hybrid + Graph + Ontology	taobao	88.59%	**89.51%**
Hybrid + Transition + Ontology	taobao	88.45%	89.21%

Besides, we tested our model in a special tiny dataset (chatting-style), which shows better performance. It may suggest this model can work better for tiny dataset learning.

Future work

Here are some defects in our model and benchmark yet:

1. A more serious benchmark is required.
   - more testing required
   - train and test our model in some more widely-used datasets (like PTB and CTB)
   - introduce external embedding for a fully benchmark (especially for Chinese)

2. Some structural defects of the neural network bring the trouble of inefficiency.

3. Try to find a better normalization method to feed the loss calculated by the two different evaluation algorithm to the classifier.

上海科技大学
ShanghaiTech University

Thank you!