Tag-Enhanced Tree-Structured Neural Networks for Implicit Discourse Relation Classification

Yizhong Wang, Sujian Li, Jingfeng Yang,
Xu Sun, Houfeng Wang

北京大學
PEKING UNIVERSITY

Outline

- Background: Discourse Relation Classification
- Motivation: Parse tree should be considered!
- Methods:
    - Tree-Structured LSTM/GRU
    - Enhance them with constituent tags
- Experiments: SOA performance on PDTB
- Conclusion: Parse tree does help!

Background: Discourse Relations

[Prasad et al. 2008]

Definition: How two segments of discourse are **logically connected.**

Background: Discourse Relations

[Prasad et al. 2008]

Definition: How two segments of discourse are **logically connected.**

- **Explicit:**

[I love presentation] but [preparing for it is bothersome!]

Arg1 Connective Arg2 (Comparison)

PEKING UNIVERSITY 北京大學

Background: Discourse Relations

[Prasad et al. 2008]

Definition: How two segments of discourse are **logically connected.**

- **Explicit:**

[I love presentation] but [preparing for it is bothersome!]

Arg1 Connective Arg2

(Comparison)

- **Implicit:**

Arg1: [The economic plans came into effect.]

Implicit=Therefore

Arg2: [Global economy returns to growth.]

(Cause)

北京大学 PEKING UNIVERSITY

Implicit Discourse Relation Classification

- Classify the discourse relation $\tilde{y}$ given two arguments r_1, r_2 without connective information:

$$\arg\max_{\tilde{y}} P(\tilde{y}|r_1, r_2)$$

- Even challenging for human!

Arg1: [But a few funds have taken other defensive steps. Some have raised their cash positions to record levels.]

Arg2: [High cash positions help buffer a fund when the market falls.]

Restatement or Cause?

Encoding the Arguments

- Traditional feature-based methods:
  - Lin et al. (2009)
  - Pitler et al. (2009)
  - Rutherford and Xue (2014)

Fast and significant improvement!

- Neural network models:
  - Ji and Eisenstein (2014)
  - Zhang et al. (2015)
  - Liu and Li (2016)
  - Qin et al. (2016, 2017)

Syntax hasn't been fully exploited.

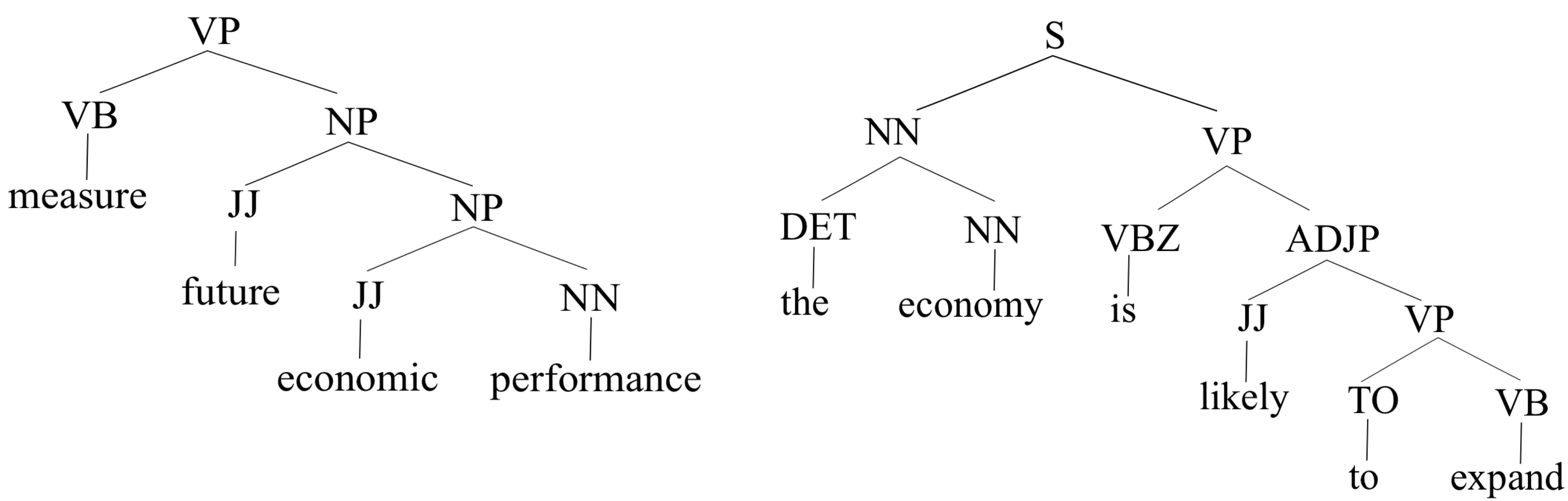
Motivation: Parse Tree Might Help

- Naturally capture the important phrases along the parse tree

Arg1: The index is intended to measure future economic performance

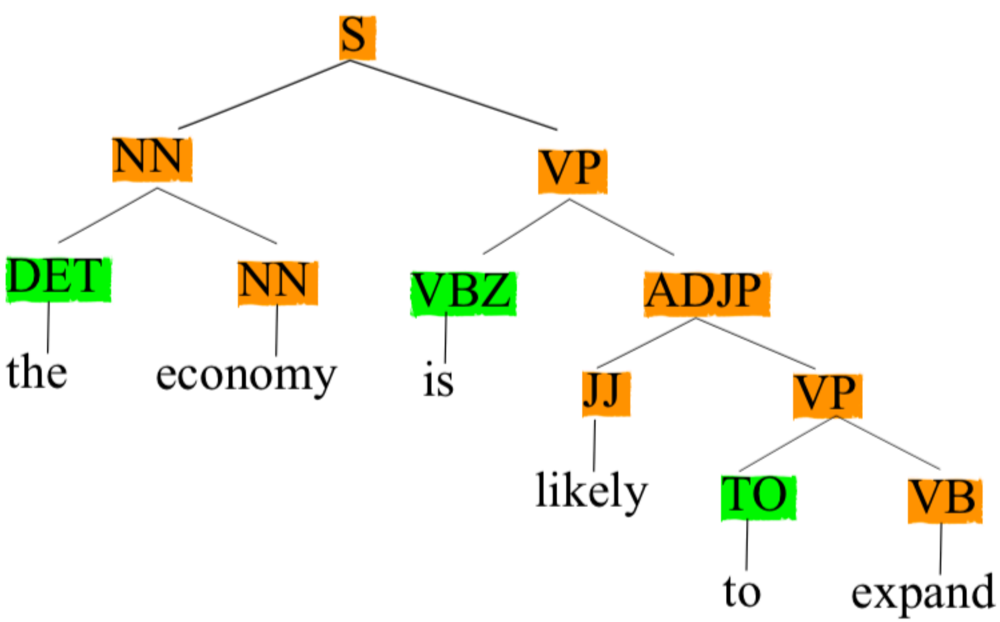
Arg2: A figure above 50 indicates the economy is likely to expand

(Expansion.Restatement.Specification, wsj_0233)



Parse Tree Might Help (cont.)

- Constituent Tags provide another important signal:
  - Defined by Bies et al. (1995), including clause-level tags (SBAR, SINV, SQ …), phrase-level tags (NP, VP, PP…) and word-level tags (NN, VP, JJ …).
  - describe the syntactic role of a constituent.



- Orange nodes are more informative.
- Green nodes are less important.

Method: Tree-Structured Neural Nets

- Encoder encodes the two arguments with tree-structured neural networks:
  - Tree-LSTM (Tai et al. ,2015)
  - Tree-GRU (Kokkinos and Potamianos, 2017)
  - Tag-Enhanced Tree-LSTM
  - Tag-Enhanced Tree-GRU
- Relation classifier classifies the relation type based on the encoding:

$$\hat{y} = softmax\left(W^{(\hat{y})}\left[r_1, r_2\right] + b^{(\hat{y})}\right)$$

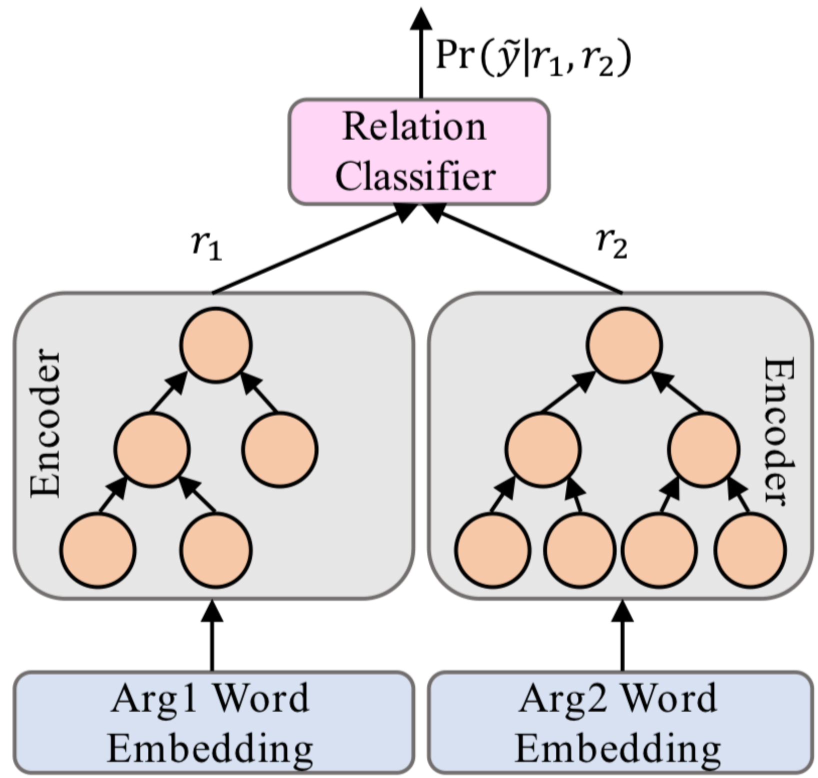


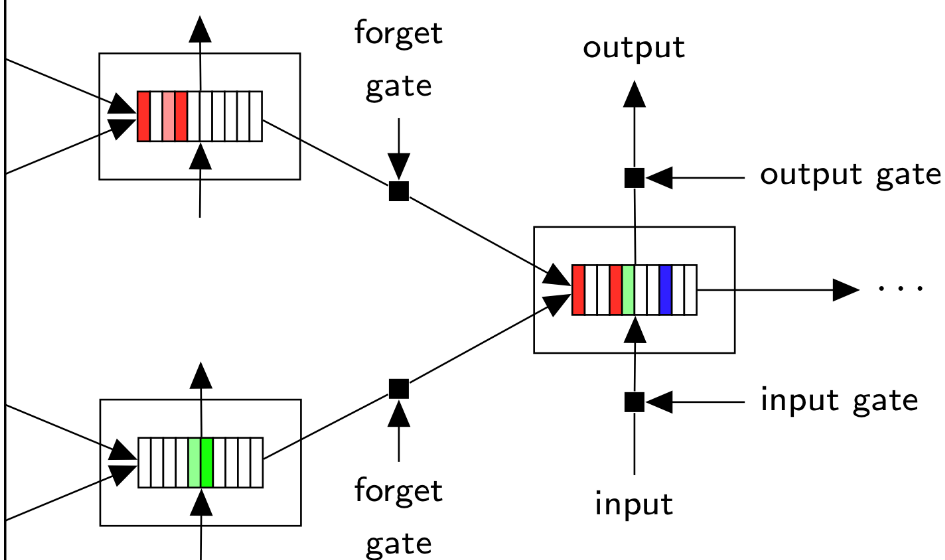
Figure 2: Architecture of our discourse relation classification model. Layers with the same color share the same parameters.

Tree-LSTM and Tree-GRU Models

[Tai et al. ,2015; Kokkinos and Potamianos, 2017]

- In the Tree-LSTM/GRU models, a similar LSTM/GRU unit is applied to each node in the tree in a bottom-up manner.
- At each step, the unit considers the current input and the state information from two preceding nodes.

The Tree-LSTM Unit



$$i_j = \sigma\left(W^{(i)}x_j + U^{(i)}\left[h_j^L, h_j^R\right]\right)$$

$$f_j = \sigma\left(W^{(f)}x_j + U^{(f)}\left[h_j^L, h_j^R\right]\right)$$

$$o_j = \sigma\left(W^{(o)}x_j + U^{(o)}\left[h_j^L, h_j^R\right]\right)$$

$$u_j = \tanh\left(W^{(u)}x_j + U^{(u)}\left[h_j^L, h_j^R\right]\right)$$

$$c_j = i_j \odot u_j + f_j \odot c_j^L + f_j \odot c_j^R$$

$$h_j = o_j \odot \tanh\left(c_j\right)$$

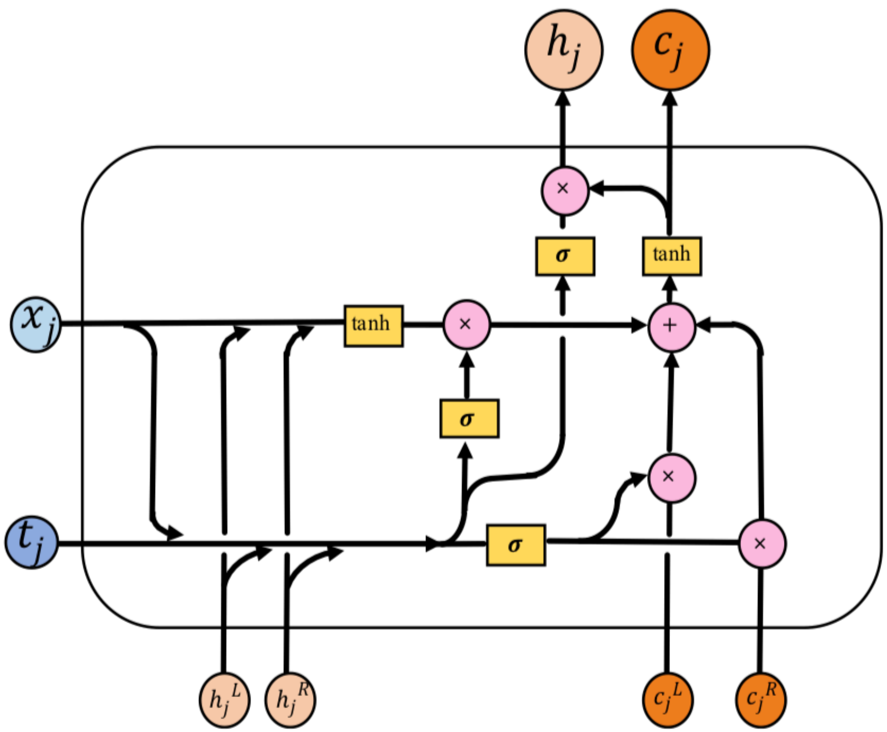
What More Can We Do?

- Linguistic view: Constituent tags reflect the importance of the constituents.
- Model vies: Gates control the flow of information along the tree and thus determine the semantic composition;

Leveraging the tags to control the flow of information and enhance the Tree-LSTM/GRU.

Tag-Enhanced Tree-LSTM Unit

- Embed the constituent tag at node j as t_j and consider it when computing the the input/forget/output gates:



$$i_j = \sigma\left(W^{(i)}x_j + M^{(i)}t_j + U^{(i)}\left[h_j^L, h_j^R\right]\right)$$

$$f_j = \sigma\left(W^{(f)}x_j + M^{(f)}t_j + U^{(f)}\left[h_j^L, h_j^R\right]\right)$$

$$o_j = \sigma\left(W^{(o)}x_j + M^{(o)}t_j + U^{(o)}\left[h_j^L, h_j^R\right]\right)$$

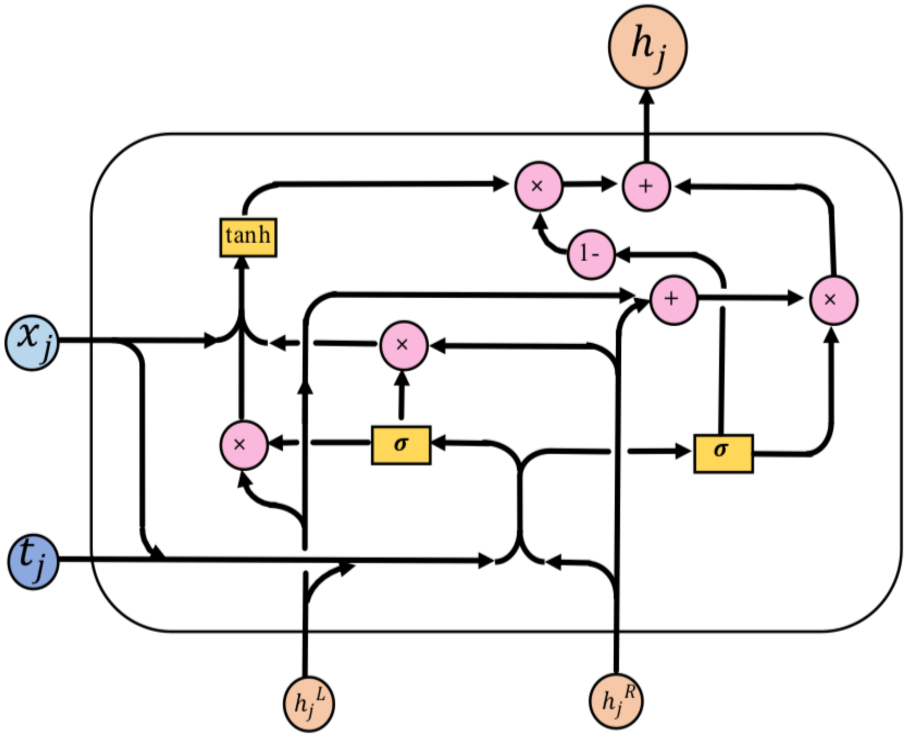
$$u_j = \tanh\left(W^{(u)}x_j + U^{(u)}\left[h_j^L, h_j^R\right]\right)$$

$$c_j = i_j \odot u_j + f_j \odot c_j^L + f_j \odot c_j^R$$

$$h_j = o_j \odot \tanh\left(c_j\right)$$

Tag-Enhanced Tree-GRU Unit

- Similarly, use tag embedding to compute new reset and update gates:



$$\boxed{r_j} = \sigma\left(W^{(r)}x_j + \boxed{M^{(r)}t_j} + U^{(r)}\left[h_j^L, h_j^R\right]\right)$$

$$\boxed{z_j} = \sigma\left(W^{(z)}x_j + \boxed{M^{(z)}t_j} + U^{(z)}\left[h_j^L, h_j^R\right]\right)$$

$$\tilde{h}_j = \tanh\left(W^{(h)}x_j + U^{(h)}\left[h_j^L \odot r_j, h_j^R \odot r_j\right]\right)$$

$$h_j = z_j \odot \tilde{h}_j + (1 - z_j) \odot \left(h_j^L + h_j^R\right)$$

北京大学 PEKING UNIVERSITY

Experiments

- We train and test our models on Penn Discourse Treebank (Prasad et al., 2008)
- Training objective:

$$J(\theta) = \frac{1}{N} \sum_{k=1}^{N} E(\hat{y}, y) + \frac{\lambda}{2} \|\theta\|^2$$

$$E(\hat{y}, y) = -\sum_{j=1}^{n} y_j \times \log \hat{y}_j$$

- We experiment on the multi-class classification of both the Level-1 classes (4 labels) and the Level-2 types (11 labels)

北京大學 PEKING UNIVERSITY

Experimental Results

- Tree-LSTM/GRU achieve better performance than the their sequential correspondings.
- The tag information is effective for both models.

Models	Level-1 Classification		Level-2 Classification	
	Dev	Test	Dev	Test
Bi-LSTM	55.10	56.88	35.02	42.44
Bi-GRU	55.21	57.01	35.34	42.46
Tree-LSTM	56.04	58.89	35.76	43.02
Tree-GRU	55.36	58.98	36.09	43.78
Tag-Enhanced Tree-LSTM	**56.97**	**59.85**	35.92	**45.21**
Tag-Enhanced Tree-GRU	56.63	59.75	**36.93**	44.55

Table 1: The accuracy score of multi-class classification

Comparison with Previous Work

Systems	Accuracy
Zhang et al. (2015)	55.39
Rutherford and Xue (2014)	55.50
Rutherford and Xue (2015)	57.10
Liu et al. (2016)	57.27
Liu and Li (2016)	57.57
Ji et al. (2016)	59.50
Tag-Enhanced Tree-LSTM	**59.85**
Tag-Enhanced Tree-GRU	59.75

Table 3: Accuracy (%) for Level-1 multi-class classification on the test set, compared with other state-of-the-art systems.

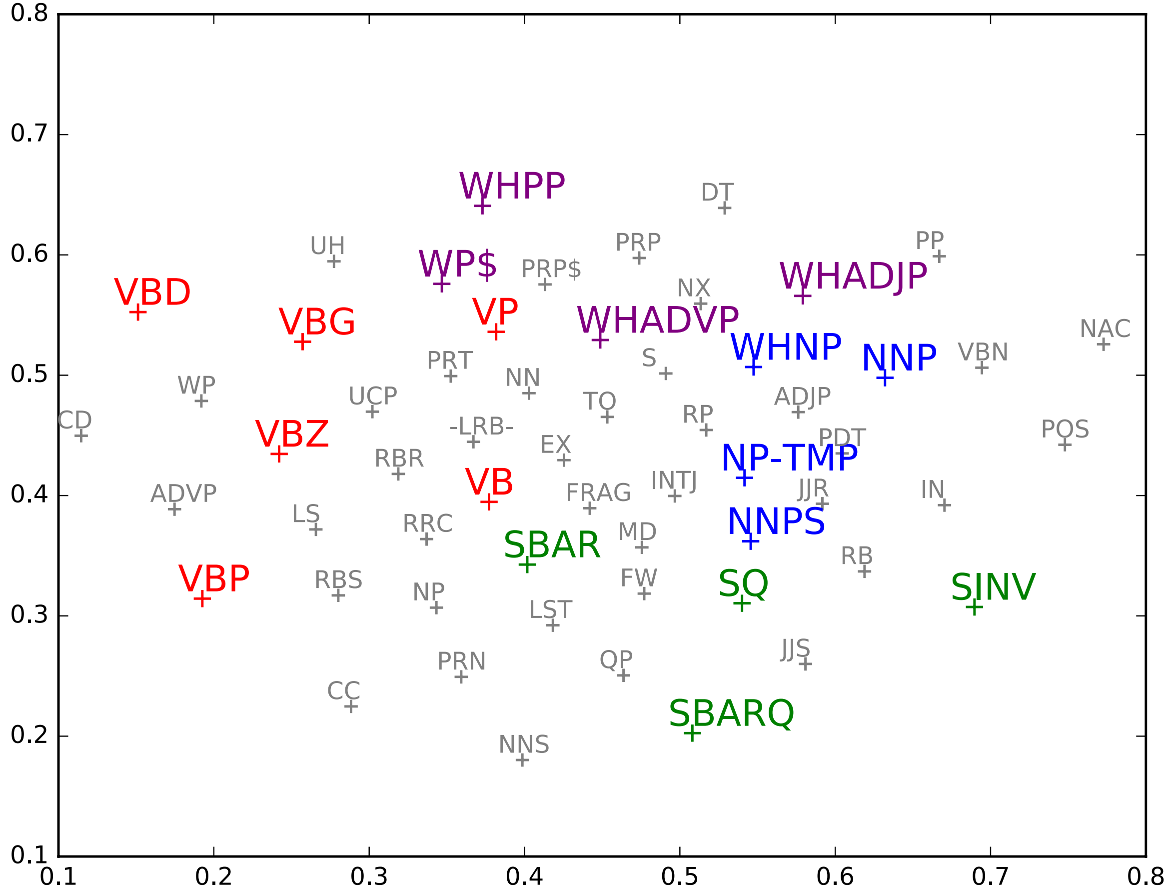
Comparison with Previous Work (cont.)

Systems	Accuracy
Lin et al. (2009)	40.66
Ji and Eisenstein (2014)	44.59
Qin et al. (2016)	45.04
Qin et al. (2017)	**46.23**
Tag-Enhanced Tree-LSTM	45.21
Tag-Enhanced Tree-GRU	44.55

Table 4: Accuracy (%) for Level-2 multi-class classification on the test set, compared with other state-of-the-art systems.

Analysis of the Constituent Tags

- We visualize the embeddings of constituent tags using t-SNE method (Maaten and Hinton, 2008).



Conclusions

1. Two latest tree-structured neural networks are applied to the discourse relation classification task.
2. The syntactic parse tree are exploited from two aspects:
   - The tree structure are used to recursively compose semantics in a bottom-up manner;
   - The constituent tags are used to control the semantic composition process via embedding and gating.
3. State-of-the-art performance.
4. Future work: more tasks, more syntactic information

Thank you!

Contact: yizhong@pku.edu.cn

Code is available:

https://github.com/EastonWang/TagNN-PDTB

北京大學
PEKING UNIVERSITY