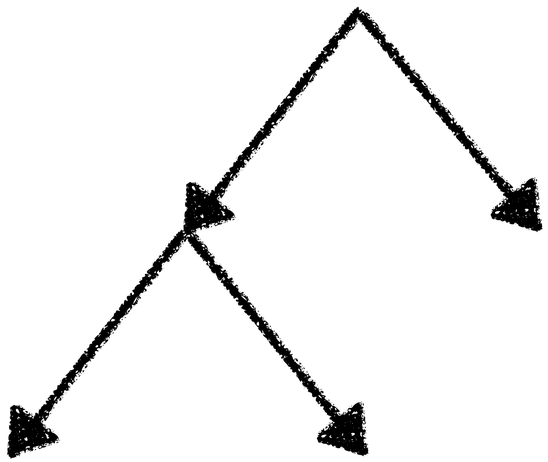
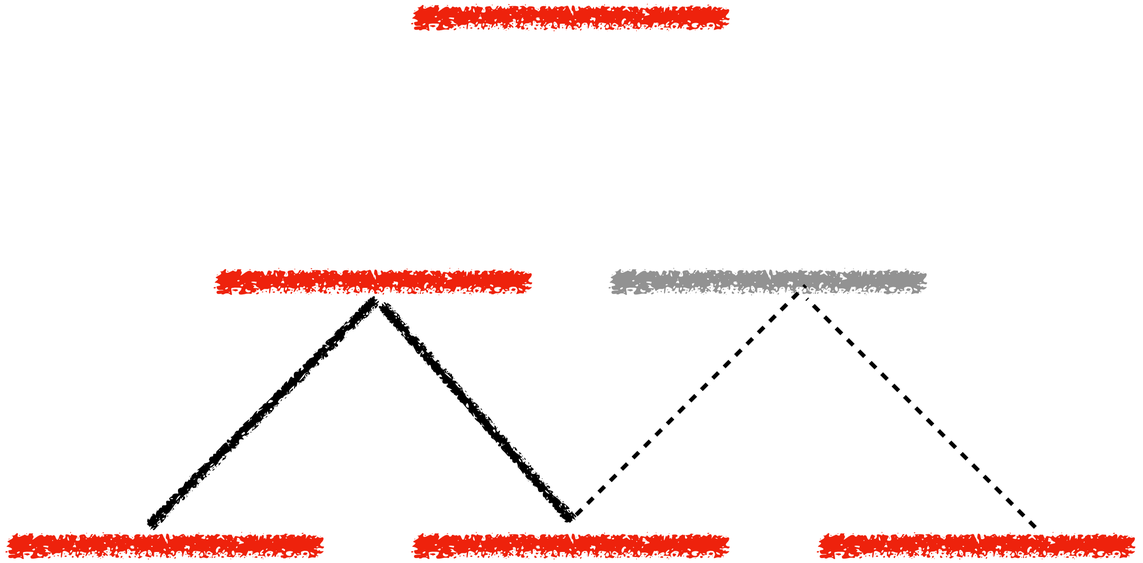
A Span-based Linearization for Constituent Trees

Yang Wei, Yuanbin Wu, Man Lan

East China Normal University
School of Computer Science and Technology
godweiyang@gmail.com

CKY-style

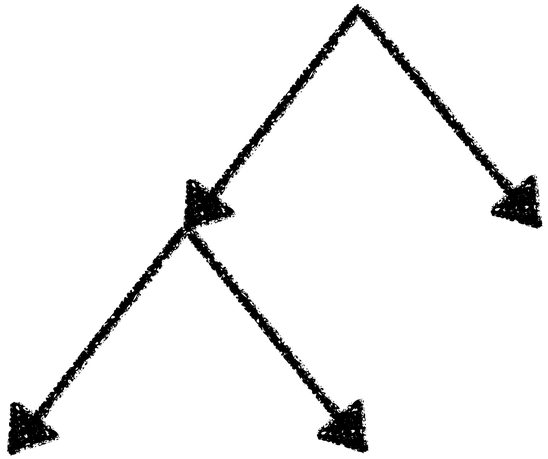




Accuracy

Efficiency

Transition-based

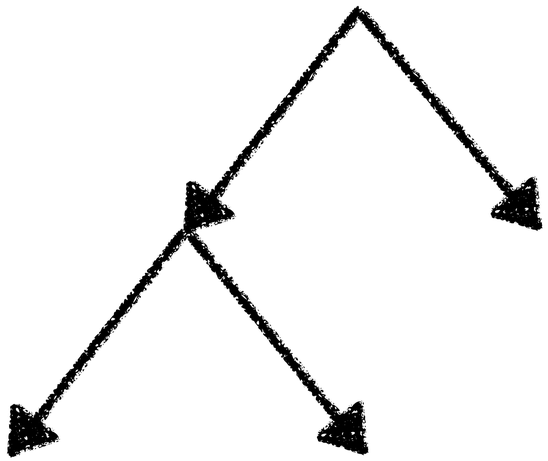


shift shift reduce shift reduce

Accuracy

Efficiency

Sequence Labeling/Seq2seq

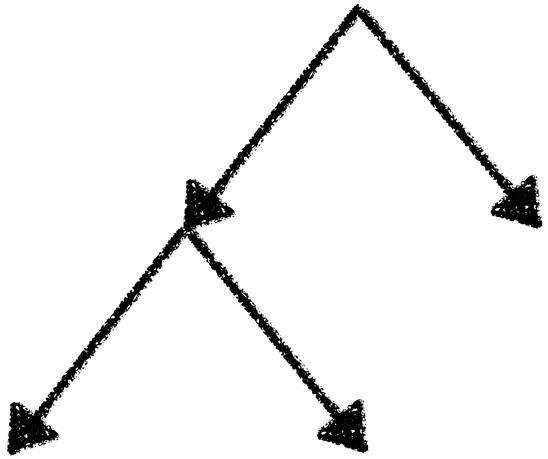


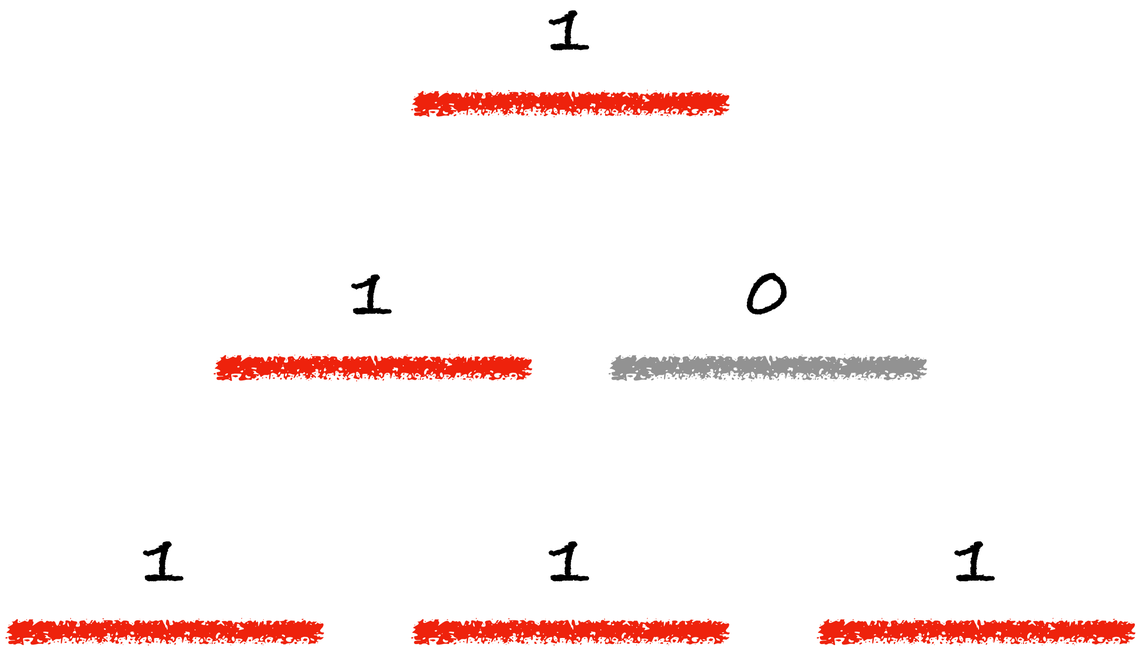
((( ) ( )) ( ))

Accuracy

Efficiency

Other Local Models

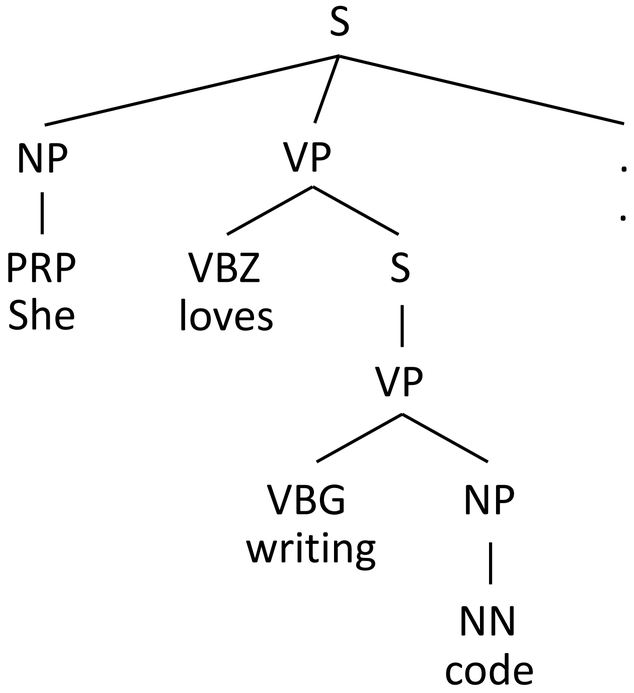




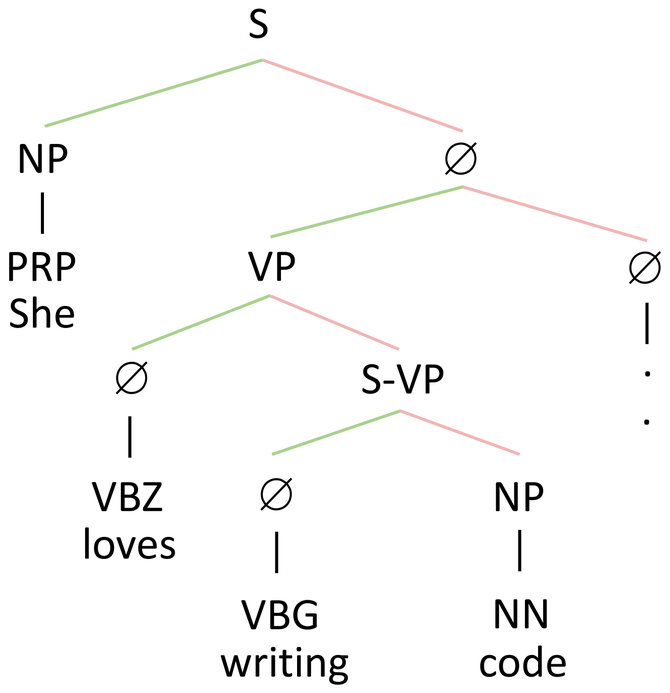
Accuracy

Efficiency

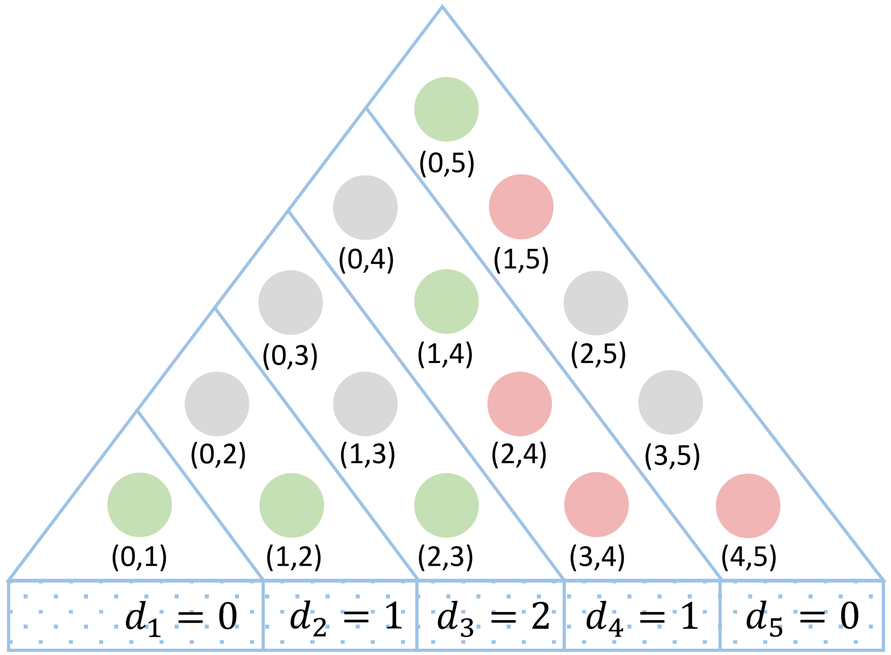
Our Linearization



(a) Original parsing tree.



(b) Right binary tree.

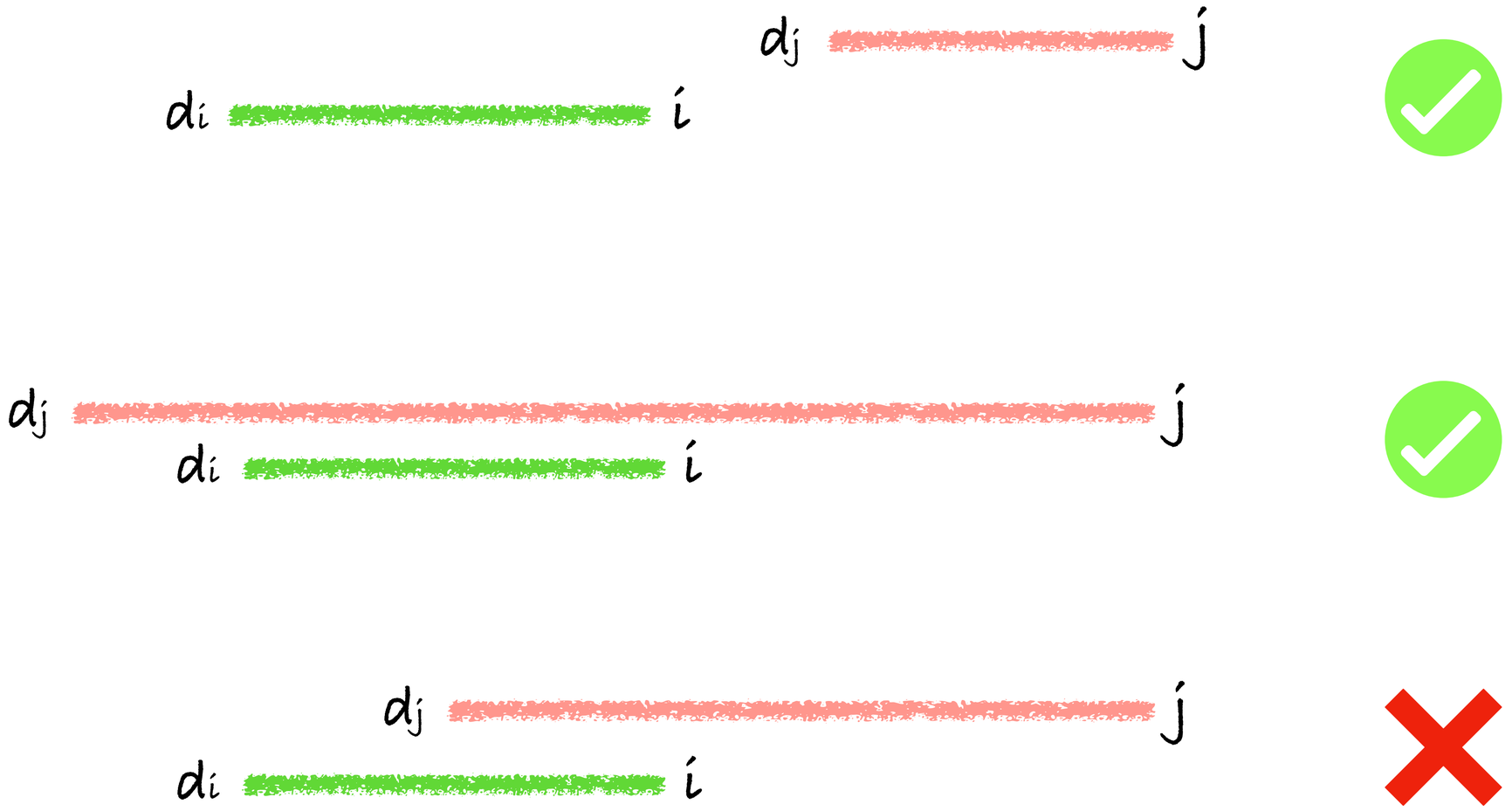


(c) Span table and linearization.

Accuracy

Efficiency

Tree Reconstruction

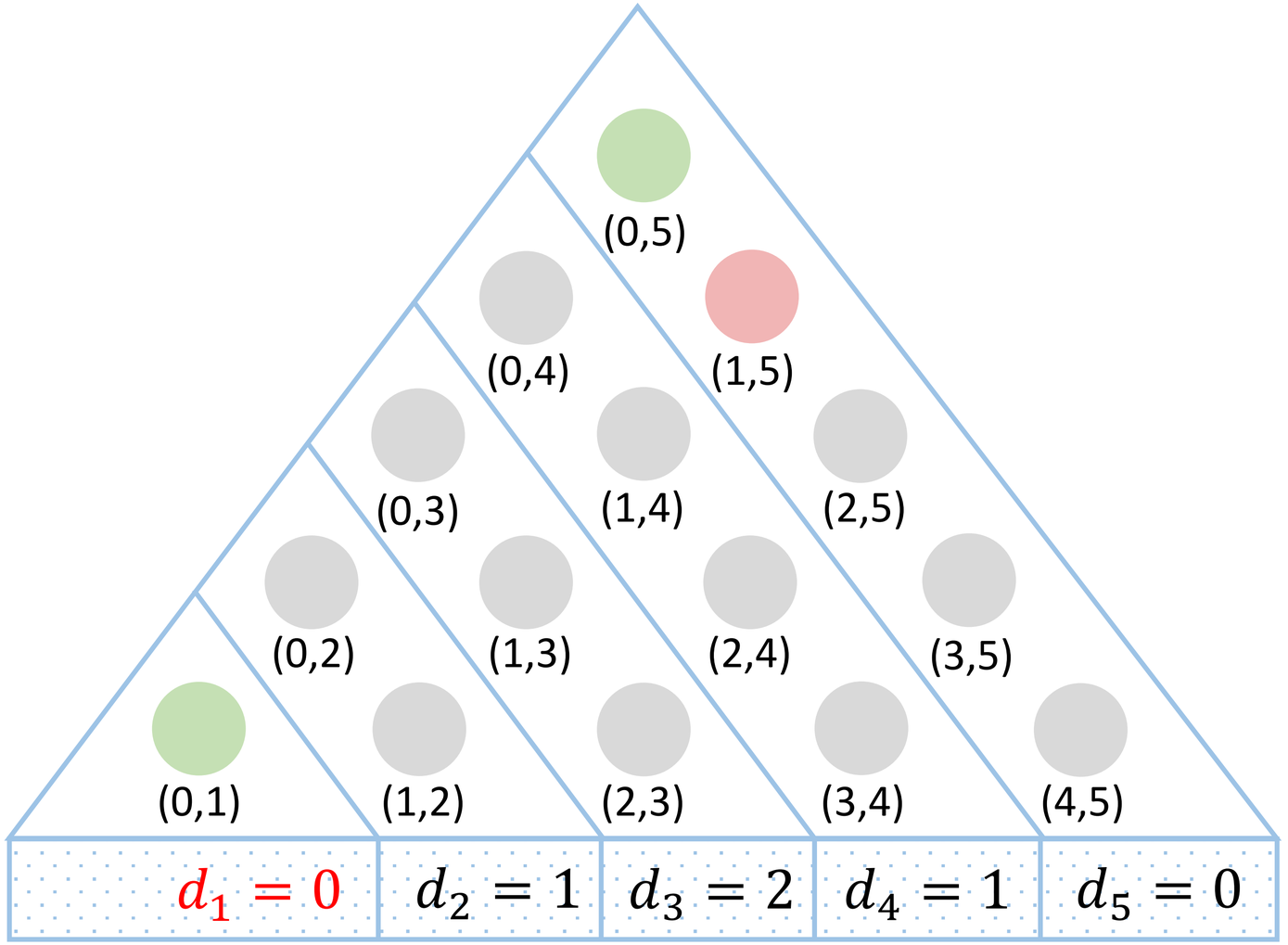


Tree Reconstruction

Algorithm 2 Tree reconstruction.

function TREE($i, j, \mathcal{D}$)
    if $i + 1 = j$ **then**
        node $\leftarrow$ Leaf($w_j, \ell(i, j)$)
    else
        $k \leftarrow \max\{k' \mid d_{k'} = i, i < k' < j\}$
        child$_l$ $\leftarrow$ TREE($i, k, \mathcal{D}$)
        child$_r$ $\leftarrow$ TREE($k, j, \mathcal{D}$)
        node $\leftarrow$ Node(child$_l$, child$_r$, $\ell(i, j)$)
    end if
    return node
end function

Example



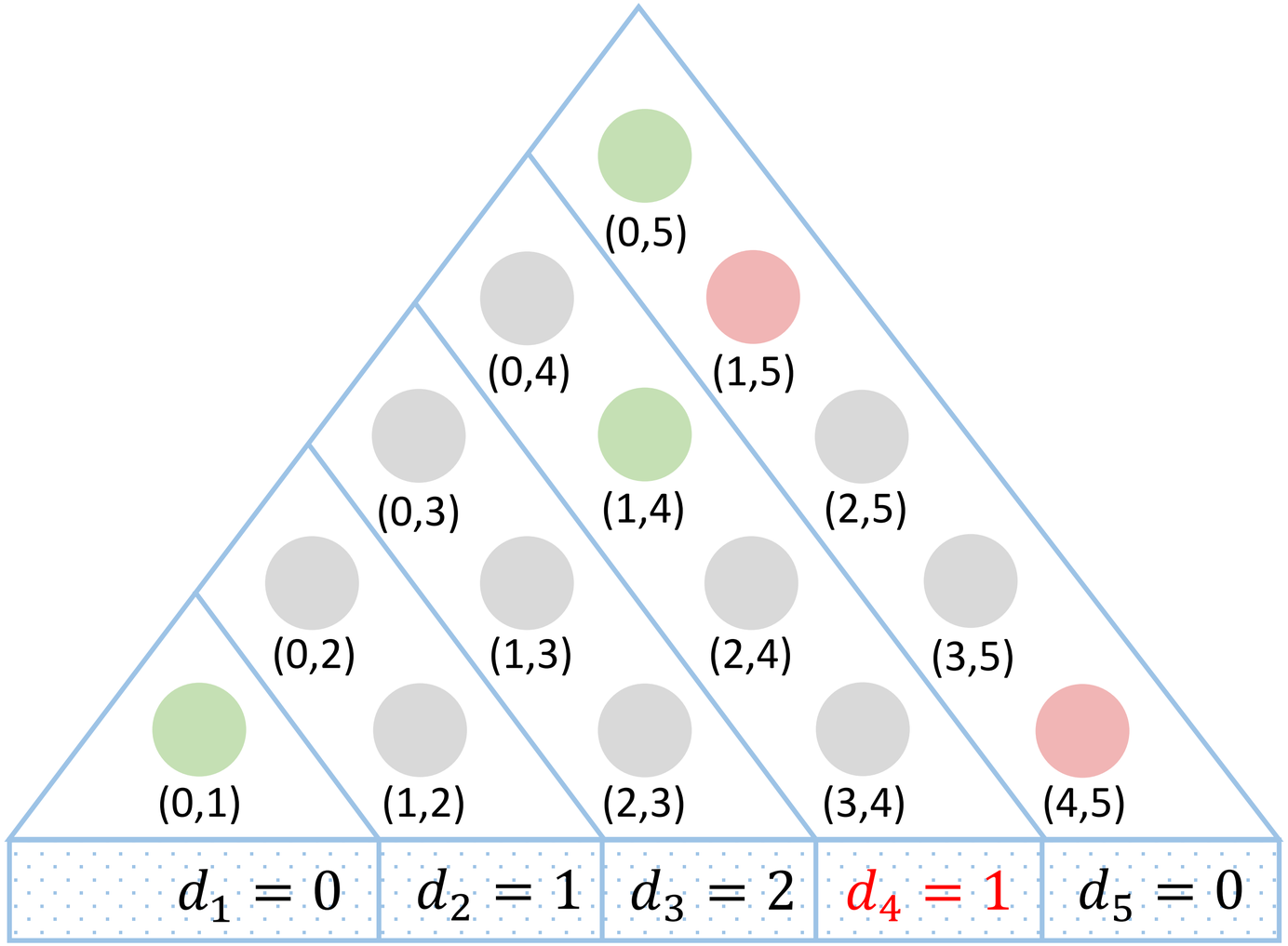
parent: (0, 5)

di: 0, 1, 2, 1

i: 1, 2, 3, 4

(0, 1) (1, 5)

Example



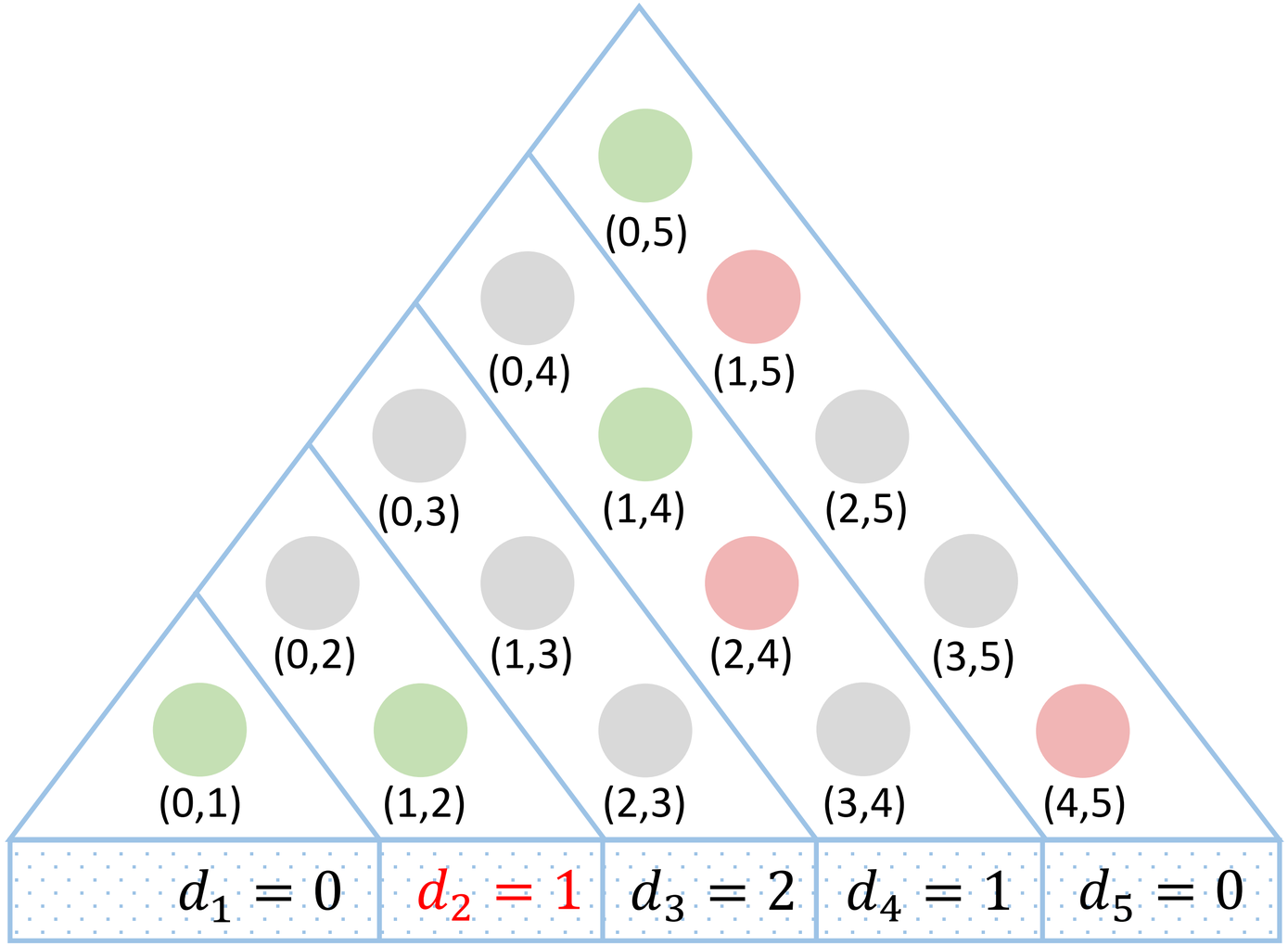
parent: (1, 5)

di: 0, 1, 2, 1

i: 1, 2, 3, 4

(1, 4) (4, 5)

Example



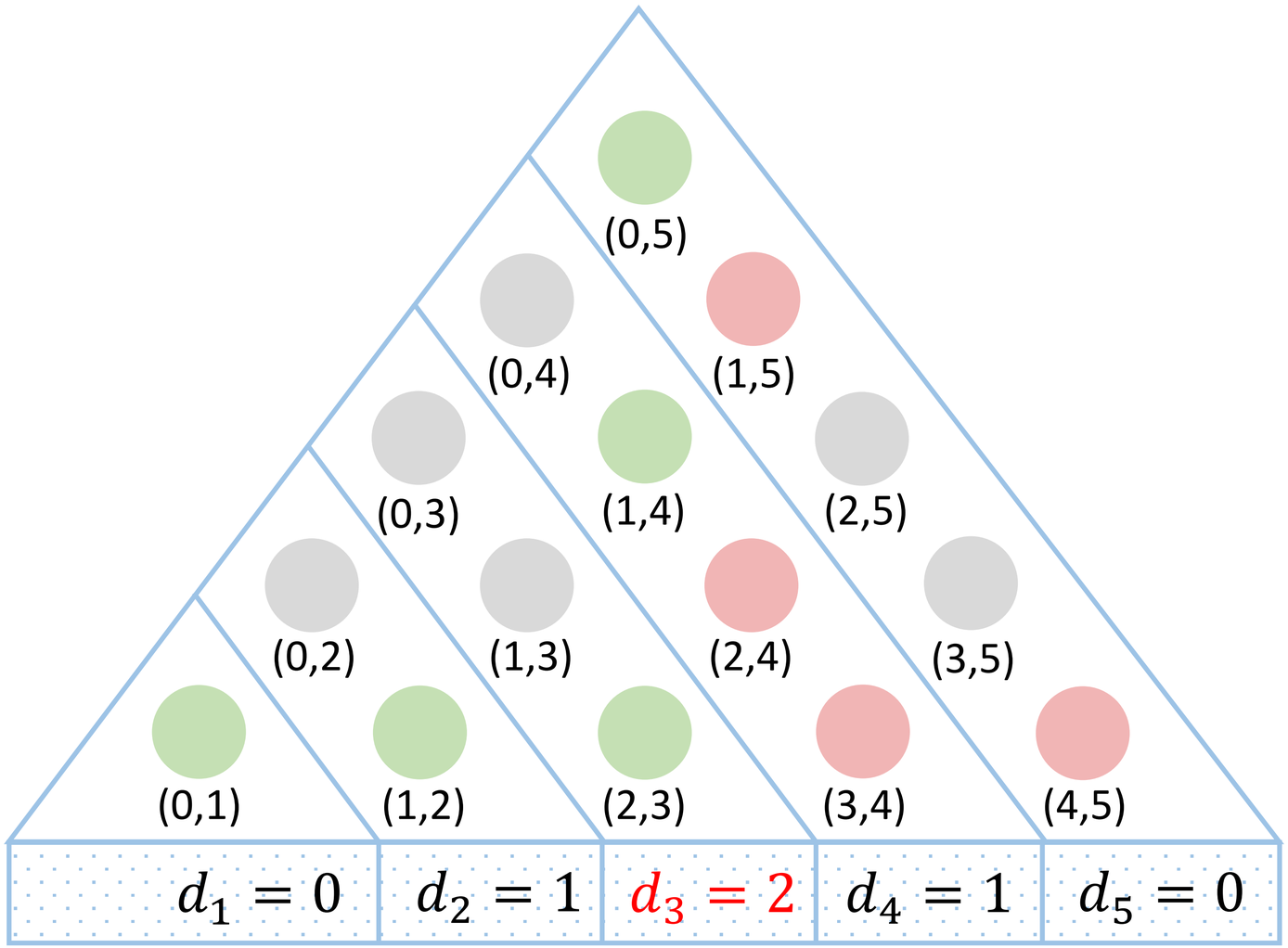
parent: (1, 4)

di: 0, 1, 2, 1

i: 1, 2, 3, 4

(1, 2) (2, 4)

Example



parent: (2, 4)

di: 0, 1, 2, 1

i: 1, 2, 3, 4

(2, 3) (3, 4)

Parser

- Encoder

$$\boldsymbol{x}_i = [\boldsymbol{e}_i; \boldsymbol{c}_i; \boldsymbol{p}_i].$$

$$\boldsymbol{h}_i = [\overrightarrow{\boldsymbol{h}}_i; \overleftarrow{\boldsymbol{h}}_{i+1}].$$

- Decoder

$$\boldsymbol{l}_i = \text{MLP}_l(\boldsymbol{h}_i), \quad \boldsymbol{r}_i = \text{MLP}_r(\boldsymbol{h}_i).$$

$$\alpha_{ij} = \boldsymbol{l}_i^\top \mathbf{W} \boldsymbol{r}_j + \boldsymbol{b}_1^\top \boldsymbol{l}_i + \boldsymbol{b}_2^\top \boldsymbol{r}_j,$$

$$P(i|j) = \text{Softmax}_i(\alpha_{ij}), \forall i < j.$$

$$d_j = \arg\max_i P(i|j), \forall i < j.$$

$$P(\ell|i,j) = \text{Softmax}(\text{MLP}_{\text{label}}([\boldsymbol{l}_i; \boldsymbol{r}_j]))_\ell.$$

Parser

- Training Objective

$$\mathcal{L} = -\frac{1}{n}(\sum_{i=1}^{n} \log P(d_i|i) + \sum_{(i,j,\ell)\in\mathcal{T}} \log P(\ell|i,j)).$$

structural loss

label loss

Tree Inference

- CKY-style decoding

$$\mathcal{G}(i,j) = \max\{P(i|k) \times \mathcal{G}(k,j) \mid i < k < j\},$$

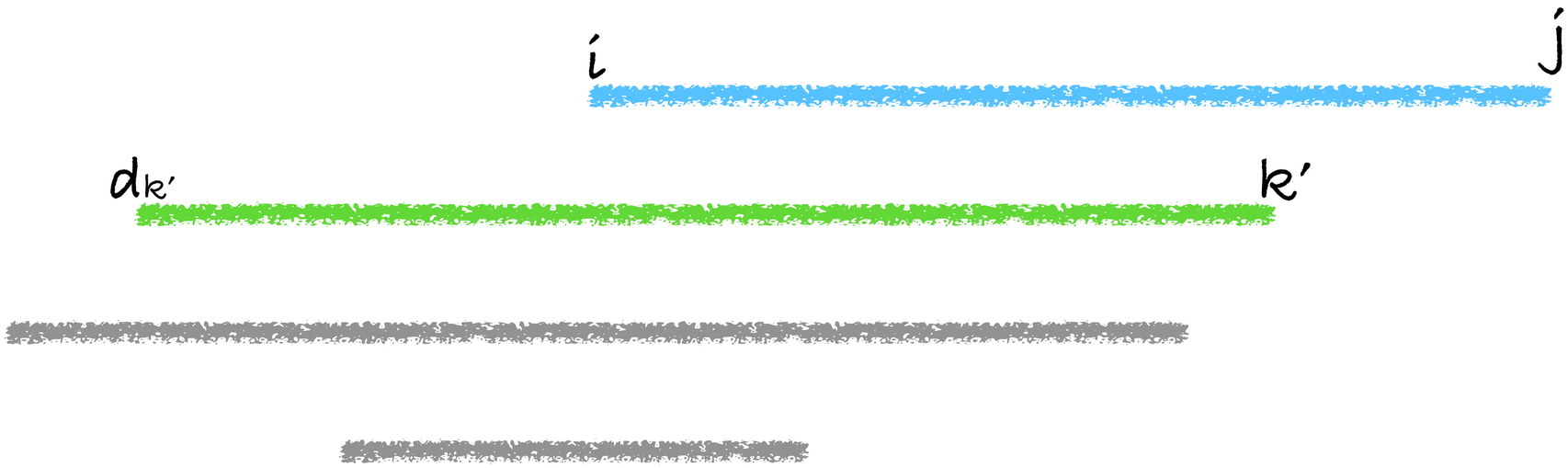
- Greedy decoding

$$k \leftarrow \max\{k' \mid d_{k'} \leq i, i < k' < j\}.$$

$$k \leftarrow \arg\min_{k'} d_{k'}.$$

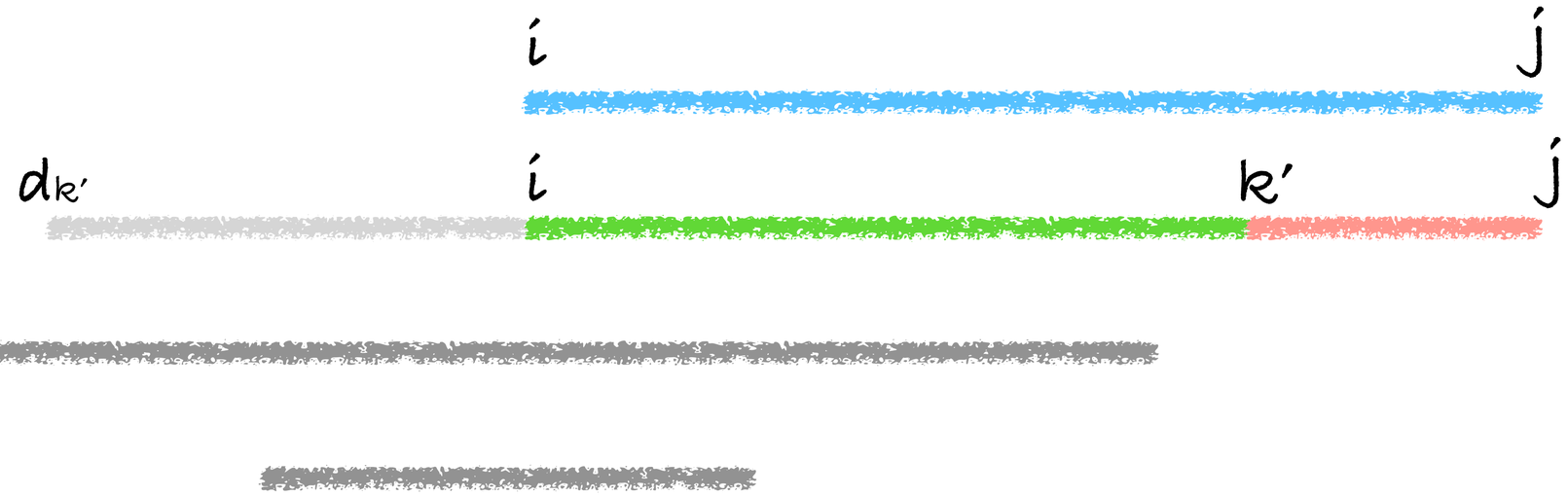
Greedy Decoding

$$k \leftarrow \max\{k' \mid d_{k'} \leq i, i < k' < j\}.$$



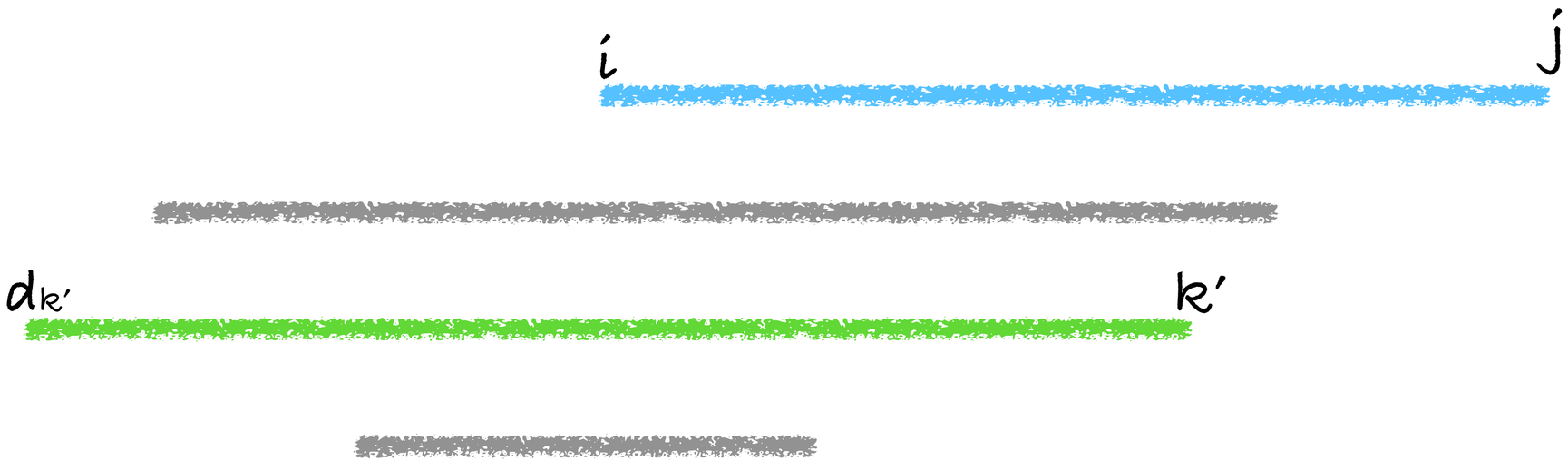
Greedy Decoding

$$k \leftarrow \max\{k' \mid d_{k'} \leq i, i < k' < j\}.$$



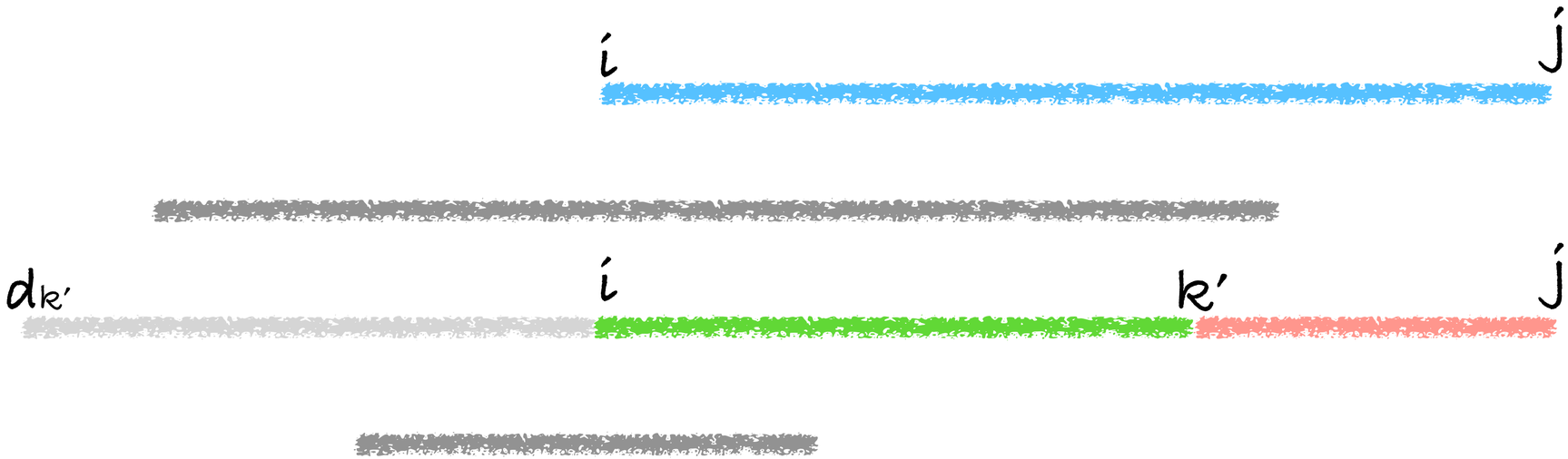
Greedy Decoding

$$k \leftarrow \arg\min_{k'} d_{k'}.$$

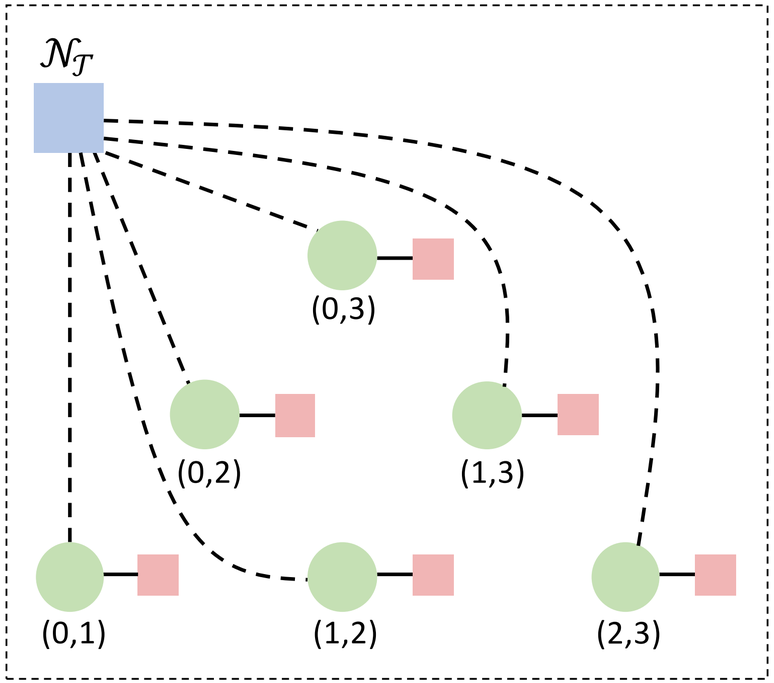


Greedy Decoding

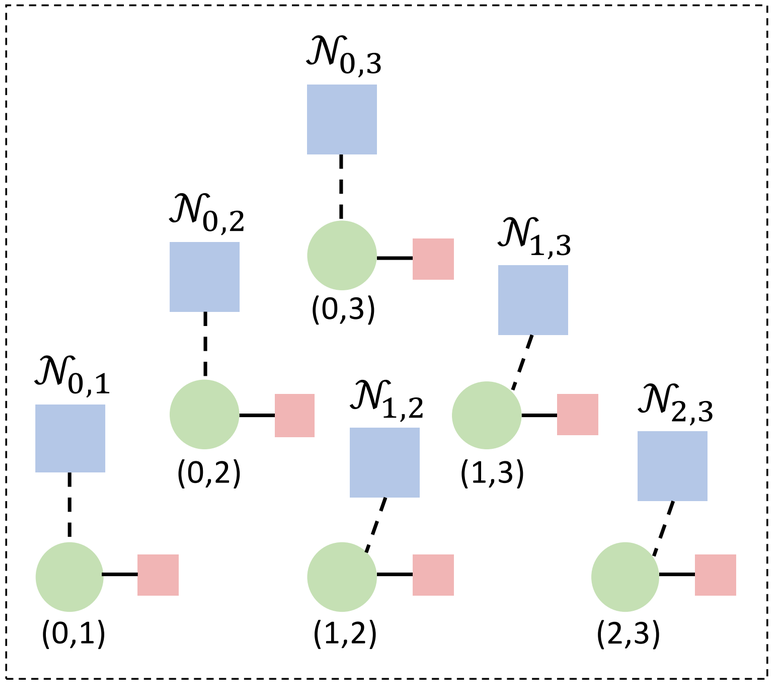
$$k \leftarrow \arg\min_{k'} d_{k'}.$$



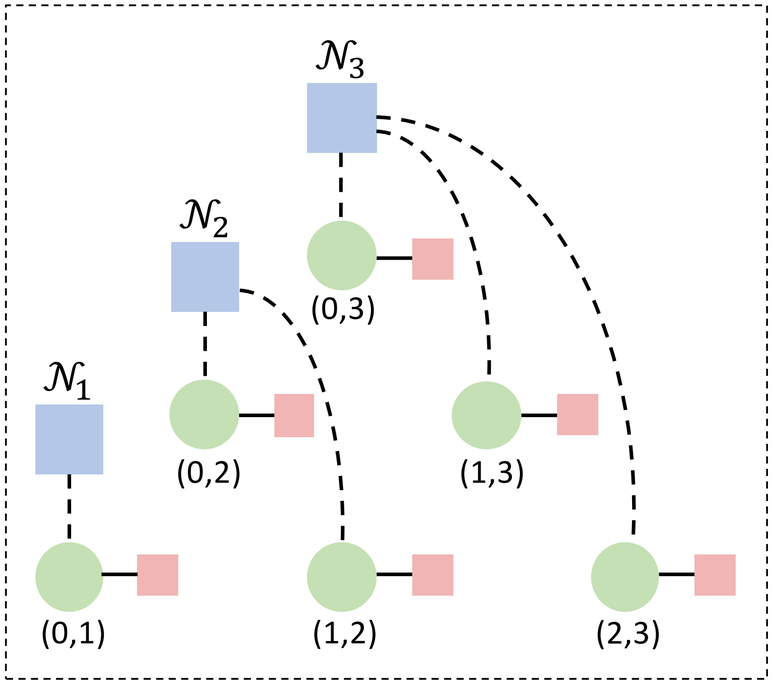
Normalization



(a) Global normalization.



(b) Local normalization.



(c) Our normalization.

Results on PTB

Model	LR	LP	F1
Global Model			
Stern et al. (2017a)	90.6	93.0	91.8
Gaddy et al. (2018)	91.8	92.4	92.1
Kitaev and Klein (2018a)♠	93.2	**93.9**	93.6
Zhou and Zhao (2019)♠	**93.6**	**93.9**	**93.8**
Local Model			
Vilares et al. (2019)	-	-	90.6
Liu et al. (2018)	-	-	91.2
Ma et al. (2017)	-	-	91.5
Shen et al. (2018)	91.7	92.0	91.8
Liu and Zhang (2017a)	-	-	91.8
Hong and Huang (2018)	91.5	92.5	92.0
Teng and Zhang (2018)	92.2	92.5	92.4
Dyer et al. (2016)♡	-	-	92.4
Stern et al. (2017b)♡	92.6	92.6	92.6
Our Model	92.3	92.9	92.6
Our Model♠	**93.3**	**94.1**	**93.7**
Pre-training/Ensemble/Re-ranking			
Liu et al. (2018)	-	-	92.3
Choe and Charniak (2016)	-	-	93.8
Liu and Zhang (2017a)	-	-	94.2
Fried et al. (2017)	-	-	94.7
Kitaev and Klein (2018a)♠	94.9	95.4	95.1
Kitaev and Klein (2018b)♠	95.5	95.7	95.6
Zhou and Zhao (2019)♠	**95.7**	96.0	**95.8**
Our Model (+BERT)	95.6	96.0	**95.8**
Our Model (+BERT)♠	95.5	**96.1**	**95.8**

Parsing Speeds

Model	sents/sec
Global Model	
Stern et al. (2017a)	20
Kitaev and Klein (2018a)♠ (w. `Cython`)	150
Zhou and Zhao (2019)♠ (w. `Cython`)	159
Local Model	
Teng and Zhang (2018)	22
Stern et al. (2017a)	76
Liu and Zhang (2017b)	79
Shen et al. (2018)	111
Shen et al. (2018) (w/o tree inference)	351
Vilares et al. (2019)	942
Our Model	220
Our Model♠	155

In the Future

- Linguistic interpretation.
- Combine the two different linearization methods (left and right).
- Apply GAT to enhance the representation, similar to [Ji et al., 2019].

Contact

WeChat 

Homepage antnlp.org

Blog godweiyang.com