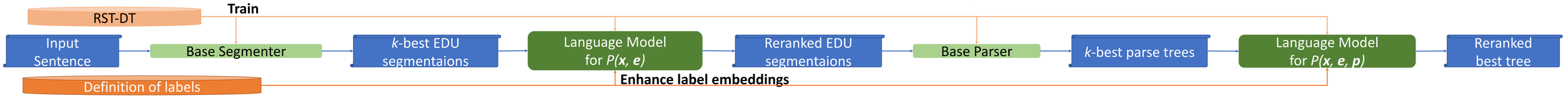
A Language Model-based Generative Classifier for Sentence-level Discourse Parsing

東京工業大学
Tokyo Institute of Technology

Ying Zhang Hidetaka Kamigaito Manabu Okumura

{ zhang, kamigaito, oku } @ lr.pi.titech.ac.jp

Language Model-based Generative Classifier (LMGC) for the sentence-level parsing w/ auto segmentation



Motivation

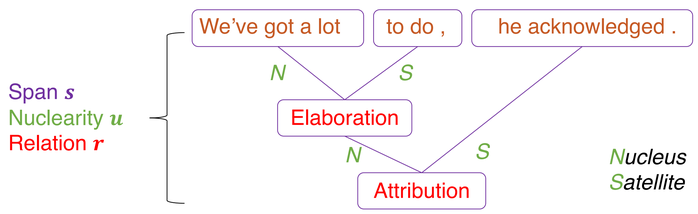
Discourse segmentation : detect elementary discourse units (EDUs) boundaries in a given text.

Text x

We've got a lot to do, he acknowledged.

EDUs e: We've got a lot (e_1) | to do , (e_2) | he acknowledged . (e_3)

Discourse parsing : link detected EDUs.



Most prior work is on the basis of discriminative models $P(\boldsymbol{y}|\boldsymbol{x})$, which cannot consider mapping from predictable labels $\boldsymbol{y}$ to input texts $\boldsymbol{x}$ to exploit more label information.

Experiments

RST Discourse Treebank (RST-DT) corpus

Task	Train	Valid	Test
(a) Segmentation	6,768	905	991
(b) Sentence-level parsing w/ gold segmentation	4,524	636	602
(c) Sentence-level Parsing w/ auto segmentation	-	861	951

Results for the sentence-level parsing w/ auto segmentation.

Model	Seg	Parse		
		Span	Nuclearity	Relation
Pointer-networks*	-	91.75	86.38	77.52
Oracle$_{seg}$	98.24	-	-	-
Base segmenter	93.92	-	-	-
GPT2LM$_e$	95.03	-	-	-
LMGC$_e$	96.51	-	-	-
Enhance$_e$	**96.79**	-	-	-
Extend$_e$	96.48	-	-	-
Oracle	-	93.95	91.25	85.93
Base parser	-	93.53	88.08	78.75
GPT2LM$_r$	-	92.02	84.20	74.49
LMGC$_s$	-	93.96‡	88.46	79.25
Enhance$_s$	-	**94.00**†	88.50	79.33
LMGC$_u$	-	93.96†	**89.90**†	80.33†
Enhance$_u$	-	93.92‡	89.74†	80.22†
LMGC$_r$	-	93.65	89.08†	80.57†
Enhance$_r$	-	93.73	89.16†	**81.18**†

†, ‡ : the F_1 score is significantly superior to the base parser with a p-value < 0.01 and < 0.05, respectively.

Results for the segmentation.

Model	Precision	Recall	F_1
Oracle	97.73	98.67	98.20
Pointer-networks*	93.34	97.88	95.55
Base segmenter	92.22	95.35	93.76
GPT2LM$_e$	94.05	95.72	94.88
LMGC$_e$	95.31	97.56	96.43†
Enhance$_e$	**95.54**	**97.93**	**96.72**†
Extend$_e$	95.05	97.86	96.44†

† : the F_1 score is significantly superior to GPT2LM with a p-value < 0.01.

Results for the sentence-level parsing w/ gold segmentation.

Model	Span	Nuclearity	Relation
Oracle	98.67	95.88	90.07
Pointer-networks*	97.44	91.34	81.70
Base parser	97.92	92.07	82.06
GPT2LM$_r$	96.35	88.11	77.86
LMGC$_s$	98.23‡	92.31	82.22
Enhance$_s$	98.27‡	92.39	82.42
LMGC$_u$	**98.31**‡	**94.00**†	83.63†
Enhance$_u$	**98.31**†	93.88†	83.56†
LMGC$_r$	98.00	93.09†	83.99†
Enhance$_r$	98.12	93.13†	**84.69**†

†, ‡ : the F_1 score is significantly superior to the base parser with a p-value < 0.01 and < 0.05, respectively.

Methods

1. Linearize Tree $\boldsymbol{z} = (z_1, \ldots, z_a) = (\boldsymbol{x}, \boldsymbol{y})$

(Attribution (Elaboration (N (N *We've got a lot*)$_N$ (S *to do ,*)$_S$)$_N$)$_{Elaboration}$ (S *he acknowledged .*)$_S$)$_{Attribution}$

- with EDU boundary labels $(\boldsymbol{x}, \boldsymbol{e})$

  e_1 [EDU] e_2 [EDU] e_3 [EDU]

- with span labels $(\boldsymbol{x}, \boldsymbol{e}, \boldsymbol{s})$

  (Span (Span e_1)$_{Span}$ (Span e_2)$_{Span}$)$_{Span}$ (Span e_3)$_{Span}$

- with nuclearity labels $(\boldsymbol{x}, \boldsymbol{e}, \boldsymbol{u})$

  (N (N e_1)$_N$ (S e_2)$_S$)$_N$ (S e_3)$_S$

- with relation labels $(\boldsymbol{x}, \boldsymbol{e}, \boldsymbol{r})$

  (Span (Span e_1)$_{Span}$ (Elaboration e_2)$_{Elaboration}$)$_{Span}$ (Attribution e_3)$_{Attribution}$

2. Joint Probabilities

$$\log P(\boldsymbol{z}; \theta) \approx PLL(\boldsymbol{z}; \theta) \approx \sum_{t=1}^{a} \log P(z_t | z_{<t}, z_{>t}, M_t; \theta)$$

M_t: [MASK]

Joint sequence $\boldsymbol{z}$: *We* ? *got a lot* [EDU] *to do ,* [EDU] *he acknowledged .* [EDU]

3. Label Embedding

elementary discourse units are the minimal building blocks of a discourse tree

Definition of [EDU]

- *LMGC – Enhance* :
  Embedding of [EDU] = Ave (Embedding of elementary, Embedding of discourse, ... , Embedding of tree)
- *LMGC – Extend* :
  Joint sequence $\boldsymbol{z}$: *We've got a lot* [EDU] *to do ,* [EDU] *he acknowledged .* [EDU] [EDU] : elementary discourse units are the minimal building blocks of a discourse tree

4. Objective Function

all possible sequences $Z(x)$ ⊃ top-k sequences $Z'(x)$

$\boldsymbol{z}_g \in Z(x)$: correct joint sequence
O_a : all permutations of set $\{1, 2, \ldots, a\}$
c : the number of non-predicted tokens
θ: model parameter

For $\boldsymbol{z} \in Z'(x) \cup \{\boldsymbol{z}_g\}$, we maximize the expectation

$$\mathbb{E}_{o \in O_a} \sum_{t=c+1}^{a} [I_z \log P(z_{o_t} | z_{o<t}, M_{o>c}; \theta) + (1 - I_z) \log(1 - P(z_{o_t} | z_{o<t}, M_{o>c}; \theta))]$$

where I_z is the indicator function, defined as $I_z := \begin{cases} 1 & if\ \boldsymbol{z} = \boldsymbol{z}_g, \\ 0 & if\ \boldsymbol{z} \neq \boldsymbol{z}_g \end{cases}$