The 61st Annual Meeting of the Association for Computational Linguistics

Holographic CCG Parsing

Ryosuke Yamaki[1], Tadahiro Taniguchi[1], Daichi Mochihashi[2]

[1]Ritsumeikan University, [2]The Institute of Statistical Mathematics

RITSUMEIKAN

Research Organization of Information and Systems
The Institute of Statistical Mathematics

Motivation

➢ Explicit modeling of phrase structure

- Recent CCG supertagging and parsing models demonstrate high performance yet rely on **non-explicit modeling** of dependencies between words through neural networks.

⇩

☞ **Explicit modeling** of phrase structure with neural networks.

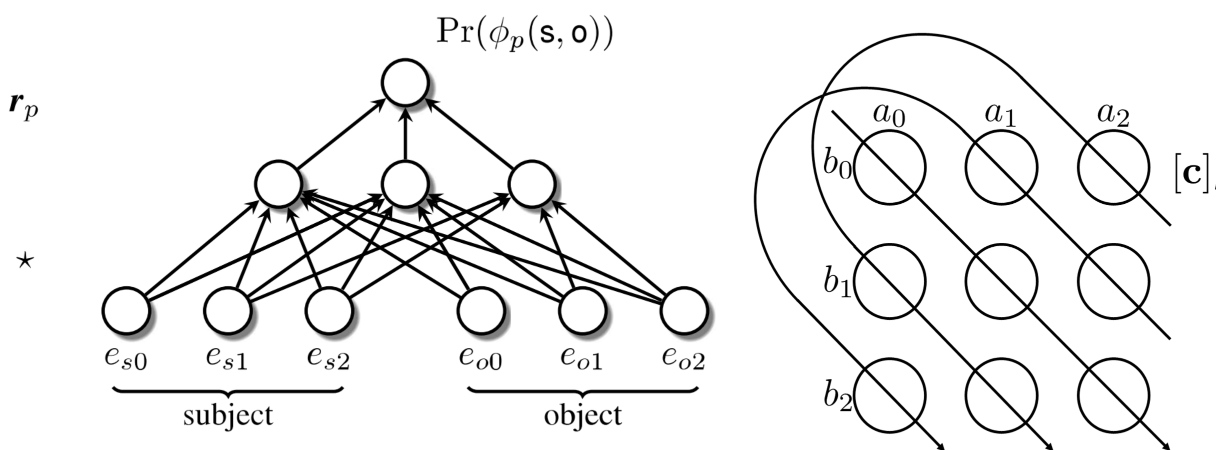
➢ Syntactic phrase-level representation

☞ Compose **syntactically rich phrase-level representations** while maintaining training efficiency.

Holographic CCG

➢ Holographic Embeddings (Nickel et al., 2016)

- Embedding knowledge graphs into vector space for statistical modeling
- Vector composition using **circular correlation** to capture dependencies between entities
- **Similarity of knowledge graphs and phrase structures that need to capture dependencies between components**



$$\mathbf{c} = \mathbf{a} \star \mathbf{b}$$

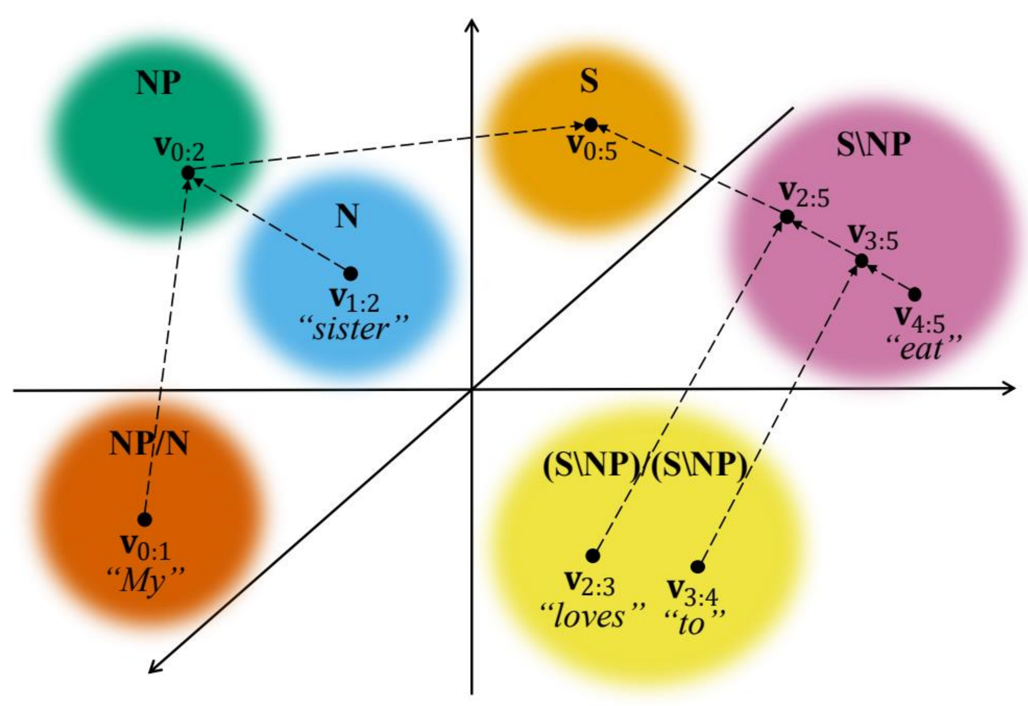
$$[\mathbf{c}]_k = [\mathbf{a} \star \mathbf{b}]_k = \sum_{i=0}^{d-1} a_i b_{(k+i)} \mod d$$

$$c_0 = a_0 b_0 + a_1 b_1 + a_2 b_2$$
$$c_1 = a_0 b_1 + a_1 b_2 + a_2 b_0$$
$$c_2 = a_0 b_2 + a_1 b_0 + a_2 b_1$$

(Nickel et al., 2016)

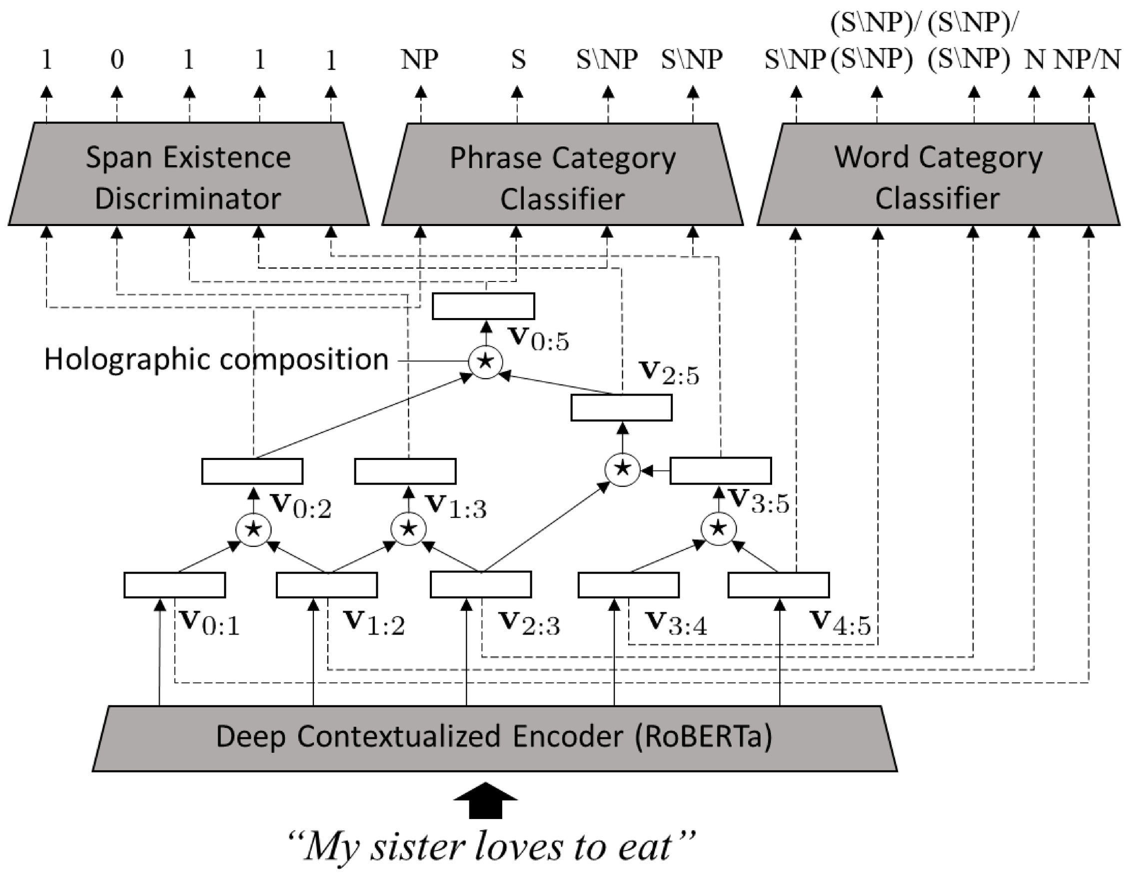
➢ Holographic CCG (Hol-CCG)

- Formulate CCG as a **recursive compositional operation** between distributed representations in a vector space.
- Applicable to **Supertagging** and **Span-based Parsing**.



➢ Model Structure

1. Encode word sequence into distributed representations.
2. Recursively compose phrase-level representations.
3. Predict CCG categories and span existence.



➢ Span-based Parsing

1. Store word-level representations.
2. Recursively compose phrase-level representations.
3. Directly evaluates category assignment to phrases.

Algorithm 1: Span-based CKY parsing

$\mathbf{v}_{0:1}, \mathbf{v}_{1:2}, \cdots, \mathbf{v}_{n-1:n} = Encode(w_1, w_2, \cdots, w_n)$;
for $i = 0, \cdots, n-1$ **do**
  $P_w(i, i+1) = SM(\mathbf{Q}_w \sigma(LN(\mathbf{U}_w \mathbf{v}_{i:i+1} + \mathbf{b}_w)) + \mathbf{c}_w)$; ▷ Equation (11)
  for $C \in \{X | P_w(i, i+1)[X] > t_w = 0.1\}$ **do**
    $prob[i, i+1, C] = \log P_w(i, i+1)[C]$;
    $vector[i, i+1, C] = \mathbf{v}_{i:i+1}$;
for $\ell = 2, \cdots, n$ **do**
  for $i = 0, \cdots, n-\ell$ **do**
    $j = i + \ell$;
    for $k = i+1, \cdots, j-1$ **do**
      for $C_1 \in \{X | prob[i, k, X] > 0\}$ **do**
        $\mathbf{v}_{i:k} = vector[i, k, C_1]$;
        for $C_2 \in \{X | prob[k, j, X] > 0\}$ **do**
          $\mathbf{v}_{k:j} = vector[k, j, C_2]$;
          for $C \in \{X | C_1 C_2 \to X \in R\}$ **do**
            $\mathbf{v}_{i:j} = \mathbf{v}_{i:k} \star \mathbf{v}_{k:j}$; ▷ Equations (4) and (5)
            $P_s(i, j) = SM(\mathbf{Q}_s \sigma(LN(\mathbf{U}_s \mathbf{v}_{i:j} + \mathbf{b}_s)) + \mathbf{c}_s)$; ▷ Equation (13)
            if $P_s(i, j)[e] > t_s = 0.01$ **then**
              $P_p(i, j) = SM(\mathbf{Q}_p \sigma(LN(\mathbf{U}_p \mathbf{v}_{i:j} + \mathbf{b}_p)) + \mathbf{c}_p)$; ▷ Equation (12)
              if $P_p(i, j)[C] > t_p = 0.01$ **then**
                $p = \log P_p(i, j)[C] + \log P_s(i, j)[e] + prob[i, k, C_1] + prob[k, j, C_2]$;
                if $p > prob[i, j, C]$ **then**
                  $prob[i, j, C] = p$;
                  $backpointer[i, j, C] = (k, C_1, C_2)$;
                  $vector[i, j, C] = \mathbf{v}_{i:j}$;

Experiment

- Dataset: CCGbank (Hockenmaier et al., 2007)
- Calculate the model's prediction error by cross entropy
  - ➢ Category assignment to words and phrases: $\mathcal{L}_w, \mathcal{L}_p$
  - ➢ Existence of span: $\mathcal{L}_s$
- Compare models by changing the combination of back-propagating errors ($\mathcal{L}$)
  - ➢ Baseline: $\mathcal{L} = \mathcal{L}_w$
  - ➢ Hol-CCG: $\mathcal{L} = \mathcal{L}_w + \mathcal{L}_p + \mathcal{L}_s$
- Supertagging by Baseline and Hol-CCG
- Parsing using C&C Parser (Clark and Curran, 2007) and Hol-CCG's span-based parsing

Result & Discussion

- Hol-CCG outperforms baseline.
- Span-based Parsing outperforms C&C Parser.
- **Explicit modeling of word/phrase dependencies** through composition of phrase representations is effective for both supertagging and parsing.

Training Objectives	Parser	Acc	LF
$\mathcal{L}_w$(baseline)	C&C	96.41 ± 0.03	91.77 ± 0.03
$\mathcal{L}_w + \mathcal{L}_p + \mathcal{L}_s$(Hol-CCG)	C&C	$\underline{\mathbf{96.59 \pm 0.02}}$	$\underline{\mathbf{92.03 \pm 0.04}}$
	Span	-	$\underline{\mathbf{92.61 \pm 0.03}}$

- Hol-CCG achieved SoTA in supertagging accuracy and LF with C&C Parser.
- Hol-CCG's span-based parsing is competitive with current SoTA performance.

Model	Parser	Acc	LF
Vaswani et al., 2016	C&C	94.5	88.32
Bhargava and Penn, 2020	C&C	96.00	90.9
Prange et al., 2021	C&C	96.22	90.91
Clark, 2021	C&C	-	91.9
	Span	-	<u>**92.9**</u>
Hol-CCG	C&C	<u>**96.60**</u>	<u>**92.12**</u>
	Span	-	92.67

Nickel, Maximilian, Lorenzo Rosasco, and Tomaso Poggio. "Holographic embeddings of knowledge graphs." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 30. No. 1. 2016.
Julia Hockenmaier, and Mark Steedman. "CCGbank: a corpus of CCG derivations and dependency structures extracted from the Penn Treebank." Computational Linguistics 33.3 (2007): 355–396.
Stephen Clark, and James R. Curran. "Wide-coverage efficient statistical parsing with CCG and log-linear models." Computational Linguistics 33.4 (2007): 493–552.
Vaswani, Ashish, et al. "Supertagging with lstms." Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016.
Bhargava, Aditya, and Gerald Penn. "Supertagging with CCG primitives." Proceedings of the 5th Workshop on Representation Learning for NLP. 2020.
Prange, Jakob, Nathan Schneider, and Vivek Srikumar. "Supertagging the long tail with tree-structured decoding of complex categories." Transactions of the Association for Computational Linguistics 9 (2021): 243–260.
Clark, Stephen. "Something old, something new: Grammar-based CCG parsing with transformer models." arXiv preprint arXiv:2109.10044 (2021).