School of Information Science and Technology
ShanghaiTech University

Structured Mean-Field Variational Inference for Higher-Order Span-Based Semantic Role Labeling

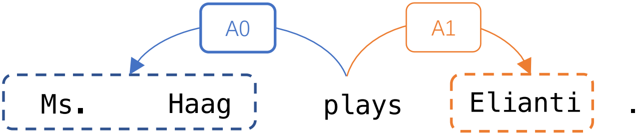
Wei Liu   Songlin Yang   Kewei Tu

July 5, 2023

Agenda

Dependency Graph Parsing & Latent Structure

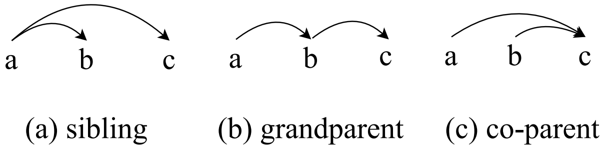
- Many natural language processing (NLP) tasks can be formulated as **dependency graph parsing** problems such as **span-based semantic role labeling (SRL)**.
- Spans are often constituents in a constituency tree in certain NLP tasks, including **span-based semantic role labeling (SRL)**, which can be considered as a type of **latent structure**.



Span-SRL

CRF & Higher-Order Inference

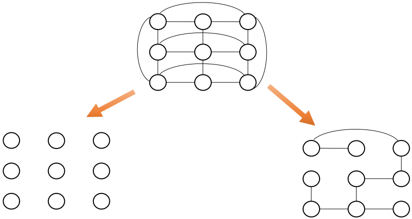
- We use a **conditional random field (CRF)** based dependency parser to predict all predicate-argument pairs.
- **Higher-order inference** has been widely used in CRF-based dependency parsers, but it is intractable in a general dependency graph.



Three kinds of higher-order relations

CRF & Higher-Order Inference

- **Mean-Field Variational Inference (MFVI)** is a commonly used approximation algorithm that enables higher-order inference in a general graph by constraining the original distribution to a **subset of distribution**.



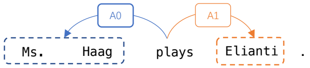
Mean-Field Variational Inference (MFVI)

Decomposed Dependency Graph

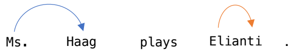
- We focus on **second-order inference**.
- Performing second-order inference in a sentence requires the enumeration of two spans and one token (the predicate), which results in a complexity of $O(n^5)$.

Decomposed Dependency Graph

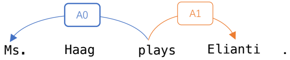
- Therefore, we decompose the dependency graph into three dependency edges named **predicate-to-head (P2H)**, **predicate-to-tail (P2T)** and **head-to-tail (H2T)**.
- Reducing complexity for second-order infernce to $O(n^3)$.



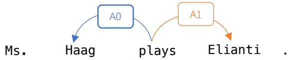
(a) Original span-based SRL



(b) H2T edge after decomposition



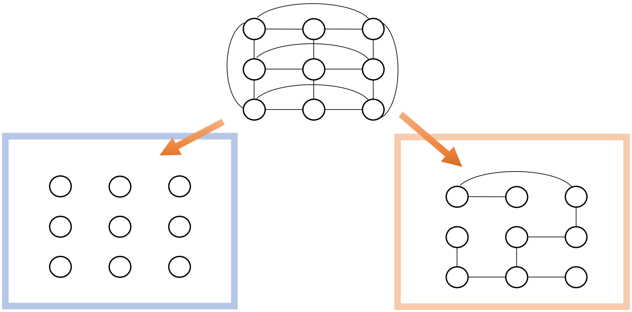
(c) P2H edge after decomposition



(d) P2T edge after decomposition

Structured Mean-Field Variational Inference (Structured-MFVI)

- Previous works often treat a subset of distribution as fully factorized, disregarding the rich structural information, such as the aforementioned tree structure, present within sentences.



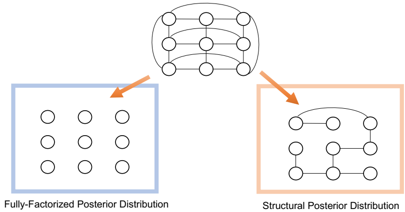
Fully-Factorized Posterior Distribution

Structural Posterior Distribution

Structured-MFVI

Structured Mean-Field Variational Inference (Structured-MFVI)

- We propose a new algorithm for **structured mean-field variational inference (Structured-MFVI)** that uses a partially observed **TreeCRF (PO-TreeCRF)** distribution instead of the over-simplified **fully factorized distribution**.
- **The H2T edge** allows structural information to propagate to the other two edges via structured mean-field variational inference (Structured-MFVI).
- Please refer to our paper for detailed information about our algorithm (Structured-MFVI).



Structured-MFVI

Main Results

	CoNLL05-WSJ			CoNLL05-Brown			CoNLL12		
	P	R	F1	P	R	F1	P	R	F1
				w/o gold predicates					
He et al, 2017	80.20	82.30	81.20	67.60	69.60	68.50	78.60	75.10	76.80
He et al, 2018 + ELMO	84.80	87.20	86.0	73.90	78.40	76.10	81.90	84.00	82.90
Jia et al, 2022 + BERT	–	–	86.70	–	–	78.58	–	–	84.22
Zhou et al, 2022 + BERT	87.15	88.44	87.79	79.44	80.85	80.14	83.91	85.61	84.75
zhang et al, 2022 + BERT	87.00	**88.76**	87.87	79.08	81.50	80.27	84.53	**86.41**	**85.45**
1O + BERT	87.11	87.40	87.25	79.89	79.93	79.91	84.76	84.42	84.59
Ours + BERT	**88.05**	88.61	**88.33**	**81.13**	**81.58**	**81.36**	**84.95**	85.85	85.40
				w/ gold predicates					
He et al, 2017	85.00	84.30	84.60	74.90	72.40	73.60	83.50	83.30	83.40
He et al, 2018 + ELMO	–	–	87.40	–	–	80.40	–	–	85.50
Shi and Lin, 2019 + BERT	88.60	**89.00**	88.80	81.90	**82.10**	82.00	85.90	87.00	86.50
Conia and Navigli, 2020 + BERT	–	–	–	–	–	–	86.90	87.70	87.30
Blloshmi et al, 2021 + BART	–	–	–	–	–	–	87.80	86.80	87.30
Liu et al, 2022 + SpanBERT	–	–	–	–	–	–	–	–	87.50
Jia et al, 2022 + BERT	–	–	88.25	–	–	81.90	–	–	87.18
Zhou et al, 2022 + BERT	89.03	88.53	88.78	83.22	81.81	82.51	87.26	87.05	87.15
Zhang et al, 2022 + BERT	89.00	**89.03**	89.02	82.81	82.35	82.58	87.52	**87.79**	87.66
1O + BERT	89.09	87.57	88.32	83.30	79.49	81.35	87.45	86.75	87.10
Ours + BERT	**89.77**	88.46	**89.11**	**83.96**	**81.76**	**82.85**	**88.10**	87.38	**87.74**

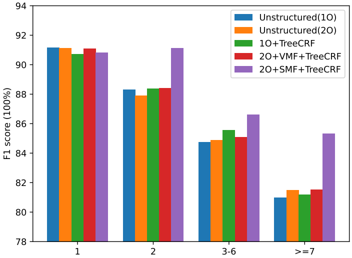
Comparison of our model and other models on test sets of CoNLL05-WSJ, CoNLL05-Brown, and CoNLL12.

Ablations

Model	P	R	F1
Unstructured(1O)	87.11	87.40	87.25
Unstructured(2O)	87.21	88.34	87.77
$1O_{+TreeCRF}$	87.79	87.57	87.68
$2O_{VMF+TreeCRF}$	87.53	88.26	87.90
$2O_{SMF+TreeCRF}$ (Final)	**88.05**	**88.61**	**88.33**

Ablation studies on CoNLL05-WSJ dataset. VMF indicates vanilla mean-field and SMF indicates structured mean-filed.

Ablations



F1 score regarding to different argument span length. The x-axis denotes the length of argument spans. The y-axis denotes the F1 score.

Thank you for your listening!