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

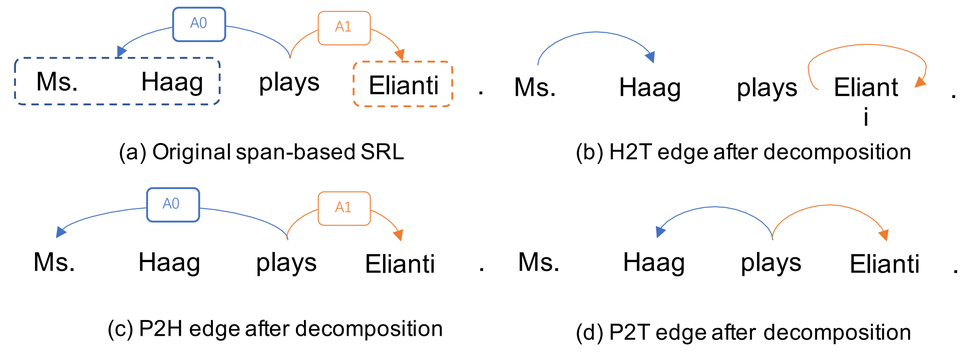
Wei Liu, Songlin Yang, Kewei Tu*
School of Information Science and Technology, ShanghaiTech University
Shanghai Engineering Research Center of Intelligent Vision and Imaging

Highlights

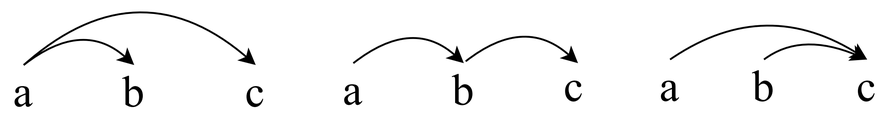
- High-order Inference. Many natural language processing (NLP) tasks can be formulated as dependency graph parsing problems such as span-based semantic role labeling (SRL). One commonly used method to enhance dependency graph parsers is named High-order Inference.

- Latent Structure. 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.

- Desiderata. How do we combine the best of two worlds? Let's think step by step 🤔😄



Span-based SRL & Decomposed Dependency Parsing Graph



(a) sibling (b) grandparent (c) co-parent

We focus on second-order modeling in this work. Three kinds of high-order modeling we adopt.

👣The First Step: Decomposition & Mean-Field Variational Inference (MFVI)

- Model. We use **a conditional random field (CRF)** based dependency parser to predict all predicate-argument pairs.

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

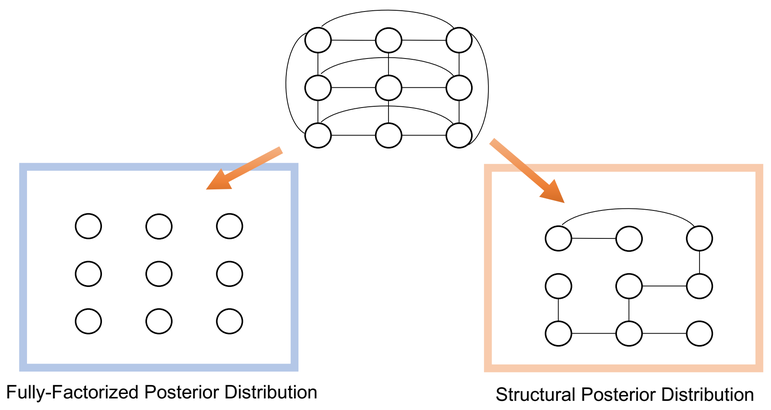
- Decomposition 😌. Therefore, we decompose the dependency graph into three dependency edges named **predicate-to-head (P2H)**, **predicate-to-tail (P2T)** and **head-to-tail (H2T)**.

- Intractability 😵. However, performing high-order inference in a general graph is intractable!

- Mean-Field Variational Inference (MFVI) 😌. **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**.

👣The Second Step: Structured Mean-Field Variational Inference (MFVI)

- Strong Independent Assumption 😵. 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.
- Structured Mean-Field Variational Inference (MFVI) 😌. Structured mean-field variational inference (Structured-MFVI) that uses **a partially observed TreeCRF (PO-TreeCRF)** distribution instead of the over-simplified fully factorized distribution. Please refer to our paper for algorithm details.
- **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 more details.



Vanilla MFVI vs. Structural MFVI

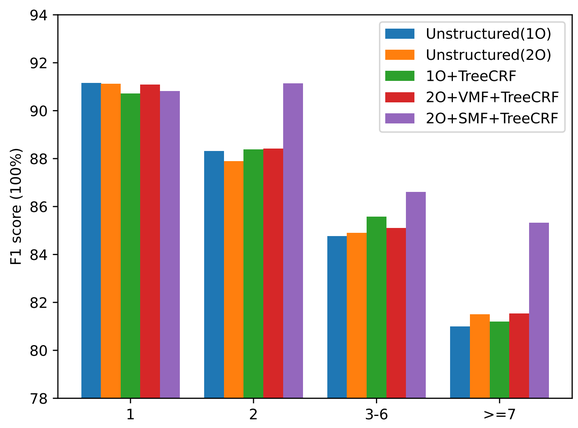
Experiment Results

- Our proposed model clearly outperforms the baseline, obtaining state-of-the-art performances (when using Bert-large-cased) on five out of six evaluation metrics

- Ablation studies for our Structured MFVI show the effectiveness of our algorithm

- Performance regarding to different argument span lengths demonstrates the superiority of our algorithm for long spans

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**



liuwei4@shanghaitech.edu.cn

https://github.com/VPeterV/Structured-MFVI

https://vpeterv.github.io/
https://sustcsonglin.github.io/
https://faculty.sist.shanghaitech.edu.cn/faculty/tukw/

上海科技大学 ShanghaiTech University