Frame Semantic Role Labeling Using Arbitrary-Order Conditional Random Fields

Chaoyi Ai, Kewei Tu*

School of Information Science and Technology, ShanghaiTech University
Shanghai Engineering Research Center of Intelligent Vision and Imaging
{aichy,tukw}@shanghaitech.edu.cn

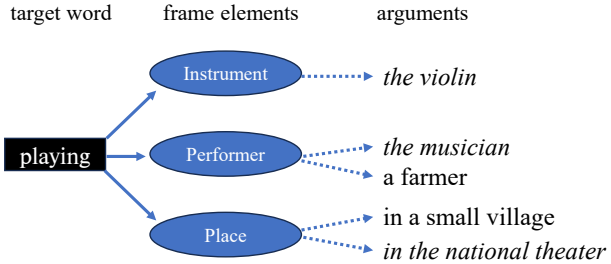
Abstract

This paper presents an approach to frame semantic role labeling (FSRL), a task in natural language processing that identifies semantic roles within a text following the theory of frame semantics. Unlike previous approaches which do not adequately model correlations and interactions amongst arguments, we propose arbitrary-order conditional random fields (CRFs) that are capable of modeling full interaction amongst an arbitrary number of arguments of a given predicate. To achieve tractable representation and inference, we apply canonical polyadic decomposition to the arbitrary-order factor in our proposed CRF and utilize mean-field variational inference for approximate inference. We further unfold our iterative inference procedure into a recurrent neural network that is connected to our neural encoder and scorer, enabling end-to-end training and inference. Finally, we also improve our model with several techniques such as span-based scoring and decoding. Our experiments show that our approach achieves state-of-the-art performance in FSRL.

Introduction

Frame semantic role labeling (FSRL) is a task grounded in the theory of frame semantics (Litkowski 2004), aiming to identify and assign semantic roles, known as frame elements (FEs), to the arguments of each predicate in a sentence. In this context, each predicate triggers a specific semantic frame defined with a unique set of frame elements. The FrameNet lexical resource (Baker, Fillmore, and Lowe 1998) provides comprehensive descriptions for frames and frame elements, which encapsulate a wide range of events, relationships, objects, and situations. By extracting frame semantic structures from text, FSRL serves as an invaluable tool for a multitude of downstream applications, including but not limited to information extraction (Surdeanu et al. 2003), summary generation (Trandabăț 2011), machine translation (Liu and Gildea 2010; Marcheggiani, Bastings, and Titov 2018), question answering (Shen and Lapata 2007; Eckert and Neves 2018; Khashabi et al. 2018), and reading comprehension (Wang et al. 2015; Guo et al. 2020).

Over the past several years, a series of approaches have been proposed for FSRL that reach increasingly high accuracy (Kshirsagar et al. 2015; Swayamdipta et al. 2017; Bastianelli, Vanzo, and Lemon 2020; Zheng et al. 2022; Zheng, Wang, and Chang 2023). However, most of these approaches still do not adequately model high-order correlations and interactions amongst arguments which are prevalent in FSRL. Take the sentence in Figure 1 for example: "*We watch the musician, raised in a small village by a farmer, playing the violin in the national theater.*" The frame for this sentence is denoted as **Performance** with the target word being "`playing`". In this example, each frame element presents multiple plausible candidate arguments. Determining the correct arguments by observing them in isolation, i.e., using first-order inference, proves challenging. Even taking pairwise interactions between arguments into consideration is still not enough: "a farmer" and "in a small village" are quite compatible as frame elements of "playing". To ascertain the correct frame semantic structure, it is helpful to evaluate the full combinations of all arguments through high-order inference. In our example, the combination "*the musician*", "*the violin*", and "*in the national theater*" as frame elements related to the target word "*playing*" appears more probable than others such as "*a farmer*", "*the violin*", and "*in a small village*". Recent researches in FSRL (Zheng et al. 2022; Zheng, Wang, and Chang 2023) have typically used first-order inference or only modeled pairwise argument interactions via GNN or cross-encoder. Consequently, FSRL exploring high-order modeling and inference amongst all arguments remains untapped.



We watch *the musician*, raised in a small village by a farmer, playing the violin *in the national theater*.

Figure 1: An example input sentence for FSRL. Italicized texts are the correct arguments.

* Corresponding Author

In this paper, we propose an FSRL approach with high-order inference based on arbitrary-order conditional random fields (CRFs). Specifically, we learn a different CRF for each frame which contains one variable for each frame element of the frame representing the corresponding argument and a single factor connecting to all the variables, thus modeling high-order interaction amongst all the arguments. Given that the number of frame elements can vary, the factor defined in the corresponding CRF may manifest as an arbitrary-order tensor. Unfortunately, the computational complexity of representing and performing inference on such a CRF grows exponentially with the frame element number. Therefore, we propose to decompose the factor using canonical polyadic decomposition (CPD) (Rabanser, Shchur, and Günnemann 2017), reducing the space complexity of the CRF to be linear in the frame element number. To facilitate tractable inference, we further employ mean-field variational inference (MFVI) (Xing, Jordan, and Russell 2012) for approximate inference on the CRF, which has a linear time complexity in the frame element number.

Drawing inspiration from previous research (Zheng et al. 2015), we unfold MFVI into a recurrent neural network, which interfaces seamlessly with our neural encoding and scoring modules, enabling end-to-end training and inference. Note that our approach to high-order inference diverges significantly from previous work in other fields such as dependency parsing (McDonald and Pereira 2006; Carreras 2007; Koo and Collins 2010; Wang, Huang, and Tu 2019; Wang and Tu 2020) which is limited to second-order or third-order inference modeling local interactions to avoid high computational complexity. In contrast, our approach models global interactions amongst all arguments and utilizes techniques such as CPD and MFVI to achieve tractability.

In addition to high-order inference, we also make several improvements to the encoding, scoring and decoding modules in comparison with previous state-of-the-art method (Zheng, Wang, and Chang 2023). For example, we propose a span-based method for scoring and decoding arguments, instead of separate scoring and greedy decoding of start and end positions of arguments, i.e., pointer network (Vinyals, Fortunato, and Jaitly 2015). We show that these improvements significantly increase the accuracy of our FSRL approach.

Our contributions can be summarized as follows:

- We address FSRL using arbitrary-order CRFs, which directly model full interactions amongst all arguments.
- We employ both CPD and MFVI to facilitate tractable representation and inference for the arbitrary-order CRFs.
- We also make several additional improvements to the encoding, scoring and decoding modules that are empirically beneficial.
- In our empirical evaluation, our approach delivers state-of-the-art performance on FSRL. [1]

[1]Code: https://github.com/aichy98/FrameSRL-AAAI24

Problem Definition

Frame semantic role labeling (FSRL) has a goal of identifying arguments of frame-evoking targets in a sentence and labeling them with frame elements. Consider a sentence denoted as $S = w_1, \ldots, w_n$ where a target word w_{tar} elicits a frame f. In the context of our study, both w_{tar} and f are provided. Denote the arguments for the target word w_{tar} by $a_1, \ldots, a_k$. The FSRL task is to pinpoint the start and end positions s_i and e_i for each argument $a_i = w_{s_i}, \ldots, w_{e_i}$. Subsequently, each argument a_i is assigned a semantic role $role_i$ which belongs to $\mathcal{R}_f$, the set of frame elements of frame f. Note that in the context of FSRL, roles and frame elements are used interchangeably. Alternatively, for frame f and all frame elements $\mathcal{R}_f = \{role_1, \ldots, role_m\}$, the FSRL task is to assign each frame element $role_i$ with a particular argument a_i within sentence S. It is important to note that there may be no corresponding span in S for $role_i$, in which case, $role_i$ is assigned with a null span. We also note that the second formulation of FSRL prohibits two or more arguments to have the same role. Nevertheless, multiple arguments sharing a role is infrequent (less than 0.5% in FN1.5 and FN1.7) and hence we adopt the second formulation in this paper.

Method

Our proposed model is depicted in Figure 2(a), incorporating three main components: a cross-encoder, a unary scorer, and an arbitrary-order scorer and decoder.

Feature Extraction via Cross-Encoder

Our encoder design follows that of the AGED model (Zheng, Wang, and Chang 2023), where frame definitions are treated as templates and their frame elements are considered as slots. First, we represent frame f with a textual description, as follows:

$$D_f = \texttt{frame name|raw def|FE list} \quad (1)$$

where `frame name` represents the name of frame f, `raw def` denotes the textual definition of the frame, and `FE list` is a collection of all frame elements associated with the frame excluding any frame elements that are already embedded in `raw def`. This revision was made in response to the observation that some frame elements, particularly non-core ones, are not always explicitly referred to in the frame definitions. D_f is designed in a way to ensure an exhaustive representation of frame, where each frame element is coupled with a unique slot in the description. When a frame element is mentioned multiple times in the description, we consider the leftmost occurrence as its slot. Figure 3 shows an example.

We then enter text S and description D_f into a pretrained language model (PLM) using the following structure:

$$\texttt{[CLS]}\ S\ \texttt{[SEP]}\ D_f\ \texttt{[SEP]} \quad (2)$$

The PLM serves as a cross-encoder, generating contextualized representation for each token within the text and description. Cross-encoding facilitates the learning of alignments between arguments present in the text and frame elements in the description by leveraging the self-attention

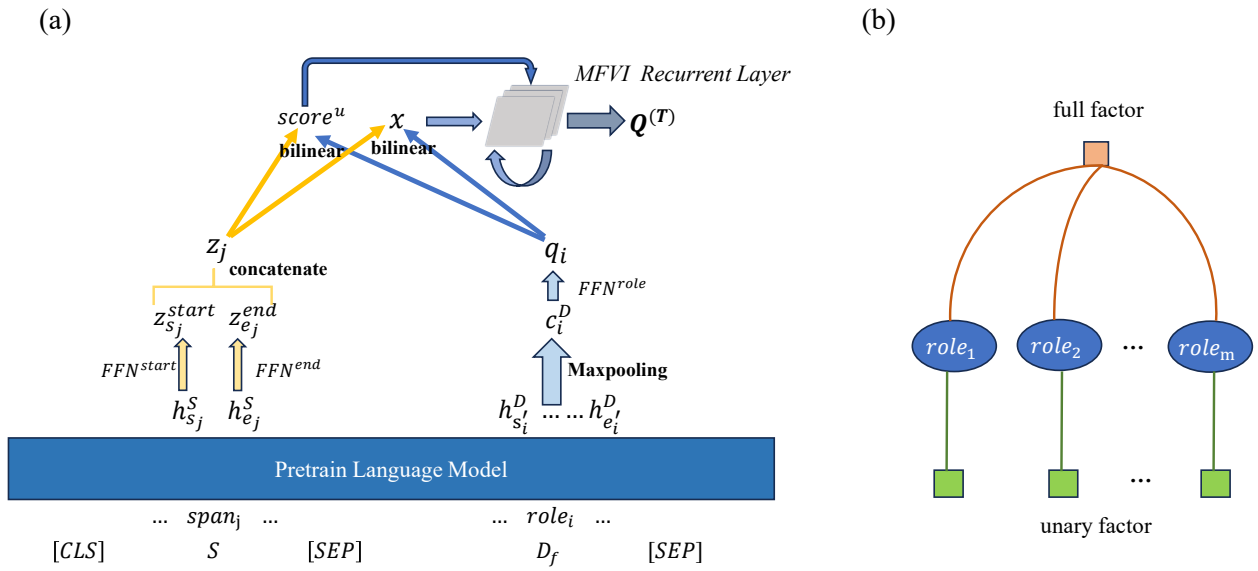


Figure 2: (a) Overall architecture of our model. (b) The factor graph of our arbitrary-order CRF.

mechanism inherent in PLMs. Consequently, it fosters the modeling of semantics of frame element labels and bolsters the interactions amongst arguments. Furthermore, it enables efficient extraction of all arguments from the text from a holistic perspective, guided by the description D_f.

To guide the focus of the PLM towards targets and frame element mentions, special tokens are added into the text-description pair. `<t>` and `</t>` are employed to encircle the target word w_{tar} in S, `<f>` and `</f>` are used to enclose `frame name`, and `<r>` and `</r>` encapsulate all frame element mentions in D_f even if occurring many times.

Unary Scoring

Consider all the m frame elements of frame f, represented as $role_1, \ldots, role_m$. Each frame element $role_i$ corresponds to a frame element slot in D_f denoted as $w^D_{s'_i}, \ldots, w^D_{e'_i}$, where s'_i and e'_i denote the start and end positions of the slot within definition D_f. To generate a query vector q_i for frame element $role_i$, we employ the maxpooling operation and a single-layer feedforward neural network (FNN):

$$c^D_i = \text{Maxpooling}\left(h^D_{s'_i}, \ldots, h^D_{e'_i}\right) \quad (3)$$

$$q_i = \text{FNN}^{role}\left(c^D_i\right) \quad (4)$$

where h^D_k represents the contextualized representation of the k-th token w^D_k within description D_f.

We also compute embeddings for all the $l = \frac{n(n+1)}{2}$ spans within sentence S if the length of the sentence S equals n. For the j-th span denoted as $span_j = w_{s_j}, \ldots, w_{e_j}$, where s_j and e_j refer to the start and end positions of $span_j$ in sentence S respectively, the span embedding is defined as follows:

$$z_j = [z^{start}_{s_j}, z^{end}_{e_j}] \quad (5)$$

$$= [\text{FNN}^{start}(h^S_{s_j}), \text{FNN}^{end}(h^S_{e_j})] \quad (6)$$

where FNN^{start} and FNN^{end} are two distinct single-layer feedforward neural networks used to generate the start and end embeddings, denoted by z^{start} and z^{end}, respectively. The term h^S_k corresponds to the contextualized representation of the k-th token w_k in sentence S and $[\cdot, \cdot]$ signifies the concatenation operation.

We set the output dimension of the FNNs above such that $q_i \in \mathbb{R}^d$ and $z_j \in \mathbb{R}^{2d}$ where d is a hyper-parameter. A bilinear operation is used to obtain a unary score, representing how likely $span_j$ can take $role_i$ as an argument,

$$score^u_{i,j} = z_j^\top \cdot U \cdot q_i \quad (7)$$

where $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, l$ and $U \in \mathbb{R}^{2d \times d}$.

Note that it is possible that $role_i$ has no corresponding argument, i.e., $role_i$ corresponds to the null span $span_0$. We designate a special unary score $score^u_{i,0}$ for this possibility with a fixed value of 0:

$$score^u_{i,0} = 0 \quad (8)$$

Define the unary score vector $score^u_i \in \mathbb{R}^{l+1}$ for $role_i$ as follows:

$$score^u_i := [score^u_{i,0}, score^u_{i,1}, \ldots, score^u_{i,l}]$$

Then the probability distribution over possible spans of $role_i$ is computed as follows:

$$\Pr(role_i | S, f, tar) = \text{Softmax}\left(score^u_i\right) \quad (9)$$

where tar denotes the index of target word w_{tar} within the text S.

S

He also said that his visit will mainly focus on the humanitarian situation of Iraq , which has been under crippling U.N. sanctions since its 1990 invasion of Kuwait, the primacord of the 1991 Gulf War.

D_f

Revenge | This frame concerns the infliction of punishment in return for a wrong suffered. An *Avenger* performs a *Punishment* on an *Offender* as a consequence of an earlier action by the *Offender*, the *Injury* . The *Avenger* inflicting the *Punishment* need not be the same as the *Injured Party* who suffered the *Injury*, but the *Avenger* does have to share the judgment that the *Offender* 's action was wrong. The judgment that the *Offender* had inflicted an *Injury* is made without regard to the law. | *Degree Depictive Instrument Manner Place Purpose Result Time Duration*

Figure 3: An example of text S and description D_f. The target word is "`sanctions`" which evokes the frame "*Revenge*". The arguments in S are color-coordinated with the corresponding frame elements in D_f. For a frame element that appears for multiple times in D_f, only the leftmost appearance is employed as the slot.

After obtaining unary scores, we may proceed with first-order inference as done in previous studies (Zheng, Wang, and Chang 2023). The predicted argument of the i-th frame element is simply the span with the highest probability:

$$a_i = \arg\max_{0 \le j \le l} \Pr(role_i = span_j | S, f, tar) \tag{10}$$

Note that our first-order inference is span-based, with scoring and decoding at the span granularity, which is different from previous work that separately and sequentially scores and selects start and end positions of spans. We will show the empirical advantage of our span-based inference in the analysis section.

Arbitrary-Order Scoring & Inference

To model interactions amongst arguments of different frame elements, we introduce arbitrary-order conditional random fields (CRFs). For each frame, we design a CRF as depicted as a factor graph in Figure 2(b). Each variable in the CRF denotes the argument of one frame element, whose domain is the set of all spans of the sentence plus a null span, thus having a size of $l+1$. Each variable is connected to a unary factor parameterized with unary scores defined earlier, representing the likelihood of the variable taking different spans as its value if evaluated in isolation. A full factor connects to all the variables, modeling high-order interactions among them. Since the number of frame elements varies across frames, the CRF may have an arbitrary number of variables and thus the full factor can have an arbitrary order, which is why we call our model an arbitrary-order CRF. Note that while we define CRFs of different frames separately, they are interconnected by sharing the same encoder and scorer.

In the CRF, the potential function ϕ_i^u for the unary factor of i-th frame element is defined as the exponential of corresponding unary scores $score_i^u$, and the potential function ϕ^f of the full factor similarly defined based on a full score tensor $score^f$.

$$\phi_i^u\left(span_j\right) = exp\left(score_{i,j}^u\right) \tag{11}$$

$$\begin{aligned}\phi^f\left(role_1 = span_{j_1}, \cdots, role_m = span_{j_m}\right) \\ = exp\left(score^f_{j_1,\cdots,j_m}\right)\end{aligned} \tag{12}$$

The full score tensor $score^f$ exhibits a size of $(l+1)^m$, wherein m is the number of frame elements and l is the number of spans. To ensure tractable representation and inference, we assume that $score^f$ is in the Kruskal form, which is closely related to canonical polyadic decomposition (CPD) of tensors.

$$score^f = \sum_{r=1}^{R} x_r^1 \circ x_r^2 \circ \cdots \circ x_r^m \tag{13}$$

where R signifies the rank of CPD, $\circ$ symbolizes outer product, and $x_r^i \in \mathbb{R}^{(l+1)}$ is computed in a manner akin to unary scores:

$$x_r^i = Z^\top \cdot W^r \cdot q_i \tag{14}$$

where $W^r \in \mathbb{R}^{2d \times d}$ is initialized as 0, $Z = [z_{CLS}, z_1, z_2, \ldots, z_l] \in \mathbb{R}^{2d \times (l+1)}$ and z_{CLS} is the embedding of the null span, derived in the same fashion as Eq.(5) and Eq.(6) based on h^S_{CLS}, the contextualized representation of `[CLS]`.

Considering that $\{W^r | r = 1, 2, \cdots, R\}$ are all initially set at zero, the full score $score^f$ is also zero. Consequently, only the unary scores are present, indicating that the learning process effectively begins from the first-order model.

Although it is now tractable to represent the full factor, performing exact arbitrary-order inference over the CRF is still NP-hard. Therefore, we leverage mean-field variational inference (MFVI) (Xing, Jordan, and Russell 2012) for tractable approximate inference. MFVI successively refines an approximate posterior marginal distribution $Q_i^{(t)}(role_i)$ for each variable $role_i$ by drawing upon messages from all connecting factors at the t-th iteration. Below we abuse the notation and denote the probability vector of $Q_i^{(t)}(role_i)$ with $Q_i^{(t)} \in \mathbb{R}^{l+1}$.

Before the first iteration of MFVI, we initialize $Q_i^{(0)}$ by normalizing exponentiated unary scores:

$$Q_i^{(0)} = \text{Softmax}\left(score_i^u\right) \tag{15}$$

At iteration t of MFVI, the aggregated message for every

variable $role_i$ from the full factor is computed as follows:

$$F_i^{(t-1)} = score^f_{1,2,\cdots,m} Q_1^{(t-1)} Q_2^{(t-1)} \cdots Q_{i-1}^{(t-1)} Q_{i+1}^{(t-1)} \cdots Q_m^{(t-1)} \tag{16}$$

$$= \sum_{r=1}^{R} \left[\left(\prod_{j \neq i} \left(x_r^{j\top} Q_j^{(t-1)} \right) \right) x_r^i \right] \tag{17}$$

Notably, Eq.(16) uses Einstein summation (einsum) notations (Stover and Weisstein 2023) and $\left(x_r^{j\top} Q_j^{(t-1)}\right)$ is a scalar.

We then update $Q_i^{(t)}$ as follows:

$$Q_i^{(t)} = \text{Softmax}\left(score_i^u + F_i^{(t-1)}\right) \tag{18}$$

After T iterations, we regard $Q_i^{(t)}$ as our final prediction distribution of the i-th frame element:

$$\Pr(role_i | S, f, tar) = Q_i^{(T)}(role_i) \tag{19}$$

then we can predict the argument of $role_i$ by Eq.(10).

The computational complexity of each MFVI iteration is $O(m^2 Rl)$. We can implement cache optimization for Eq.17 to further reduce the computational complexity. Specifically, at iteration t, we first calculate and cache $\prod_j \left(x_r^{j\top} Q_j^{(t-1)}\right)$, and then when computing $F_i^{(t-1)}$ for each i, we simply divide $x_r^{i\top} Q_i^{(t-1)}$ from the cached product instead of computing the product from scratch. In this way, the computational complexity becomes $O(mRl)$, which is linear with respect to both m, the number of frame elements, and l, the number of argument spans.

Furthermore, we find it beneficial to incorporate the regularization method recommended by regularized Frank-Wolfe (Lê-Huu and Alahari 2021). Specifically, we introduce two hyper-parameters: $\tau > 0$, a convex regularization term which modulates the smoothness of the distribution, and $\alpha \in [0, 1]$, which determines the step size in each update cycle. The update procedure expressed in Equation (18) now takes the following form:

$$Q_i^{(t)} = \alpha \text{Softmax}\left(\frac{score_i^u + F_i^{(t-1)}}{\tau}\right) + (1-\alpha) Q_i^{(t-1)} \tag{20}$$

Finally, we draw inspiration from (Zheng et al. 2015) and unfold the MFVI iterations as a recurrent neural network and connect it with our neural encoder and scorer, as shown in Figure 2(a), facilitating end-to-end training and inference.

Training Objective

We deploy cross-entropy as the training loss function:

$$\mathcal{L} = -\frac{1}{m} \sum_{i=1}^{m} \log \Pr(role_i = \hat{a}_i | S, f, tar) \tag{21}$$

where $\hat{a}_i$ denotes the ground truth argument span of $role_i$.

	frame	FE	train	dev	test	exemplar
FN 1.5	1019	9634	17143	2333	4458	153952
FN 1.7	1221	11428	19875	2309	6722	192461

Table 1: Comparison of FrameNet 1.5 and FrameNet 1.7 versions.

Experiment

Datasets

We used the benchmark datasets FrameNet versions 1.5 and 1.7[2], hereafter referred to as FN1.5 and FN1.7, respectively, to evaluate the effectiveness of our models. FN1.5 is widely used in previous research. FN1.7 is more comprehensive than FN1.5 and is known for its extended semantic content. We adhered to the train/dev/test split used in prior work (Peng et al. 2018). We utilized FrameNet's exemplar sentences, annotations linked to frames and their lexical units as supplemental training data, a practice frequently adopted in preceding researches (Chen, Zheng, and Chang 2021; Bastianelli, Vanzo, and Lemon 2020; Zheng et al. 2022; Zheng, Wang, and Chang 2023). The respective statistics of the FN1.5 and FN1.7 datasets are highlighted in Table 1.

Hyper-parameters

In our model, we employ `bert-base-uncased`[3] as the pretrained language model. We do grid search for hyper-parameter tuning and details of our hyper-parameter settings can be found in the supplementary material.

Setup

We divide the experiments into two categories: those performed without the incorporation of exemplar instances as supplementary training data (termed *w/o exemplar*) and those that integrate these instances (denoted as *w/ exemplar*). In the latter category, preliminary training was conducted on exemplar sentences, followed by continuous training on the standard training set. This strategy was adopted in light of the domain gap between exemplar instances and actual training instances, as discussed by Kshirsagar et al. (2015).

For fair comparison, we only compare our method with previous FSRL methods that use PLMs. For *w/o exemplar*, we compare with semi-CRF (Swayamdipta et al. 2017), Lin, Sun, and Zhang (2021), Kalyanpur et al. (2020), and AGED (Zheng, Wang, and Chang 2023). For *w/ exemplar*, we compare with Chen, Zheng, and Chang (2021), Bastianelli, Vanzo, and Lemon (2020), KID (Zheng et al. 2022), and AGED (Zheng, Wang, and Chang 2023). The AGED model is the previous state-of-the-art. We attempted to replicate the results of the AGED model using its official code base but could not achieve the performance reported in its paper. Therefore, we managed to enhance its performance by integrating additional multi-layer perceptrons (MLPs), applying gradient clipping, and tuning hyper-parameters.

[2]https://framenet.icsi.berkeley.edu/fndrupal/

[3]https://huggingface.co/bert-base-uncased

Model	FN 1.5			FN 1.7		
	Precision	Recall	F1-score	Precision	Recall	F1-score
w/o exemplar:						
semi-CRF (2017)	-	-	73.56	-	-	72.22
Lin, Sun, and Zhang (2021)	-	-	73.28	-	-	72.06
Kalyanpur et al. (2020)	-	-	-	71	73	72
AGED† (2023) w/o exemplar	71.93	76.78	74.28	74.02	75.46	74.73
AGED‡ (2023) w/o exemplar	71.81	76.65	74.15	74.99	75.21	75.09
ours (first-order) w/o exemplar	72.47	76.48	74.42*	75.24	**75.38**	75.31*
ours (arbitrary-order) w/o exemplar	**72.53**	**76.94**	**74.67**§	**75.60**	75.33	**75.46**§
w/ exemplar:						
Chen, Zheng, and Chang (2021)	69.27	75.39	72.20	-	-	-
Bastianelli, Vanzo, and Lemon (2020)	74.23	76.94	75.56	-	-	-
KID (2022)	71.7	79.0	75.2	74.1	77.3	75.6
AGED† (2023) w/ exemplar	73.06	79.84	76.30	75.84	77.87	76.84
AGED‡ (2023) w/ exemplar	73.93	79.07	76.41	75.72	77.55	76.62
ours (first-order) w/ exemplar	74.49	79.01	76.68*	76.29	77.53	76.90*
ours (arbitrary-order) w/ exemplar	**74.61**	**79.19**	**76.83**§	**76.40**	**77.72**	**77.06**§

Table 2: Main results on the test sets of FN 1.5 and FN 1.7. †: reported results in their paper; ‡: our reproduced results. $\star$ indicates that our first-order model achieves significantly stronger F1-score than AGED with $p < 0.05$ on ASD; § indicates that our arbitrary-order model achieves significantly stronger F1-score than our first-order model with $p < 0.05$ on ASD.

Evaluation

Evaluation of performance was carried out using the micro-F1 score [4] as the standard metric. We used exact match of argument spans in micro-F1 computation, which requires the start and end positions as well as the frame element of a predicted argument must be in full alignment with the ground truth.

Main Results

Table 2 shows the results of our experiments. For our methods and reproduction of previous work, we report the averaged results from four runs with different random seeds. We also apply Almost Stochastic Dominance (ASD) (Dror, Shlomov, and Reichart 2019) to do significance test on the F1 scores. The results indicate that regardless of whether exemplars are used as supplementary training data or not, our first-order method significantly outperforms AGED and the other previous methods. In the next section, we will empirically analyze the source of the improvements of our first-order method over AGED. Further, it can be seen that our arbitrary-order model significantly outperforms our first-order model, which shows the advantage of directly modeling full interaction among all arguments.

Analysis

Ablation Study of Our First-Order Model

Our first-order model is very similar to AGED, the previous state-of-the-art FSRL method, and differs only in the following aspects:

- "MLPs", denotes using MLPs, i.e., FNN^{role}, FNN^{start}, and FNN^{end}.
- "GC", stands for gradient clipping.
- "Zero", indicates that unary scores of the null spans are set to 0, rather than deriving scores from z_{CLS}.
- "Span", corresponds to using span-based method instead of pointer network.

[4] https://www.cs.cmu.edu/~ark/SEMAFOR/eval/

Model	F1
AGED	74.76
AGED w/ MLPs	74.84
AGED w/ GC	74.94
AGED w/ MLPs & GC	75.09
AGED w/ MLPs & GC & Zero	75.17
AGED w/ MLPs & GC & Span	75.28
First-order	75.31

Table 3: Ablation study on the differences of our first-order model vs. AGED on FN1.7 without exemplar. All variants maintain consistent hyper-parameters, yet differ in their network architectures. "AGED" is the model via their office code. "AGED w/ MLPs & GC" denotes our reproduced baseline. The term "First-order" corresponds to AGED combined with MLPs, gradient clipping, span-based method, and the aforementioned Zero setting.

We perform an ablation study over the above differences on FN1.7 without exemplar. All the variants use the same set of hyper-parameters. The results are presented in Table 3. It is evident that every difference leads to some improvement in the F1-score, and span-based method contributes the largest improvements, so does gradient clipping.

Case Study

Table 4 presents two examples from the test set of FrameNet that illustrate the enhancements made by our arbitrary-order

Sentence	First-Order vs. Arbitrary-Order
There are many stories of refugees who (Employee) arrived with nothing in their pockets , set up a small sidewalk stall , worked (target word) diligently (Manner) until they had their own store (Duration) , and then expanded it into a modest chain .	Frame : Being_employed In the first-order model, the frame element Employee is attributed to "many stories of refugees who", while the arbitrary-order model adjusts the frame element Employee to "refugees who".
There are (target word) still (Time) bars and clubs (Entity) here (Place) , but the area has become almost mainstream , and office towers are replacing many of the sinful old premises .	Frame : Existence In the first-order model, the frame element Entity is corresponded to "still bars and clubs" and the frame element Time is assigned with *the null span*. In the arbitrary-order model, the frame element Entity and the frame element Time are corrected to the right arguments.

Figure 4: Examples showing how our arbitrary-order model improves over the first-order model. In the left column, we color and label the target word and all the arguments.

model over the first-order model. From the analysis, we can see the arbitrary-order model possesses the capability to rectify the inaccuracies in the arguments by deploying the full factor, which is inherently linked to all frame elements.

Related Work

Frame Semantic Role Labeling (FSRL)

Prior studies on FSRL (Kshirsagar et al. 2015; Swayamdipta et al. 2017; Bastianelli, Vanzo, and Lemon 2020) employed a two-step methodology. They first identify potential argument spans and subsequently classify these spans into frame elements. A common oversight of these methods is the neglect of argument interactions and the disregard of label semantics within a standard role classifier. The method proposed by Zheng et al. (2022) models interaction between arguments and labels via a GNN. Furthermore, AGED (Zheng, Wang, and Chang 2023) explicitly models label semantics by using a cross-encoder that encodes text and frame definition pairs, resulting in rapid and accurate FSRL predictions.

However, AGED utilizes a pointer network (Vinyals, Fortunato, and Jaitly 2015) to individually and sequentially forecast the start and end positions of arguments. Specifically, the pointer network first predicts the start position with the maximum score, and then predicts the end position similarly to the right of the predicted start position. However, the highest scored start position may not align with the gold start position, leading to potentially incorrect predictions. In addition to this issue, the interaction amongst arguments in AGED is at best implicitly and indirectly enabled via the cross-encoder.

To overcome the above-mentioned challenges, our approach adopts a span-based method that jointly decodes the start and end positions of arguments, and employs an arbitrary-order CRF to explicitly model the interplay of all arguments. Specifically, we note that our span-based unary score can be seen as the sum of start and end position scores in AGED (with the implicit constraint that the start position is to the left of the end position):

$$
\begin{aligned}
score^{u}_{i,j} &= z_j^{\top} \cdot U \cdot q_i && (22)\\
&= \left[\ z_{s_j}^{start\top}, z_{e_j}^{end\top}\ \right] \cdot U \cdot q_i && (23)\\
&= z_{s_j}^{start\top} \cdot U \cdot q_i + z_{e_j}^{end\top} \cdot U \cdot q_i && (24)\\
&= score^{start}_{i,j} + score^{end}_{i,j} && (25)
\end{aligned}
$$

where $score^{start}_{i,j}$ and $score^{end}_{i,j}$ represent the start and end scores of the i-th frame element for the j-th span, respectively, as determined by the pointer network in AGED. By maximizing the summation instead of separately and sequentially maximizing the two scores in a greedy manner, our approach is able to avoid potential decoding errors and obtains empirical improvement shown in our ablation study.

High-Order Methods

High-order methods have been studies for a long time in the domain of dependency parsing (McDonald and Pereira 2006; Carreras 2007; Koo and Collins 2010; Wang, Huang, and Tu 2019; Wang and Tu 2020). High-order methods have also been extended to tasks such as semantic role labeling (SRL) (Jia et al. 2022; Liu, Yang, and Tu 2023) and information extraction (Jia et al. 2023). Additionally, jointly modeling the arguments in SRL can be effectively facilitated by employing Tree Kernel methods(Moschitti, Pighin, and Basili 2006, 2008). However, note that SRL differs from FSRL in that labels like `Arg0`, `Arg1`, `ArgM-LOC` do not inherently convey label semantic meaning, while frame elements in FSRL bear intrinsic lexical significance. Moreover, our FSRL method involves learning a distinct CRF for each frame, incorporating a shared backbone and parameterization, while SRL does not define frames, precluding direct application of our method to SRL. In addition, most of the above-mentioned prior high-order methods limit themselves to second-order or third-order modeling and inference due to high computational complexity. In contrast, our approach innovatively introduces a factor that interlinks all frame elements and utilizes tensor decomposition and approximate inference for tractability, facilitating arbitrary-order modeling and inference.

Conclusion

In this paper, we propose a novel arbitrary-order approach to frame semantic role labeling (FSRL). Our approach models full interactions amongst all arguments and applies canonical polyadic decomposition (CPD) and mean-field variational inference (MFVI) to ensure computational tractability. Empirical evaluations demonstrate that our approach achieves state-of-the-art performance in the task of FSRL.

References

Baker, C. F.; Fillmore, C. J.; and Lowe, J. B. 1998. The Berkeley FrameNet Project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, 86–90. Montreal, Quebec, Canada: Association for Computational Linguistics.

Bastianelli, E.; Vanzo, A.; and Lemon, O. 2020. Encoding syntactic constituency paths for frame-semantic parsing with graph convolutional networks. *arXiv preprint arXiv:2011.13210*.

Carreras, X. 2007. Experiments with a Higher-Order Projective Dependency Parser. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 957–961. Prague, Czech Republic: Association for Computational Linguistics.

Chen, X.; Zheng, C.; and Chang, B. 2021. Joint Multi-Decoder Framework with Hierarchical Pointer Network for Frame Semantic Parsing. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2570–2578. Online: Association for Computational Linguistics.

Dror, R.; Shlomov, S.; and Reichart, R. 2019. Deep Dominance - How to Properly Compare Deep Neural Models. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2773–2785. Florence, Italy: Association for Computational Linguistics.

Eckert, F.; and Neves, M. 2018. Semantic role labeling tools for biomedical question answering: a study of selected tools on the BioASQ datasets. In *Proceedings of the 6th BioASQ Workshop A challenge on large-scale biomedical semantic indexing and question answering*, 11–21. Brussels, Belgium: Association for Computational Linguistics.

Guo, S.; Li, R.; Tan, H.; Li, X.; Guan, Y.; Zhao, H.; and Zhang, Y. 2020. A Frame-based Sentence Representation for Machine Reading Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 891–896. Online: Association for Computational Linguistics.

Jia, Z.; Yan, Z.; Han, W.; Zheng, Z.; and Tu, K. 2023. Modeling Instance Interactions for Joint Information Extraction with Neural High-Order Conditional Random Field. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 13695–13710. Toronto, Canada: Association for Computational Linguistics.

Jia, Z.; Yan, Z.; Wu, H.; and Tu, K. 2022. Span-based semantic role labeling with argument pruning and second-order inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 10822–10830.

Kalyanpur, A.; Biran, O.; Breloff, T.; Chu-Carroll, J.; Diertani, A.; Rambow, O.; and Sammons, M. 2020. Open-domain frame semantic parsing using transformers. *arXiv preprint arXiv:2010.10998*.

Khashabi, D.; Khot, T.; Sabharwal, A.; and Roth, D. 2018. Question answering as global reasoning over semantic abstractions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Koo, T.; and Collins, M. 2010. Efficient Third-Order Dependency Parsers. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1–11. Uppsala, Sweden: Association for Computational Linguistics.

Kshirsagar, M.; Thomson, S.; Schneider, N.; Carbonell, J.; Smith, N. A.; and Dyer, C. 2015. Frame-Semantic Role Labeling with Heterogeneous Annotations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 218–224. Beijing, China: Association for Computational Linguistics.

Lê-Huu, D.; and Alahari, K. 2021. Regularized frank-wolfe for dense crfs: Generalizing mean field and beyond. *Advances in Neural Information Processing Systems*, 34: 1453–1467.

Lin, Z.; Sun, Y.; and Zhang, M. 2021. A Graph-Based Neural Model for End-to-End Frame Semantic Parsing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3864–3874. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Litkowski, K. 2004. Senseval-3 task: Automatic labeling of semantic roles. In *Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, 9–12. Barcelona, Spain: Association for Computational Linguistics.

Liu, D.; and Gildea, D. 2010. Semantic Role Features for Machine Translation. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 716–724. Beijing, China: Coling 2010 Organizing Committee.

Liu, W.; Yang, S.; and Tu, K. 2023. Structured Mean-Field Variational Inference for Higher-Order Span-Based Semantic Role Labeling. In *Findings of the Association for Computational Linguistics: ACL 2023*, 918–931. Toronto, Canada: Association for Computational Linguistics.

Marcheggiani, D.; Bastings, J.; and Titov, I. 2018. Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 486–492. New Orleans, Louisiana: Association for Computational Linguistics.

McDonald, R.; and Pereira, F. 2006. Online Learning of Approximate Dependency Parsing Algorithms. In *11th Conference of the European Chapter of the Association for Computational Linguistics*, 81–88. Trento, Italy: Association for Computational Linguistics.

Moschitti, A.; Pighin, D.; and Basili, R. 2006. Semantic Role Labeling via Tree Kernel Joint Inference. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, 61–68. New York City: Association for Computational Linguistics.

Moschitti, A.; Pighin, D.; and Basili, R. 2008. Tree Kernels for Semantic Role Labeling. *Computational Linguistics*, 34(2): 193–224.

Peng, H.; Thomson, S.; Swayamdipta, S.; and Smith, N. A. 2018. Learning Joint Semantic Parsers from Disjoint Data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 1492–1502. New Orleans, Louisiana: Association for Computational Linguistics.

Rabanser, S.; Shchur, O.; and Günnemann, S. 2017. Introduction to tensor decompositions and their applications in machine learning. *arXiv preprint arXiv:1711.10781*.

Shen, D.; and Lapata, M. 2007. Using Semantic Roles to Improve Question Answering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 12–21. Prague, Czech Republic: Association for Computational Linguistics.

Stover, C.; and Weisstein, E. W. 2023. Einstein Summation. From MathWorld–A Wolfram Web Resource.

Surdeanu, M.; Harabagiu, S.; Williams, J.; and Aarseth, P. 2003. Using Predicate-Argument Structures for Information Extraction. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 8–15. Sapporo, Japan: Association for Computational Linguistics.

Swayamdipta, S.; Thomson, S.; Dyer, C.; and Smith, N. A. 2017. Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold. *arXiv preprint arXiv:1706.09528*.

Trandabăț, D. 2011. Using semantic roles to improve summaries. In *Proceedings of the 13th European Workshop on Natural Language Generation*, 164–169. Nancy, France: Association for Computational Linguistics.

Vinyals, O.; Fortunato, M.; and Jaitly, N. 2015. Pointer networks. *Advances in neural information processing systems*, 28.

Wang, H.; Bansal, M.; Gimpel, K.; and McAllester, D. 2015. Machine Comprehension with Syntax, Frames, and Semantics. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 700–706. Beijing, China: Association for Computational Linguistics.

Wang, X.; Huang, J.; and Tu, K. 2019. Second-Order Semantic Dependency Parsing with End-to-End Neural Networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4609–4618. Florence, Italy: Association for Computational Linguistics.

Wang, X.; and Tu, K. 2020. Second-Order Neural Dependency Parsing with Message Passing and End-to-End Training. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 93–99. Suzhou, China: Association for Computational Linguistics.

Xing, E. P.; Jordan, M. I.; and Russell, S. 2012. A generalized mean field algorithm for variational inference in exponential families. *arXiv preprint arXiv:1212.2512*.

Zheng, C.; Chen, X.; Xu, R.; and Chang, B. 2022. A Double-Graph Based Framework for Frame Semantic Parsing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4998–5011. Seattle, United States: Association for Computational Linguistics.

Zheng, C.; Wang, Y.; and Chang, B. 2023. Query Your Model with Definitions in FrameNet: An Effective Method for Frame Semantic Role Labeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 14029–14037.

Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; and Torr, P. H. 2015. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, 1529–1537.