Span-Based Semantic Role Labeling with Argument Pruning and Second-Order Inference

Zixia jia, Zhaohui Yan, Haoyi Wu , Kewei Tu
{jiazx, yanzhh, wuhy1, tukw}@shanghaitech.edu.cn

1. Motivation

We study span-based semantic role labeling with high-order inference

Advantages of span-based SRL:
- Capable of handling multiple predicates.
- Better utilization of span-level features.

Disadvantages of span-based SRL:
- $O(n^3)$ complexity of enumerating predict-argument pairs.
- High degree of imbalance between positive and negative samples.

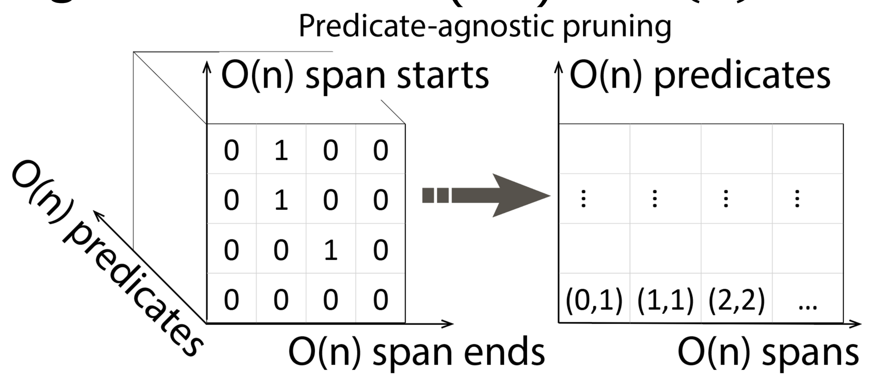
Advantage of high-order inference:
- capture interactions between multiple arguments and predicates.

Disadvantage of high-order inference:
- Exact high-order inference is intractable, even for second-order inference, there are already $O(n^5)$ parts.

2. Our approach

Step1: Pruning (corresponding to PAPN model)

Predicate-agnostic pruning, reduces candidate arguments from $O(n^2)$ to $O(n)$.



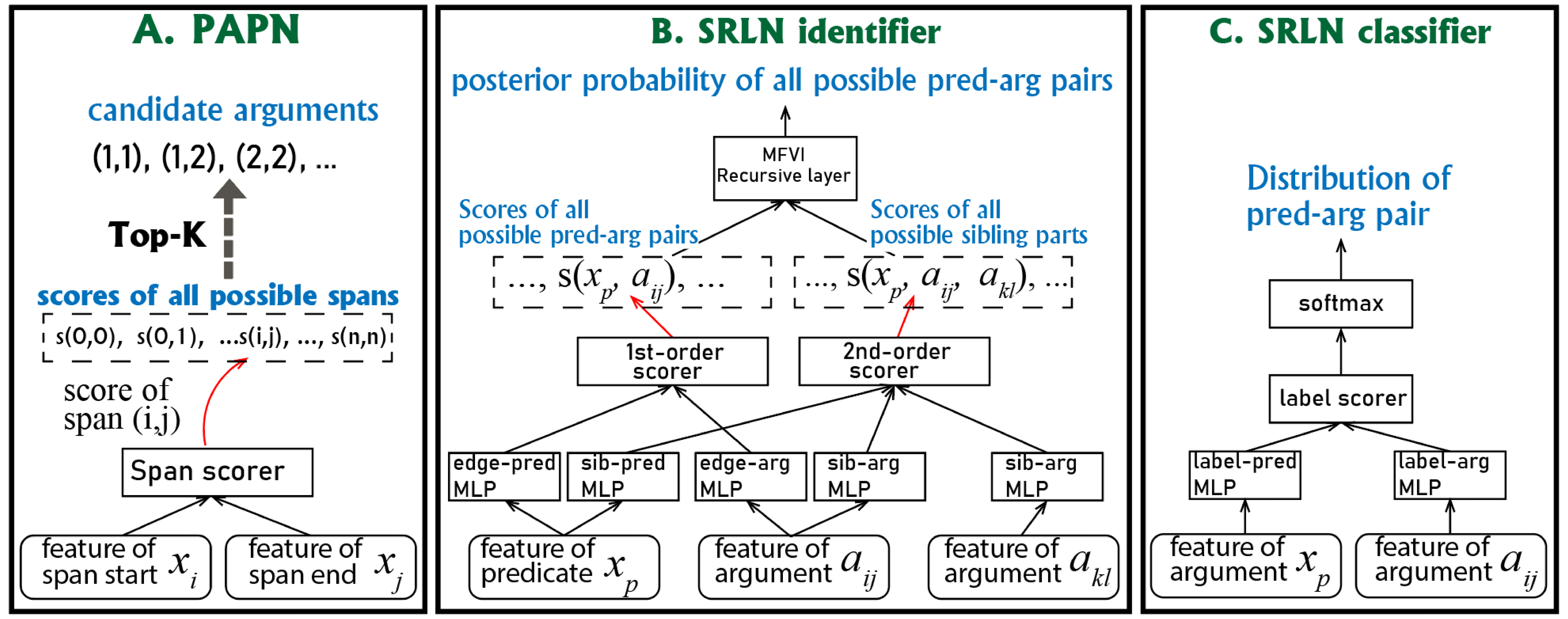
Step2: Identification (corresponding to SRLN Identifier)

Semantic Role Labeling Network with optional high-order decoder, which identified the candidate arguments of each predicate.

Step3: Classification (corresponding to SRLN Classifier)

Semantic Role Labeling Network, which labeled arguments of each predicate.

3. Our Model



A. PAPN

- We score every possible span by its start and end tokens, pick top-k spans as candidate arguments.
- We train PAPN by optimizing the loss as follows:

$$Loss = \sum_{i \le j} \text{CrossEntropy}\left(y_{ij}, \text{Sigmoid}(s_{ij})\right)$$

s_{ij}: score of span (i,j)

$$y_{ij} = \begin{cases} 1 & \text{if span(i,j) is an argument of some predicate} \\ 0 & \text{otherwise} \end{cases}$$

B. SRLN identifier

- With candidate arguments ($O(n)$) from PAPN, SRLN identifier considers the interaction between each pair of potential pred-arg pairs that share the same predicate. $\left(O(n^3)\right)$
- We use Mean Field Variational Inference to get the posterior probability of all possible pred-arg pairs.
- Dynamic programming is then used to identify all the pred-arg pairs under the constraint that there are no overlapping arguments for each predicate.

PS: for first order model, SRLN identifier only uses the first-order scorer.

C. SRLN classifier

- With the result of identified pred-arg pairs from SRLN identifier, we score each pred-arg pair and get its distribution over the SRL label set.

4. Experiment Settings

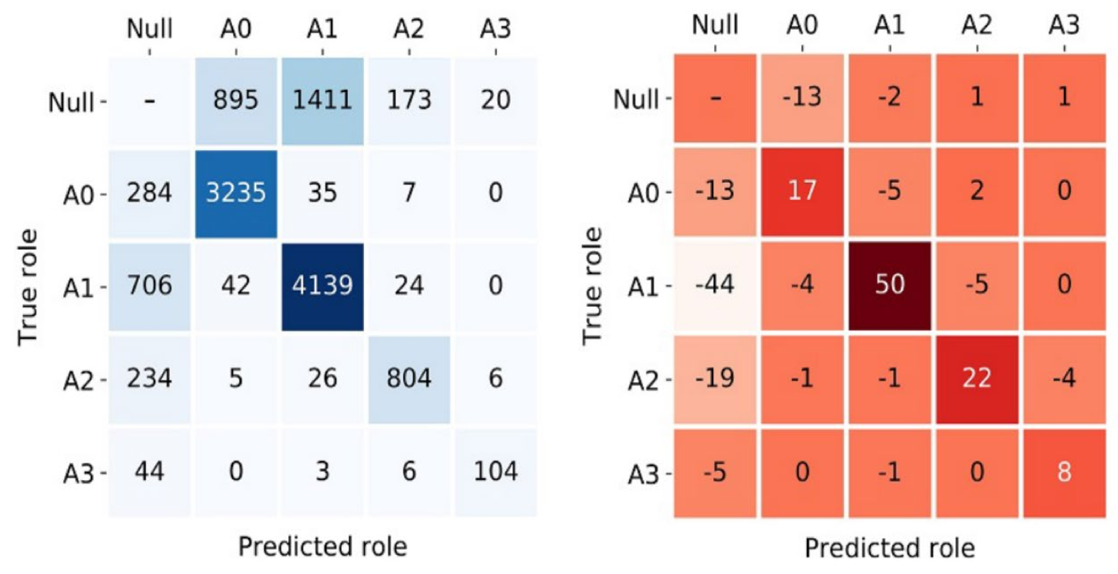
1. Datasets CoNLL2005, CoNLL2012.

2. Evaluation micro-average F1 score for correctly predicted (predicate, argument span, label) tuples.

3. Two SRL settings predicted predicates setting and gold predicates setting.

6. Experiment Analysis

1. Error correction of our second-order model over our first-order model



(a) first-order (b) error correction matrix

2. Independent and joint learning of PAPN and SRLN

ELMo	Predicted predicates		Gold Predicates	
	WSJ	Brown	WSJ	Brown
He et al. (2018)†	86.00	76.10	87.40	80.40
He et al. (2018) ‡	85.94	76.51	87.34	79.01
Li et al. (2019)†	86.30	76.40	87.70	80.50
Li et al. (2019)‡	85.69	75.48	87.21	78.05
JointL	86.15	76.84	87.78	79.83
Ours	86.56	77.29	88.00	80.23

- The comparison results are from previous SOTA models which perform joint training of pruning and SRL model.
- Our experiments show that results of independent learning is better than the results of joint learning.

5. Experiment Results

Predicted Predicates

Predicted predicate	WSJ	Brown	CoNLL12	Avg.
GloVe embedding	F1	F1	F1	
He et al. (2017)	81.2	68.5	76.8	75.5
He et al. (2018)	82.5	70.8	79.8	77.7
Li et al. (2019)	83.0	-	-	-
Ours (1st order)	84.29	71.78	81.61	79.23
Ours (2nd order)	**84.50**	**72.70**	**81.63**	79.57
Strubell et al. (2018)* (LISA)	83.61	71.91	80.70	78.73
Zhou, Li, and Zhao (2020)*	84.56	72.55	-	-
Contextualized embedding				
He et al. (2018)▷	86.0	76.1	82.9	81.7
Li et al. (2019)▷	86.3	76.4	83.1	81.9
Ours (1st order) + ELMo	86.56	77.29	83.98	82.61
Ours (2nd order) + ELMo	86.61	77.45	83.87	82.64
Ours (1st order) + BERT	86.63	78.75	84.16	83.18
Ours (2nd order) + BERT	86.70	78.58	84.22	83.17
Ours (1st order) + RoBERTa	86.57	79.44	84.17	83.45
Ours (2nd order) + RoBERTa	**87.03**	**79.71**	**84.33**	83.69
Strubell et al. (2018)*▷ (LISA)	86.55	78.05	83.12	82.57
Zhou, Li, and Zhao (2020)*◇	87.62	80.34	-	-
Mohammadshahi et al. (2021)*◇	87.57	80.53	-	-

7. Conclusion

- We propose a novel graph-based approach to span-based SRL which consists of a predicate-agnostic argument pruning network and an SRL network.
- The pruning network reduces the number of candidate arguments to $O(n)$
- The SRL network can perform both first-order decoding and second-order decoding using recurrent neural networks unfolded from mean-field variational inference.
- Experiments show effectiveness of our model and The second-order decoding achieves better result than the first-order.