Sequence-to-Graph Transduction with Second-Order Edge Inference for Cross-Framework Meaning Representation Parsing

Xinyu Wang · Yixian Liu · Zixia Jia · Chengyue Jiang · Kewei Tu

School of Information Science and Technology, ShanghaiTech University

Paper:

Code:

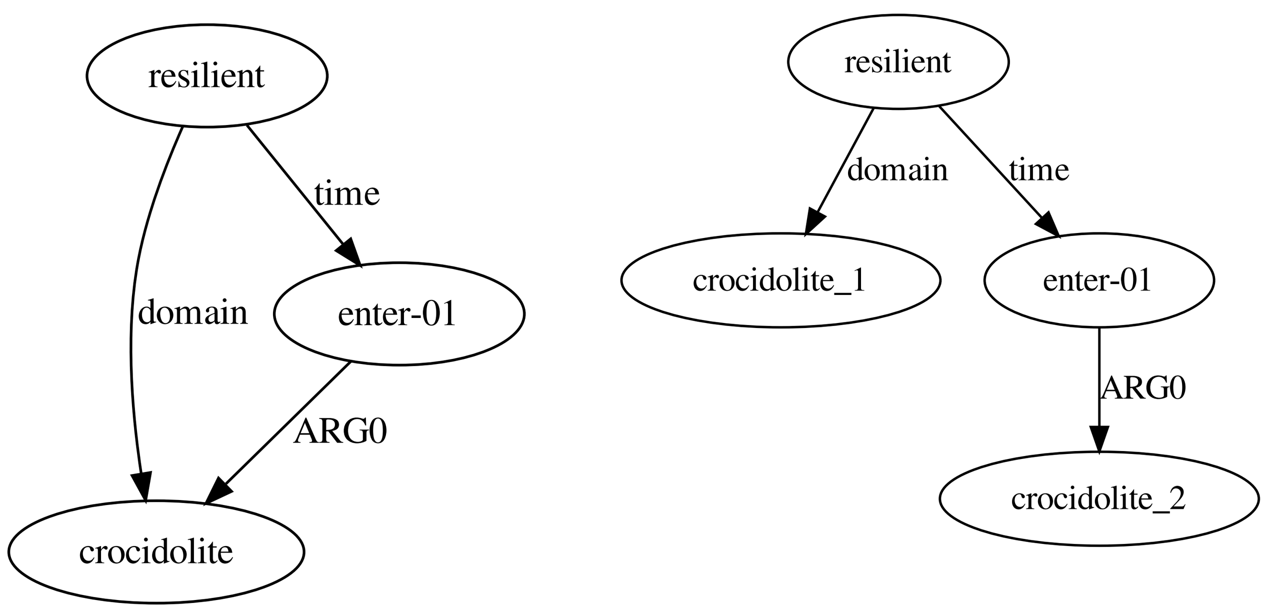
Task Definition and Our Approach

The task combines five different frameworks of graph-based meaning representation.

- DELPH-IN MRS Bi-Lexical Dependencies (DM)
- Prague Semantic Dependencies (PSD)
- Elementary Dependency Structures (EDS)
- Universal Conceptual Cognitive Annotation (UCCA)
- Abstract Meaning Representation (AMR)

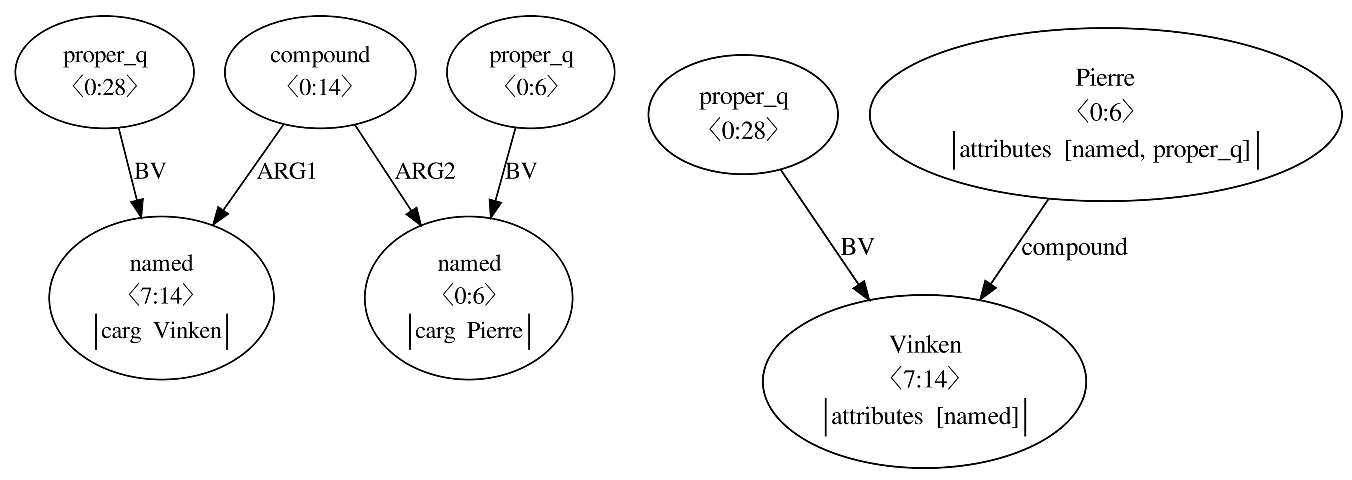
Our system is a graph-based method which combines an extended pointer-generator network introduced by Zhang et al. (2019) to generate nodes for EDS, UCCA and AMR graphs and a second-order mean field variational inference module introduced by Wang et al. (2019) to predict edges for all the frameworks.

Graph Pre-processing



(a) Before conversion. (b) After conversion.

Figure 1. An example of converting AMR graphs into tree structures. This is a sub-graph of sentence *#20003002*.



(a) Before reduction. (b) After reduction.

Figure 2. An example of EDS reduction. This is a sub-graph of sentence *#20001001*.

Overall Structure

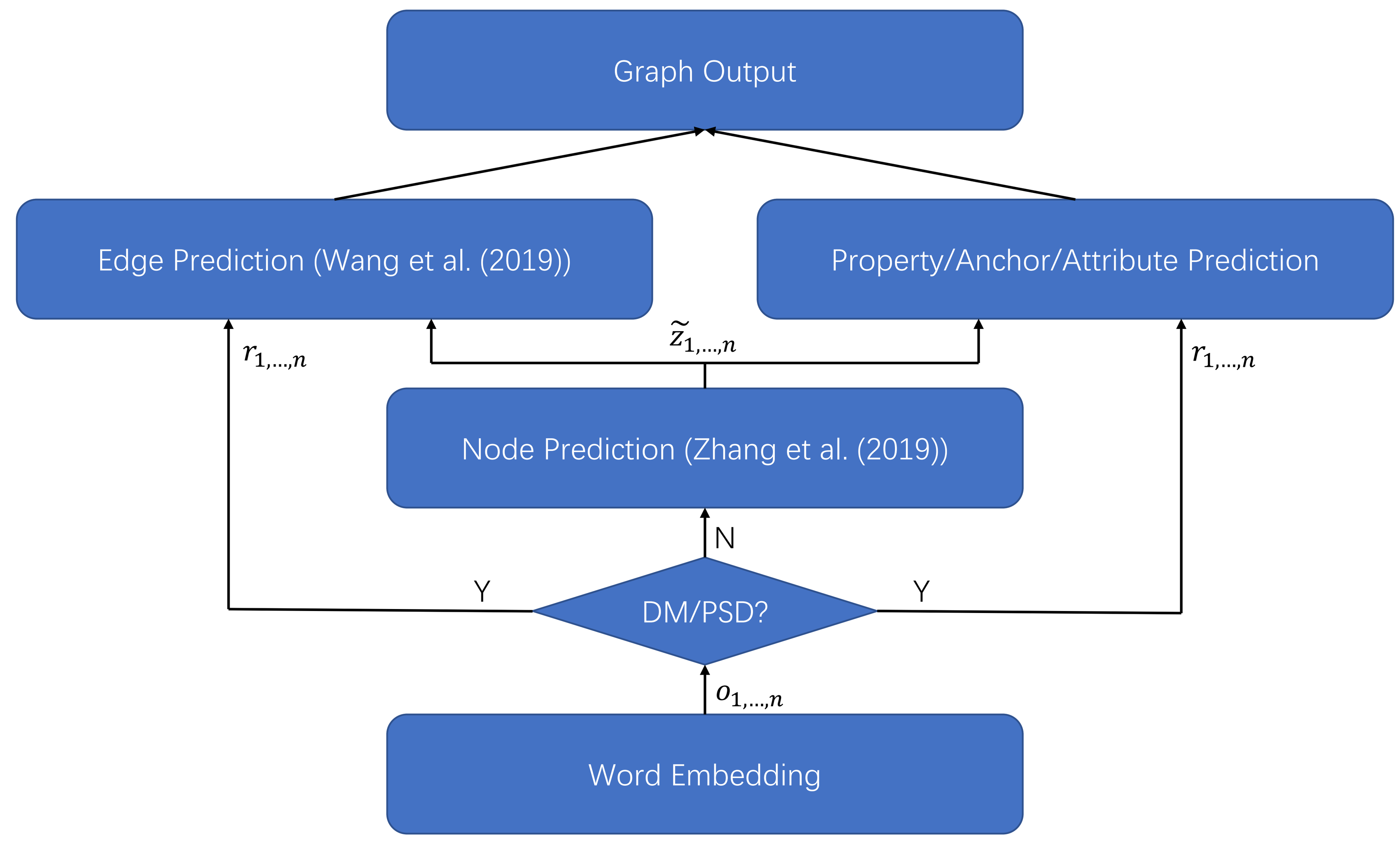


Figure 3. Illustration of our system architecture

Second-Order Edge Inference

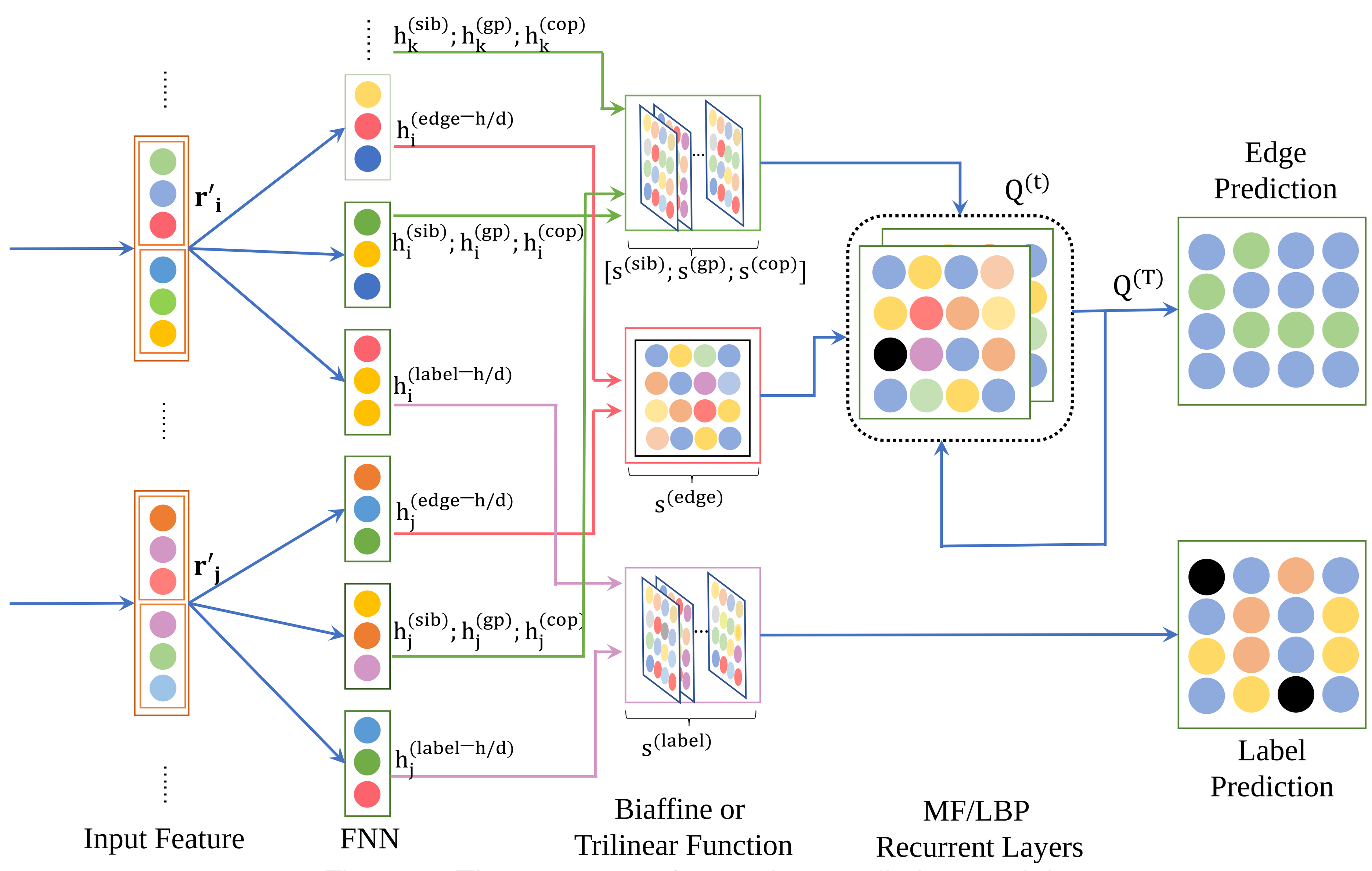


Figure 4. The structure of our edge prediction module.

Overall Results

	DM	PSD	EDS	UCCA	AMR
Ours-all	94.88	89.49	86.90	-	63.59
Best-all	**95.50**	**91.28**	**94.47**	**81.67**	**73.38**
Ours-lpps	94.28	85.22	87.49	-	66.82
Best-lpps	**94.96**	**88.46**	**92.82**	**82.61**	**73.11**

Table 1. Comparison of cross-framework F1 scores achieved by our system and best scores of other teams for each metric. *all* represents the F1 score over the full test set for each framework. *lpps* represents a 100-sentence sample from the little prince containing graphs over all the frameworks.

Our system ranks the 1st and 2nd for DM and PSD on in-framework scores. The detailed comparison is in Table 2.

	DM		PSD		Avg	
	all	lpps	all	lpps	all	lpps
Ours	**92.98**	**94.46**	81.61	**81.91**	**87.30**	**88.19**
Best	92.52	93.68	**81.66**	81.47	87.09	87.58

Table 2. Comparison of in-framework labeled F1 scores on DM and PSD by our system and the best scores from the other teams. Note that the Best scores are from multiple systems instead of a single system.

Reference:

Sheng Zhang, Xutai Ma, Kevin Duh, and Benjamin Van Durme. 2019. AMR Parsing as Sequence-to-Graph Transduction. In *ACL 2019*.

Xinyu Wang, Jingxian Huang, and Kewei Tu. 2019. Second-order semantic dependency parsing with end-to-end neural networks. In *ACL 2019*.

Analysis

Table 3 and 4 show detailed comparison for each evaluation component for DM and PSD. For DM, our system outperforms systems of the other teams on tops, properties and edges prediction and is competitive on anchors. For PSD, our system is also competitive on all the components except labels. There is a large gap in the performance of node label prediction between our system and the best one on both DM and PSD, we believe adding an MLP layer for label prediction would diminish this gap.

	tops	labels	properties	anchors	edges	average
Ours-all	**93.68**	90.51	**95.16**	98.38	**92.32**	94.32
Best-all	93.23	**96.34**	94.93	**98.74**	92.08	**94.76**
Ours-lpps	**99.00**	87.26	**94.53**	**99.36**	**93.92**	94.03
Best-lpps	96.48	**94.82**	94.36	99.04	93.28	**94.64**

Table 3. Comparison of cross-framework F1 scores achieved by our system and best scores of the other teams for each evaluation component on DM.

	tops	labels	properties	anchors	edges	average
Ours-all	95.68	84.79	91.83	97.66	**79.50**	88.77
Best-all	**95.83**	**94.68**	**92.38**	**98.35**	79.44	**90.76**
Ours-lpps	96.00	76.72	84.73	97.61	**79.80**	85.22
Best-lpps	**96.40**	**92.04**	**86.00**	**98.46**	79.18	**88.40**

Table 4. Comparison of cross-framework F1 scores achieved by our system and best scores of the other teams for each evaluation component on PSD.