A Double-Graph Based Framework for Frame Semantic Parsing

Anonymous ACL submission

Abstract

Frame semantic parsing is a fundamental NLP task, which consists of three subtasks: frame identification, argument identification and role classification. Most previous studies tend to neglect relations between different subtasks and arguments and pay little attention to ontological frame knowledge defined in FrameNet. In this paper, we propose a Knowledge-guided Incremental semantic parser with Double-graph (KID). We first introduce Frame Knowledge Graph (FKG), a heterogeneous graph containing both frames and FEs (Frame Elements) built on the frame knowledge so that we can derive knowledge-enhanced representations for frames and FEs. Besides, we propose Frame Semantic Graph (FSG) to represent frame semantic structures extracted from the text with graph structures. In this way, we can transform frame semantic parsing into an incremental graph construction problem to strengthen interactions between subtasks and relations between arguments. Our experiments show that KID outperforms the previous state-of-the-art method by up to 1.7 F1-score on two FrameNet datasets.

1 Introduction

The frame semantic parsing task (Gildea and Jurafsky, 2002; Baker et al., 2007) aims to extract frame semantic structures from sentences based on the lexical resource FrameNet (Baker et al., 1998). As shown in Figure 1, given a target in the sentence, frame semantic parsing consists of three subtasks: frame identification, argument identification and role classification. Frame semantic parsing can also contribute to downstream NLP tasks such as machine reading comprehension (Guo et al., 2020), relation extraction (Zhao et al., 2020) and dialogue generation (Gupta et al., 2021).

FrameNet is a lexical database, which defines more than one thousand hierarchically-related frames to represent situations, objects or events, and nearly 10 thousand FEs (Frame Elements) as frame-specific semantic roles with more than 100,000 annotated exemplar sentences. In addition, FrameNet defines ontological frame knowledge for each frame such as frame semantic relations, FE mappings and frame/FE definitions. The frame knowledge plays an important role in frame semantic parsing. Most previous approaches (Kshirsagar et al., 2015; Yang and Mitchell, 2017; Peng et al., 2018) only use exemplar sentences and ignore the ontological frame knowledge. Recent researches (Jiang and Riloff, 2021; Su et al., 2021) introduce frame semantic relations and frame definitions into the subtask frame identification. Differ from previous work, we construct a heterogeneous graph named Frame Knowledge Graph (FKG) based on frame knowledge to model multiple semantic relations between frames and frames, frames and FEs, as well as FEs and FEs. Furthermore, we apply FKG to all subtasks of frame semantic parsing, which can fully inject frame knowledge into frame semantic parsing. The knowledge-enhance representations of frames and FEs are learned in a unified vector space and this can also strengthen interactions between frame identification and other subtasks.

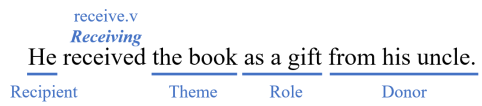

Figure 1: An example of the frame semantic structure. Given the target **receive** in this sentence, the frame identification is to identify the frame ***Receiving*** evoked by it; the argument identification is to find the arguments (*He*, *the book*, ...) of this target; the role classification is to assign frame elements (**Recipient**, **Theme**, ...) as semantic roles to these arguments.

Most previous systems neglect interactions between subtasks, they either focus on one or two subtasks (Hermann et al., 2014; FitzGerald et al.,

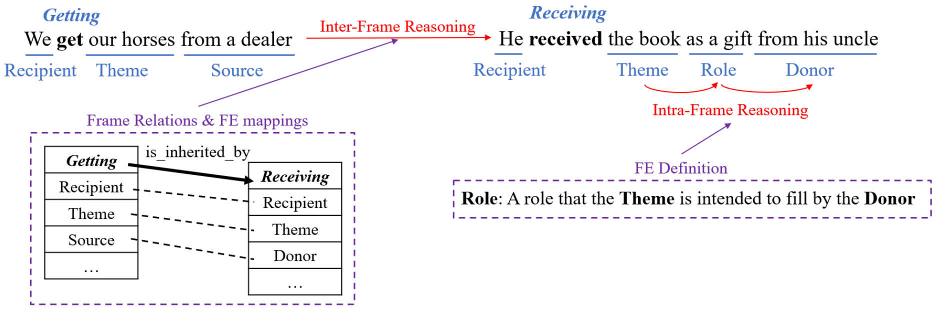

Figure 2: An example of how frame knowledge contributes to frame semantic parsing. The frame semantic relations and FE mappings guide inter-frame reasoning (from left sentence to right); and the FE definitions help with intra-frame reasoning (**Theme** to **Role** and **Role** to **Donor**)

2015; Marcheggiani and Titov, 2020) of frame semantic parsing or treat all subtasks independently (Das et al., 2014; Peng et al., 2018). Furthermore, in argument identification and role classification, previous approaches process each argument separately with sequence labeling strategy (Yang and Mitchell, 2017; Bastianelli et al., 2020) or span-based graphical models (Täckström et al., 2015; Peng et al., 2018). In this paper, we propose Frame Semantic Graph (FSG) to represent frame semantic structures and treat frame semantic parsing as a process to construct this graph incrementally. With graph structure, historical decisions of parsing can guide the current decision of argument identification and role classification, which highlights interactions between subtasks and arguments.

Based on two graphs mentioned above, we propose our framework KID (**K**nowledge-guided **I**ncremental semantic parser with **D**ouble-graph). FKG provides a static knowledge background for encoding frames and FEs while FSG represents dynamic parsing results in frame semantic parsing and highlights relations between arguments.

Overall, our contributions can be summarized as follow:

- We build FKG based on the ontological frame knowledge in FrameNet. FKG incorporates frame semantic parsing with structured frame knowledge, which can get knowledge-enhanced representations of frames and FEs.
- We propose FSG to represent the frame semantic structures. We treat frame semantic parsing as a process to construct the graph incrementally. This graph focuses on the target-argument and argument-argument relations.

We evaluate the performance of KID on two FrameNet datasets: FN 1.5 and FN 1.7, the results show that the KID achieves state-of-the-art on these datasets by increasing up to 1.7 points on F1-score. Our extensive experiments also verify the effectiveness of these two graphs.

2 Ontological Frame Knowledge

Frame semantics relates linguistic semantics to encyclopedic knowledge and advocates that one cannot understand the semantic meaning of one word without essential frame knowledge related to the word (Fillmore and Baker, 2001). For a frame, the frame knowledge of it contains frame/FE definitions, frame semantic relations and FE mappings. FrameNet defines 8 kinds of frame semantic relations such as **Inheritance**, **Perspective_on** and **Using**; for any two related frames, the FrameNet defines FE mappings between their FEs. For example, the frame ***Receiving*** inherits from ***Getting*** and the FE **Donor** of ***Receiving*** is mapped to the FE **Source** of ***Getting***. Each frame or FE has its own definition and may mention other FEs in the same frame.

we propose two ways of reasoning about frame semantic parsing: inter-frame reasoning and intra-frame reasoning in Figure 2. Frame knowledge mentioned above can guide both ways of reasoning. The frame semantic relation and FE mappings between ***Receiving*** and ***Getting*** allow us to learn experience from the left sentence when parsing the right sentence because similar argument spans of two sentences will have related FEs as their roles. The FE definitions reflect dependencies between arguments. The definition of **Role** in frame ***Receiving*** mentions **Theme** and **Donor**, which reflects dependencies between argument *the book* and argument *as a gift*.

3 Task Formulation

Given a target t in the sentence $S = w_0, \ldots, w_{n-1}$, the frame semantic parsing aims to extract the frame semantic structure of t from S. Suppose that there are k arguments of t in S: $a_0, \ldots, a_{k-1}$, all subtasks can be formulated as follow:

- Frame identification: finding an $f \in F$ evoked by t, where F denotes the set of all frames in the FrameNet.
- Argument identification: finding the boundaries i_τ^s and i_τ^e for each argument $a_\tau = w_{i_\tau^s}, \ldots, w_{i_\tau^e}$.
- Role classification: assigning an FE $r_\tau \in R_f$ to each a_τ, where R_f denotes the set of all FEs of frame f.

4 Method

Overall, KID first encodes all frames and FEs to knowledge-enhanced representations via frame knowledge graph encoder (section 4.1). For a sentence with a target, contextual representations of tokens are derived from the sentence encoder (section 4.2). The target span will be embedded into the vector space of FKG and then a frame will be identified with a scoring module (section 4.3). The target and frame evoked by it will be combined into the initial node of FSG. Frame semantic graph decoder (section 4.4) will expand FSG incrementally from the target node to complete FSG. Based on the representation of current snapshot of partial FSG, KID identifies a new argument with a pointer network (section 4.4.1) and assigns it an FE (section 4.4.2).

4.1 Frame Knowledge Graph Encoder

FKG is an undirected multi-relational heterogeneous graph, and Figure 3 shows a subgraph of FKG. Its nodes contain both frames and FEs and there are four kinds of relations in FKG: frame-FE, frame-frame, inter-frame FE-FE and intra-frame FE-FE relations. The following will show how we extract these relations from frame knowledge:

Frame-FE: we connect a frame with its FEs. With this relation, we can learn representations of frames and FEs in a unified vector space to strengthen interactions between frame identification and other subtasks.

Frame-frame and **inter-frame FE-FE**: these two kinds of relations are frame semantic relations and

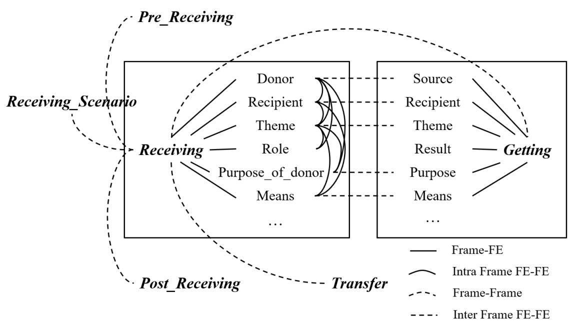

Figure 3: A subgraph of FKG. We only show intra-frame and inter-frame FE-FE relations in the frame ***Receiving***. Inside the solid rectangular box are a frame and its FEs.

FE mappings respectively and here we ignore relation types of frame semantic relations. They can both guide inter-frame reasoning.

Intra-frame FE-FE: If the definition of an FE mentions another FE in the same frame, they will have intra-frame FE-FE relations with each other. This relation can help with intra-frame reasoning.

The frame knowledge graph encoder aims to get knowledge-enhanced representations of nodes in FKG via an RGCN (Schlichtkrull et al., 2018) module. We use F to represent all frames in FrameNet and R_f to represent all FEs of frame f. In addition, we use $R = \bigcup_{f \in F} R_f$ to represent all FEs in the FrameNet. Let $0, \ldots, |F| - 1$ denote all frames and $|F|, \ldots, |F| + |R| - 1$ denote all FEs. Moreover, we introduce a special dummy node into FKG: $|F| + |R|$. So the vectors $y_0, \ldots, y_M \in \mathbb{R}^{d_n}$ denotes the representations of all nodes in FKG, where $M = |F| + |R|$.

For each node i, we take a randomly initialized embedding $y_i^{(0)} \in \mathbb{R}^{d_n}$ as the input feature of the RGCN module. Then we can get representations of all frames and FEs:

$$y_0, \ldots, y_M = \mathrm{RGCN}\left(y_0^{(0)}, \ldots, y_M^{(0)}\right) \quad (1)$$

The RGCN module models four kinds of relations: Frame-FE, intra-frame FE-FE, frame-frame and inter-frame FE-FE. We also use name information to get better representations for FEs. The FEs whose names are the same will have same embeddings: for $i, j \geq |F|$, $y_i^{(0)} = y_j^{(0)}$ if the name of i is the same as j. With this trick, we can fuse both name information and structure information into the representations of FEs.

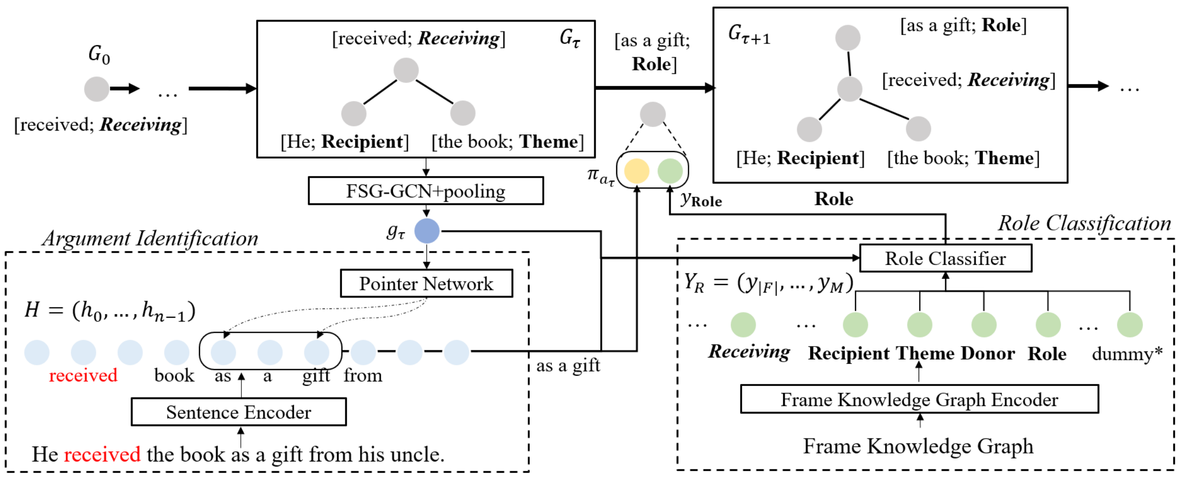

Figure 4: Based on the representation g_τ of partial FSG G_τ, frame semantic graph decoder identifies new argument *as a gift* with pointer networks, and label it with FE **Role**. G_τ will be updated to $G_{\tau+1}$ with (*as a gift*, **Role**).

4.2 Sentence Encoder

The sentence encoder converts tokens of the sentence $S = w_0, \ldots, w_{n-1}$ to their representations $h_0, \ldots, h_{n-1} \in \mathbb{R}^{d_h}$.

We use LSTM (Hochreiter and Schmidhuber, 1997) and GCN (Kipf and Welling, 2016) to model both sequential structure and dependency structure:

$$\alpha_0, \ldots, \alpha_{n-1} = \text{BiLSTM}(e_0, \ldots, e_{n-1}) \quad (2)$$

$$\beta_0, \ldots, \beta_{n-1} = \text{GCN}^{dep}(\alpha_0, \ldots, \alpha_{n-1}) \quad (3)$$

e_i denotes the embedding of word w_i. We get contextual representations h_i by adding β_i to α_i.

The GCN uses the graph structure of the dependency tree of S. The dependency tree is regarded as an undirected graph. We follow previous studies (Marcheggiani and Titov, 2020; Bastianelli et al., 2020) to use syntax structures here because syntax structure is proved beneficial to semantic parsing.

Furthermore, we use boundary information to represent spans like $s = w_i, \ldots, w_j$ based on token representations because we need to embed spans into the vector space of FKG in frame identification and role classification:

$$Q(i,j) = \text{FFN}((h_j - h_i) \oplus (h_j + h_i)) \quad (4)$$

The dimension of $Q(i,j)$ is d_n. The $\oplus$ denotes concatenation operation.

4.3 Frame Identification

A frame $f \in F$ will be identified based on the target t, representations of tokens $h_0, \ldots, h_{n-1}$ and representations of frames $y_0, \ldots, y_{|F|-1}$ with a scoring module. The target $t = w_{i_t^s}, \ldots, w_{i_t^e}$ will be embedded to the vector space of all frames as $\gamma_t \in \mathbb{R}^{d_n}$. We can calculate dot product scores between γ_t and all frames $Y_f = (y_0, \ldots, y_{|F|-1}) \in \mathbb{R}^{d_n \times |F|}$ to get the probability distribution of f. Let π_t denotes $Q(i_t^s, i_t^e)$ and FFN denotes Feed Forward Network:

$$\gamma_t = \tanh(\text{FFN}(\pi_t)) \quad (5)$$

$$P(f|S,t) = \text{softmax}\left(Y_f^\top \cdot \gamma_t\right) \quad (6)$$

4.4 Frame Semantic Graph Decoder

We propose FSG to represent the frame semantic structure of t in the sentence S and we treat the frame semantic parsing as a process to construct FSG incrementally. Intermediate results of FSG are partial FSGs representing all historical decisions. Suppose that there are k arguments of target t: $a_0, \ldots, a_{k-1}$ and they have roles $r_0, \ldots, r_{k-1}$. For τ-th snapshot of FSG G_τ, it contains $\tau + 1$ nodes: one target node (t, f) and τ argument nodes (if exist) $(a_0, r_0), \ldots, (a_{\tau-1}, r_{\tau-1})$. The target node will be connected with all argument nodes. The indices of nodes in G_τ depend on the order they are added in the graph, 0 denotes the target node and $1, \ldots, \tau$ denotes $(a_0, r_0), \ldots, (a_{\tau-1}, r_{\tau-1})$. Based on the representation g_τ of each snapshot G_τ, KID uses pointer networks to find boundaries i_τ^s, i_τ^e of a_τ and then embeds a_τ into the vector space of FEs to find an FE r_τ as its semantic role. The G_τ will be updated to $G_{\tau+1}$ with the new node (a_τ, r_τ) until the r_τ is the special dummy node in FKG. Figure 4 shows how to find a new node and add it into the FSG.

We first need to encode G_τ to its representation g_τ:

$$g_\tau = \text{Maxpooling}\left(z_0, \ldots, z_\tau\right) \tag{7}$$

$$z_0, \ldots, z_\tau = \text{GCN}^{\text{FSG}}\left(z_0^{(0)}, \ldots, z_\tau^{(0)}\right) \tag{8}$$

$$z_j^{(0)} = \begin{cases} \pi_t \oplus y_{i_f}, & j = 0 \\ \pi_{a_j} \oplus y_{i_{r_j}}, & j = 1, \ldots, \tau \end{cases} \tag{9}$$

Where i_f and i_{r_j} denotes indices of f and r_j in FKG, and $\pi_{a_j} = Q(i_j^s, i_j^e)$. The GCN module is to encode partial FSG.

4.4.1 Argument Identification

Based on g_τ, we need to find an argument $a_\tau = w_{i_\tau^s}, \ldots, w_{i_\tau^e}$. We use pointer networks to identify its boundaries i_τ^s and i_τ^e separately. Take i_τ^s as example:

$$\rho_\tau^s = \text{FFN}\left(g_\tau\right) \tag{10}$$

$$P\left(i_\tau^s | S, G_\tau\right) = \text{softmax}\left(H^\top \cdot \rho_\tau^s\right) \tag{11}$$

Where $H = (h_0, \ldots, h_{n-1}) \in \mathbb{R}^{d_h \times n}$ represents the output of the sentence encoder, and $\rho_\tau^s \in \mathbb{R}^{d_h}$ is used to find the start position of argument span a_τ.

4.4.2 Role Classification

Based on g_τ and a_τ, we embed a_τ into the vector space of FEs as $\gamma_{a_\tau} \in \mathbb{R}^{d_n}$. Similar to frame identification, we calculate dot product scores between γ_{a_τ} and all FEs $Y_R = (y_{|F|}, \ldots, y_{|F|+|R|}) \in \mathbb{R}^{d_n \times (|R|+1)}$ to get the conditional probability distribution of r given a_τ and G_τ.

$$\gamma_{a_\tau} = \text{FFN}(\pi_{a_\tau} \oplus g_\tau) \tag{12}$$

$$P\left(r_\tau | S, G_\tau, a_\tau\right) = \text{softmax}\left(Y_R^\top \cdot \gamma_{a_\tau}\right) \tag{13}$$

5 Training and Inference

5.1 Training

We train our model with all subtasks jointly. As the results of other subtasks are highly dependent on frame identification, we use the target with its gold frame as the initial node of FSG, $G_0 = (t, f^{gold})$. The frame semantic graph decoder is autoregressive, so the decoder expands FSG with node predicted by itself instead of using gold node.

$$\mathcal{L}^f = -\log P(f = f^{\text{gold}} | S, t) \tag{14}$$

$$\mathcal{L}_{s/e}^a = -\sum_{\tau=0}^{k-1} \log P(i_\tau^{s/e} = I_\tau^{s/e} | S, G_\tau) \tag{15}$$

$$\mathcal{L}^r = -\sum_{\tau=0}^{k} \log P(r_\tau = r_\tau^{\text{gold}} | S, G_\tau, a_\tau) \tag{16}$$

$$\mathcal{L} = \lambda_1 \mathcal{L}^f + \lambda_2(\mathcal{L}_s^a + \mathcal{L}_e^a) + \lambda_3 \mathcal{L}^r \tag{17}$$

Where $f^{\text{gold}}, I_\tau^s, I_\tau^e, r_\tau^{\text{gold}}$ are gold labels, and $G_{\tau+1} = G_\tau + (\hat{a}_\tau, \hat{r}_\tau)$. r_k^{gold} is "Dummy", indicating the end of the parsing.

5.2 Inference

KID predicts frame and all arguments with their roles in a sequential way. We use probabilities above with some constraints: 1. We use lexicon filtering strategy: for a target t, we can use the lemma l_t of it to find a subset of frames $F_{l_t} \subset F$ so that we can reduce the searching space; 2. Similarly, we take $R_{\hat{f}}$ instead of R as the set of candidate FEs; 3. In argument identification, we will mask spans that are already selected as arguments, and i_τ^e should be no less than i_τ^s.

6 Experiment

6.1 Datasets

We evaluate KID on two FrameNet datasets: FN 1.5 and FN 1.7.[1] FN 1.7 is an extension version of FN 1.5, including more fine-grained frames and more instances. FN 1.5 defines 1019 frames and 9634 FEs while FN 1.7 defines 1221 frames and 11428 FEs. We use the same splits of datasets as Peng et al. (2018), and we also follow Kshirsagar et al. (2015) to use exemplar instances to pretrain our model because these exemplar sentences extremely enrich the corpus. Table 1 shows the numbers of instances in two datasets.

	#exemplar	#train	#dev	#test
FN 1.5	153952	17143	2333	4458
FN 1.7	192461	19875	2309	6722

Table 1: Number of instances in two datasets.

6.2 Models

We compare KID with following baselines:

[1] https://FN.icsi.berkeley.edu/fndrupal/about

Model	Full structure Precision	Recall	F1-score	Arg only (with gold frame) Precision	Recall	F1-score
SEMAFOR (2014)	-	-	-	65.6	53.8	59.1
SEMAFOR (HI) (2015)	-	-	-	67.2	54.8	60.4
Hermann et al. (2014)	74.3	66.0	69.9	-	-	-
Täckström et al. (2015)	**75.4**	65.8	70.3	-	-	-
FitzGerald et al. (2015)	74.8	65.5	69.9	-	-	-
open-SESAME (2017)	71.0	67.8	69.4	**69.4**	60.5	64.6
KID (GloVe)	73.8	**76.8**	**75.3**	64.6	**68.2**	**66.4**
SEMAFOR (HI + exemplar) (2015)	-	-	-	66.0	60.4	63.1
Swayamdipta et al. (2018)	-	-	-	67.8	66.2	67.0
Marcheggiani and Titov (2020)	-	-	-	69.8	68.8	69.3
Peng et al. (2018)	**79.2**	71.7	75.3	-	-	-
Chen et al. (2021)	75.1	76.9	76.0	-	-	-
Yang and Mitchell (2017)	77.3	71.2	74.1	**70.2**	60.2	65.5
KID (GloVe + exemplar)	75.5	**80.1**	**77.7**	66.8	**73.7**	**70.1**
Bastianelli et al. (2020) (JL)	-	-	-	**74.6**	74.4	74.5
Chen et al. (2021) (BERT)	78.2	82.4	80.2	-	-	-
KID (BERT)	**79.3**	**84.2**	**81.7**	71.7	**79.0**	**75.2**

Table 2: Empirical results on FN 1.5. All models are single-task, non-ensemble. The upper block lists models trained without exemplar instances, the lower block lists models with pretrained language models. KID outperforms other models under all conditions. KID gets much higher recall because of the incremental strategy to identify arguments.

SEMAFOR: a widely-used open-resource statistical model proposed by Das et al. (2010, 2014).

SEMAFOR (HI): an improved version of **SEMAFOR** using exemplar instances and hierarchy features (FE mappings) proposed by Kshirsagar et al. (2015)

Hermann et al. (2014): a neural network-based model learning representations of words and frames.

Täckström et al. (2015): identifying arguments with a global graphical model.

FitzGerald et al. (2015): an extension of Täckström et al. (2015) learning neural representations of frames and FEs.

open-SESAME: a syntax-free open-resource semantic parser proposed by Swayamdipta et al. (2017).

Swayamdipta et al. (2018): an extension version of **open-SESAME** with multi-task and exemplar instances.

Yang and Mitchell (2017): a joint model integrating both sequential and relational models.

Peng et al. (2018): a joint model using latent structure variables.

Chen et al. (2021): a joint encoder-decoder model predicting arguments and roles sequentially.

Marcheggiani and Titov (2020): a GCN-based model over constituency trees.

Bastianelli et al. (2020) (JL) : a GCN-based model encoding syntactic constituency path. JL denotes joint learning on all subtasks of frame semantic parsing.

Kalyanpur et al. (2020): a T5-based model treating frame semantic parsing as a sequence-to-sequence generation task.

6.3 Empirical Results

We compare KID with models mentioned above on FN 1.5 and FN 1.7. We focus on two metrics: full structure F1 and arg F1.[2] Full structure F1 shows the performance of models on extracting full frame semantic structures from text and arg F1 evaluates the results of argument identification and role classification with gold frames.

Table 2 shows results on FN 1.5 and Table 3, 4 shows results on FN 1.7. For a fair comparison, we divide models into three parts: the first part of models do not use exemplar instances as training data; the second part of models use exemplar instances without any pretrained language models; the third part of models use both exemplar instances and pretrained language models. KID (GloVe) uses GloVe (Pennington et al., 2014) as word embeddings and KID (BERT) uses pretrained language models BERT (Devlin et al., 2019) and fine-tunes BERT to encode word representations. KID achieves state-of-the-art of two metrics on both datasets under all circumstances.

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

Model	Precision	Recall	F1-score
Peng et al. (2018)	78.0	72.1	75.0
KID (GloVe)	77.0	79.8	78.4
KID (BERT)	81.1	83.3	82.2

Table 3: Full structure F1 on FN 1.7.

Model	Precision	Recall	F1-score
open-SESAME (2017)	62	55	58
KID (GloVe)	69.2	73.3	71.2
Kalyanpur et al. (2020)	71	73	72
KID (BERT)	74.1	77.3	75.6

Table 4: Arg F1 on FN 1.7. Results of other models are obtained from Kalyanpur et al. (2020).

Model	Full structure F1	Arg F1
KID (GloVe)	75.28	66.35
w/o FSG	74.43	64.99
w/o FKG	74.60	64.96
w/o double-graph	73.34	63.41
KID (BERT)	79.44	71.59
w/o double-graph	77.77	68.77

Table 5: Ablation study on double-graph architecture. w/o denotes "without". w/o FSG uses LSTM as its decoder and w/o FKG does not use RGCN to encode frames and FEs. We also test the influence of double-graph architecture for KID (BERT).

Model	Full structure F1	Arg F1
KID (GloVe)	75.28	66.35
w/o frame-FE	74.84	65.70
w/o frame-frame	74.97	66.06
w/o intra-frame FE-FE	75.10	66.60
w/o inter-frame FE-FE	75.13	65.87
FI w/o FKG	75.00	65.61
w/o FKG	74.60	64.96

Table 6: Ablation study on structures of FKG. We remove each kind of relations of FKG and all get a drop of full structure F1. FI w/o FKG denotes not using FKG in frame identification (FI). w/o FKG uses input vectors of frame knowledge encoder directly.

It is worth noting that KID achieves much higher recall than other models. We attribute this to the incremental strategy of building FSG. By constructing FSG incrementally, KID can capture relations between arguments and identify arguments that are hard to find in other models.

6.4 Ablation Study

To prove the effectiveness of double-graph architecture, we conduct further experiments with KID on FN 1.5. Table 5 shows ablation study on double-graph architecture. w/o FSG uses LSTM instead of our frame semantic graph decoder. It takes a sequence of arguments and their roles that have already been identified as input to predict the next argument. FSG performs better than LSTM because it captures target-argument and argument-argument relations and can model long-distance dependencies. w/o FKG directly uses input vectors of frame knowledge graph encoder, and results also show that knowledge-enhanced representations are better than randomly initialized embeddings.

FKG is a multi-relational heterogeneous graph. The ablation study on structures of FKG is shown in Table 6. In addition, we evaluate the performance of FI w/o FKG, which identifies frames with a simple linear classification layer instead of FKG, and the results prove that FKG strengthens interactions between frame identification and role classification.

6.5 Transfer learning ability of FKG

As we have discussed in Figure 2, if frame B is related to frame A, a sentence with frame A can contribute to paring another sentence with frame B by inter-frame reasoning. Frame-frame and inter-frame FE-FE relations of FKG can guide KID to learn experience from other frames.

To confirm that FKG has ability of transfer learning, we design zero (few)-shot learning experiments on FN 1.7. Target word *get* can evoke multiple frames in FrameNet, and we choose instances including target *get* with three frames (***Arriving***, ***Getting*** and ***Transition_to_state***) as test instances. We only add few (or zero) instances including other targets with these frames in train and development sets and compare performance of KID with KID w/o FKG. If FKG has ability of transfer learning, KID with FKG can learn experience from other related frames like ***Receiving*** and its performance will not be influenced so much by the sparsity of labels.

Table 7 shows the results of our experiments. K=0 indicates zero-shot learning while K={4,16,32} indicates few-shot learning. KID without FKG performs much worse in zero-shot learning. As the number of instances that can be seen in training grows up, the performance of KID with FKG gets a steady increase while the performance of KID without FKG increases rapidly. Results verify our assumption that even with few

Model	K				
	0	4	16	32	full
KID (GloVe)	56.26	63.28	65.32	65.95	70.32
w/o FKG	0.00	50.70	56.40	57.59	63.94
Δ	56.26	12.58	8.92	8.36	6.38

Table 7: Experiments on confirming transfer learning ability of FKG. K denotes the number of instances of each frame in training set. Full means adding all instances of these frames except those including target *get* in train and development sets. Lack of labeled instances has much less impact on Arg F1 performance of KID with FKG, which confirms our assumption.

train instances FKG can guide inter-frame reasoning with its structure and allow models to learn experience from other seen frames.

7 Related Work

Frame semantic parsing has catched wide attention since it was released on SemEval 2007 (Baker et al., 2007). The task is to extract frame structures defined in FrameNet (Baker et al., 1998) from text. From then on, a large amount of systems are applied on this task, ranging from traditional machine learning classifiers (Johansson and Nugues, 2007; Das et al., 2010) to fancy neural models like recurrent neural networks (Yang and Mitchell, 2017; Swayamdipta et al., 2017) and graph neural networks (Marcheggiani and Titov, 2020; Bastianelli et al., 2020).

A lot of previous systems neglect interactions between subtasks and relations between arguments. They either focus on one or two subtasks (Hermann et al., 2014; FitzGerald et al., 2015; Marcheggiani and Titov, 2020) of frame semantic parsing or treat all subtasks independently (Das et al., 2014; Peng et al., 2018). Täckström et al. (2015) propose an efficient global graphical model, so they can enumerate all possible argument spans and treat the assignment as the Integer Linear Programming problem. Later systems like FitzGerald et al. (2015); Peng et al. (2018) follow this method. Swayamdipta et al. (2017); Bastianelli et al. (2020) use sequence-labeling strategy, and Yang and Mitchell (2017) integrate these two methods with a joint model. Only few approaches like Chen et al. (2021) model interactions between subtasks, which use the encoder-decoder architecture to predict arguments and roles sequentially. However, the sequence modeling of Chen et al. (2021) does not consider structure information and is not good at capturing long-distance dependencies. We use graph modeling to enhance structure information and strengthen interactions between target and argument, argument and argument.

Only a few systems utilize linguistic knowledge in FrameNet. Kshirsagar et al. (2015) use FE mappings to share information in FEs. In frame identification, Jiang and Riloff (2021) encode definitions of frames and Su et al. (2021) use frame identification and frame semantic relations. However, they do not utilize ontological frame knowledge in all subtasks while we construct a heterogeneous graph containing both frames and FEs. Besides, our model does not need extra encoders to encode definitions, which reduces parameters of the model.

Some systems also treat constituency parsing or other semantic parsing tasks like AMR as a graph construction problem. Yang and Deng (2020) use GCN to encode intermediate constituency tree to generate a new action on the tree. Cai and Lam (2019, 2020) construct AMR graphs with the Transformer (Vaswani et al., 2017) architecture.

8 Conclusion

In this paper, we incorporate knowledge into frame semantic parsing by constructing Frame Knowledge Graph. FKG provides knowledge-enhanced representations of frames and FEs and can guide intra-frame and inter-frame reasoning. We also propose frame semantic graph to represent frame semantic structures. We regard frame semantic parsing as an incremental graph construction problem. The process to construct FSG is structure-aware and can utilize relations between arguments. Our framework Knowledge-guided Incremental semantic parser with Double-graph (KID) achieves state-of-the-art on FrameNet benchmarks. However, how to utilize linguistic knowledge better is still to be resolved. Future work can focus on better modeling of ontological frame knowledge, which will be useful for frame semantic parsing and transfer learning in frame semantic parsing.

References

Collin Baker, Michael Ellsworth, and Katrin Erk. 2007. SemEval-2007 task 19: Frame semantic structure extraction. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pages 99–104, Prague, Czech Republic. Association for Computational Linguistics.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Emanuele Bastianelli, Andrea Vanzo, and Oliver Lemon. 2020. Encoding syntactic constituency paths for frame-semantic parsing with graph convolutional networks. arXiv preprint arXiv:2011.13210.

Deng Cai and Wai Lam. 2019. Core semantic first: A top-down approach for AMR parsing. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3799–3809, Hong Kong, China. Association for Computational Linguistics.

Deng Cai and Wai Lam. 2020. AMR parsing via graph-sequence iterative inference. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1290–1301, Online. Association for Computational Linguistics.

Xudong Chen, Ce Zheng, and Baobao Chang. 2021. Joint multi-decoder framework with hierarchical pointer network for frame semantic parsing. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 2570–2578, Online. Association for Computational Linguistics.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2014. Frame-semantic parsing. Computational Linguistics, 40(1):9–56.

Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic frame-semantic parsing. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 948–956, Los Angeles, California. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Charles J Fillmore and Collin F Baker. 2001. Frame semantics for text understanding. In Proceedings of WordNet and Other Lexical Resources Workshop, NAACL, volume 6.

Nicholas FitzGerald, Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. Semantic role labeling with neural network factors. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 960–970, Lisbon, Portugal. Association for Computational Linguistics.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. Computational linguistics, 28(3):245–288.

Shaoru Guo, Ru Li, Hongye Tan, Xiaoli Li, Yong Guan, Hongyan Zhao, and Yueping Zhang. 2020. A frame-based sentence representation for machine reading comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 891–896, Online. Association for Computational Linguistics.

Prakhar Gupta, Jeffrey Bigham, Yulia Tsvetkov, and Amy Pavel. 2021. Controlling dialogue generation with semantic exemplars. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3018–3029, Online. Association for Computational Linguistics.

Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. 2014. Semantic frame identification with distributed word representations. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1448–1458, Baltimore, Maryland. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural computation, 9(8):1735–1780.

Tianyu Jiang and Ellen Riloff. 2021. Exploiting definitions for frame identification. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 2429–2434, Online. Association for Computational Linguistics.

Richard Johansson and Pierre Nugues. 2007. LTH: Semantic structure extraction using nonprojective dependency trees. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pages 227–230, Prague, Czech Republic. Association for Computational Linguistics.

Aditya Kalyanpur, Or Biran, Tom Breloff, Jennifer Chu-Carroll, Ariel Diertani, Owen Rambow, and Mark Sammons. 2020. Open-domain frame semantic parsing using transformers. arXiv preprint arXiv:2010.10998.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Meghana Kshirsagar, Sam Thomson, Nathan Schneider, Jaime Carbonell, Noah A. Smith, and Chris Dyer. 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), pages 218–224, Beijing, China. Association for Computational Linguistics.

Diego Marcheggiani and Ivan Titov. 2020. Graph convolutions over constituent trees for syntax-aware semantic role labeling. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 3915–3928, Online. Association for Computational Linguistics.

Hao Peng, Sam Thomson, Swabha Swayamdipta, and Noah A. Smith. 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), pages 1492–1502, New Orleans, Louisiana. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In European semantic web conference, pages 593–607. Springer.

Xuefeng Su, Ru Li, Xiaoli Li, Jeff Z. Pan, Hu Zhang, Qinghua Chai, and Xiaoqi Han. 2021. A knowledge-guided framework for frame identification. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 5230–5240, Online. Association for Computational Linguistics.

Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. 2017. Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold. ArXiv, abs/1706.09528.

Swabha Swayamdipta, Sam Thomson, Kenton Lee, Luke Zettlemoyer, Chris Dyer, and Noah A. Smith. 2018. Syntactic scaffolds for semantic structures. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3772–3782, Brussels, Belgium. Association for Computational Linguistics.

Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. Efficient inference and structured learning for semantic role labeling. Transactions of the Association for Computational Linguistics, 3:29–41.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008.

Bishan Yang and Tom Mitchell. 2017. A joint sequential and relational model for frame-semantic parsing. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1247–1256, Copenhagen, Denmark. Association for Computational Linguistics.

Kaiyu Yang and Jia Deng. 2020. Strongly incremental constituency parsing with graph neural networks. arXiv preprint arXiv:2010.14568.

Hongyan Zhao, Ru Li, Xiaoli Li, and Hongye Tan. 2020. Cfsre: Context-aware based on frame-semantics for distantly supervised relation extraction. Knowledge-Based Systems, 210:106480.