Exploiting Definitions for Frame Identification

Tianyu Jiang and **Ellen Riloff**
School of Computing
University of Utah
Salt Lake City, UT 84112
{tianyu, riloff}@cs.utah.edu

Abstract

Frame identification is one of the key challenges for frame-semantic parsing. The goal of this task is to determine which frame best captures the meaning of a target word or phrase in a sentence. We present a new model for frame identification that uses a pre-trained transformer model to generate representations for frames and lexical units (senses) using their formal definitions in FrameNet. Our frame identification model assesses the suitability of a frame for a target word in a sentence based on the semantic coherence of their meanings. We evaluate our model on three data sets and show that it consistently achieves better performance than previous systems.

1 Introduction

Research on *frame semantics* has grown within the fields of natural language processing and cognitive science since the 1970s as the study of how we associate words and phrases with cognitive structures called *frames*, which characterize a small abstract scene or situation (Fillmore, 1976, 1982). The Berkeley FrameNet project (Baker et al., 1998) provides an online lexical database for frame semantics together with a corpus of annotated documents. Frame semantic parsing is the task of automatically extracting frame semantic structures from sentences. The process typically consists of three steps: *target identification*, which identifies frame-evoking predicates in the sentence; *frame identification*, which identifies the evoked frame for each target; and *argument identification*, which identifies arguments of a frame and labels them with semantic roles (frame elements). In this work, we focus on the frame identification problem.

FrameNet 1.7 contains more than 13,000 lexical units (a word lemma with a sense), each associated with a semantic frame. A polysemous word is associated with multiple lexical units (one for each sense), and is therefore linked to multiple frames. The frame identification task requires a system to identify the most relevant frame for a target word or phrase based on its sentence context. Here is an example:

> The pandemic has *sparked* a lot of problems for the economy.

Given the target word *sparked*, the goal is to determine which frame should be triggered. The word lemma *spark* has two senses in FrameNet: "with obj. ignite" and "provide the stimulus for". The former sense is associated with the *Setting_fire* frame and the latter one is associated with the *Cause_to_start* frame. The *Setting_fire* frame is defined as "this frame describes the creation of a flame...", and the *Cause_to_start* frame is defined as "a cause, animate or inanimate, causes a process, the effect, to begin". So *Cause_to_start* is the correct frame for this sentence.

Previous work has shown the success of using feature engineering with linear classification models (Johansson and Nugues, 2007) and discriminative probabilistic models (Das et al., 2010), which were later improved by applying distributed word representations and deep neural network models (Hermann et al., 2014). Syntactic information, typically dependency paths, has consistently played an important role in frame identification (Das et al., 2014; Peng et al., 2018).

Our work is motivated by the rich lexicographic information about frames and lexical units provided by the FrameNet database, which has not been fully utilized for the frame identification task. Recent advances in large pre-trained transformer models (Devlin et al., 2019) have demonstrated the ability to capture semantic meaning in dictionary definitions for the related problem of word sense disambiguation (Huang et al., 2019; Blevins and Zettlemoyer, 2020).
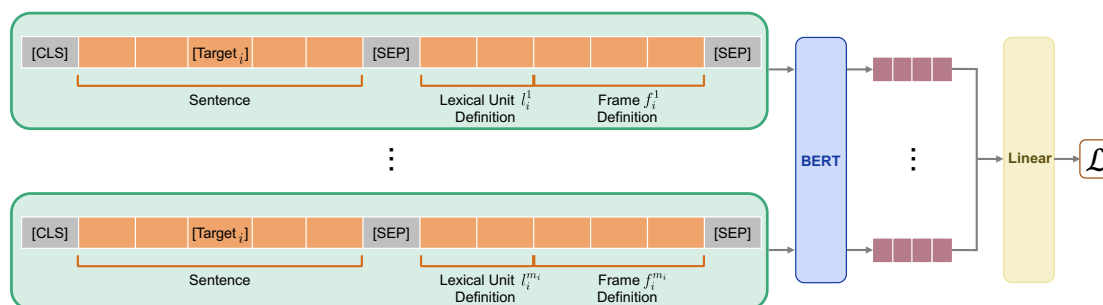
2429

Figure 1: Overview of the FIDO architecture. Each green block represents a different candidate pair (lexical unit, frame) for the same Target_i.

Our model uses the definitions of frames and lexical units in FrameNet as a source of knowledge to help assess the semantic coherence between the target word and candidate frames. Specifically, we utilize the contextual embeddings produced by the BERT (Devlin et al., 2019) model to determine if a candidate lexical unit and frame express the same meaning as the target word in the given context. Our model achieves state-of-the-art performance on two FrameNet datasets and a FrameNet-annotated dataset based on Yahoo! Answers. Our code is open-source and available online[1].

2 Related Work

There has been considerable work on the frame identification problem with respect to FrameNet, especially since the SemEval 2007 shared task (Baker et al., 2007). Johansson and Nugues (2007) used a SVM classifier to disambiguate frames with hand-crafted features. Das et al. (2010) applied feature-based discriminative probabilistic (log-linear) models for frame identification. Hermann et al. (2014) presented a method using distributed representations of predicates and their syntactic context by mapping input representations and frame representations to a common latent space using the WSABIE algorithm (Weston et al., 2011). Hartmann et al. (2017) built a simplified model based on Hermann et al. (2014) and achieved comparable results. They also released a new FrameNet-annotated test set based on user-generated web text from Yahoo! Answers. Yang and Mitchell (2017) integrated a bidirectional LSTM neural network and a relational network to jointly decode frames.

More recently, Botschen et al. (2018) brought in multimodal representations grounded in images to improve frame identification. Peng et al. (2018)

proposed a joint inference formulation that learns semantic parsers from multiple datasets.

In contrast to the previous models, our model does not rely on syntactic features. We assess semantic coherence directly from the input sentence and definitions in FrameNet.

Another line of related work is learning embeddings from dictionary definitions. It has been shown that neural networks can extract semantic information from dictionary definitions (Kumar et al., 2019; Bosc and Vincent, 2018). Recent work in word sense disambiguation (Huang et al., 2019; Blevins and Zettlemoyer, 2020) has demonstrated that providing pre-trained language models with sense definitions (glosses) can be effective. Yong and Torrent (2020) also used the sense definitions of lexical units for their research on frame induction. Our model adopts a similar architecture as Huang et al. (2019), but we focus on the frame identification task and we explore the use of both lexical unit and frame definitions for this task.

3 Method

Given a sentence and a target word or phrase, the frame identification task assigns the most relevant frame to the target according to the sentence context. Figure 1 shows the framework of our model called FIDO (**F**rame **I**dentification with **D**efiniti**O**ns). Our system takes the sentence and the definitions of associated lexical units (senses) and their frames as input to the BERT model, as indicated by the green blocks. Each green block represents the target word in the sentence, one of its senses, and that sense's associated frame in FrameNet. Then we use the output vectors to produce a probability distribution over all of the candidate frames. We select the frame with the maximum probability as the answer.

---

[1] https://github.com/tyjiangU/fido

3.1 Notation

We denote the ith example ($i = 1, 2, ..., n$) consisting of a sentence and designated target word or phrase as $\langle s, t \rangle_i$, its correct frame as f_i^*, the set of lexical units associated with the target as $l_i^1, l_i^2, ..., l_i^{m_i}$, and their corresponding frames as $f_i^1, f_i^2, ..., f_i^{m_i}$ (f_i^* is among them). We seek to estimate the probability of the jth frame being the correct frame by:

$$\Pr(f_i^j | \langle s, t \rangle_i) = \frac{\exp(g(\langle s, t \rangle_i, f_i^j))}{\sum_{k=1}^{m_i} \exp(g(\langle s, t \rangle_i, f_i^k))} \quad (1)$$

where $g(\cdot)$ is a function produced by our model for scoring the assignment of a frame to the sentence and target. We use negative log likelihood as our loss function:

$$\mathcal{L} = -\sum_{i=1}^{n} \log \Pr(f_i^* | \langle s, t \rangle_i) \quad (2)$$

where n is the total number of training examples.

3.2 Modeling

FrameNet provides unique definitions for each lexical unit (LU) and frame. A LU is a pairing of a word lemma and a meaning (sense). Determining the correct LU (sense) uniquely determines the correct frame because each sense of a polysemous word is linked to a different frame. For example, the word *cut* can trigger different frames depending on its meaning (the definition sentences follow the bold lexical unit or frame names), as shown in Table 1.

Lexical Unit	Associated Frame
cut.n: the way or style in which a garment or the hair is cut	**Hair_configuration**: temporary or permanent styles and configurations of hair
cut.v: divide into pieces with a knife or other sharp implement	**Cutting**: an agent cuts an item into pieces using an instrument

Table 1: Examples of lexical units and their associated frames.

We use the BERT (Devlin et al., 2019) model as the base of our architecture to produce the function $g(\cdot)$ as described in Eq (1). For each target, first we extract LUs from FrameNet that have the same lemma and their corresponding frames to form a set of candidate (LU, Frame) pairs. Our goal is to predict whether the target in the sentence has the same meaning as the definitions of a candidate LU and its associated frame.

As input to the BERT model, we use the sentence as the first sequence and concatenate a LU definition and frame definition as the second sequence. Each definition starts with the LU name or frame name and a colon, followed by the definition description.

Instead of using the output vector of the `[CLS]` token as is typical, we use the last hidden vector of the target word as output (if there is more than one token, we only use the first one). By passing the output vector through a linear layer, we then get a score for assigning a candidate frame to the sentence and target. Finally the scores for all candidate frames are passed through the softmax function to get the probabilities in Eq (1).

4 Experiments

4.1 Datasets

FrameNet: To compare FIDO with previous systems, we evaluate our model on FrameNet (FN) 1.5 using the same train/dev/test data split as Das et al. (2014). We also evaluate our model on FN 1.7 which has been available since 2016 and contains nearly 20% more gold annotated data than FN 1.5. We use the same data split as Swayamdipta et al. (2017) for FN 1.7. Table 2 shows the number of examples in each split.

	FN 1.5	FN 1.7	YAGS
Train	15,017	19,391	-
Dev	4,463	2,272	1,000
Test	4,457	6,714	2,093

Table 2: Dataset sizes.

YAGS: YAGS (Hartmann et al., 2017) is a FrameNet-annotated test set based on question answering data from Yahoo! Answers[2], a community-driven question-and-answer forum. The annotations are based on FN 1.5. We train on FN 1.5 and evaluate on the YAGS test set to compare results with Hartmann et al. (2017).

---

[2] https://webscope.sandbox.yahoo.com/

Model	Accuracy
Hermann et al. (2014)	88.4
Hartmann et al. (2017)	87.6
Yang and Mitchell (2017)	88.2
Open-SESAME (2017)	86.9
Botschen et al. (2018)	88.8
Peng et al. (2018)	90.0
FIDO	**91.3**

Table 3: Accuracy on FN 1.5.

Dataset	Model	Accuracy
FN 1.7	Peng et al. (2018)	89.1
	FIDO	**92.1**
YAGS	Hartmann et al. (2017)	62.5
	FIDO	**70.5**

Table 4: Accuracy on FN 1.7 and YAGS.

Model	FN 1.5	FN 1.7
FIDO	**91.3**	**92.1**
FIDO (FRdef only)	90.1	91.1
FIDO (LUdef only)	88.9	90.7
FIDO (NO def)	80.3	79.4
FIDO (CLS)	89.3	90.5

Table 5: Ablation study on FN 1.5 and FN 1.7.

4.2 Training Details

We use the pre-trained uncased BERT$_{\mathrm{BASE}}$ model with the same settings as Devlin et al. (2019) and fine-tune on our training data. We set the max sequence length as 300, batch size as 16, learning rate started at 2e-5, and train for 5 epochs. All reported results are averaged over 3 runs with random seeds.

4.3 Results

Table 3 compares our model with previous methods on the FN 1.5 dataset. Hermann et al. (2014), Hartmann et al. (2017), and Open-SESAME (Swayamdipta et al., 2017) use distributed representations and syntactic features with neural networks. Botschen et al. (2018) extends Hartmann et al. (2017) with visual embeddings. Yang and Mitchell (2017) integrates a sequential and relational network for joint learning. Peng et al. (2018) has achieved the best prior results on frame identification using a multitask approach to learn semantic parsers from disjoint corpora. It is worth noting that besides the FN 1.5 training set, they also use 153,952 exemplar sentences for training, which is more than 10 times the size of our training data. FIDO achieves better performance than all of the prior systems.

Table 4 shows our results compared to Peng et al. (2018) on FN 1.7. FIDO achieves a 3.0% absolute accuracy gain on this data set. The YAGS data set (Hartmann et al., 2017) contains *unknown* targets

that do not have related LUs in FN 1.5 and also *unlinked* targets (i.e., the provided gold frame does not belong to the set of frames associated with this target in FN). Our model is not able to make a correct prediction for these cases based on its design. There are 122 unknown or unlinked targets in the test set, on which our model will get a zero score. Despite this limitation, our model still outperforms Hartmann et al. (2017), which demonstrates its ability to generalize across text genres.

4.4 Analysis

We performed an ablation study to assess the contributions of each part of our model. In Table 5, the first row shows the results for our complete FIDO model. Rows 2-3 show results when using only the definitions of frames (FRdef only) or LUs (LUdef only). We see that the frame definitions contribute the most to performance. Using the LU definitions alone on FN 1.7 also achieves quite good results. But combining both definitions together yields better results than either one alone.

In order to tease apart the impact of the definitions from the impact of BERT, we did an experiment replacing each definition simply with the name of the frame or LU. These results appear in the FIDO (NO def) row. Removing the definitions results in a large performance drop. The definitions clearly play a major role.

In the bottom row, we show the results of experiments using the output vector of the `[CLS]` token (all other settings the same), which did not perform as well as using the target token. This is not surprising as `[CLS]` aggregates the entire sequence representation rather than focusing on the target.

Previous work also reported accuracy on ambiguous cases (i.e., when the target word is associated with multiple frames), which more directly shows the model's ability to disambiguate frames. However, the set of ambiguous targets is different across papers. To avoid comparing apples and oranges,

Dataset	Model	Amb1	Amb2
FN 1.5	Peng et al. (2018)	78.0	-
	FIDO	**81.0**	83.6
FN 1.7	Peng et al. (2018)	77.5	-
	FIDO	**83.8**	85.9

Table 6: Accuracy on ambiguous cases.

we report accuracy on two different sets of ambiguous targets. In Table 6, the **Amb1** column follows Peng et al. (2018), which uses the gold LU's part-of-speech (POS) tag to form the candidate frame list. In this setting, if a target has just one sense when its POS is known, it is not considered to be ambiguous. Our model outperforms Peng et al. (2018) on both FN 1.5 and FN 1.7 datasets. The **Amb2** column shows the accuracy of ambiguous targets using only the lemma of the target (i.e., not relying on gold POS tags). We encourage future work to articulate which setting is used.

We also analyzed whether unseen frames and unseen targets were a major source of errors for our model. On FN 1.7, our FIDO model achieved 92.1% accuracy, so it mislabeled 7.9% of the test cases. We found that 1.4% of the test cases were mislabeled and had an unseen frame (i.e., the gold frame was not seen with the target in the training data), and 0.52% of the test cases were mislabeled and had an unseen target (i.e., the target was not seen in the training data). Therefore only about 1/4 of the FIDO errors were due to unseen frames and unseen targets. We conclude that even for frames and targets that appear in the training data, there is still substantial room for improvement on this task.

5 Conclusion

We tackled the frame identification problem by assessing the semantic coherence between the meaning of a target word in a sentence, and a candidate frame. Our model exploits the frame and lexical unit definitions provided by FrameNet and a pre-trained transformer model to generate semantic representations. The experiments show that this model achieves better performance than previous systems on two versions of FrameNet data and the YAGS dataset. Our work has demonstrated that a relatively simple architecture that brings together pre-trained language models with frame and sense definitions can produce a highly effective system for frame identification.

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 *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (ACL-COLING 1998), Volume 1*, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Terra Blevins and Luke Zettlemoyer. 2020. Moving down the long tail of word sense disambiguation with gloss informed bi-encoders. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, pages 1006–1017, Online. Association for Computational Linguistics.

Tom Bosc and Pascal Vincent. 2018. Auto-encoding dictionary definitions into consistent word embeddings. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*, pages 1522–1532, Brussels, Belgium. Association for Computational Linguistics.

Teresa Botschen, Iryna Gurevych, Jan-Christoph Klie, Hatem Mousselly-Sergieh, and Stefan Roth. 2018. Multimodal frame identification with multilingual evaluation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018)*, pages 1481–1491, New Orleans, Louisiana. Association for Computational Linguistics.

Dipanjan Das, Desai Chen, André FT 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 (NAACL 2010)*, 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 (NAACL 2019)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Charles J Fillmore. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280 (1), pages 20–32.

Charles J Fillmore. 1982. Frame semantics. In *Linguistics in the Morning Calm, The Linguistic Society of Korea (ed.)*, pages 111–137. Seoul: Hanshin Publishing Company.

Silvana Hartmann, Ilia Kuznetsov, Teresa Martin, and Iryna Gurevych. 2017. Out-of-domain FrameNet semantic role labeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)*, pages 471–482, Valencia, Spain. 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 (ACL 2014)*, pages 1448–1458, Baltimore, Maryland. Association for Computational Linguistics.

Luyao Huang, Chi Sun, Xipeng Qiu, and Xuanjing Huang. 2019. GlossBERT: BERT for word sense disambiguation with gloss knowledge. 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 2019)*, pages 3509–3514, Hong Kong, China. 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.

Sawan Kumar, Sharmistha Jat, Karan Saxena, and Partha Talukdar. 2019. Zero-shot word sense disambiguation using sense definition embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, pages 5670–5681, Florence, Italy. 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 (NAACL 2018)*, pages 1492–1502, New Orleans, Louisiana. 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 preprint arXiv:1706.09528*.

Jason Weston, Samy Bengio, and Nicolas Usunier. 2011. Wsabie: Scaling up to large vocabulary image annotation. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011)*, page 2764–2770. AAAI Press.

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 (EMNLP 2017)*, pages 1247–1256, Copenhagen, Denmark. Association for Computational Linguistics.

Zheng Xin Yong and Tiago Timponi Torrent. 2020. Semi-supervised deep embedded clustering with anomaly detection for semantic frame induction. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 3509–3519, Marseille, France. European Language Resources Association.