Utilizing Pseudo Dialogue in Conversational Semantic Frame Analysis

Shiho Matta, Yin Jou Huang, Hirokazu Kiyomaru, Sadao Kurohashi
Kyoto University
{matta, huang, kiyomaru, kuro}@nlp.ist.i.kyoto-u.ac.jp

Abstract

Semantic frame analysis aims to extract structural knowledge from unstructured texts. A semantic frame analyzer is usually trained using manually annotated dialogue data. However, it is costly to gather and annotate dialogue data. In this paper, we propose a method to create pseudo data in addition to real training data to improve the performance of semantic frame analysis. Experiments showed that compared to only using real training data, our method successfully improved the performance of semantic frame analysis on a cooking dialogue corpus by incorporating pseudo data.

1 Introduction

When we try to understand what important events are mentioned in a dialogue context, we need to identify words that represent those events. A semantic frame is a knowledge structure that represents a specific event mentioned in the context [1]. The main action of the event is indicated by a **trigger**, and details about the action are supplemented by **arguments**.

In this paper, we focus on the task of semantic frame analysis [2] to extract events from dialogues in Japanese. Figure 1 shows an utterance from a dialogue with its semantic frames. In this piece of dialogue, the trigger is "休ませて," and it has three arguments: "タルト生地," "最低でも 1 時間," and "冷蔵庫" with their types labeled respectively. Semantic frame analysis aims at identifying these components in a given context.

One of the main challenges of semantic frame analysis is the paucity of data. However, collecting a large amount of dialogue data costs more money and time compared to monologue and written-style data, not to mention we have to properly label the data with semantic frames to use them as training data.

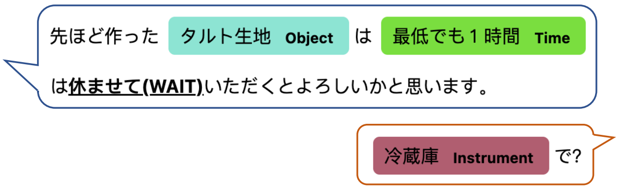

先ほど作った タルト生地 Object は 最低でも１時間 Time
は休ませて(WAIT)いただくとよろしいかと思います。
冷蔵庫 Instrument で?

Figure 1: An example of dialogue annotated with semantic frames.

To solve the aforementioned problem, we propose a method to generate pseudo dialogues with semantic frame annotation as extra training data (Figure 2). Our proposed method contains three key components: the frame analyzer, the pseudo dialogue generator, and the pseudo annotation mechanism. First, the frame analyzer extracts semantic frames from written-style texts. Second, the pseudo dialogue generator is given semantic frames and generates pseudo dialogue based on them. Last, we apply pseudo annotation on the pseudo dialogue by identifying the location of triggers and arguments.

To apply our method, we first train the frame analyzer and the pseudo dialogue generator with a moderate-sized annotated dataset so they could learn how to extract semantic frames from texts and generate pseudo dialogues based on them. Once the models are trained, we feed texts into this pipeline and obtain pseudo annotated data. We refer to this synthesized data as **silver data** and the data used to initialize the method as **gold data**.

Experiments show that compared to the baseline model trained only on gold data, a performance gain is achieved by incorporating annotated pseudo dialogues as silver data.

2 Related Work

Our method can be seen as a data augmentation method as it increases the number and diversity of training examples without explicitly collecting new data [3]. Dai et al. [4] propose to generate pseudo dialogues as data augmentation

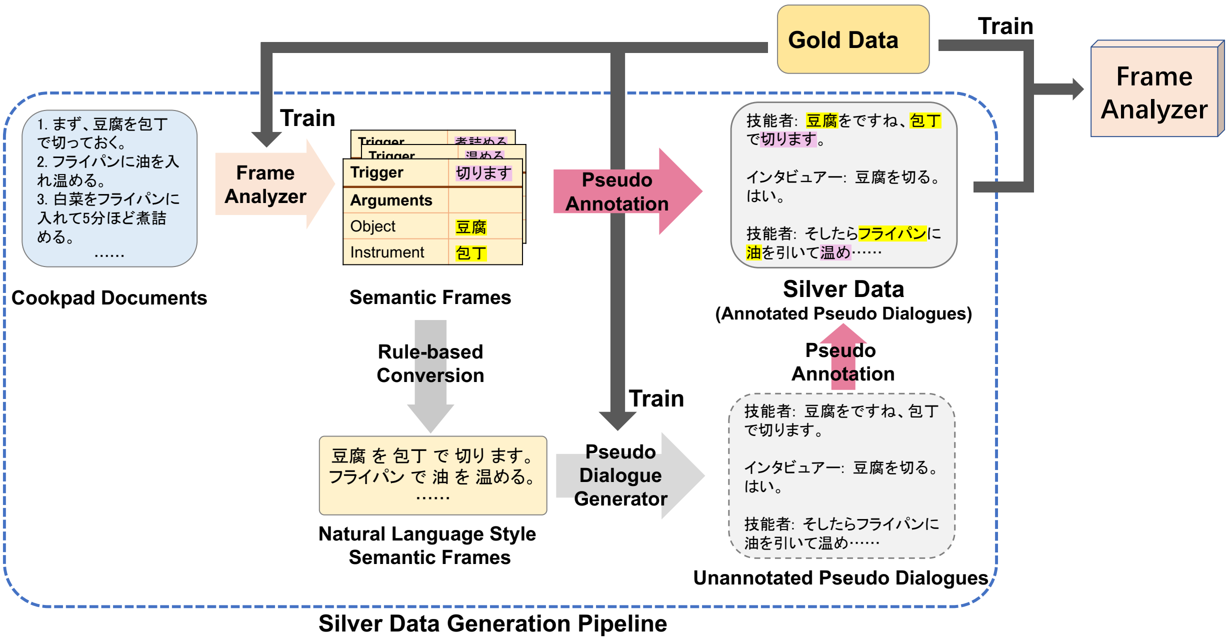

Gold Data
Train
Frame Analyzer
Train
1. まず、豆腐を包丁で切っておく。
2. フライパンに油を入れ温める。
3. 白菜をフライパンに入れて5分ほど煮詰める。
……
Cookpad Documents
Frame Analyzer
Trigger 煮詰める
Trigger 温める
Trigger 切ります
Arguments
Object 豆腐
Instrument 包丁
Semantic Frames
Pseudo Annotation
技能者: 豆腐をですね、包丁で切ります。
インタビュアー: 豆腐を切る。はい。
技能者: そしたらフライパンに油を引いて温め……
Silver Data
(Annotated Pseudo Dialogues)
Rule-based Conversion
Train
Pseudo Annotation
豆腐 を 包丁 で 切り ます。
フライパン で 油 を 温める。
……
Natural Language Style Semantic Frames
Pseudo Dialogue Generator
技能者: 豆腐をですね、包丁で切ります。
インタビュアー: 豆腐を切る。はい。
技能者: そしたらフライパンに油を引いて温め……
Unannotated Pseudo Dialogues
Silver Data Generation Pipeline

Figure 2: Overview of the proposed method.

for conversational question answering. However, due to the difference in the target tasks, it is not directly applicable to semantic frame analysis. Ding et al. [5] represent labeled data by intertwining words and word-level tags and generate synthetic labeled data for sequence labeling tasks. However, the labeled sequence is drastically different from natural language text, making it difficult to take full advantage of the pre-trained language models [6].

3 Proposed Method

As shown in Figure 2, our pipeline for producing silver data consists of three parts: the frame analyzer, the pseudo dialogue generator, and the pseudo annotation mechanism. In this section, we explain how we construct pseudo data using this pipeline, as well as how we use the pseudo data to enhance the frame analyzer's performance.

3.1 Semantic Frame Analysis

Semantic frame analysis is the task of locating triggers and arguments in a given text and identifying their types. The semantic frame analysis component, which we refer to as the frame analyzer, takes a piece of text as input. We first locate triggers in the text (trigger detection), then for each trigger, we locate its arguments (argument detection). Both trigger detection and argument detection are formulated as sequence labeling problems with the BIO tagging scheme. For the baseline, we train a RoBERTa [7] base model on gold data in which the dialogues are annotated with triggers and arguments.

Once the frame analyzer has been trained, we feed it with written-style monologue texts, which are abundant, and obtain a large number of semantic frames. These semantic frames serve as seeds for pseudo dialogues.

3.2 Pseudo Dialogue Generation

For the pseudo dialogue generation part, we fine-tune a BART model [8] using semantic frame and dialogue pairs from the gold data so it takes semantic frames as the input and outputs pseudo dialogue. As semantic frames are structural data, we serialize them to feed them into the model by a rule-based conversion method. We show an example in Figure 2 (see Natural Language Style Semantic Frames). Once it is fine-tuned, we feed it with the semantic frames extracted by the frame analyzer and generate pseudo dialogues. Note that at this point, the pseudo dialogues are plain text in dialogue form and do not contain any annotation.

3.3 Pseudo Label Assignment

We need to apply pseudo annotations to pseudo dialogues to use them as silver data for the frame analyzer. This involves identifying the triggers and arguments present in the dialogues, which are components of the semantic frames used to generate them.

Trigger To align components in semantic frames and tokens in pseudo dialogues, considering the various conjugation forms of Japanese predicates (triggers), we use a two-step method: parsing triggers with a Japanese morphological analyzer and obtaining its base form, then parsing the pseudo dialogue also using the tool and checking if any token's base form matches the trigger's base form. If none match, we discard the semantic frame as the trigger is at the core of a semantic frame.

Argument To locate the arguments of a trigger in the pseudo dialogue that may have undergone slight modifications during the generation process, we perform a string similarity search in a certain span before and after the trigger for each argument. The nearest candidate to the trigger with the highest similarity score will be selected. If the score is below a threshold, the candidate is discarded to avoid matching to something completely irrelevant.

3.4 Pseudo Data Construction and Incorporation

To mass-produce silver data, we feed written-style monologue texts into the frame analyzer trained on gold data, generate pseudo dialogues and apply pseudo annotation to them. Instead of feeding an entire dialogue as the input into the frame analyzer, we divide them by utterance for trigger detection. For argument detection, we divide them by the range in which we search for arguments around a trigger. Note that we also divide the gold data into segments using this method.

To utilize silver data to enhance the frame analyzer's performance, we first train it on silver data until the validation loss stops improving, then we continue training it using the gold data.

4 Experiment

We conduct experiments on the culinary interview dialogue corpus dataset to investigate the effectiveness of the proposed method.

4.1 Dataset

We use the Culinary Interview Dialogue Corpus (CIDC) dataset [9] as gold data. The CIDC dataset contains dialogues between an interviewer and an expert exchanging information on how to make a dish. A piece of data in the dataset contains utterances from both people, and the semantic frames inside it are manually annotated. The CIDC dataset contains 308 dialogues with an average length of 2,061 words. There are 11 types of triggers and 5 types of arguments defined in the dataset.

As an external source to acquire a large number of data for pseudo dialogue generation, we use the Cookpad recipes [10], aiming to produce semantic frames with diversity. Each recipe has the "steps" section that explains how the dish is made. We feed the part into the frame analyzer. Unlike the CIDC dataset containing dialogues, Cookpad recipes are written-style texts. Even though the frame analyzer is trained only on dialogue data, it works with Cookpad recipes without major drawbacks because written texts are usually more straightforward than spoken language. We have up to 1.6 million Cookpad recipes. 100k recipes can eventually be converted to roughly 493k pieces of training data for trigger detection and 273k for argument.

4.2 Experimental Settings

We used a Japanese RoBERTa base model[1] as the backbone of the frame analyzer. We split the gold data, which contains 308 dialogues, and used 278 dialogues for training, 15 for validation, and 15 for testing. After being divided into smaller pieces for input (see Section 3.4), the gold training data contains 57k for trigger detection and 19k for argument detection. We prepared silver data of different sizes, starting from 1 time the size of the gold data up to 16 times to see to what extent the frame analyzer could benefit. For the pseudo dialogue generator, we fine-tuned a Japanese BART[2] large model on the CIDC dataset. We segmented the dialogues in it using a heuristic method, which resulted in 1,391 dialogue sessions.

We compared the baseline model trained on gold data only and our proposed model that is first trained on silver data, then on gold data. We evaluated them by calculating the weighted f1-score of trigger detection and argument detection.

4.3 Result and Analysis

Table 1 and Table 2 show the result of semantic frame analysis using different sizes of silver data. We noticed that incorporating even a small size of silver data can bring

[1] https://huggingface.co/nlp-waseda/roberta-base-japanese

[2] https://nlp.ist.i.kyoto-u.ac.jp/?BART%E6%97%A5%E6%9C%AC%E8%AA%9EPretrained%E3%83%A2%E3%83%87%E3%83%AB

Table 1: Performance of trigger detection.

Training data	Weighted F1 (±std)
Gold (baseline)	0.598 ± 0.023
Gold + Silver (1x)	0.634 ± 0.010
Gold + Silver (2x)	0.632 ± 0.014
Gold + Silver (4x)	0.640 ± 0.014
Gold + Silver (8x)	**0.644** ± 0.010
Gold + Silver (16x)	0.637 ± 0.010

Table 2: Performance of argument detection.

Training data	Weighted F1 (±std)
Gold (baseline)	0.508 ± 0.013
Gold + Silver (1x)	0.534 ± 0.013
Gold + Silver (2x)	0.533 ± 0.011
Gold + Silver (4x)	0.544 ± 0.012
Gold + Silver (8x)	**0.544** ± 0.010
Gold + Silver (14x)	0.543 ± 0.009

reasonable performance gain. The best weighted f1-score was achieved when we used silver data 8 times the size of gold data.

5 Case Study

We looked into detection results by the frame analyzer.

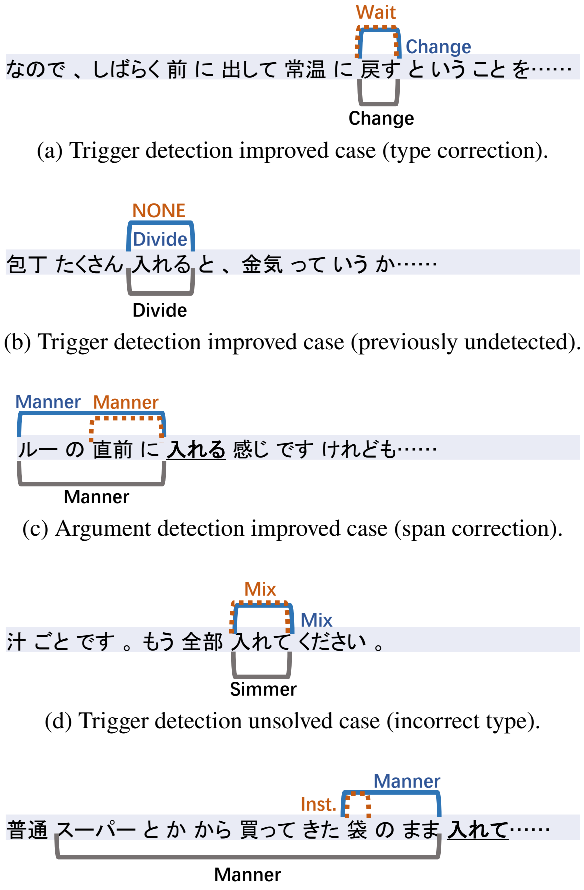
Improved Cases As shown in Figure 3a, the proposed model predicted the trigger with a proper type label. This example suggests that our model looked further into the context of the sentence.

We found that the proposed model can recognize targets that the baseline model cannot. In Figure 3b, given the context, "入れる" in "包丁を入れる" can be interpreted as "to cut something with a knife," so it does indicate a cooking event. The proposed model also recognizes more instances of argument types **TIME**, **TEMPERATURE**, and **MANNER**.

We noticed that both models struggle and make discontinuous or partially incomplete predictions when it comes to target arguments that have spans of more than three tokens. Nevertheless, our proposed model does slightly better in long-span targets (Figure 3c).

Unsolved Cases An unsolved case of trigger detection is shown in Figure 3d. In this case, although **SIMMER** is the right type considering the context of the whole dialogue, the input is too short for the frame analyzer to make the right decision since trigger detection is done on the utterance level. We plan to extend the input length for trigger detection to provide further information to the frame analyzer to resolve this problem.

As mentioned above, long-span targets are harder tasks. In the example in Figure 3e, both models could not solve the case perfectly.


Wait
Change
なので 、しばらく 前 に 出して 常温 に 戻す と いう こと を……
Change
NONE
Divide
包丁 たくさん 入れる と 、金気 って いう か……
Divide
Manner
Manner
ルー の 直前 に 入れる 感じ です けれども……
Manner
Mix
Mix
汁 ごと です 。もう 全部 入れて ください 。
Simmer
Manner
Inst.
普通 スーパー と か から 買って きた 袋 の まま 入れて……
Manner

(a) Trigger detection improved case (type correction).

(b) Trigger detection improved case (previously undetected).

(c) Argument detection improved case (span correction).

(d) Trigger detection unsolved case (incorrect type).

(e) Argument detection unsolved case (incorrect type and span).

Figure 3: Improved and unsolved cases. The blue line is the baseline and the orange dotted line is the proposed model. The gray line is the true label.

6 Conclusion

In this paper, we proposed a method to create pseudo data in addition to real data to help improve the performance of semantic frame analysis. Experimental results showed that we successfully enhanced the overall performance of the frame analyzer by incorporating pseudo data. We believe this kind of data augmentation will benefit tasks with low quantity of training materials.

Acknowledgement

This paper is based on results obtained from a project, JPNP20006, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

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

[2] Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. Frame-semantic parsing. Computational Linguistics, Vol. 40, No. 1, pp. 9–56, March 2014.

[3] Steven Y. Feng, Varun Gangal, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura, and Eduard Hovy. A survey of data augmentation approaches for NLP. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 968–988, Online, August 2021. Association for Computational Linguistics.

[4] Zhuyun Dai, Arun Tejasvi Chaganty, Vincent Y Zhao, Aida Amini, Qazi Mamunur Rashid, Mike Green, and Kelvin Guu. Dialog inpainting: Turning documents into dialogs. In Proceedings of the 39th International Conference on Machine Learning, pp. 4558–4586. PMLR, July 2022.

[5] Bosheng Ding, Linlin Liu, Lidong Bing, Canasai Kruengkrai, Thien Hai Nguyen, Shafiq Joty, Luo Si, and Chunyan Miao. DAGA: Data augmentation with a generation approach for low-resource tagging tasks. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 6045–6057, Online, November 2020. Association for Computational Linguistics.

[6] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. OpenAI blog, Vol. 1, No. 8, p. 9, 2019.

[7] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

[8] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics.

[9] Taro Okahisa, Ribeka Tanaka, Takashi Kodama, Yin Jou Huang, and Sadao Kurohashi. Constructing a culinary interview dialogue corpus with video conferencing tool. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 3131–3139, Marseille, France, June 2022. European Language Resources Association.

[10] Jun Harashima, Michiaki Ariga, Kenta Murata, and Masayuki Ioki. A large-scale recipe and meal data collection as infrastructure for food research. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 2455–2459, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).