AISFG: Abundant Information Slot Filling Generator

Yang Yan[*], **Junda Ye**[*], **Zhongbao Zhang, Liwen Wang**
Beijing University of Posts and Telecommunications, Beijing, China
{yanyang42,jundaye,zhongbaozb,w_liwen}@bupt.edu.cn

Abstract

As an essential component of task-oriented dialogue systems, slot filling requires enormous labeled training data in a certain domain. However, in most cases, there is little or no target domain training data is available in the training stage. Thus, cross-domain slot filling has to cope with the data scarcity problem by zero/few-shot learning. Previous researches on zero/few-shot cross-domain slot filling focus on slot descriptions and examples while ignoring the slot type ambiguity and example ambiguity issues. To address these problems, we propose **A**bundant **I**nformation **S**lot **F**illing **G**enerator (**AISFG**), a generative model with a novel query template that incorporates domain descriptions, slot descriptions, and examples with context. Experimental results show that our model outperforms state-of-the-art approaches in zero/few-shot slot filling task.[1]

1 Introduction

Slot filling is a critical part of downstream tasks in natural language understanding (NLU) such as dialogue systems. Recently, some supervised slot filling models have achieved state-of-the-art results within deep learning (Mesnil et al., 2013; Yao et al., 2013; Louvan and Magnini, 2018; Kim et al., 2019).Nonetheless, these methods have a strong dependency on the domain-specific labels, which is not capable of transferring to new domains that always contain little or no data.

To alleviate the problem of resource gap between source and target domains, cross-domain zero-shot has become an important research direction in slot filling. However, most researches on slot filling utilize token-level classification frameworks, which means they either convert it to a BIO tag labeling task (Shah et al., 2019; Bapna et al., 2017; Du et al., 2021) or predict the start and end position as QA task in the sentence (Du et al., 2021; Yu et al., 2021a), to extract spans. Recently, prompt learning methods reformulate the downstream tasks to a similar form with the pre-training tasks, which can fully utilize the knowledge encoded in PLMs and improve the performance of the downstream tasks in the scenarios of data scarcity. Motivated by this, we consider making the slot filling task consistent with PLMs pre-training tasks.

In this paper, we propose a generative template-based zero-shot slot filling framework named Abundant Information Slot Filling Generator (AISFG), which utilizes pre-trained generative model as the backbone and generates responses in natural language style. Thus, the slot filling task is consistent with the pre-training task of the PLM. In particular, we notice that shared cross-domain slot types sometimes refer to totally different entities across different domains. For example, in the SNIPS dataset (Coucke et al., 2018), domains *RateBook* and *SearchScreeningEvent* share a common slot type *object_type*, but it refers to book type and movie schedule, respectively. We call this slot type ambiguity issue. Moreover, we argue that incorporating slot examples only is far from fully utilizing the example information. For instance, for sentence *give 5 out of 6 stars to creatures of light and darkness*, we give the slot description to find *best rating* with examples *like 6 and 5*, the model produces the wrong answer 5 but not 6. We conjecture that, the model just learn that predicting a number is satisfying from these two examples but does not understand what is the best rating means. We call this example ambiguity issue. Thus, we attach our attention to domain-specific descriptions and examples with context to alleviate the above two issues. Specifically, we design the query template by incorporating domain descriptions, slot descriptions and examples with context.

The contribution of this paper can be summarized in three aspects:

[*]Contributed equally.
[1]The source code is available at https://github.com/realyanyang/AISFG.

Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 4180 - 4187
July 10-15, 2022 ©2022 Association for Computational Linguistics

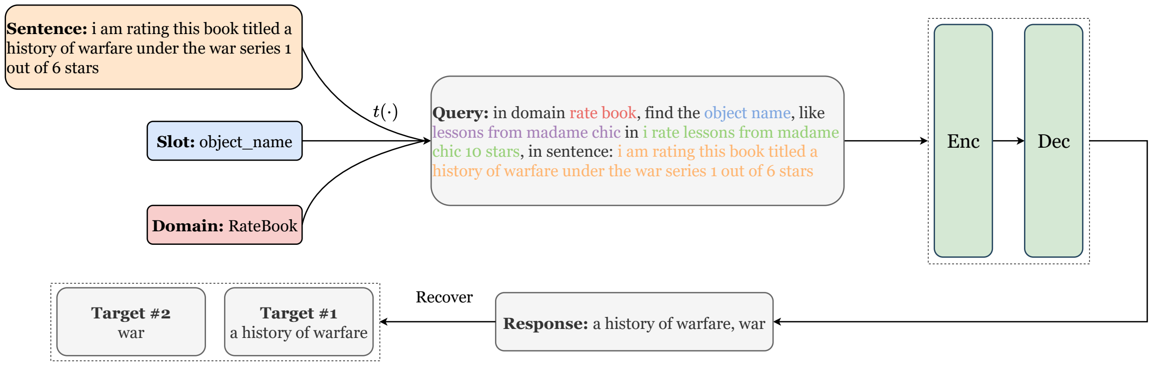

Figure 1: Illustration of our proposed AISFG.

- We propose a generative template-based zero-shot slot filling framework. To the best of our knowledge, we are the first to apply the generative framework to perform the zero-shot cross-domain slot filling task.
- We focus on slot type ambiguity and example ambiguity, which is ignored by previous researches. We incorporate domain-specific descriptions and examples with context to the query template to handle these two issues.
- The experimental results show that AISFG achieves better performance than the existing methods on the setting of zero/few-shot.

2 Methodology

2.1 Slot Filling as Generation

Given a sentence x from domain $d \in \mathcal{D}$, slot filling aims to predict a set of (slot type, span) pairs (s, y), where $s \in \mathcal{S}$ is a specific entity type from a fixed set of slot types, and $y = \{y_1, y_2, \ldots, y_i\}$ is a set of spans in sentence x. Most work on slot filling utilizes token-level classification frameworks, which means they either convert it to BIO tag labeling or predict the start and end position in the sentence, to extract spans.

In contrast to this convention, we frame slot filling as a conditional sequence generation task and solve it in a sequence-to-sequence manner. As shown in Figure 1, given a sentence x, domain d, slot type s and target y, we construct natural language query $q = t_q(x, d, s)$ and response $r = t_r(y)$ based on the predefined template. The generative model takes q as input and directly predicts r in a generation manner.

Leveraging the rich knowledge in the pre-trained generative model is important for cross-domain slot filling, especially in zero/few-shot setting. Thus, we construct query and response as natural language sentences to naturally utilize them.

Query Construction To solve the slot type ambiguity and example ambiguity, we synthesize the query by incorporating domain descriptions, slot descriptions and examples with context. Specifically, query construction template is formulated as:

in domain ①, find the ②, like ③
in ④, in sentence: x.

where the blank ①, ②, ③, ④ are filled with domain description, slot description, example entity and context for example, respectively. To avoid data leakage, we construct examples and contexts manually to ensure the example entities are not appeared in the dataset. The predefined specific mappings between domain/slot and description/example/context are reported in Appendix A.1 and A.3. Moreover, a query example is illustrated in Figure 1.

Response Construction Ideally, the response should be as simple as possible to alleviate the generation difficulty. Besides, an explicit template is easily converted to the original format for compatibility. Thus, we construct response by concatenating target entities with commas:

$$y_1, y_2, \ldots, y_i.$$

where y_i is the i-th entity span in sentence x. A response example is illustrated in Figure 1.

2.2 Train and Inference

The sequence-to-sequence model is instantiated as a pre-trained generative language model, such as BART (Lewis et al., 2020). In the training stage, for the input sentence x with multiple slots, we build a training pair, which consists of a query and a response based on the templates above for each slot. We further fine-tune the parameters based on these training pairs to maximize the log likelihood for predicting gold responses just like in an ordinary generation task. In the inference stage, we

Corpus	Training Setting	Zero-shot					Few-shot on 20 samples				Few-shot on 50 samples			
	Domain↓ Model→	CT	RZT	Coach	QASF	AISFG	CT	RZT	Coach	AISFG	CT	RZT	Coach	AISFG
SNIPS	AddToPlaylist	38.82	42.77	50.90	**57.57**	56.20	58.36	63.18	62.76	**81.64**	68.69	74.89	74.68	**83.51**
	BookRestaurant	27.54	30.68	34.01	48.75	**65.94**	45.65	50.54	65.97	**78.06**	54.22	54.49	74.82	**84.60**
	GetWeather	46.45	50.28	50.47	61.27	**67.66**	54.22	58.86	67.89	**82.68**	63.23	58.87	79.64	**83.73**
	PlayMusic	32.86	33.12	32.01	38.54	**50.12**	46.35	47.20	54.04	**77.59**	54.32	59.20	66.38	**78.79**
	RateBook	14.54	16.43	22.06	36.51	**41.05**	64.37	63.33	74.68	**79.06**	76.45	76.87	84.62	**92.85**
	SearchCreativeWork	39.79	44.45	46.65	60.82	**67.46**	57.83	63.39	57.19	**71.95**	66.38	67.81	64.56	**76.00**
	SearchScreeningEvent	13.83	12.25	25.63	27.72	**35.05**	48.59	49.18	67.38	**73.91**	70.67	74.58	83.85	**91.29**
	Average F1	30.55	32.85	37.39	47.31	**54.78**	53.62	56.53	64.27	**77.84**	64.85	66.67	75.51	**84.39**
ATIS	AirlineTravel	2.14	2.86	1.64	-	**35.17**	26.05	41.37	54.91	**64.04**	35.87	51.80	66.99	**75.31**

Table 1: F1-scores (%) on SNIPS and ATIS for different target domains under zero-shot and few-shot learning settings. **Bold** indicates the best results. '-' represents the result is missed in published papers.

also build a query for each candidate slot, and the response $\hat{r}$ is generated in an auto-regressive manner, which means selecting the token as the next token with the highest probability over the vocabulary set at each time step. Note that, although we do not explicitly restrict the response tokens should originate from the input sentence x, AISFG can always do it (but there are exceptions, we show some cases in Section 3.4). Then, the response $\hat{r}$ is split by commas to recover to the original format $\hat{y}$ for evaluation.

3 Experiments

3.1 Setup

Dataset We evaluate our method on SNIPS (Coucke et al., 2018), a public spoken language understanding dataset which contains 7 domains and 39 slots. To simulate the cross-domain scenarios, we choose one domain as the target domain for test and the left six domains as the source domains for training following the setup of Liu et al. (2020).

However, domains in SNIPS are not completely independent with each other. To achieve a real cross-domain scenario, we use another commonly used dataset namely ATIS (Price, 1990) as the target domain to test our model which is trained on SNIPS.

Baselines We compare our method against a number of representative baselines as follows:

- **Concept Tagger (CT)** A slot-filling framework proposed by Gobbi et al. (2018), which utilizes original slot descriptions to generalize to unseen slot types.
- **Robust Zero-shot Tagger (RZT)** Besides leveraging slot descriptions like CT, RZT (Shah et al., 2019) further introduces examples to improve the robustness of zero-shot slot filling.
- **Coarse-to-fine Approach (Coach)** Coach (Liu et al., 2020) is a two-step coarse-to-fine model for slot-filling, which performs coarse-grained BIO labeling task in the first step and performs fine-grained slot type classification task in the second step. It also encodes slot descriptions to help recognize unseen slot types.
- **QA-driven slot filling (QASF)** QASF (Du et al., 2021) uses a linguistically motivated question generation strategy for converting slot descriptions and example values into natural questions and solves the slot filling by extracting spans from utterances with a span-based QA model.

3.2 Main Results

Cross-Domain Slot Filling The cross-domain slot filling results are reported in Table 1. In SNIPS dataset, AISFG outperforms the state-of-the-art models (QASF for zero-shot and Coach for few-shot) by 7.47% on the average F1 under zero-shot setting, 13.57% under 20-shot setting and 8.88% under 50-shot setting, which demonstrates the superiority of our method. We train AISFG on SNIPS dataset and test it on the ATIS dataset for simulating a real cross-domain scenario, results are reported in the bottom of Table 1. AISFG consistently outperforms the existing state-of-the-art approaches, especially in zero-shot setting, where AISFG achieves 32.31% F1 score improvement. The significant performance improvement proves that utilizing rich domain information to prompt knowledge in PLMs may be a shortcut for solving data scarcity issue.

Analysis on Seen versus Unseen Slots We divide the samples in each target domain into "seen" and "unseen" categories in the SNIPS dataset for further understanding the transferring ability of our model. Following Liu et al. (2020), an example is categorized as "unseen" as long as the slot does not

Training Setting	Zero-shot					Few-shot on 50 samples				
Domain↓ Model→	SD	SD+DD	SD+E	SD+EC	AISFG	SD	SD+DD	SD+E	SD+EC	AISFG
AddToPlaylist	54.57	54.94	55.03	**56.22**	56.20	80.84	83.49	82.84	83.23	**83.51**
BookRestaurant	63.48	63.85	62.82	65.93	**65.94**	81.11	83.15	83.16	83.86	**84.60**
GetWeather	61.09	61.35	64.73	65.56	**67.66**	83.35	83.50	83.17	83.66	**83.73**
PlayMusic	44.41	44.99	46.51	47.48	**50.12**	78.62	78.50	78.06	78.68	**78.79**
RateBook	30.67	30.61	33.38	**41.44**	41.05	92.04	92.08	91.23	92.22	**92.85**
SearchCreativeWork	65.60	64.02	66.14	66.45	**67.46**	75.78	**76.36**	74.19	74.89	76.00
SearchScreeningEvent	30.33	30.18	32.81	34.61	**35.05**	89.53	90.25	90.49	90.82	**91.29**
Average F1	50.02	49.99	51.63	53.95	**54.78**	83.03	83.90	83.30	83.90	**84.39**

Table 2: F1-scores (%) on SNIPS for different target domains under zero-shot and few-shot learning settings. **Bold** indicates the best results.

Setting→	0 sample		50 samples	
Model↓	unseen	seen	unseen	seen
CT	3.38	37.23	52.65	65.66
RZT	2.19	40.99	50.28	61.63
Coach	9.31	46.22	68.59	74.55
QASF	41.73	56.23	-	-
AISFG	**63.96**	**76.11**	**85.09**	**88.77**

Table 3: Averaged F1-scores (%) over all target domains on SNIPS dataset for "seen" and "unseen" slots. '-' represents the result is missed in published papers

exist in the remaining six source domains. Otherwise, it is tagged as "seen".

Table 3 shows the average F1 results on seen and unseen slots in target domains under zero-shot and 50-shot settings. From this table, AISFG achieves the best results in both seen and unseen slots, especially on unseen slots in zero-shot scenario, which indicates our method is more effective for transferring knowledge from source to the target domain. Besides, approaches leverage PLMs (*i.e.*, QASF and AISFG) have significant advantages over others, which demonstrates utilizing knowledge encoded in PLMs can enhance the transferring ability.

3.3 Ablation Studies

To solve the slot type ambiguity issue and example ambiguity issue, we incorporate domain description and context for example to the query template, respectively. To understand the influence of these components in the query template, we perform ablation studies on the SNIPS dataset and report results in Table 2. In this table, SD, DD, E, EC refers to Slot Description, Domain Description, Example and Example with Context, respectively. For example, "SD+EC" represents the query template is built by domain description plus example with context.

The Effect of Domain Description Comparing SD versus SD+DD and SD+EC versus AISFG, we observe that incorporating domain descriptions can improve the performance under few-shot setting but not for zero-shot setting. We conjecture the reason is that, the domain description we used (shown in Table 4) is simply converted from the domain name and contains limited information, which is not enough to provide domain knowledge. However, when some target domain examples are given (*i.e.*, few-shot), the meaning of domain descriptions can be enhanced by these training examples. Then, the domain description becomes a domain indicator and can be used to distinguish the slot types shared by different domains, resulting in improved performance.

The Effect of Context Example Comparing SD versus SD+EC and SD+DD versus AISFG, we observe that incorporating examples with contexts can further boost the performance in both zero-shot and few-shot settings, which demonstrates these examples with contexts are helpful for cross-domain slot filling. Moreover, SD+EC outperforms SD+E in both zero-shot and few-shot settings. Especially, for the zero-shot setting in domain *RateBook*, which contains plenty of ambiguity examples, SD+EC achieves 8.06% improvement over SD+E, which indicates contexts are useful to alleviate example ambiguity.

3.4 Error Analysis and Case Study

To better understand our proposed model, we analyze the error predictions in all cross-domain experiments on SNIPS dataset (Table 1) and categorize them into three types: boundary mistakes, vocabu-

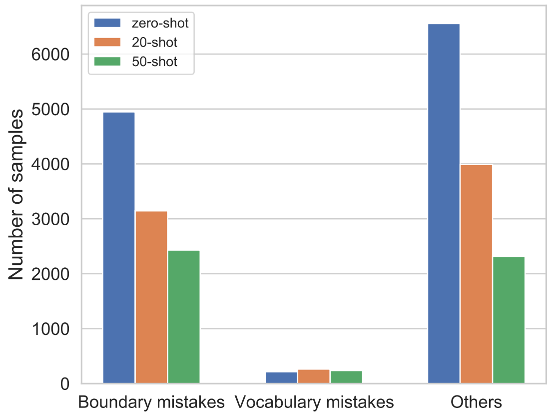

Figure 2: Number of samples against mistake types.

Boundary mistake example
Sentence: i want the complete global albums collection.
Slot: object_name
Gold Label: the complete global albums collection.
Prediction: global albums collection.

Vocabulary mistake example (synonym)
Sentence: find a picture called blink of an eye.
Slot: object_name
Gold Label: blink of an eye.
Prediction: wink of an eye.

Vocabulary mistake example (tense)
Sentence: show schedule for loved by thousands.
Slot: object_name
Gold Label: loved by thousands.
Prediction: love by thousands.

Figure 3: Examples for boundary mistake and vocabulary mistake.

lary mistakes and others. Boundary mistakes indicate that AISFG predicts the main body of the gold label while missing/adding other words, which may lead to an inaccurate boundary (top example in Figure 3). Vocabulary mistakes refer to the predictions that i) some words in the gold label are replaced by synonyms (middle example in Figure 3); ii) some words in the gold label are substituted by different tenses (bottom example in Figure 3).

As shown in Figure 2, the most common mistakes are the boundary mistakes. Especially, in 50-shot experiment, the amount of boundary-mistake predictions reaches almost half of all mistakes. We find that the model often struggles to detect the exact same span as the ground truths, as shown in the top example in Figure 3. When the shot number increases from 0 to 20 and from 20 to 50, the amount of boundary mistakes reduces by 36.4% and 22.7%, respectively, indicating that increasing shot number is beneficial for addressing the boundary mistakes.

Vocabulary mistakes such as synonyms or tenses replacement are the least common. Since the proposed AISFG is a generative model, we carry no restriction on what the model produces, therefore it may come with the vocabulary mistakes. However, most of these mistakes can be solved by post-processing the outcome of the model (*i.e.*, comparing the prediction with the original sentence and replacing the wrong words), which is one of the improvements of our future AISFG. Moreover, as shown in Figure 2, increasing shot number is useless for tackling the vocabulary mistakes.

4 Related Work

The main challenge of cross-domain slot filling is to handle domain-specific slot types which have few or no supervision signals during the training stage. To handle the unseen slots, previous methods introduce slot descriptions (Lee and Jha, 2019; Liu et al., 2020; Bapna et al., 2017) and slot examples (Guerini et al., 2018; Shah et al., 2019) to capture the semantic relationship between unseen slots and input sentences. Based on these researches, we incorporate more information, like domain description and example context, to boost the performance on unseen slots.

Typically, traditional methods formulate the slot filling task as a token-level classification task (Liu et al., 2020; Bapna et al., 2017; Shah et al., 2019; Wang et al., 2021). To leverage the knowledge encoded in pre-trained language models, RCSF (Yu et al., 2021b) utilizes BERT (Devlin et al., 2019) to predict the start/end position of the target slot entity. However, the downstream task (*i.e.*, token/position prediction) is inconsistent with the pre-training task (*i.e.*, Masked LM and NSP for BERT), which may limit the expressiveness of the PLM (Mehri and Eskenazi, 2021). To bridge the gap, we treat the slot filling as a generative task and utilize a PLM whose pre-training task is also a generative task.

5 Conclusion & Discussion

In this paper, we introduce a novel zero-shot cross-domain slot filling model named AISFG, which can adapt to unseen domains seamlessly with the help of domain and slot description and together with full context examples. Experiments show that our model significantly outperforms existing zero-shot cross-domain slot filling approaches. Moreover, we conduct error analysis and case study to better understand our proposed model and leave solving boundary mistakes and vocabulary mistakes as our future work.

Acknowledgments

This work was partially supported by National Natural Science Foundation under Grants U1936103 and 61921003.

References

Ankur Bapna, Gökhan Tür, Dilek Hakkani-Tür, and Larry P. Heck. 2017. Towards zero-shot frame semantic parsing for domain scaling. In *Interspeech 2017, 18th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, August 20-24, 2017*, pages 2476–2480. ISCA.

Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, Maël Primet, and Joseph Dureau. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *CoRR*, abs/1805.10190.

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.

Xinya Du, Luheng He, Qi Li, Dian Yu, Panupong Pasupat, and Yuan Zhang. 2021. QA-driven zero-shot slot filling with weak supervision pretraining. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 654–664, Online. Association for Computational Linguistics.

Jacopo Gobbi, Evgeny A. Stepanov, and Giuseppe Riccardi. 2018. Concept tagging for natural language understanding: Two decadelong algorithm development. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Torino, Italy, December 10-12, 2018*, volume 2253 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Marco Guerini, Simone Magnolini, Vevake Balaraman, and Bernardo Magnini. 2018. Toward zero-shot entity recognition in task-oriented conversational agents. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 317–326, Melbourne, Australia. Association for Computational Linguistics.

Kunho Kim, Rahul Jha, Kyle Williams, Alex Marin, and Imed Zitouni. 2019. Slot tagging for task oriented spoken language understanding in human-to-human conversation scenarios. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 757–767, Hong Kong, China. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Sungjin Lee and Rahul Jha. 2019. Zero-shot adaptive transfer for conversational language understanding. In *The Thirty-Third AAAI Conference on Artificial Intelligence, 2019*, pages 6642–6649. AAAI Press.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. 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*, pages 7871–7880, Online. Association for Computational Linguistics.

Zihan Liu, Genta Indra Winata, Peng Xu, and Pascale Fung. 2020. Coach: A coarse-to-fine approach for cross-domain slot filling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 19–25, Online. Association for Computational Linguistics.

Samuel Louvan and Bernardo Magnini. 2018. Exploring named entity recognition as an auxiliary task for slot filling in conversational language understanding. In *Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI*, pages 74–80, Brussels, Belgium. Association for Computational Linguistics.

Shikib Mehri and Maxine Eskenazi. 2021. GenSF: Simultaneous adaptation of generative pre-trained models and slot filling. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 489–498, Singapore and Online. Association for Computational Linguistics.

Grégoire Mesnil, Xiaodong He, Li Deng, and Yoshua Bengio. 2013. Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding. In *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013*, pages 3771–3775. ISCA.

Patti J. Price. 1990. Evaluation of spoken language systems: the ATIS domain. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, USA, June 24-27, 1990*. Morgan Kaufmann.

Darsh Shah, Raghav Gupta, Amir Fayazi, and Dilek Hakkani-Tur. 2019. Robust zero-shot cross-domain slot filling with example values. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5484–5490, Florence, Italy. Association for Computational Linguistics.

Liwen Wang, Xuefeng Li, Jiachi Liu, Keqing He, Yuanmeng Yan, and Weiran Xu. 2021. Bridge to target domain by prototypical contrastive learning and label confusion: Re-explore zero-shot learning for slot filling. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9474–9480, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Kaisheng Yao, Geoffrey Zweig, Mei-Yuh Hwang, Yangyang Shi, and Dong Yu. 2013. Recurrent neural networks for language understanding. In *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013*, pages 2524–2528. ISCA.

Mengshi Yu, Jian Liu, Yufeng Chen, Jinan Xu, and Yujie Zhang. 2021a. Cross-domain slot filling as machine reading comprehension. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 3992–3998. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Mengshi Yu, Jian Liu, Yufeng Chen, Jinan Xu, and Yujie Zhang. 2021b. Cross-domain slot filling as machine reading comprehension. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 3992–3998. ijcai.org.

A Appendix

A.1 Domain Descriptions

Domain	Domain Description
AddToPlaylist	add to playlist
BookRestaurant	book restaurant
GetWeather	get weather
PlayMusic	play music
RateBook	rate book
SearchCreativeWork	search creative work
SearchScreeningEvent	search screening event

Table 4: Domain descriptions for different domains.

A.2 Implementation Details

For a fair comparison with QASF, which is based on BERT (Du et al., 2021), we instantiate the sequence-to-sequence model as BART-base, which contains a similar parameter size as BERT. Following Liu et al. (2020), we reserve 500 samples in target domain as the validation set and regard the rest as the test set. We use Adam optimizer (Kingma and Ba, 2015) with a learning rate 2e-5 to fine-tune all parameters. The batch size is set to 16 and the decoding beam search size for BART is set to 2. The early stop strategy with patience of 5 is used to save the best checkpoint based on validation set.

Similar to the baselines, we use F1 score as the evaluation metric. For a given query, the true positive is the prediction with exactly matched entity extraction boundaries.

A.3 Illustration of Slot Description, Context and Example

Domain	Slot	Slot Description	Context	Example
AddToPlaylist	music_item	music item	funk outta here please add a piece to my playlist	piece
	playlist_owner	playlist owner	add the album to sebastian s ejercicio playlist	sebastian
	entity_name	entity name	i am m going to add love letter to my list of parties	love letter
	playlist	playlist	add this song to my stay happy playlist	stay happy
	artist	artist	i am going to include the taylor swift track in my bass gaming playlist	taylor swift
BookRestaurant	city	city	make a reservation at the best pub in shanghai	shanghai
	facility	facility	in washington reserve a tavern with baby chair	baby chair
	timeRange	time range	i need a reservation for the chinese food in cuba in 10 minutes	in 10 minutes
	restaurant_name	restaurant name	i am going to kfc together with my friends	kfc
	country	country	at 9 am reserve a restaurant in jerusalem for 8 persons	jerusalem
	cuisine	cuisine	i d like to reserve a table for one at a spanish restaurant in wesley	spanish
	restaurant_type	restaurant type	i d like to reserve a buffet for my family	buffet
	served_dish	served dish	find a restaurant serves hot-pot and make a reservation	hot-pot
	party_size_number	number	make a six-person reservation at an alpine wine bar	six-person
	poi	position	i d want to reserve a restaurant near my hotel	near my hotel
	sort	type	get a highly regarded sandwich shop in colombia	highly regarded
	spatial_relation	spatial relation	book a restaurant for 2 that s 10 minutes walk from here	10 minutes walk
	state	state	we d want to go to a brasserie in omaha that serves sicilian cuisine	omaha
	party_size_description	person	book a table for sebastian perez and leclerc	sebastian perez and leclerc
GetWeather	city	city	is the temperature going down to 2 in shanghai	shanghai
	state	state	check the weather in omaha	omaha
	timeRange	time range	what is the forecast for haidian in next half an hour	in next half an hour
	current_location	current location	will it rain in my present local street on 11/10/2023	present local street
	country	country	what s the weather like in jerusalem right now	jerusalem
	spatial_relation	spatial relation	is it going to rain within 10 minutes bus distance	within 10 minutes bus distance
	geographic_poi	geographic position	in west lake park, how cold will it be tomorrow	west lake park
	condition_temperature	temperature	will it get hotter in 2 hours	hotter
	condition_description	weather	will israel be hit by a snow storm	snow storm
PlayMusic	genre	genre	find me a lullaby in netease cloud music	lullaby
	music_item	music item	funk outta here please add a piece to my playlist	piece
	service	service	find me a lullaby in netease cloud music	netease cloud music
	year	year	play the most popular song in 2021	2021
	playlist	playlist	add this song to my stay happy playlist	stay happy
	album	album	play the album getting ready by eason chan	getting ready
	sort	type	play shall we talk by eason chan	shall we talk
	artist	artist	i am going to include the taylor swift track in my bass gaming playlist	taylor swift
RateBook	object_part_of_series_type	series	i rate the sequel 0 point	sequel
	object_select	this current	the book deserves a 5 star	the
	rating_value	rating value	the book deserves a 5 star	5
	object_name	object name	lessons from madame chic is a fine pick for anyone interested in fashion style	lessons from madame chic
	object_type	object type	i am too timid to read horror literatures	horror
	rating_unit	rating unit	this book deserves a 5 star	star
	best_rating	best rating	the highest rating for this book is 10	10
SearchCreativeWork	object_name	object name	paris baguette is a bakery chain based in south korea owned by the spc group	paris baguette
	object_type	object type	paris baguette is a bakery chain based in south korea owned by the spc group	bakery
SearchScreeningEvent	timeRange	time range	the movie starts at half past eight pm	half past eight pm
	movie_type	movie type	i want to see a comedy like green book	comedy
	object_location_type	location type	the castro theatre is the closest movie house showing green book	movie house
	object_type	object type	show me the movie poster of green book	movie poster
	location_name	location name	the castro theatre is the closest movie house showing green book	the castro theatre
	spatial_relation	spatial relation	the castro theatre is the closest movie house showing green book	closest
	movie_name	movie name	the castro theatre is the closest movie house showing green book	green book

Table 5: Illustration of slot description, context and example.