Incorporating Instructional Prompts into A Unified Generative Framework for Joint Multiple Intent Detection and Slot Filling

Yangjun Wu, Han Wang, Dongxiang Zhang*, Gang Chen, Hao Zhang
Zhejiang University
{yjwu, zh}@zjuici.com, {22021066, zhangdongxiang,cg }@zju.edu.cn

Abstract

The joint multiple Intent Detection (ID) and Slot Filling (SF) is a significant challenge in spoken language understanding. Because the slots in an utterance may relate to multi-intents, most existing approaches focus on utilizing task-specific components to capture the relations between intents and slots. The customized networks restrict models from modeling commonalities between tasks and generalization for broader applications. To address the above issue, we propose a **U**nified **G**enerative framework (UGEN) based on a prompt-based paradigm, and formulate the task as a question-answering problem. Specifically, we design 5-type templates as instructional prompts, and each template includes a question that acts as the driver to teach UGEN to grasp the paradigm, options that list the candidate intents or slots to reduce the answer search space, and the context denotes original utterance. Through the instructional prompts, UGEN is guided to understand intents, slots, and their implicit correlations. On two popular multi-intent benchmark datasets, experimental results demonstrate that UGEN achieves new SOTA performances on full-data and surpasses the baselines by a large margin on 5-shot (28.1%) and 10-shot (23%) scenarios, which verify that UGEN is robust and effective. Our code will be publicly available at https://github.com/Young1993/UGEN

1 Introduction

In task-oriented dialogue systems, spoken language understanding (SLU) is a crucial component that aims to understand users' queries and use a semantic frame to represent users' requirements. The semantic frame usually contains intents and slot names (Tur and De Mori, 2011). Recently, multiple intent SLU has attracted lots of attention (Liu and Lane, 2016; E et al., 2019; Weld et al., 2021; Gangadharaiah and Narayanaswamy, 2019) due to the wide variety of practical application scenarios.

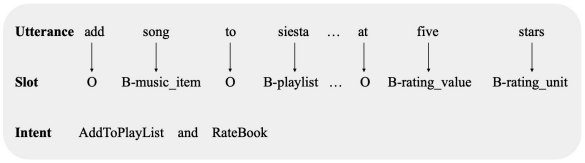

Figure 1: The semantic frame. An example from MixSNIPS dataset(Coucke et al., 2018; Qin et al., 2020).

Considering the example shown in Figure 1, the models are expected to identify the intents (AddToPlayList and RateBook) and the slot values with tags for the utterance. Current works (Qin et al., 2019, 2020; Ding et al., 2021; Qin et al., 2021; Chen et al., 2021a) usually treat Intent Detection (ID) as a classification task and Slot Filling (SF) as a sequence labeling task. The task-specific components are employed by current works to capture the connection or interaction between ID and SF, which achieve fine-grained multi-intent information integration for slot filling and obtain remarkable success.

In this paper, we're interested in exploiting a united paradigm to handle the task instead of customized networks. Prompt-learning (Liu et al., 2021; Jin et al., 2022) is a novel paradigm, which replaces the "pre-train, fine-tune" procedure with "pre-train, prompt, and predict" analogous to original pre-training language models (PLMs). With the help of a prompt template, prompt-learning benefits from fully exploiting the latent knowledge in PLMs while relieving the dependency on annotated data. Thus, prompt-based PLMs perform excellently in different tasks (classification, NER, summarization, etc.) and the few-shot setting.

To this end, we treat the joint multiple ID_SF as a question-answering problem and present a simple unified generative framework (UGEN) based on instructional prompts. Briefly, we first define 5-type descriptive templates (shown in Figure 2)

* Corresponding author

Proceedings of the 29th International Conference on Computational Linguistics, pages 7203–7208
October 12–17, 2022.

as inputs. Per template contains one context that refers to the original utterance, one question (e.g., "what are the intents of the sentence according to options?") as the driver to direct UGEN to realize the paradigm, and the corresponding options (e.g., play music, rate book) to restraint the answer search space. Through these instructional prompts, UGEN is directed to acquire the ability to capture the relationship between intents and slots. Then the correct intents and slots are predicted as the final answer (e.g., "add to playlist, rate book").

Experiments on two multi-intent benchmarks show that UGEN outperforms the baselines and achieves new SOTA performances. Remarkably, UGEN exceeds the comparison models by a large margin (28.1%, 23%, and 5.1%) in the 5/10-shot settings and 10% training data. The further analyses demonstrate that our approach has a strong ability of robustness and generalization. Meanwhile, it has the advantage of fast adaptation to practical scenario with limited annotation data and easy reproduction without task-specific components.

2 Related Work

Prompt-based Learning. With the release of GPT-3 (Brown et al., 2020), prompt-based learning methods have attracted more and more attention (Gu et al., 2021; Jin et al., 2022). The new paradigm can utilize the pre-trained language models with the form of cloze-style template, such as "I love this movie. It was a [Z] movie", and the model generates the probability of the [Z] in (good/bad). Hence, it directly models the probability of text $P(x|\theta)$ itself and uses the probability to predict y instead of the $P(y|x;\theta)$ [1] like traditional methods, which can narrow down the gap between pre-training and fine-tuning.

Few-shot Learning (FSL) with PLMs. FSL aims to absorb experience from only a few samples and make a great adaptation to the new problem(Wang et al., 2019). Usually, the models for FSL are trained on one accessible set of source domains and then evaluated on another set of unseen target domains. As the pre-trained models become more and more powerful, prompt-based methods with PLMs have achieved substantial improvements compared to those fine-tuned in low-resource settings, which displays promising prospects for few-shot learning in natural language tasks (Han et al., 2021; Li et al., 2021; Chen et al., 2021b).

[1]Here, we take the input x, learn the model parameters θ , and predict the output y.

3 Methodology

In this section, we briefly illustrate the problem definition of multiple ID_SF and main architecture. Then, we discuss the design of instruction-based templates and how to convert the ID_SF to the generation task.

3.1 Problem definition

The task of multiple ID_SF aims to classify all the possible intents and identify the slot values with the corresponding slot names in a given sentence. Given the input sentence $X = \{w_1, w_2, ..., w_n\}$, n is the length of X. The candidate intents $I = \{i_1, i_2, ..., i_m\}$, and m is the number of categories. The slot names $S = \{s_1, s_2, ..., s_k\}$, and k is the number of slot types.

To pursue simple model architecture (shown in Figure 2), in this work, we employ T5 (Raffel et al., 2020) as our backbone to model the probability of text $P(X|\theta)$. The answers Y are generated by UGEN, which contain intents (e.g., i_1, i_k) or slots (e.g., $\{w_1, w_2\}$ is one s_2), split by comma.

3.2 Instructional templates

To formulate the joint ID_SF as a question-answering problem and better exploit the knowledge learned in the PLMs, we design 5-type templates in line with QA and the pre-training-style tasks. Specifically, each template is defined to comprise three units: (1) **Context**, the original sentence X to express users' queries. (2) **Question**, the role of question Q is to guide the model to understand the paradigm and then generate the corresponding answer for the given Context X. In this study, the questions involve 5 types (shown in Figure 2): Question-1 is about the intents classification while the others are slot-related. For instance, question-1, "What are the intents of the sentence according to options?" is directed to intents labels. (3) **Options** O list all the intents labels or slot names as the candidate choices, and they act as a constraint to teach the model to select words in limited space (template's content).

Since the number of slot types are usually far larger than intents', we introduce 4-type questions to enhance the attention for slots. Specifically, Question-2 (e.g. Which words are the slot values in the sentence? for the context "Add this track to

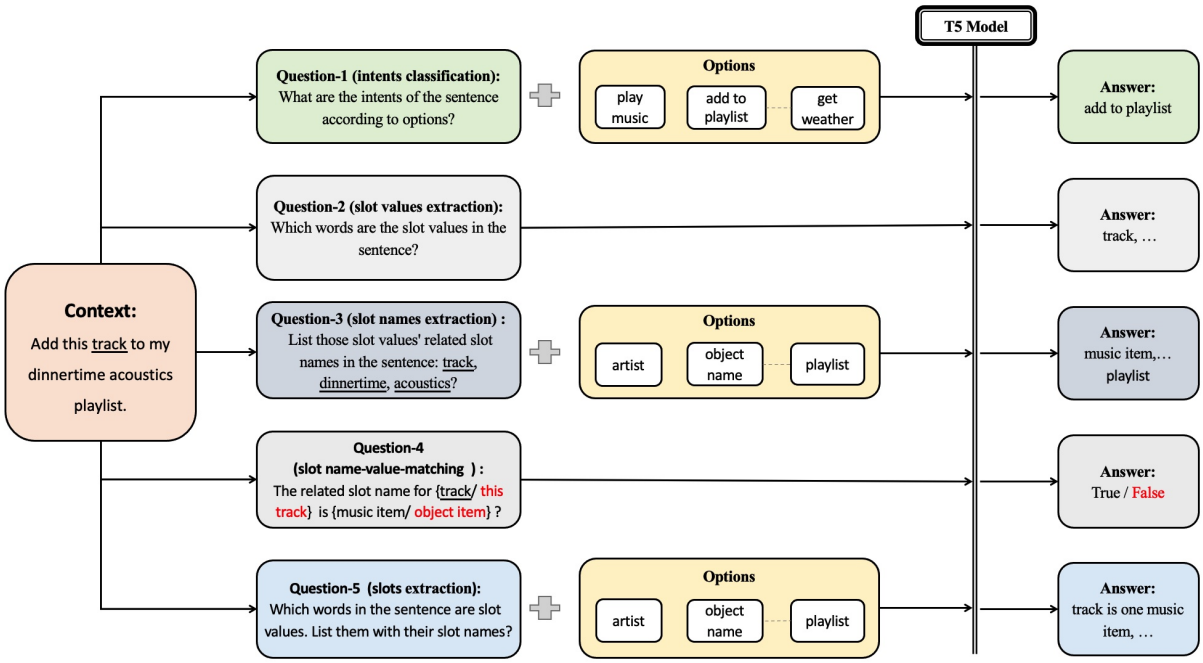

Figure 2: UGEN architecture with 5-type prompt templates based on combination of context, question, and options. For the Question-4, those words marked in red are negative samples.

my dinnertime acoustics playlist.") leads UGEN to extract the words that are exactly the slot values. Question-3 lists the slot values in X and steers the UGEN to select their slot names from options. To recognize the connection between slot values and their names, Question-4 is synthesized through a slot value with its slot name (positive) or random span in X with a slot name (negative). Question-5 is the most challenging, requiring UGEN to induce all the slot values and their names by the given question and options. Here, Questions 2-4 act as the auxiliary drivers and encourage the model to capture the links between slot names and their mentions in the Context X.

To simulate the pre-training manner, the input X is converted to "<s> Context: X </s> Question:Q </s> Options: O </s>". Here, the special tokens <s> and </s> are used to separate context, question and options. For the intents, the output Y is "i_1, i_k", i_k is one of the intent labels and generally, $i_k \leq 3$. On the slots side, the output Y (e.g., "$\{w_1, w_2\}$" is one "s_2") consists of slot values $\{w_1, w_2\}$ or slot values with slot names s_2, such as "track is one music item".

At the training stage, we first pre-process the original utterances with all the 5-type templates and shuffle the processed samples, then feed them into the UGEN to direct the model to understand the implicit correlations between intents and slots. The questions 2 to 4 are only used in the training phase and act more like auxiliary tasks. In the evaluation phase, only question-1 and question-5 are used to generate the intents and slot values with slot names, respectively.

4 Experiments

4.1 Experiment Setup

Dataset We compare our method with the baselines on two popular multi-intent SLU datasets, *MixSNIPS* and *MixATIS*. *MixSNIPS* is constructed from *SNIPS* dataset (Coucke et al., 2018) which comprises 39,776/2,198/2,199 utterances for training, validation and testing, separately. *MixATIS* is collected from *ATIS* (Hemphill et al., 1990), which contains 13,161/759/828 utterances for training, validation and testing, respectively. In addition, both of datasets are the cleaned version, and the proportion of sentences with $1 \sim 3$ intentions is [0.3, 0.5, 0.2].

We train and test all the models on the 32GB Tesla V100. For full-volume data, we set batch size to 20. The learning rate with Adam optimizer is set to $3e-5$, and beam search size is set to 3. In the few shot setting (5/10, and 10% training data), we set batch size to 16. In addition, we exploit the T5-base [2] as the backbone model.

[2] https://huggingface.co/t5-base

Baselines We compare UGEN with existing top-performing multi-intent approaches:

Joint Multiple ID-SF (JM) (Gangadharaiah and Narayanaswamy, 2019) proposes a multi-task framework and utilizes an attention-based model to identify intents and produce slot labels at the token-level.

Stack-Propagation (SP) (Qin et al., 2019) adopts a joint model with Stack-Propagation to use the intent information as input for slot filling and performs the token-level intent detection to alleviate the error propagation.

AGIF (Qin et al., 2020) presents an Adaptive Graph-Interactive Framework for joint multiple intent detection and slot filling, and it extracts the intents information for token-level slot prediction.

GL-GIN (Qin et al., 2021) proposes a Global-Locally Graph Interaction Network which explores a non-autoregressive model for joint multiple intent detection and slot filling.

SDJN (Chen et al., 2021a) introduces a novel self-distillation model which formulates multiple intent detection as a weakly supervised problem and designs an auxiliary loop to decode the intents and slots.

Model	MixSNIPS			MixATIS		
	S-F1	I-Acc	O-Acc	S-F1	I-Acc	O-Acc
JM	90.6	95.1	62.9	84.6	73.4	36.1
SP	94.2	96.0	72.9	87.8	72.1	40.1
AGIF	94.2	95.1	74.2	86.7	74.4	40.8
GL-GIN	94.9	95.6	75.4	88.3	76.3	43.5
SDJN	94.4	96.5	75.7	88.2	77.1	44.6
UGEN	**95.0**	**96.9**	**78.8**	**89.2**	**83.0**	**55.3**

Table 1: Overall results on the *MixSNIPS* and *MixATIS* sets with full-data. S-F1, I-Acc,O-Acc refer to the slot F1, intent-accuracy, and overall accuracy (both intents and slots need to be right), respectively. The highest numbers are in bold.

4.2 Overall Results

Table 1 reports the test results of UGEN compared to existing top-performing models on *MixSNIPS* and *MixATIS*. To the time of writing, UGEN outperforms the comparison models in all the metrics and obtains the new SOTA. For slot F1, our method leads to slight improvements (0.1% and 0.9%) compared to the GL-GIN, which validates that UGEN is more effective while extracting the slot values with their names. Turning to intent accuracy, UGEN exceeds SDJN (the previous SOTA) by 0.4% and 5.9%, respectively. It proves that UGEN has a strong ability to identify intents. Moreover, UGEN surpasses SDJN by 3.1% and 10.7% on overall accuracy (the more tough metric), which confirms that UGEN is more powerful in understanding the implicit correlations between intents and slots. The improvements align with our design and verify that the question-driven instructions are effective.

4.3 Few shot setting

Table 2 reports the results in the setting of 5/10-shot and 10% training data. We find that UGEN can consistently exceed the comparison models by a large margin in all the metrics. For instance, not only can UGEN increase by 23.5, 13.8, and 1.5 points in slot F1, but it leads to 28.1, 23.0, and 5.1 improvements in overall accuracy. The remarkable results validate that UGEN is more robust and can effectively exploit the implicit intent-slot correlations even with limited samples.

4.4 Ablation study

To explore the contribution of instructional prompts, we first remove the auxiliary instructions (Questions 2-4). The results drop a lot (e.g., 42.2% and 49.0% for overall accuracy) in the 5/10-shot, which demonstrates the auxiliary questions-driven templates are absolutely significant. Second, we only remove options in templates but keep all the questions. Every result under 5/10-shot and 10% training data is extremely low, sharply falling 34.7%, 31.2%, and 1.6%. The results confirm that options can effectively restrain the search space while predicting the answers. All the results are reported in Table 2.

5 Conclusion

In this work, we present a novel unified generative framework (UGEN) to treat the joint multiple intent detection and slot filling as a question-answering problem. To leverage the knowledge learned in the PLMs, we define 5-type prompt templates as the drivers to lead UGEN to grasp the prompt paradigm and capture the implicit correlations between intents and slots. On two multi-intent benchmark datasets, our approach accomplishes the new state-

Model	5-shot			10-shot			10%		
	S-F1	I-Acc	O-Acc	S-F1	I-Acc	O-Acc	S-F1	I-Acc	O-Acc
SP	58.7	78.2	11.9	71.5	88.3	24.8	90.3	93.5	58.4
AGIF	60.7	77.8	14.4	73.6	86.3	27.5	91.2	93.0	62.8
GL-GIN	54.3	86.1	10.1	69.5	90.2	23.9	92.1	95.3	66.6
UGEN - auxiliary instructions	32.3	18.4	0.3	37.2	34.6	1.5	92.6	95.4	67.5
UGEN - options	61.9	49.3	7.8	72.7	71.2	19.3	93.1	95.5	70.1
UGEN	**84.2**	**92.4**	**42.5**	**87.4**	**93.3**	**50.5**	**93.6**	**96.0**	**71.7**

Table 2: Results on the *MixSNIPS* set in the few shot settings. Because Joint Multiple ID-SF (JM) and SDJN are not publicly available, we can only compare the other baselines. S-F1, I-Acc,O-Acc refer to the slot F1, intent-accuracy, and overall accuracy (both intents and slots need to be right), respectively.

of-the-art performances in all the metrics, which validates that our design is effective. Meanwhile, UGEN leads to 28.1%, 23.0%, and 5.1% improvements in the 5/10-shot and 10% training data settings, which verify that UGEN is robust with limited annotation data.

6 Acknowledgements

The authors gratefully acknowledge Kebin Fang, Yingrong Wang, and Yao Zhao for giving valuable suggestions on this study. Our thanks also go to all the anonymous reviewers for their positive feedback.

References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*.

Lisong Chen, Peilin Zhou, and Yuexian Zou. 2021a. Joint multiple intent detection and slot filling via self-distillation. *CoRR*, abs/2108.08042.

Xiang Chen, Ningyu Zhang, Lei Li, Xin Xie, Shumin Deng, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2021b. Lightner: A lightweight generative framework with prompt-guided attention for low-resource ner. *ArXiv*, abs/2109.00720.

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. *ArXiv*, abs/1805.10190.

Zeyuan Ding, Zhihao Yang, Hongfei Lin, and Jian Wang. 2021. Focus on interaction: A novel dynamic graph model for joint multiple intent detection and slot filling. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 3801–3807. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Haihong E, Peiqing Niu, Zhongfu Chen, and Meina Song. 2019. A novel bi-directional interrelated model for joint intent detection and slot filling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5467–5471, Florence, Italy. Association for Computational Linguistics.

Rashmi Gangadharaiah and Balakrishnan Narayanaswamy. 2019. Joint multiple intent detection and slot labeling for goal-oriented dialog. 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)*. Association for Computational Linguistics.

Xiaodong Gu, Kang Min Yoo, and Sang-Woo Lee. 2021. Response generation with context-aware prompt learning. *ArXiv*, abs/2111.02643.

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2021. Ptr: Prompt tuning with rules for text classification. *ArXiv*, abs/2105.11259.

Charles T. Hemphill, John J. Godfrey, and George R. Doddington. 1990. The ATIS spoken language systems pilot corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27,1990*.

Feihu Jin, Jinliang Lu, Jiajun Zhang, and Chengqing Zong. 2022. Instance-aware prompt learning for language understanding and generation. *ArXiv*, abs/2201.07126.

Bin Li, Fei Xia, Yixuan Weng, Xiusheng Huang, Bin Sun, and Shutao Li. 2021. PSG: prompt-based sequence generation for acronym extraction. *CoRR*, abs/2111.14301.

Bing Liu and Ian R. Lane. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. In *INTERSPEECH*.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *CoRR*, abs/2107.13586.

Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen, and Ting Liu. 2019. A stack-propagation framework with token-level intent detection for spoken language understanding. In *EMNLP/IJCNLP (1)*, pages 2078–2087.

Libo Qin, Fuxuan Wei, Tianbao Xie, Xiao Xu, Wanxiang Che, and Ting Liu. 2021. GL-GIN: fast and accurate non-autoregressive model for joint multiple intent detection and slot filling. *CoRR*, abs/2106.01925.

Libo Qin, Xiao Xu, Wanxiang Che, and Ting Liu. 2020. AGIF: An adaptive graph-interactive framework for joint multiple intent detection and slot filling. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, Online. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*.

Gokhan Tur and Renato De Mori. 2011. *Spoken language understanding: Systems for extracting semantic information from speech*. John Wiley & Sons.

Yaqing Wang, Quanming Yao, James Tin-Yau Kwok, and Lionel M. Ni. 2019. Generalizing from a few examples: A survey on few-shot learning. *arXiv: Learning*.

Henry Weld, X. Huang, Siqu Long, Josiah Poon, and Sooji Han. 2021. A survey of joint intent detection and slot-filling models in natural language understanding. *ArXiv*, abs/2101.08091.