SPM: A Split-Parsing Method for Joint Multi-Intent Detection and Slot Filling

Sheng Jiang[1]*, **Su Zhu**[1]*, **Ruisheng Cao**[2], **Qingliang Miao**[1] and **Kai YU**[1,2]
[1]AI Speech Co, .Ltd., Suzhou, China
[2] Shanghai Jiao Tong University, Shanghai, China

Abstract

In a task-oriented dialogue system, joint intent detection and slot filling for multi-intent utterances become meaningful since users tend to query more. The current state-of-the-art studies choose to process multi-intent utterances through a single joint model of sequence labelling and multi-label classification, which cannot generalize to utterances with more intents than training samples. Meanwhile, it lacks the ability to assign slots to each corresponding intent. To overcome these problems, we propose a Split-Parsing Method (SPM) for joint multiple intent detection and slot filling, which is a two-stage method. It first splits an input sentence into multiple sub-sentences which contain a single-intent, and then a joint single intent detection and slot filling model is applied to parse each sub-sentence recurrently. Finally, we integrate the parsed results. The sub-sentence split task is also treated as a sequence labelling problem with only one entity-label, which can effectively generalize to a sentence with more intents unseen in the training set. Experimental results on three multi-intent datasets show that our method obtains substantial improvements over different baselines.

1 Introduction

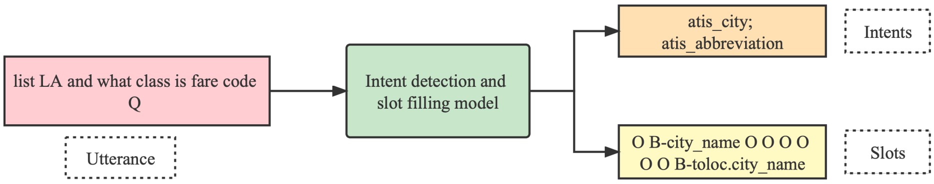
With the development of natural language technologies, the task-oriented dialogue system has become a significant practical application. It is widely applied in many industrial scenarios. One critical component in the task-oriented dialogue system is Spoken Language Understanding (SLU) (Young et al., 2013), which is further decomposed into two sub-tasks, namely intent detection and slot filling (Tur and De Mori, 2011). The slot filling task aims to convert the user utterance into a BIO label sequence of equivalent length. As for intent detection, it is essentially a sentence classification task which may have one or more labels. State-of-the-art studies tend to solve these two sub-tasks through a joint model (Goo et al., 2018; Liu et al., 2019), since slots and intents are highly correlated.

Previous literature in SLU mainly focuses on utterances with a single intent. Although classic models (Qin et al., 2019) achieve remarkable performances on those single-intent datasets, they neglect the realistic situation where the user utterance may contain multiple intents. Recently, researchers switch their attention to multi-intent benchmarks, such as MixATIS and MixSNIPS (Qin et al., 2020, 2021; Xing and Tsang, 2022a). An intuitive solution is to replace the original multi-class intent detection module into multi-label classification, see Figure 1(a). More advanced methods attempt to improve upon this backbone model. For example, Qin et al. (2020, 2021) proposes AGIF and GL-GIN, which both integrate the correlation between slots and intents into model design. Nonetheless, the separate prediction of multiple intents and slots leads to the failure of assigning appropriate slots to each intent. This mis-allocation of slots to intents may cause execution errors in a practical task-oriented dialogue system. Furthermore, the generalization capability of previous models is less investigated regarding the number of intents. For instance, if the model is merely trained on samples containing 1-3 intents, it would perform poorly on utterances with more than 3 intents.

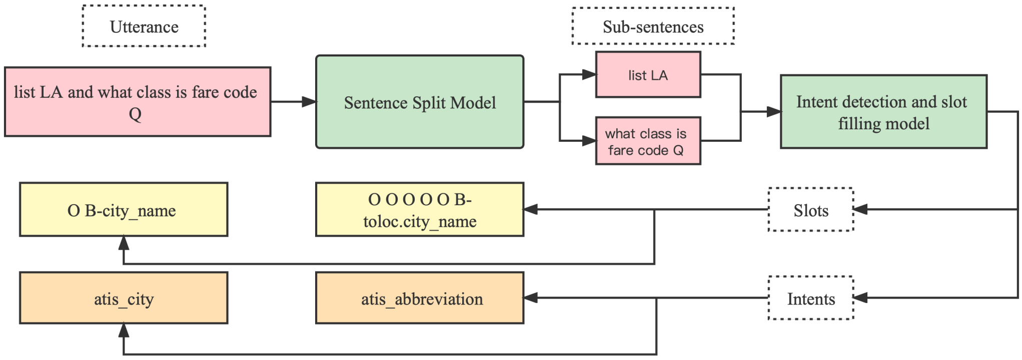
To this end, we propose a Split-Parsing Method (SPM) for joint multi-intent detection and slot filling. SPM is a two-stage SLU system. At the first stage, the utterance is split into sub-sentences, each containing exactly one intent. These sub-sentences are independent and together constitute the complete semantic representation. At the second stage, each sub-sentence is parsed by a traditional SLU model designed for single-intent. In this way, each slot is automatically aligned to their superior intent in the sub-sentence. Eventually, all parsing

* Sheng Jiang and Su Zhu are co-first authors and contribute equally to this work.

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics
Volume 5: Industry Track, pages 668–675
July 10-12, 2023 ©2023 Association for Computational Linguistics



(a) One-stage SLU



(b) SPM, two-stage SLU

Figure 1: Architectures of (a) the previous one-stage SLU system and (b) our proposed two-stage Split-Parsing Method (SPM). The utterance, "*list LA and what class is fare code Q*", is selected from MixATIS dataset (Qin et al., 2020).

results are aggregated through post-processing. As illustrated in Figure 1(b), the utterance "*list LA and what class is fare code Q*" is firstly split into two sub-sentences. Next, each sub-sentence is fed into the joint model of intent detection and slot filling to obtain the corresponding intent and slots. The slot-value pair "`city_name=`*LA*" is directly assigned to the sub-sentence "*list LA*" with intent "`atis_city`". Evidently, this method can effectively generalize to complicated utterances with more intents.

The proposed SPM is evaluated on two public English datasets (MixATIS and MixSNIPS, Qin et al., 2020), and a customized Chinese dataset which is collected from an in-vehicle dialog system. Experimental results demonstrate that the SPM can 1) achieve nearly perfect performances on the sub-sentence split task at the first stage, 2) attain stable improvements compared to one-stage method regardless of the model choice at the second stage, and 3) generalize better towards examples containing more intents unseen during training.

2 Related Work

From Single to Multiple Intents To deal with utterances with a single intent, most previous works (Liu and Lane, 2016; Hakkani-Tür et al., 2016; Zhang and Wang, 2016; Goo et al., 2018; Qin et al., 2019; Liu et al., 2019; Wang et al., 2018; Chen et al., 2019; Zhu et al., 2020) choose to tackle the intent detection and slot filling tasks in a multi-tasking manner. For sentences with multiple intents, several works (Gangadharaiah and Narayanaswamy, 2019; Qin et al., 2020, 2021; Xing and Tsang, 2022a,b) introduce a multi-label classifier to individually predict each possible intent. Recently, (Qin et al., 2021; Xing and Tsang, 2022a) proposed to model relationships between intents and slots, which takes into account the interaction between these two sub-tasks. However, previous literature fails to predict the alignment between intents and slots. Thus, it cannot determine which intent to assign for each slot. At the same time, in practical application scenarios, we need to design non-aligned slots. If we use a joint slot tagger, it is impossible to align non-aligned slots in multi-intent with their corresponding intents. In this work, we propose a two-stage pipelined SLU system to tackle the slot-intent assignment problem.

Generalization to More Intents The transfer performances in more intents is rarely studied.

Meng et al. (2022) proposed to use a sequence-to-sequence model (Dialo-USR) to generate all sub-sentences for joint multi-intent detection and slot filling. However, it also suffers from the poor generalization capability when confronted with more intents. Moreover, restricted by the auto-regressive decoding process, such a generative model introduces more overheads especially in the inference speed. Thus, it is impractical to be deployed in industrial scenarios. In contrast, we exploit a token-level sequence labeling model to act as the sentence splitting model, It shows better performances in both the accuracy and inference speed at the first stage.

3 Approach

The differences between the one-stage SLU system and our proposed two-stage SLU system (SPM) are illustrated in Figure 1. In the upper part, the user's multi-intent utterance is directly passed into a joint model of multi-label intent detection and slot filling, which is trained on multi-intent data. A token-level slots sequence and multiple intents are predicted, while it is not possible to assign each slot to the corresponding intent, since alignments between slots and intents are not modeled in this method. The below sub-figure of Fig. 1 illustrates our proposed method, where a multi-intent sentence is first split into sub-sentences by our split model (§3.1). These sub-sentences will be fed into a joint model of intent detection and slot filling separately. Thus, we can catch slot results for each individual intent. Meanwhile, the joint model of intent detection and slot filling exploited in the one-stage SLU can be applied into SPM without any change, which is much portable and easy-to-use.

3.1 Split Model

As shown in Fig. 2, to split a multi-intent sentence into sub-sentences, we regard it as a sequence labeling problem. We treat each sub-sentence as a separate slot (named as "snt"), and represent the output sequence in the way of BIO tags. For example, the annotation result of "list LA and what class is fare code Q" should be "B-snt I-snt O B-snt I-snt I-snt I-snt I-snt" in Fig. 2. It should be noted that annotations of conjunctions in multi-intent sentences at the token level are assigned with "O".

The split model can be implemented as any sequence labelling model, such as bidirectional

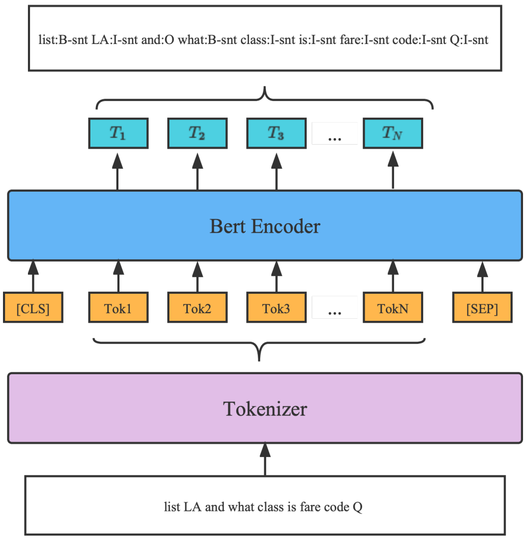


Figure 2: The sequence labeling model for the sentence splitting task. The outputs only include 3 labels, namely B-snt, I-snt and O.

[空调:对象][调热一点:value] => 调节内容:温度 => intent:车身控制

[Heat up:value] the [air–conditioner:object] a bit => adjustment:temperature => intent:car_control

Figure 3: Example for non-aligned slots in the Chinese dataset

LSTM (Graves, 2012), Bert model (Devlin et al., 2019). These sequence labelling models could potentially perform better for longer multi-intent sentences than those sentences in the training set.

4 Experiments

In this section, our SPM is trained and tested in three datasets and compared with different baseline models. From our experimental results in English dataset, adding split models can improve slot performance and our models shows better generalization ability in more intents. The evaluation results in Chinese dataset shows our SPM is faster and generalize better compared with other two-stage SLU system.

4.1 Datasets and Metrics

Regarding the English dataset, we conduct our experiments on MixATIS and MixSNIPS (Qin et al., 2020). For Chinese, we experiment on our customized dataset from realistic production scenarios. Detailed statistics are provided in Table 1. It is worth mentioning that in our Chinese dataset, due to the needs of real scenarios, we have designed some non-aligned slots. As shown in Fig. 3, the slot “adjustment:temperature” is non-aligned and necessary.

Datatset	Language	Train	Validation	Test
MixATIS	English	13,162	756	828
MixSNIPS	English	39,776	2,198	2,199
Ours	Chinese	800,000	50,000	20,000

Table 1: Statistics of multi-intent SLU benchmarks.

The evaluation is based on multiple intent detection accuracy, sub-sentence accuracy, F1 score for slot filling, and overall accuracy for the sentence-level semantic frame parsing. Notably, to align slots and intents, the slot F1 score in Chinese dataset is base on slot intent index. However, following previous works, the slot F1 score is not based on index in English dataset.

4.2 Implementation Details

In our two-stage SLU system, the basic task is to train the sentence split model. Since our split data is labeled based on token level, any sequence labeling models can be trained directly as the split model.

Split Labels Generation: Since the multi-intent sentences in MixATIS and MixSNIPS are actually generated by the combination of single-intent sentences in ATIS (Hemphill et al., 1990) and SNIPS (Coucke et al., 2018). We extracted and labeled all sub-sentences from MixATIS and MixSNIPS in token level BIO tags. For example, when the utterance is “list LA and what class is fare code Q", the output token-level tags should be “B-snt I-snt O B-snt I-snt I-snt I-snt I-snt I-snt".

During the construction of Chinese multi-intent dataset, we first extracted a certain number of single-intent utterances from production scenarios which are mainly in-car instructions.

Split Model: In our experiment, we use Bi-Model (Wang et al., 2018) and Bert(Devlin et al., 2019) as the split models in MixATIS and MixSNIPS. To compare with the sequence-to-sequence split model (Meng et al., 2022) in Chinese utterances, we also train Bert-based models(MiniRBT-h256 (Cui and Yang, 2022), Bert-wwm (Cui et al., 2021)) and mT5-small (Xue et al., 2020) for sentence split.

Intent Detection and Slot Filling Model: For the task of intent detection and slot filling, we directly use the open source model weights of AGIF (Qin et al., 2020) and GL-GIN (Qin et al., 2021). Also, we finetune the Bert model in the English datasets and our Chinese dataset for intent detection and slot filling.

4.3 Baselines

In one-stage SLU systems, we compare our our SPM with the following baselines: Bi-Model (Wang et al., 2018), AGIF (Qin et al., 2020), GL-GIN(Qin et al., 2021), ReLa-Net (Xing and Tsang, 2022b), Co-guiding Net (Xing and Tsang, 2022a) and Bert (Devlin et al., 2019).

To assign slots to corresponding intents in one-stage model, we also trained Bert in index labeling method. Slot labels based on index will have a suffix (__MI_X) to indicate the intent number. For instance, the index-based slots of the utterance “list LA and what class is fare code Q" should be “O B-city_name__MI_1 O O O O O O B-toloc.cityname__MI_2". Therefore, the slot can be aligned with the intent through the suffix of the slot.

4.4 Results in English Datasets

Using our method, adding a split model before intent detection and slot filling can help the original slot models have better performance. In this sub-section, we have evaluated our two-stage SLU system in the test sets of MixATIS and MixSNIPS, compared with different baselines. The first evaluation is in 1-3 intents and the second is in 3-5 intents.

Table 2 shows the intent detection and slot filling results of our two-stage SLU systems and baseline models in the original test sets of MixATIS and MixSNIPS. From Table 2, we observe that:

1. Adding split models improves the performance of baseline slot models to a certain extent. For instance, in MixATIS, the one-stage AGIF achieves 41.8 in overall accuracy, while it achieves 43.1 when we add the split model ad Bi-Model.

2. Our two-stage models are still competitive compared with the one-stage models with the best performance (Rela-Net and Co-guiding Net). Even in MixSNIPS, the system with Bert-base (split model and slot model) gets the best slot F1 96.0 and overall accuracy 83.2.

Since the test utterances in Table 2 only contain 1-3 intents, we also want to verify whether our SPM can achieve good performance in utterances with more intents unseen in the training set. Based on MixATIS and MIXSNIPS, we construct another

Split Model	Slot Model	MixATIS			MixSNIPS		
		Slot F1	Intent Acc	Overall Acc	Slot F1	Intent Acc	Overall Acc
-	Bi-Model	85.5	72.3	39.1	86.8	95.3	53.9
	AGIF	87.8	75.6	41.8	93.3	96.3	70.0
	GL-GIN	88.3	76.3	43.5	93.8	95.6	71.0
	ReLa-Net	**90.1**	78.5	**52.2**	94.7	97.6	76.1
	Co-guiding Net	89.8	79.1	51.3	95.1	**97.7**	77.5
	Bert_index	85.5	**82.6**	46.1	95.4	95.4	81.7
Bi-Model	Bi-Model	86.7	75.0	42.3	90.7	94.0	61.3
	AGIF	88.3	77.3	43.1	93.0	95.5	68.9
	GL-GIN	88.4	77.1	43.7	93.9	94.8	71.4
	Bert-base	86.3	77.4	49.2	94.8	96.4	77.5
Bert-base	Bi-Model	86.7	75.2	42.4	89.5	93.7	57.7
	AGIF	88.3	77.4	43.2	94.2	95.1	73.8
	GL-GIN	88.4	77.2	43.7	95.1	94.2	76.2
	Bert-base	86.3	77.9	49.3	**96.0**	95.9	**83.2**

Table 2: Results on the original test sets of MixATIS and MixSNIPS.

Split Model	Slot Model	MixATIS			MixSNIPS		
		Slot F1	Intent Acc	Overall Acc	Slot F1	Intent Acc	Overall Acc
-	AGIF	87.6	48.4	20.5	90.5	39.3	20.5
	GL-GIN	88.8	39.0	17.6	92.5	28.4	16.7
	Bert_index	88.0	14.9	7.5	93.9	16.7	12.4
Bi-Model	Bi-Model	88.4	62.9	24.9	86.6	49.8	20.4
	AGIF	88.4	66.9	27.7	91.8	50.8	27.6
	GL-GIN	89.1	66.4	27.3	92.9	50.3	29.5
	Bert-base	88.5	68.6	29.6	92.6	50.9	33.7
Bert-base	Bi-Model	88.5	80.4	31.2	89.4	90.7	38.1
	AGIF	88.5	85.5	34.8	94.5	92.1	58.5
	GL-GIN	**89.2**	85.0	34.3	95.0	91.6	62.1
	Bert-base	88.7	**87.3**	**37.8**	**96.1**	**93.5**	**72.8**

Table 3: Results on MixATIS and MixSNIPS with more intents (3-5).

Split Model	Intents	MixATIS Sub-sentence Acc	MixSNIPS Sub-sentence Acc
mT5-base	1-3	95.1	74.9
mT5-large		97.6	88.8
mT5-xl		98.1	98.6
Bi-Model		99.4	99.4
Bert-base		**99.7**	**99.5**
mT5-small	3-5	34.9	54.0
Bi-Model		86.4	67.4
Bert-base		**99.2**	**99.6**

Table 4: Sub-sentence accuracy results of different split models in MixATIS and MixSNIPS with 1-3 intents and 3-5 intents

test set with 3-5 intents. The experimental results in Table 3 show that our method has better generalization ability than the one-stage models in more intents.

Table 3 shows the intent detection and slot filling results on the multi-intent transfer test sets. The multi-intent transfer test sets only contain the utterances of 3-5 intents, which have never been experienced during the training of models. In Table 3, we can observe that the one-stage SLU systems perform poorly in the accuracy of slots and intents. For instance, GL-GIN only gets 16.7 overall accuracy in MixSNIPS, while the two-stage SLU with split Bi-Model and slot GL-GIN achieves 29.5. At the same time, it can be seen from Table 3 that the split model can still complete the sentence split task to a certain extent on the utterances with more intents. Especially when we use Bert as split and slot models, the intent accuracy in MixATIS and MixSNIPS are 87.3 and 93.5.

Interestingly, we found that the use of the pretraining model as the split model has better performance in both 1-3 intents and more intents. Therefore, we evaluated the sub-sentence accuracy of the split models. As shown in Table 4, sequence labeling models (Bi-Model and Bert) all achieve higher sub-sentence accuracy than sequence-to-sequenec

Split Model	Slot Model	Params of Split Model	1-5 Intents		6-10 Intents		Speed (split) (ms/sentence)	Speed (split + slot) (ms/sentence)
			Slot F1	Overall Acc	Slot F1	Overall Acc		
-	MiniRBT-index	-	97.74	84.51	74.57	5.33	-	166.60
	Bert-wwm-index		98.24	88.72	71.57	4.10	-	290.48
mT5-small	MiniRBT	300M	98.46	93.61	95.71	78.79	41.10	330.06
MiniRBT		10.4M	99.46	95.76	98.99	88.35	0.99	291.01
Bert-wwm		110M	99.50	95.92	99.17	89.26	2.04	290.00

Table 5: Evaluation results and inference speed on Chinese multi-intent dataset.

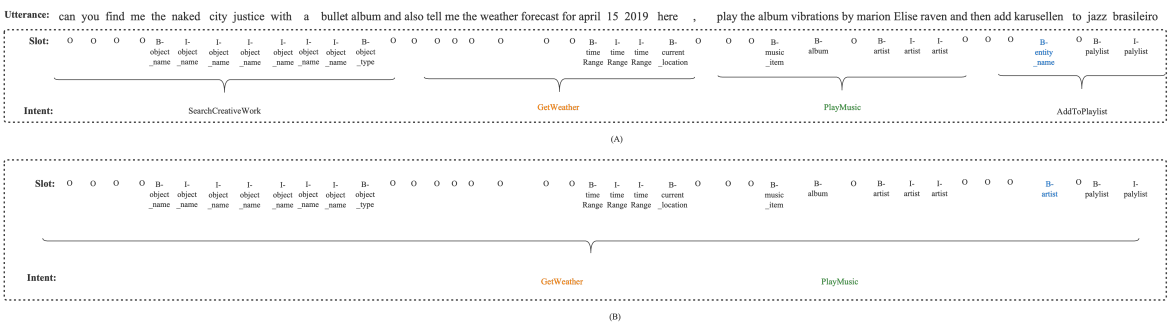


Figure 4: Case study of two-stage SLU with slot model AGIF (A) and one-stage SLU with slot model AGIF (B), blue denotes different slots, orange and green denote same intents

models (mT5). And this is more obvious in the utterance with more intents, like Bert-base achieves 99.2 sub-sentence accuracy but mT5-small only gets 34.9.

4.5 Results in Chinese Datasets

Our SPM still have advantages compared the other two stage SLU system. Here we evaluate the split model as the sequence-to-sequence model (Meng et al., 2022)) and the sequence labeling model (Bert) in slots and inference speeds. The sequence-to-sequence model used in (Meng et al., 2022) is mT5 (Xue et al., 2020). And the sequence labeling models we used for split are MiniRBT (Cui and Yang, 2022) and Bert-wwm (Cui et al., 2021). All split models in Table 5 are trained in 1-5 intents.

Table 5 shows the intent detection and slot filling results on utterances with 1-5 intents and 6-10 intents. We can note that the model using sequence labeling is higher in slot filling F1 and overall accuracy than the model using sequence-to-sequence, even mT5-small has much more parameters than MiniRBT and Bert-wwm.

In terms of the generalization ability in more intents, the use of sequence labeling as split model is more advantageous. From Table 5, the slot F1 and overall accuracy when using mT5-small as split model are 95.71 and 78.79, while for MiniRBT, they are 98.99 and 88.35.

When comparing the inference speed, our models are also faster. Table 5 shows the time required by different models to complete a complete slot filling task on a multi-intent utterance, that is, the time of sentence split and the slot filling for all sub-sentences. In Table 5, each utterance needs 40.10 ms to split and slot filling if the split model is mT5-small, however, for MiniRBT and Bert-wwm, the time is only 0.99 ms and 2.04 ms. What's more, the added time of our method is negligible in the whole process as shown in the right end of Table 5.

4.6 Case Study

To demonstrate how our two-stage SLU system outperforms one-stage SLU systems, we present the results of intent detection and slot filling of a case with 4 intents in Fig. 4. Fig. 4 (A) shows the correct slot labels and intent labels. The utterance is split into 4 sub-sentences and each slot is aligned with its intent. Meanwhile, as the intent increases, the sentence length also grows. This would make one stage model hard to detect enough intents and accurate slots. For example, the detected intents of Fig. 4 (B) miss the real first and last intents.

5 Future Work

Due to split-parse approach, when there are errors in sentence segmentation, cumulative errors are inevitable. The next step is to optimize for the situation of cumulative errors. There are two directions for optimization. Firstly, the accuracy of the sentence segmentation model can be further improved to reduce the probability of cumulative errors. Secondly, in the slot filling model, the model can be designed to support utterances with 1-2 intents. When the sentence split model incorrectly splits multiple sub-sentences into one, a slot filling model that supports multiple intents can correctly tag slots and detect intents.

6 Conclusion

In our paper, we propose a two-stage SLU system based on the split-parsing method. With plugging our split model into the original SLU system, the performance can be improved. Compared with the commonly used one-stage SLU systems, our method can better generalize in more intents unseen in training. Meanwhile, the split-parsing method can effectively align slots with their corresponding intents in the segmented sentences. And compared with other two-stage SLU systems using sequence-to-sequence as the split model, our model can achieve better performance of intent and slot filling detection with higher inference speed.

Acknowledgements

This research was funded by Jiangsu Key R&D Plan (Industry Foresight and Key Core Technologies) 2022 under Grant BE2022059-1, and Scientific and Technological Innovation 2030 under Grant 2021ZD0110900.

References

Qian Chen, Zhu Zhuo, and Wen Wang. 2019. BERT for joint intent classification and slot filling. *CoRR*, abs/1902.10909.

Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, et al. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190*.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Yiming Cui and Ziqing Yang. 2022. Minirbt. `https://github.com/iflytek/MiniRBT`.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc. NAACL*, pages 4171–4186.

Rashmi Gangadharaiah and Balakrishnan Narayanaswamy. 2019. Joint multiple intent detection and slot labeling for goal-oriented dialog. In *NAACL*, pages 564–569.

Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen. 2018. Slot-gated modeling for joint slot filling and intent prediction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 753–757.

Alex Graves. 2012. *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer Berlin Heidelberg.

Dilek Hakkani-Tür, Gokhan Tur, Asli Celikyilmaz, Yun-Nung Chen, Jianfeng Gao, Li Deng, and Ye-Yi Wang. 2016. Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM. In *Proc. INTERSPEECH*, pages 715–719.

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*.

Bing Liu and Ian Lane. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. In *Proc. INTERSPEECH*, pages 685–689.

Yijin Liu, Fandong Meng, Jinchao Zhang, Jie Zhou, Yufeng Chen, and Jinan Xu. 2019. Cm-net: A novel collaborative memory network for spoken language understanding. *arXiv preprint arXiv:1909.06937*.

Haoran Meng, Zheng Xin, Tianyu Liu, Zizhen Wang, He Feng, Binghuai Lin, Xuemin Zhao, Yunbo Cao, and Zhifang Sui. 2022. Dialogusr: Complex dialogue utterance splitting and reformulation for multiple intent detection. *arXiv preprint arXiv:2210.11279*.

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. *arXiv preprint arXiv:1909.02188*.

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. *arXiv preprint arXiv: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. *arXiv preprint arXiv:2004.10087*.

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

Yu Wang, Yilin Shen, and Hongxia Jin. 2018. A bi-model based rnn semantic frame parsing model for intent detection and slot filling. *arXiv preprint arXiv:1812.10235*.

Bowen Xing and Ivor Tsang. 2022a. Co-guiding net: Achieving mutual guidances between multiple intent detection and slot filling via heterogeneous semantics-label graphs. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 159–169.

Bowen Xing and Ivor W Tsang. 2022b. Group is better than individual: Exploiting label topologies and label relations for joint multiple intent detection and slot filling. *arXiv preprint arXiv:2210.10369*.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2020. mt5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*.

Steve Young, Milica Gašić, Blaise Thomson, and Jason D Williams. 2013. Pomdp-based statistical spoken dialog systems: A review. *Proceedings of the IEEE*, 101(5):1160–1179.

Xiaodong Zhang and Houfeng Wang. 2016. A joint model of intent determination and slot filling for spoken language understanding. In *Proc. IJCAI*, pages 2993–2999.

Su Zhu, Ruisheng Cao, and Kai Yu. 2020. Dual learning for semi-supervised natural language understanding. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:1936–1947.