Exploiting Topic Information for Joint Intent Detection and Slot Filling

Anonymous ACL submission

Abstract

Intent detection and slot filling are two important basic tasks in natural language understanding. Actually, there are multiple intents in an utterance. How to map different intents to corresponding slot becomes a new challenge for recent research. Existing models solve this problem by using neural layers to adaptively capture related intent information for each slot, which the process of intent selection is not clear enough. It is observed that there is strong consistency between intents and topics of a sentence, thus we exploit topic information for joint intent detection and slot filling via a topic fusion mechanism, where token-level topic information take the place of intent information to guide slot prediction. In addition, sentence-level topic information is also utilized to enhance the intent detection. Experiment results show explicit improvements on two public datasets, where provide 4.8% improvement in sentence accuracy on MixATIS and 0.7% improvement in intent detection on MixSNIPS.

1 Introduction

Natural language understanding (NLU) is an essential component in spoken dialogue system, which typically consists of intent detection and slot filling. These two tasks focus on capturing user' intent and extracting critical constituents via annotating the utterance. There is an example from SNIPS dataset shows below. The utterance "`i want to play music from iheart`" is supposed to be identified by the intent label "`PlayMusic`" on a sentence-level as well as the slot label "`B-service`" for the value "`iheart`" on a word-level.

Intent detection and slot filling are naturally defined as two separate tasks (Tur and De Mori, 2011). Intent detection can be treated as a classification problem, while slot filling can be seen as a sequence labeling task. These two tasks are easy to proceed separately via pipeline approaches, but such frameworks may cause error propagation. To solve the problems caused by pipeline manners, joint learning methods are introduced to identify the intent of the utterance and extract the slot information simultaneously. Some previous works on joint models utilize neural networks to share utterance-level representations between the two tasks (Guo et al., 2014; Hakkani-Tur et al., 2016; Chen et al., 2016). Furthermore, recent studies attempt to establish relationship between intent and slots (Goo et al., 2018; Qin et al., 2019) to enhance the performance of joint models. Most aforementioned approaches focus on single intent prediction, while users usually indicate multiple intents in real-world scenario. How to leverage multiple intents to guide corresponding slot prediction becomes a new challenge for recent studies.

In order to utilize multiple intents to lead slot prediction, Gangadharaiah and Narayanaswamy (2019) first propose an attention-based neural network model for the joint tasks, while each token is provided with the same multiple intents information. In addition, an adaptive graph interactive framework is introduced to map fine-grained intent information to slot filling on each token (Qin et al., 2020, 2021). However, we consider that aforementioned methods do not verity that if the correct related intent information works on the corresponding token, since the fine-grained intent information captured by graph interaction layer is not explicit. Therefore, we attempt to apply external knowledge to definitely guide slot prediction on each token.

In this paper, we apply topic information into joint multiple intent detection and slot filling via a topic fusion mechanism. Recent studies have shown significant improvement on exploiting syntactic knowledge into NLU tasks (Wang et al., 2020). Inspired by Wang et al. (2020), we find that there is strong consistency between intents and topics of an utterance so that we make an attempt to apply topic information into the joint tasks. To this end, a topic fusion mechanism is introduced to combine the topic information with middle layer of intent detection as well as each token's hidden state of slot filling encoder. Such a fusion mechanism is utilized to reinforce the intent detection on sentence level and guide each token for slot prediction.

Our contributions are as follows:

- To the best of our knowledge, we are the first to utilize topic information to support joint multiple intent detection and slot filling, where a topic fusion mechanism is explored to enhance the intent detection on sentence level and guide slot filling on token level.
- We conduct experiments on two public datasets: MixATIS and S-nips, which achieve 4.8% F1 score improvement in sentence accuracy on MixATIS and 0.7% F1 score improvement in intent detection on MixSNIPS.

2 Approach

In this section, we will discuss our proposed model in detail. Figure 1 gives an overview of our approach. We can see that the intent detection and slot filling are transformed to multi-label classification task and sequence labeling task respectively. Following that, we first introduce the Topic Information Extractor(3.1) and Topic Fusion Mechanism(3.2), which utilized in our framework. Then we discuss a topic fusion mechanism applied into intent detection(3.3) and slot filling(3.4). Last a joint learning scheme(3.5) is utilized to optimize the two tasks simultaneously.

2.1 Topic Information Extractor

Latent Dirichlet Allocation (Blei et al., 2003) is a popular topic modeling technique, which maps high dimensional word space to low dimensional topic space while reserving the implicit connection. In our framework, we use LDA model to acquire topic information of input sequence $\{x_1, x_2, x_3, ..., x_T\}$. In the corpus D, each document d_m includes N_m words and can be denoted by a K-dimensional "document-topic" distribution θ_m^k. And each topic k containing V words, is denoted by a V-dimensional "topic-word" distribution ϕ_k^t.

We follow Blei et al. (2003) to use *Collapsed Gibbs Sampling* to learn the "document-topic" distribution θ_m^k and the "topic-word" distribution ϕ_k^t. The process of *Collapsed Gibbs Sampling* can be written as:

$$\phi_k^t = \frac{n_k^t + \beta}{\sum_{t=1}^{V}(n_k^t + \beta)} \quad (1)$$

$$\theta_m^k = \frac{n_m^k + \alpha}{\sum_{k=1}^{K}(n_m^k + \alpha)} \quad (2)$$

where n_k^t represents the number of times that word t has been assigned to topic k and n_m^k denotes the number of times that topic k has been assigned to a word of the document d_m.

In each iteration, the topic assignment for word $w \in D$ is updated alternatively by sampling from a multinomial distribution $P = [p_1, ..., p_k, ..., p_K]$.

$$p_k \propto \phi_k^t \cdot \theta_m^k \quad (3)$$

where p_k denotes the probability that topic k is sampled. After the given S iterations, the 'document-topic" distribution θ_m^k and "topic-word" distribution ϕ_k^t can be obtained.

Instead of directly utilizing the distribution θ_m^k and ϕ_k^t, we design a method to extract sentence-level topic information E_S^L and token-level topic information E_i^L. In particular, $E_S^L = \phi^{emb}(s_1, s_2, ..., s_q)$ is the set of sentence-level topic words embedding, where $(s_1, s_2, ...s_q)$ is obtained from ϕ_k^t and θ_m^k.$E_i^L = \phi^{emb}(w_1, w_2, ..., w_p)$ is the set of token-level topic words embedding, where $(w_1, w_2, ..., w_p)$ is obtained from ϕ_k^t and θ_m^k

2.2 Topic Fusion Mechanism

In our model, we use Factorization Machine (FM) to fuse the topic information with the context. FM is produced by Rendle (2010) to interact features for recommendation system. Different from existing efforts, which utilizes FM to compute the cost of Neural Network, we apply it as a fusion layer to learn the features interactions of topic information and context. The basic FM algorism is defined as follows:

$$H^{FM} = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j \quad (4)$$

where w_0 is the global bias, w_i is the trainable parameter and $\langle v_i, v_j \rangle = \sum_{f=1}^{n} v_{i,f} v_{j,f}$.

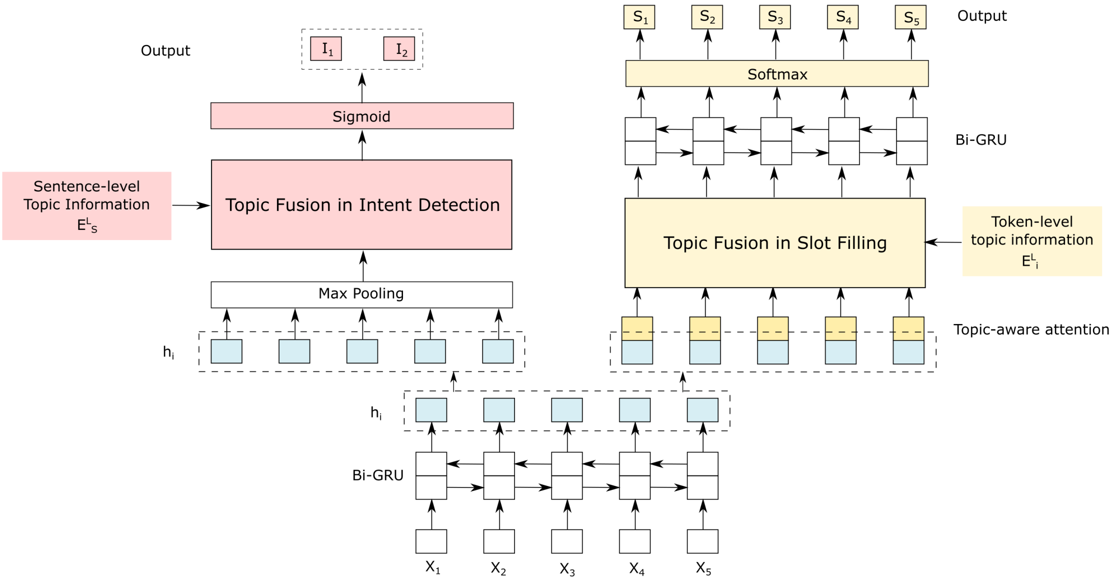
Output
I1
I2
Sigmoid
Sentence-level Topic Information E^L_S
Topic Fusion in Intent Detection
Max Pooling
h_i
h_i
Bi-GRU
X1
X2
X3
X4
X5
S1
S2
S3
S4
S5
Output
Softmax
Bi-GRU
Topic Fusion in Slot Filling
Token-level topic information E^L_i
Topic-aware attention

Figure 1: Framework of topic information fusion model

2.3 Intent Detection

Input Representation Layer The GRU (Graph Recurrent Unit) Network is first proposed by Cho et al. (2014) to consider sequence labeling tasks. We utilize bidirectional GRU to read the input sequence $\{e_1, e_2, ..., e_T\}$, where $e_i = \phi^{emb}(x_1, x_2, ...x_T)$ and ϕ^{emb} is the embedding function combining word-level and character-level embedding. Then we get the hidden state of bidirectional GRU $H = \{h_1, h_2, ..., h_T\}$.

Topic Fusion in Intent Detection Normally Intent Detection is treated as a classification problem. Recent models utilize deep learning frameworks to solve this task (Xia et al., 2018; Yolchuyeva et al., 2019; Okur et al., 2019; Tian and Gorinski, 2020). Some of them apply attention mechanism (Bahdanau et al., 2014) to focus on partial features. We find that topic information has strong connection with intents of an utterance, thus a topic fusion mechanism is suggested showed in Eq.(4). In intent detection, topic fusion mechanism is utilized to combine sentence-level topic information with the context. The formulation is written as follows:

$$h^I = Maxpooling(h_i) \tag{5}$$

$$h^{I,L} = h_1^{FM} + h_2^{FM} \tag{6}$$

where h_i is the hidden state of bidirectional GRU and $h^{I,L}$ is the sentence-level topic information fusion layer that modified from Eq.(4). Since the equation is too long, we decompose is into Eq.(7) and Eq.(8):

$$h_1^{FM} = W_{F0}(E_S^L, h^I) + b_{F0} \tag{7}$$

$$h_2^{FM} = \frac{1}{2}((v_0(E_S^L, h^I))^2 - v_0^2(E_S^L, h^I)^2) \tag{8}$$

where E_S^L denotes the sentence-level topic information, W_{F0} and v_0 are the trainable matrix parameters.

Since the intent detection is treated as a multi-label task, we use sigmoid function to give the probability distribution y_I over intent labels:

$$y^I = \sigma(W_I h^{I,L} + b_I) \tag{9}$$

where σ represents the sigmoid activation function.

2.4 Slot filling

Topic-aware Mechanism Inspired by Sutskever et al. (2014), we modifies the traditional attention algorism to learn related topic information for each token. The topic-attention output is computed as:

$$\alpha_i = softmax(E_i^L U W_{topic}^T) \tag{10}$$

$$c_i^L = \alpha_i W_{topic} \tag{11}$$

$$h_i' = [c_i^L, h_i] \tag{12}$$

where E_i^L is the token-level topic information, c_i^L provides additional topic information for each token, which concatenates with the hidden state of context h_i.

Topic Fusion in Slot Filling Similar to intent detection, topic fusion mechanism is leveraged to combine token-level topic information with each token of an utterance. $h_i^{S,L}$ is the token-level topic information layer, which is decomposed into Eq.(14) and Eq.(15). Then the Bidirectional GRU reads it forwardly and backwardly:

$$h_i^{S,L} = h_{1i}^{FM} + h_{2i}^{FM} \tag{13}$$

$$h_{1i}^{FM} = W_{F1}(E_i^L, h_i') + b_{F1} \tag{14}$$

$$h_{2i}^{FM} = \frac{1}{2}((v_1(E_i^L, h_i'))^2 - v_1^2(E_i^L, h_i')^2) \tag{15}$$

$$h_i^{S,L'} = BiGRU(h_i^{S,L}) \tag{16}$$

The softmax activation function is applied to predict the probability distribution of slot labels:

$$y_i^S = softmax(W_S h_i^{S,L'} + b_S) \tag{17}$$

2.5 Joint Training

To learn intent detection and slot filling jointly, we adopt a joint training model to consider the two tasks and update parameters simultaneously. The cross-entropy loss for intent detection and slot filling is computed as:

$$L_I = -\sum_{m=1}^{M} \hat{y^I} \log(y^I) \tag{18}$$

$$L_S = -\frac{1}{T} \sum_{i=1}^{T} \sum_{c=1}^{C} \hat{y_i^S} \log(y_i^S) \tag{19}$$

where M is the number of intent labels, T is the number of words in an utterance and C is the number of slot labels. We use $\hat{y^I}$ and $\hat{y_i^S}$ to denote the ground truth label of intent and slot.

The training target of the model is to minimize the united loss function. Finally, the joint objective is defined as:

$$Loss = \gamma L_I + (1 - \gamma) L_S \tag{20}$$

where γ is a hyper-parameter to adjust the importance of the two tasks.

3 Experiment and Analysis

In this section, we first introduce the dataset used in the experiments. Then an analysis about our model according to the experimental results will be mentioned.

3.1 Dataset

We use the two public datasets, the MixATIS dataset (Tur et al., 2010; Qin et al., 2021) and MixSNIPS dataset, to conduct our experiments. All datasets are annotated with intent and entity labels. The data division we used is the same as Qin et al. (2021), where the MixATIS consists of 13162 utterances for training, 756 utterances for validation and 828 utterance for testing. Another dataset MixSNIPS includes 39776, 2198, 2199 utterances for training, validating and testing.

3.2 Baselines

To confirm the effectiveness of our framework, we compared it with some published state-of-the-art models, which are shown as follows:

- **Attention-based** (Liu and Lane, 2016) develops an attention-based RNN models for joint intent detection and slot filling. The model uses an attention mechanism to extract features from utterance context for the prediction of slot and intent.
- **Slot-gated Full Atten** (Goo et al., 2018) leverages attention mechanism to combine intent detection with slot filling task, which enables intent information to apply into the process of slot prediction via a slot-gated algorism.
- **Bi-Model** (Wang et al., 2018) proposes a RNN semantic frame parsing model to consider cross-impact between intent and slots.
- **SF-ID Network SF-First with CRF** (E et al., 2019) utilizes a SF-ID network to establish interrelated relations for slot filling and intent detection, in which the two subtasks promote each other simultaneously via attention mechanism.
- **Stack-Propagation** (Qin et al., 2019) adopts a joint model which can directly incorporate intent information to guide slot filling.
- **Joint Multiple ID-SF** (Gangadharaiah and Narayanaswamy, 2019) investigates an attention-based neural network for multi-label intent detection and slot filling.

Table 1: Comparison with published results of joint models on the MixATIS and MixSnips dataset

Model	MixATIS			MixSnips		
	Slot(F1)	Intent(Acc)	Sentence(Acc)	Slot(F1)	Intent(Acc)	Sentence(Acc)
Attention-based (Liu and Lane, 2016)	86.4	74.6	39.1	89.4	95.4	59.5
Slot-gated Full Atten (Goo et al., 2018)	87.8	63.9	35.5	87.9	94.6	55.4
Bi-Model (Wang et al., 2018)	83.9	70.3	34.4	90.7	95.6	63.4
SF-ID Network SF-First with CRF (E et al., 2019)	87.4	66.2	34.9	90.6	96.0	59.9
Stack-Propagation (Qin et al., 2019)	87.8	72.1	40.1	94.2	96.0	72.9
Joint Multiple ID-SF (Gangadharaiah and Narayanaswamy, 2019)	84.6	73.4	36.15	90.6	95.1	62.9
AGIF (Qin et al., 2020)	86.7	74.4	40.8	94.2	95.1	74.2
GL-GIN(Qin et al., 2021)	88.3	**76.3**	43.5	**94.9**	95.6	**75.4**
Our model	**88.7**	73.0	**48.3**	94.4	**96.3**	69.8

Table 2: Results of ablation study on MixATIS dataset

	Slot (F1)	Intent (Acc)	Sentence (Acc)
Our model	**88.67**	**73.0**	**48.32**
Our model (no token-L topic)	88.56	71.73	46.69
Our model (no sentence-L topic)	88.50	68.91	43.72
Our model (no both component)	88.43	67.80	43.23

- **AGIF** (Qin et al., 2020) suggests an adaptive graph-interactive framework to learn the strong relationship between the slot and intents.
- **GL-GIN** (Qin et al., 2021) explores a non-autoregressive model for joint intent detection and slot filling to achieve more fast and accurate.

3.3 Training Details

In our experiments, the embedding layer merges word embedding and character embedding. We use pre-trained word vectors via FastText (Mikolov et al., 2018) and the character vectors are randomly initialized. Both the vectors are fine-tuned during training. The number of the bidirectional GRU units is set to 450, which is equal to the sum of dimensions of the word embedding and character embedding. Besides, the batch size is 64. Cross entropy is used as loss function and optimization is Adam (Kingma and Ba, 2014). To reduce the over-fitting, we apply dropout rate 0.2 to the bidirectional GRU. The iteration will be terminated after the F1 score of slot filling stop increasing 5 iterations continuously.

3.4 Results and Analysis

Evaluation Method. To evaluate the performance of our model, we adopt F1 score and accuracy compared with five state-of-the-art models. Following previous works, the F1 score is calculated from Precision (P) and Recall (R).We score a slot as correct if both the entity boundaries and the entity type are correct. An utterance is considered as correct if both the slots and intent are correct. The experimental results are shown in Table 1.

Main Results. We compare our model with current published joint models shown in Tabel 2. It is explicit that our method outperforms other methods for joint slot filling and intent detection, which achieves state-of-the-art performance mostly on the MixATIS and MixSnips datasets. Compared to the current best model GL-GIN (Qin et al., 2021) on MixATIS dataset, our method achieves substantial improvements on F1 score of slot filling and sentence accuracy. Especially in sentence accuracy, the our model achieves 4.5% absolute gain. Similar on MixSnips dataset, our model perfroms better than GL-GIN (Qin et al., 2021) with the improvements of 0.7% on F1 score of slot filling.

It is considered that the performance gain of intent detection and slot filling is mainly because the effectiveness of our topic fusion mechanism. The results verify the statement that topic information is beneficial to intent detection and slot filling. As mentioned above, existing joint models try to connect fine-grained multiple intent information for slot filling on each token. But we think that topic information is more explicit for the two tasks because there is strong consistency between topic information and intent. Furthermore, it is obvious that the improvement of sentence accuracy is prominent on MixATIS dataset. This may credit to our topic fusion mechanism is more efficient to dataset which contains more classification of labels.

3.5 Ablation Study

To demonstrate the effectiveness of each component in our joint intent detection and slot filling, we also conduct ablation experiment to understand the impact of each component on the whole model. Particularly, we investigate the topic fusion mechanism on sentence-level and token-level. The results are summarized in Table 2.The second line contains the results of our complete model and three extra experiments are performed. In the first experiment, we delete the topic fusion mechanism in slot filling, which does not utilize token-level topic information. Then we keep the token-level topic information and delete sentence-level topic information. Furthermore, we remove both of the aforementioned parts to conduct the experiment, which only apply Bidirectional GRU for joint models.

As shown in Table 2, the accuracy of intent detection increase from 71.73% to 73% when only applying sentence-level topic information. The accuracy of sentence also achieves 1.63% improvement, which verifies the benefit of sentence-level topic information to global utterance semantic comprehension. In the ablation test of token-level topic information, the accuracy of intent detection improves from 68.91% to 73% and the F1 score of slot filling improves from 88.50% to 88.67%. Thus we can conclude that topic information works effectively on the two subtasks. In addition, we find that the improvement of utilizing sentence-level topic information is more absolute than utilizing token-level topic information. This may credit to sentence-level topic information we obtained is more closed to intent information of an utterance.

4 Related Work

Traditional systems treat intent detection and slot filling as two separate tasks in a pipeline. Intent detection is usually considered as a text classification task, which relies on the methods of support vector machines (SVMs) (Haffner et al., 2003) and deep learning frameworks (Xia et al., 2018; Okur et al., 2019; Tian and Gorinski, 2020). Recently, a transformer model and universal sentence encoder based deep averaging network are utilized in intent detection task (Yolchuyeva et al., 2019). For slot filling, this task is formulated as a sequence labeling problem. Previous work on slot filling is relied on Conditional Random Field (CRF) (Lafferty et al., 2001) and maximum entropy Markov models (MEMMs) (McCallum et al., 2000). Currently, deep learning methods are combined with CRF to solve the slot filling problems. (Gong et al., 2019) proposes a deep cascade multi-task learning scheme for slot filling based on BiLSTM-CRF. It is simple to conduct these two tasks separately but pipeline methods may cause error propagation problem.

To solve the problems caused by pipeline methods, joint slot filling and intent detection models are proposed to improve the utterance semantics and solve the error propagation problem of pipeline methods (Goo et al., 2018). Prior method about joint models is to share the same text representation and utilize a joint loss function for global optimization (Guo et al., 2014; Chen et al., 2016). A convolutional neural network (CNN) for slot filling and intent detection is introduced, which extracts features through CNN layers for slot filling and shared by intent detection (Xu and Sarikaya, 2013). The RNN-LSTM architecture proposed by (Hakkani-Tur et al., 2016) enables slot filling and intent detection optimized in a single model based on bidirectional RNN with LSTM cells. In addition, (Liu and Lane, 2016) develops an attention-based RNN model for joint intent detection and slot filling. Besides, recent studies build relationship between slots and intent. Goo et al. (2018) utilizes attention mechanism to combine intent information with slot filling task. Similar to slot-gated mechanism, (Wang et al., 2018) utilizes a Bi-model based RNN semantic frame parsing network structure to establish cross-impact between intent and slots. Reference (E et al., 2019) introduces a SF-ID network to build interrelated relations for slot filling and intent detection to help them promote each other simultaneously.

Most aforementioned methods focus on single intent prediction, while users usually indicate multiple intents in real-world scenario. In order to utilize multiple intents to lead slot prediction, Gangadharaiah and Narayanaswamy (2019) first propose an attention-based neural network model for multiple intent detection and slot filling. However, it dose not map fine-grained intent information to slot filling that each token is provided with the same multiple intents information. Qin et al. (2020) indicate that incorporating the same intents information for all tokens may lead to ambiguity, thus an adaptive graph-interactive framework for joint multiple intent detection and slot filling is introduced. To achieve fine-grained multiple intent integration, they use graph attention network to connect multiple intents and slot. Furthermore, Qin et al. (2021) suggest to utilize no-autoregressive model to accelerate the process of training and inference, which has achieved promising performance.

Compared with previous works, we apply topic information into joint intent detection and slot filling. It is observed that there is strong consistency between intents and topics of an utterance. Therefore, a topic fusion mechanism is produced to combine sentence-level topic information and token-level topic information with the context, which reinforces the intent prediction on sentence level and guides each token for slot filling.

5 Conclusion

In this paper, we leverage topic information produce by LDA for joint intent detection and slot filling. To this end, a topic fusion mechanism is introduce to combine topic information with the context. Such a fusion mechanism is used to enhance the prediction of intent and guide slot filling on each token. Experimental results show effectiveness of our model and outperform previous state-of-the-art models on two public datasets mostly. In the future, we will focus on how to integrate LDA model with neural network and attempt to apply it into other NLU tasks.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. Arxiv:1409.0473.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Yun Nung Chen, Dilek Hakkani-Tur, Gokhan Tur, Jianfeng Gao, and Li Deng. 2016. End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding. In *The 17th Annual Meeting of the International Speech Communication Association (INTERSPEECH 2016)*. The International Symposium on Computer Architecture.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation.

Haihong E, Peiqing Niu, Zhongfu Chen, and Meina Song. 2019. A novel bi-directional interrelated model for joint intent detection and slot filling. In *Association for Computational Linguistics*, pages 5467–5471. Association for Computational Linguistics.

Rashmi Gangadharaiah and Balakrishnan Narayanaswamy. 2019. Joint multiple intent detection and slot labeling for goal-oriented dialog. In *NAACL-HLT (1)*, pages 564–569. Association for Computational Linguistics.

Yu Gong, Xusheng Luo, Yu Zhu, Wenwu Ou, Zhao Li, Muhua Zhu, Kenny Q. Zhu, Lu Duan, and Xi Chen. 2019. Deep cascade multi-task learning for slot filling in online shopping assistant. In *Association for the Advance of Artificial Intelligence*, pages 6465–6472. Association for the Advance of Artificial Intelligence Press.

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 *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (2)*, pages 753–757. Association for Computational Linguistics.

Daniel Guo, Gokhan Tur, Wen Tau Yih, and Geoffrey Zweig. 2014. Joint semantic utterance classification and slot filling with recursive neural networks. In *Spoken Language Technology*, pages 554–559. Institute of Electrical and Electronics Engineers.

Patrick Haffner, Gokhan Tur, and Jerry H. Wright. 2003. Optimizing svms for complex call classification. In *International Conference on Acoustics, Speech and Signal Processing (1)*, pages 632–635. Institute of Electrical and Electronics Engineers.

Dilek Hakkani-Tur, 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 *The 17th Annual Meeting of the International Speech Communication Association*, pages 715–719. The International Symposium on Computer Architecture.

Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. Arxiv:1412.6980.

John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conf. on Machine Learning*, pages 282–289.

Bing Liu and Ian Lane. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. *CoRR*, abs/1609.01454.

A. McCallum, D. Freitag, and F. Pereira. 2000. Maximum entropy Markov models for information extraction and segmentation. In *Proceedings of the International Conference on Machine Learning*.

Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. 2018. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Eda Okur, Shachi H. Kumar, Saurav Sahay, Asli Arslan Esme, and Lama Nachman. 2019. Natural language interactions in autonomous vehicles: Intent detection and slot filling from passenger utterances. *CoRR*, abs/1904.10500.

L. Qin, X. Xu, W. Che, and T. Liu. 2020. Agif: An adaptive graph-interactive framework for joint multiple intent detection and slot filling. The 2020 Conference on Empirical Methods in Natural Language Processing.

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. Association for Computational Linguistics.

Libo Qin, Fuxuan Wei, Tianbao Xie, Xiao Xu, and Ting Liu. 2021. Gl-gin: Fast and accurate non-autoregressive model for joint multiple intent detection and slot filling. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics.

Steffen Rendle. 2010. Factorization Machines. In *Proceedings of the 2010 IEEE International Conference on Data Mining*, ICDM '10, pages 995–1000. IEEE.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *NIPS*, pages 3104–3112.

Yusheng Tian and Philip John Gorinski. 2020. Improving end-to-end speech-to-intent classification with reptile. *CoRR*, abs/2008.01994.

Gokhan Tur and Renato De Mori. 2011. *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*. Wiley.

Gokhan Tur, Dilek Hakkani-Tur, and Larry P. Heck. 2010. What is left to be understood in atis? In *Spoken Language Technology Workshop*, pages 19–24. Institute of Electrical and Electronics Engineers.

J. Wang, K. Wei, M. Radfar, W. Zhang, and C. Chung. 2020. Encoding syntactic knowledge in transformer encoder for intent detection and slot filling.

Yu Wang, Yilin Shen, and Hongxia Jin. 2018. A bimodel based rnn semantic frame parsing model for intent detection and slot filling. In *NAACL-HLT (2)*, pages 309–314. Association for Computational Linguistics.

Congying Xia, Chenwei Zhang, Xiaohui Yan, Yi Chang, and Philip S. Yu. 2018. Zero-shot user intent detection via capsule neural networks. *CoRR*, abs/1809.00385.

Puyang Xu and Ruhi Sarikaya. 2013. Convolutional neural network based triangular crf for joint intent detection and slot filling. In *Automatic Speech Recognition and Understanding Workshop*, pages 78–83. Institute of Electrical and Electronics Engineers.

Sevinj Yolchuyeva, Geza Nemeth, and Balint Gyires-Toth. 2019. Self-attention networks for intent detection. In *Recent Advances in Natural Language Processing*, pages 1373–1379. INCOMA Ltd.