Joint Intent Detection and Slot Filling via CNN-LSTM-CRF

Bamba Kane, Fabio Rossi, Ophélie Guinaudeau, Valeria Chiesa, Ilhem Quénel, Stéphane Chau
Research and Innovation Direction
ALTRAN SOPHIA-ANTIPOLIS, FRANCE
{bamba.kane, fabio.rossi, ophelie.guinaudeau, valeria.chiesa, ilhem.quenel, stephane.chau}@altran.com

2020 6th IEEE Congress on Information Science and Technology (CiSt) | 978-1-7281-6646-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/CIST49399.2021.9357183

***Abstract*—Intent detection and slot filling are two main tasks in the domain of Spoken Language Understanding (SLU). The methods employed may treat the intent detection and slot filling as two independent tasks or use a joint model. Using a joint model takes into account the cross impact between the two tasks. In this article, we introduce *CoBiC* a new model combining CNN (Convolutional Neural Network), Bidirectional LSTM (Long Short-Term Memory) and CRF (Conditional Random Field) to extract the intents and the related slots. The same architecture of *CoBiC* can either be used as an independent model or joint model for intent detection and slot filling. Our method improves the state-of-the-art results on ATIS (Airline Travel Information Systems) benchmark. We also apply our model on a private dataset consisting of clients requests to a vocal assistant. The results demonstrate that *CoBiC* has strong generalization capability.**

***Index Terms*—Spoken Language Understanding, Intent detection, Slot filling, Recurrent Neural Networks, Convolutional Neural Networks**

I. Introduction

Spoken language understanding (SLU) is an emerging field in between speech and language processing, investigating human / machine and human / human communication by leveraging technologies from signal processing, pattern recognition, machine learning and artificial intelligence. SLU systems are designed to extract the meaning from speech utterances and its applications are vast, from voice search in mobile devices to meeting summarization, attracting interest from both commercial and academic sectors. SLU is a key part of dialogue systems and contains two main tasks: Intent Detection and Slot Filling. Intent detection mainly analyzes user's utterance behavior in input sentences, such as booking tickets and hotels, asking for weather information, etc. Slot filling solves the problem of labeling specific domain keywords and attributes. Table I shows an example of intent and slots extracted from the utterance *"find movies of Tarantino"*.

TABLE I
EXAMPLE OF INTENT AND SLOTS FROM A USER QUERY

Query	find	movies	of	Tarantino
Slots	O	*genre*	O	director
Intent	find_movie			

In recent years, deep neural networks have brought strong generalization capacity allowing to learn the features of input text automatically. They can also capture deeper semantic information in the process of learning and training. So, deep neural network methods give higher performances compared with traditional statistical machine learning methods in the field of natural language processing. Many of the deep learning based methods consider intent detection and slot filling as two independent tasks [1], [2]. At the same time, due to the interdependence of slot filling and intent detection, joint recognition has become a hot research topic in SLU [3]–[5]. Intent detection can be treated as a semantic utterances classification. In this context classical methods like SVM or deep neural networks may be used. Slot filling can be seen as a sequence labelling task that makes correspondence between an input word sequence $x = (x_1, x_2, ..., x_T)$ and a slot sequence $y^s = (y_1^s, y_2^s, ..., y_T^s)$. Many approaches have been used for slot filling: Maximum Entropy Markov Models (MEMMs) [6], Conditional Random Field (CRF) [7], and Recurrent Neural Network (RNN) [3], [8]. In this article, we propose a new model named *CoBiC* which first uses a Convolutional Neural Network (CNN) [9] to encode word-level information into its word-level representation. Then, we feed a bidirectional Long Short Term Memory (BiLSTM) with the word level representation to model context information of each word. Finally, we use a sequential CRF to get the slots for the whole sentence. The proposed approach *CoBiC* has three main contributions:

1) a new architecture for joint intent detection and slot filling;
2) an empirical evaluation of the model on ATIS benchmark and on a private dataset consisting of clients requests to a vocal assistant that we named *ChatData*;
3) improvement of the performances compared to the state-of-the-art methods on ATIS benchmark;

The paper is organized as following: in section II, the state-of-the-art methods for intent detection and slot filling are described. In section III, the proposed method *CoBiC* for intent detection and slot filling is detailed. We analyze *CoBiC* working in an independent way, and we focus on *CoBiC* as a joint model which is capable of working in a synchronous way. In section IV, the experimentation details are given. The results on ATIS and *ChatData* are shown.

978-1-7281-6646-9/20/$31.00 ©2020 IEEE

II. State of the art

A. Intent detection

Intent detection can be seen as semantic utterance classification. For this purpose, many approaches have been used: CNN [10], LSTM [11], Attention based CNN [12], hierarchical attention networks [13], adversarial multi-task learning [14].

B. Slot Filling

1) Traditional method for Slot Filling: There are mainly three basic traditional methods that can be listed for slot-filling purpose: in the first group, there are dictionary-based methods [15] that use string matching in order to find slots in an utterance. The main disadvantages consist in poor performances and a limited/basic library. Rule-based methods [16] are the second group. They are capable of identifying slots through rule matching: they enrich the context by adding lexical and semantic rules, but at the same time once a new slot is added, conflicts with previous rule must be handled taking time to re-write rules, causing a negative impact on the flexibility of the model. In the third group, there are statistical methods [17] that basically are trained over an annotated text and improve themselves via multiple iteration of a given objective function.

2) Deep learning for Slot Filling: When considering slot filling as an independent task from intent detection, different methods have been employed to extract slots from utterances like CNN [18], deep LSTM [8], RNN [19], encoder-labeler deep LSTM [2], joint pointer and attention [20]

C. Joint Intent Detection and Slot Filling

The joint models for intent detection and slot filling aim to take fully advantage of the existing cross impact between the intent and the slots. For this target, *Xu and al.* [21] has proposed a CNN-CRF which uses character embeddings as the inputs to CNN model and then get a character-level representation to detect the intent and the associated slots. Joint RNN-LSTM is developed by *Hakkani and al.* [4] in order to take into account the context of each word. An attention-based BiRNN model is exposed by *Liu and Lane* [3] to help the RNN to deal with long range dependencies. Since these works do not model explicitly the relationships between the intent and slots and use the same loss function, *Wang and al.* [22] have proposed a Bi-Model based RNN (two BiLSTM and with or without LSTM decoder) and two different loss functions for intent detection and slot filling. To deal also with the "problem" of using the same loss function, *Goo and al.* [23] introduced a slot-gated mechanism. In order to address data sparsity and the lack of generalization capability. Due to the lack of annotated data, different techniques like BERT (Bidirectional Encoder Representations from Transformers) [24] are proposed for training general purpose language representation models using a huge amount of unannotated text. *Chen and al.* [25] rely on BERT to propose a new model for intent detection and slot filling. In [26], it is proposed a multi-task hierarchical approach using hierarchical convolutional network and hierarchical convolutional recurrent network in order to capture the dependencies between utterances. They aim to capture the past information in order to better interpret the present utterance said by the user. Gupta and al. [27], developed a framework which explores convolutional, reccurrent and attention models in order to allow an easy interpretation of the results.

III. Methodology

In this section, we expose the different components of our new model, *CoBiC*, for intent detection and slot filling. We detail each layer of the proposed neural network before detailing *CoBiC*'s architecture. We will also motivate why each layer of *CoBiC* is used. The global architecture of *CoBiC* is depicted on figure 3.

A. CNN

CNN is used in order to get a word-level representation for each word of an utterance. In [10], it is proven that a simple CNN with one layer of convolution on top of word vectors from an unsupervised neural language model, gives very good performances for text classification. Chiu and Nicols [28] also have shown that CNN can be used efficiently to extract morphological information like suffix or prefix of a word and encode it into a neural representation. Figure 1 shows the steps of applying CNN on text inputs.

The first part of *CoBiC* consists in a CNN having as input the embedded vector for each word of an utterance in order to get the best word-level representation.

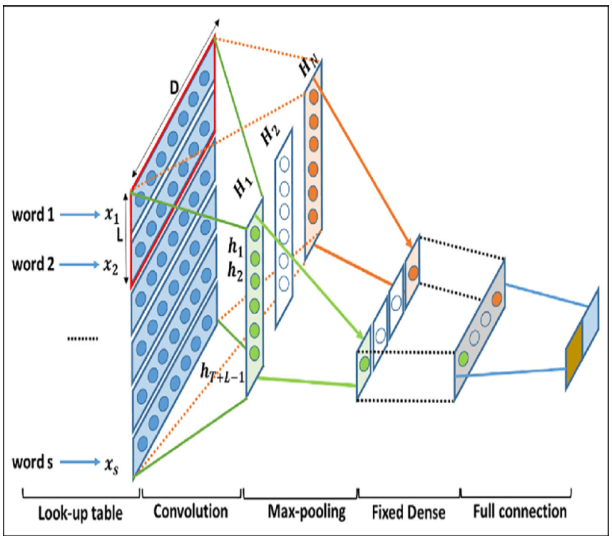

Fig. 1. CNN on text (Nguyen and al. 2019)

B. LSTM

RNN are powerful models that can capture the time dynamics and long range dependencies but they suffer from the vanishing/exploding gradient problem [29].

LSTM [30] has been proposed to solve this issue. A LSTM unit is composed of three gates: input, output and forget, which can check the proportion of information to forget or to keep for the next step.

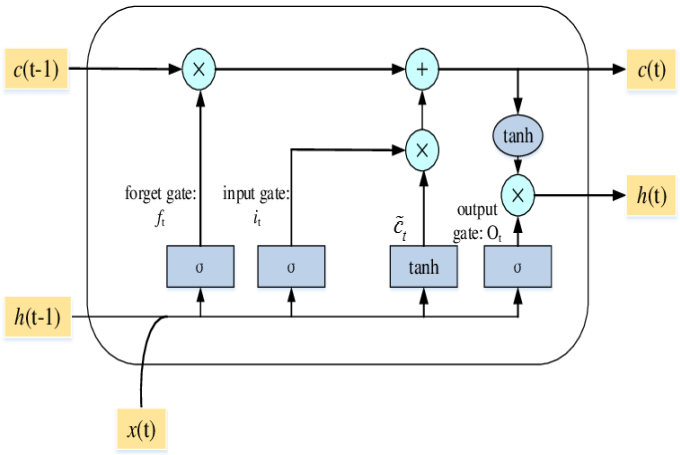

Fig. 2. LSTM unit (Yuan et al. 2019)

Figure 2 shows in details a LSTM unit with its different components.

The formulas for updating a LSTM unit at time t are:

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i)$$
$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f)$$
$$\widetilde{c}_t = \tan(W_c ht - 1 + U_c x_t + b_c)$$
$$c_t = f_t \bigodot c_{t-1} + i_t \bigodot \widetilde{c}_t$$
$$o_t = \sigma(W_0 h_{t-1} + U_0 x_t + b_0)$$
$$h_t = o_t \bigodot \tan(c_t)$$

where σ is the element-wise sigmoïd function and $\bigodot$ is the element-wise product. x_t corresponds to the input vector (like word embedding) at time t, and h_t is the hidden state (or output) vector storing all the useful information at (and before) time t. U_i, U_f, U_c, U_o denote the weight matrices of different gates for input x_t, and W_i, W_f, W_c, W_o are the weight matrices for hidden state h_t. b_i, b_f, b_c b_o denote the bias vectors.

Due to the cross impact between intent detection and slot filling, getting access to both right context (future information) and left context (past information) is very important. A LSTM unit only captures information from the past. A proven solution is the use of bidirectional LSTM (BiLSTM). It consists in showing each sequence forwards and backwards to two different hidden state in order to get the past and future information and, finally, to get the output by concatenating the two hidden states.

C. CRF

For slot filling, it is interesting to take into account the correlation between slots and then decode the best chain of slots for an utterance.

Formally, let $z = \{z_1, ..., z_n\}$ be an utterance where z_i is the vector for the i^{th} word and $y = \{y_1, ..., y_n\}$ be the generic sequence of slots for z. $\gamma(z)$ constitutes the set of possible slots for z. CRF corresponds to a family of conditional probability $P(y|z; W; b)$ over all possible slot sequences y given z and can be formulated as:

$$P(y|z; W; b) = \frac{\prod_{i=1}^{n} \psi_i(y_{i-1}, y_i, z)}{\sum_{y' \in \gamma(z)} \prod_{i=1}^{n} \psi_i(y'_{i-1}, y'_i, z)}$$

where $\psi_i(y', y, z) = exp(W^T_{y',y} z_i + b_{y',y})$ are potential functions, and $W^T_{y',y}$ and $b_{y',y}$ are the weight vector and bias corresponding respectively to label pair (y', y).

D. Proposed model: CoBiC

We name the proposed model *CoBiC* for *Convolution neural network, BiLSTM and CRF*. For *CoBiC*, we give as input to the CNN layer the embeddings of each word of the utterances. The CNN layer allows also to capture morphological information like prefix or suffix. Then, we feed a BiLSTM layer with the output of the CNN layer and the words embeddings in order to take into account the context of each word. Finally, the output vectors from BiLSLTM are given as inputs to the CRF layer to decode the best slot sequence. The hidden state of the last LSTM unit corresponds to the intent of an input sentence. Figure 3 gives details about the architecture of *CoBiC* with $w, w_2, ..., w_n$ corresponding to each word of the utterance and $s_1, s_2, ..., s_n$ the related slots.

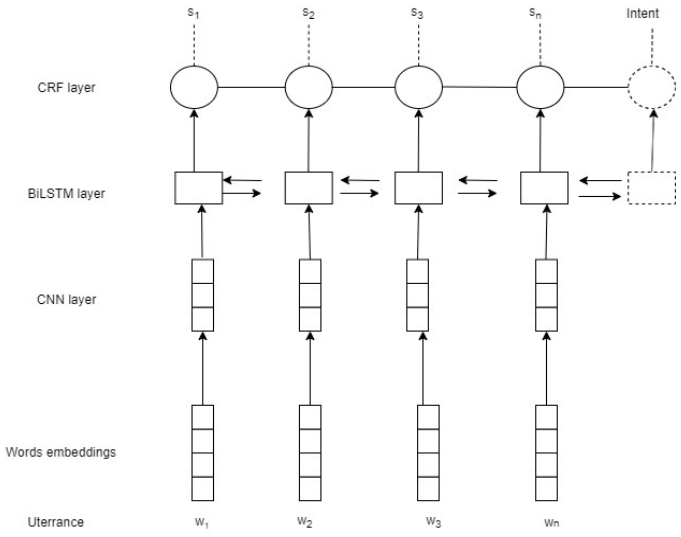

Fig. 3. Architecture of *CoBiC*

One major interest of *CoBiC* is that the same defined architecture can be used either in an independent or joint way to capture the intent and corresponding slots.

IV. Experiment

In order to evaluate the proposed model *CoBiC*, we conduct experiments on ATIS (Airline Travel Information Systems)

which is widely used in SLU. We also apply *CoBiC* on our private data *ChatData* to prove its generalization capability.

A. Datasets

The first dataset used is ATIS which contains utterances related to flight reservations. The second dataset, called *ChatData*, is provided by a private company. The utterances describe the users requests to a virtual assistant. The table II, gives details about each dataset: the vocabulary size, the number of different slots and intents, the size of the training set, the test set and the development set. For *ChatData*, we do not use a development set.

TABLE II
ATIS AND *ChatData*: DATASETS USED FOR THE EXPERIMENTS

	ATIS	***ChatData***
Vocabulary size	722	16938
#Slots	120	164
#Intents	21	296
Training set size	4478	37934
Testing set size	893	9484
Development size	500	-

Compare to ATIS which covers a single domain, *ChatData* is more challenging due to its intent diversity and very large vocabulary. On the contrary, intents in ATIS are all about flight information and the utterances have similar vocabularies. Moreover, the intents in the training set are very unbalanced with 74% of them being *atis_flight* and only one appearance for the intent *atis_cheapest*.

B. Training details

We do not use any pre-processing steps like removing the stop words because some words of the utterances may be stop words but correspond to slots (e.g. name of a song or a group, etc).

We used publicly and pre-trained available *Wikipedia2vec* [31] 300-dimensional embeddings trained on words from Wikipedia since it gives, in our context, better results than GloVe or Word2Vec. We used 50 filters with window length equal to 4 for the CNN layer. The layer size for the BiLSTM network is 200 and the number of hidden layers is 2.

C. Results

Table III shows the accuracy for intent detection and f-score for slot filling on ATIS dataset, precisely for the testing set. We compare the results between *CoBiC* as a joint model and state-of-art models. When applied on ATIS dataset, we observe that *CoBiC*, used as a joint model improves the performances compared with state-of-the-art models. Some of the models are only designed for slot filling, therefore the accuracy for intent detection is not given.

One major interest of *CoBiC* is that its architecture can also be used to capture intent and slots in a independent way. We compare in table IV the accuracy for intents and f-score for slots when *CoBiC* is used to extract them in an independent way and when they are captured in a joint manner. We observe

TABLE III
ATIS: COMPARISON OF MODELS FOR INTENT DETECTION AND SLOT FILLING: *CoBiC* OUTPERFORMS EXISTING MODELS

	Slots (F1)	**Intent(Acc)**
Recursive NN [5]	93.96	95.4
Online joint SLU [32]	94.64	98.21
Hybrid RNN [1]	95.06	-
Slot Gated [23]	95.2	94.1
RNN-EM [19]	95.25	-
CNN CRF [21]	95.35	-
Encoder-labeler Deep LSTM [2]	95.66	-
Joint GRU Model [33]	95.49	98.10
Attention Encoder-Decoder NN [3]	95.87	98.43
Attention BiRNN [3]	95.98	98.21
Joint BERT [25]	96.1	97.5
Joint BERT + CRF [25]	96.0	97.9
Bi-model without a decoder [22]	96.65	98.76
Bi-model with a decoder [22]	96.89	98.99
CoBiC	**97.82**	**99.43**

in table IV that the accuracy for intents is improved. It means that less significant intents are misclassified. Moreover, the f-score for slots is also improved meaning that we better predict the slots. These results confirm that *CoBiC* captures well the cross impact between intent and related slots.

TABLE IV
COMPARISON OF JOINT MODEL AND INDEPENDENT ONE USING *CoBiC*

	Slots (F1)	**Intent(Acc)**
Independent	97.69	97.42
Joint	97.82	99.43

D. CoBiC on ChatData

To prove the generalization capability of *CoBiC*, we apply it on *ChatData* described in the table II. The table V shows that the intents and correlated slots are very well predicted. The improvement of performances between joint and independent models proves that *CoBiC* fully takes into account the cross impact between intents and corresponding slots.

TABLE V
ACCURACY FOR INTENT AND SLOTS WITH *CoBiC* ON *ChatData*

	Slots (F1)	**Intent(Acc)**
Independent	88.71	97.85
Joint	92.97	99.56

Also, these results are the proof that *CoBiC* has strong generalization capability since it can be applied to capture intent and slots from real life data which contains diverse intents and a very large vocabulary.

V. CONCLUSION

In this paper, we propose a new model *CoBiC* for joint intent detection and slot filling. *CoBiC* combines CNN, BiLSTM and CRF layers. We test *CoBiC* on ATIS dataset and on a

private dataset *ChatData*. *CoBiC* outperforms the state-of-the-art results for intent detection and slot filling. We also show that the performances are improved when CoBiC is used as a joint model compared with its application for intent detection and slot filling in an independent way. The results prove that *CoBiC* fully takes into account the dependency between the intents and the related slots. The strong generalization capability of *CoBiC* is demonstrated by reaching likely the same values of performances for intent and slots when *CoBiC* is applied on a real life dataset depicting a deeper complexity. In order to reduce intent and slot misclassification, we plan to introduce into *CoBiC* an attention mechanism in order to identify where to put the focus in the utterances.

REFERENCES

[1] G. Mesnil, Y. Dauphin, K. Yao, Y. Bengio, L. Deng, D. Hakkani-Tur, X. He, L. Heck, G. Tur, D. Yu, and G. Zweig, "Using recurrent neural networks for slot filling in spoken language understanding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 3, pp. 530–539, Mar. 2015.

[2] G. Kurata, B. Xiang, B. Zhou, and M. Yu, "Leveraging sentence-level information with encoder LSTM for semantic slot filling," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, J. Su, X. Carreras, and K. Duh, Eds. The Association for Computational Linguistics, 2016, pp. 2077–2083.

[3] B. Liu and I. Lane, "Attention-based recurrent neural network models for joint intent detection and slot filling," in *Interspeech 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, CA, USA, September 8-12, 2016*, N. Morgan, Ed. ISCA, 2016, pp. 685–689.

[4] D. Hakkani-Tür, G. Tür, A. Çelikyilmaz, Y. Chen, J. Gao, L. Deng, and Y. Wang, "Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM," in *Interspeech 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, CA, USA, September 8-12, 2016*, N. Morgan, Ed. ISCA, 2016, pp. 715–719.

[5] D. Guo, G. Tür, W. Yih, and G. Zweig, "Joint semantic utterance classification and slot filling with recursive neural networks," in *2014 IEEE Spoken Language Technology Workshop, SLT 2014, South Lake Tahoe, NV, USA, December 7-10, 2014*. IEEE, 2014, pp. 554–559.

[6] A. McCallum, D. Freitag, and F. C. N. Pereira, "Maximum entropy markov models for information extraction and segmentation," in *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, June 29 - July 2, 2000*, P. Langley, Ed. Morgan Kaufmann, 2000, pp. 591–598.

[7] C. Raymond and G. Riccardi, "Generative and discriminative algorithms for spoken language understanding," in *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium, August 27-31, 2007*. ISCA, 2007, pp. 1605–1608.

[8] K. Yao, B. Peng, Y. Zhang, D. Yu, G. Zweig, and Y. Shi, "Spoken language understanding using long short-term memory neural networks," in *2014 IEEE Spoken Language Technology Workshop, SLT 2014, South Lake Tahoe, NV, USA, December 7-10, 2014*. IEEE, 2014, pp. 189–194.

[9] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[10] Y. Kim, "Convolutional neural networks for sentence classification," *CoRR*, vol. abs/1408.5882, 2014.

[11] S. V. Ravuri and A. Stolcke, "Recurrent neural network and LSTM models for lexical utterance classification," in *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015*. ISCA, 2015, pp. 135–139.

[12] Z. Zhao and Y. Wu, "Attention-based convolutional neural networks for sentence classification," in *Interspeech 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, CA, USA, September 8-12, 2016*, N. Morgan, Ed. ISCA, 2016, pp. 705–709.

[13] Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, "Hierarchical attention networks for document classification," in *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, K. Knight, A. Nenkova, and O. Rambow, Eds. The Association for Computational Linguistics, 2016, pp. 1480–1489.

[14] P. Liu, X. Qiu, and X. Huang, "Adversarial multi-task learning for text classification," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, R. Barzilay and M. Kan, Eds. Association for Computational Linguistics, 2017, pp. 1–10.

[15] P. Xu and R. Sarikaya, "Targeted feature dropout for robust slot filling in natural language understanding," in *INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18, 2014*, H. Li, H. M. Meng, B. Ma, E. Chng, and L. Xie, Eds. ISCA, 2014, pp. 258–262.

[16] S. Ren, H. Wang, D. Yu, Y. Li, and Z. Li, "Joint intent detection and slot filling with rules," in *Proceedings of the Evaluation Tasks at the China Conference on Knowledge Graph and Semantic Computing (CCKS 2018), Tianjin, China, August 14-17, 2018*, ser. CEUR Workshop Proceedings, S. Hu and L. Zou, Eds., vol. 2242. CEUR-WS.org, 2018, pp. 34–40.

[17] M. Surdeanu, D. McClosky, J. Tibshirani, J. Bauer, A. X. Chang, V. I. Spitkovsky, and C. D. Manning, "A simple distant supervision approach for the TAC-KBP slot filling task," in *Proceedings of the Third Text Analysis Conference, TAC 2010, Gaithersburg, Maryland, USA, November 15-16, 2010*. NIST, 2010.

[18] N. T. Vu, "Sequential convolutional neural networks for slot filling in spoken language understanding," in *Interspeech 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, CA, USA, September 8-12, 2016*, N. Morgan, Ed. ISCA, 2016, pp. 3250–3254.

[19] B. Peng and K. Yao, "Recurrent neural networks with external memory for language understanding," *CoRR*, vol. abs/1506.00195, 2015.

[20] L. Zhao and Z. Feng, "Improving slot filling in spoken language understanding with joint pointer and attention," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, I. Gurevych and Y. Miyao, Eds. Association for Computational Linguistics, 2018, pp. 426–431.

[21] P. Xu and R. Sarikaya, "Convolutional neural network based triangular CRF for joint intent detection and slot filling," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, December 8-12, 2013*. IEEE, 2013, pp. 78–83.

[22] Y. Wang, Y. Shen, and H. Jin, "A bi-model based RNN semantic frame parsing model for intent detection and slot filling," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*, M. A. Walker, H. Ji, and A. Stent, Eds. Association for Computational Linguistics, 2018, pp. 309–314.

[23] C. Goo, G. Gao, Y. Hsu, C. Huo, T. Chen, K. Hsu, and Y. Chen, "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, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*, M. A. Walker, H. Ji, and A. Stent, Eds. Association for Computational Linguistics, 2018, pp. 753–757.

[24] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds. Association for Computational Linguistics, 2019, pp. 4171–4186.

[25] Q. Chen, Z. Zhuo, and W. Wang, "BERT for joint intent classification and slot filling," *CoRR*, vol. abs/1902.10909, 2019.

[26] M. Firdaus, A. Kumar, A. Ekbal, and P. Bhattacharyya, "A multi-task hierarchical approach for intent detection and slot filling," *Knowl. Based Syst.*, vol. 183, 2019.

[27] A. Gupta, J. Hewitt, and K. Kirchhoff, "Simple, fast, accurate intent classification and slot labeling for goal-oriented dialogue systems," in *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue, SIGdial 2019, Stockholm, Sweden, September 11-13, 2019*, S. Nakamura, M. Gasic, I. Zuckerman, G. Skantze, M. Nakano, A. Papangelis, S. Ultes, and K. Yoshino, Eds. Association for Computational Linguistics, 2019, pp. 46–55.

[28] J. P. C. Chiu and E. Nichols, "Named entity recognition with bidirectional lstm-cnns," *Trans. Assoc. Comput. Linguistics*, vol. 4, pp. 357–370, 2016.

[29] Y. Bengio, P. Y. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] I. Yamada, H. Shindo, H. Takeda, and Y. Takefuji, "Joint learning of the embedding of words and entities for named entity disambiguation," in *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2016, pp. 250–259.

[32] B. Liu and I. Lane, "Joint online spoken language understanding and language modeling with recurrent neural networks," in *Proceedings of the SIGDIAL 2016 Conference, The 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 13-15 September 2016, Los Angeles, CA, USA*. The Association for Computer Linguistics, 2016, pp. 22–30.

[33] X. Zhang and H. Wang, "A joint model of intent determination and slot filling for spoken language understanding," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, S. Kambhampati, Ed. IJCAI/AAAI Press, 2016, pp. 2993–2999.