NLPCC 2019 · POSTER

Deep Multi-task Learning with Cross Connected Layer for Slot Filling

Junsheng Kong[1] , Yi Cai[1] , Da Ren[1], and Zilu Li[2]
1 South China University of Technology, Guangzhou, China
ycai@scut.edu.cn
2 Guangzhou Tianhe Foreign Language School, Guangzhou, China

Paper and datasets

Introduction

Slot filling, a sub-problem of SLU, extracts semantic constituents by using the words of input utterance to fill in predefined slots in a semantic Frame.

Motivation: This is a common scenario that different slot filling tasks from different but similar domains have overlapped sets of slots (shared slots). Consider these two sentences:

1. This sentence from computer domain:

I want to buy a computer which is about four thousand yuan and has a ram of 8 G.
Price Ram_size

2. This sentence from phone domain:

I plan to get a new mobile phone of Huawei, which costs about 3000 yuan.
Brand Price

Phone
Price
Brand
Cpu
Ram_size
Lock type
…

Camera
Price
Brand
Camera_lens
Shutter
Photorecepto
r
…

Shared slots:
Price Brand …

Non-shared slots:
Cpu Ram_size Lock type
Camera_lens Shutter
Photoreceptor …

How to share the information of shared slots between different slot filling tasks from different domains?

Dataset

We evaluate the proposed model on the datasets across multiple domains: E-commerce Computer, E-commerce Camera, E-commerce Phone. These datasets are obtained from the websites of the camera, computer and mobile phone. Then, we manually filter and tag the data to get the final datasets. These datasets are divided into three parts: train set, development set and test set. The vocabulary size of the dataset is 1189. The Table.1 shows the statistics of these datasets.

Dataset	Train	Dev	Test	Label num	Avg. length
E-commerce Computer	6145	1087	1113	47	17.20
E-commerce Camera	3408	522	521	25	21.84
E-commerce Phone	3455	626	616	37	19.00

Table 1: The statistics of the datasets

Examples:

Char	给	我	推	荐	一	台	松	下	的	相	机
Label	O	O	O	O	O	O	B-brand	I-brand	O	O	O
English	Please recommend me a panasonic camera										

Char	电	脑	坏	了	，	想	买	个	轻	薄	本
Label	O	O	O	O	O	O	O	O	B-product_type	I-product_type	I-product_type
English	The computer is broken, I want to buy a Ultrabook.										

Model

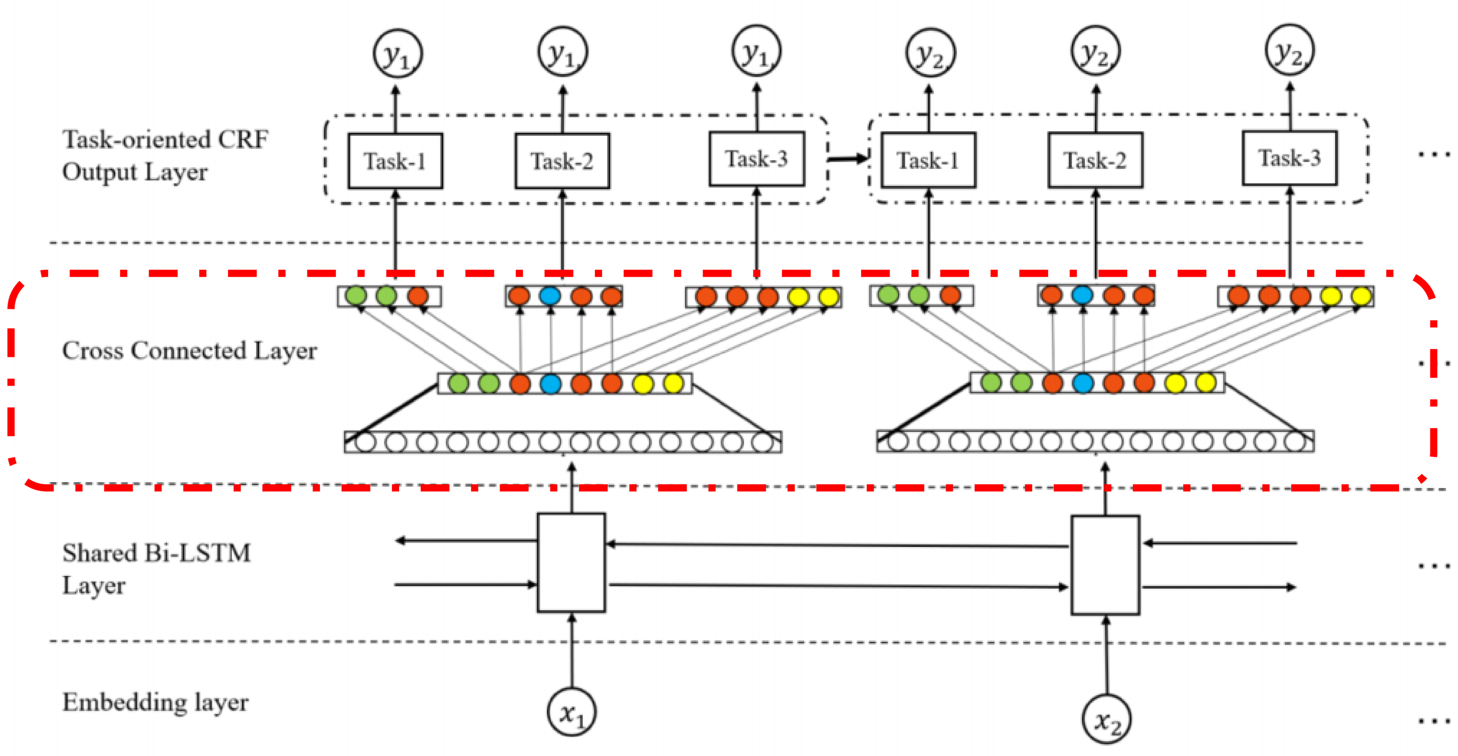

Fig. 1: The architecture of our proposed model

Cross Connected Layer（CCL）

Union process: this process transforms the hidden state h_i into the union label representation D_i on the union slots L_u.
The union slots L_u is

$$L_u = \bigcup_t L_t$$

The shared label representation D_i is

$$D_i = h_i W_{fc}$$

Here, the dimensionality of W_{fc} is $s \times l$ where s is the hidden size and l is the size of union slots L_u.

Separate process: this process convert the shared label representation Di to the task-oriented label representation $D_{(i,t)}$ for task t. For each task-oriented label representation $D_{(i,t)}$, it only use the label representation of corresponding slots. This process avoids the problem caused by that we mix all slots into one label representation. The procedure of the separate algorithm is summarized as Algorithm 1.

```
Algorithm 1: Separate slots
for t in 1, 2, ..., T do
    for (index, slot) in L_u do
        if slot in L_t then
            D_(i,t).append(D_i[index])
        end
    end
end
```

Experiments

MT-CCL models VS MT-non-CCL models

Model	Camera			Computer			Phone		
	P	R	F1	P	R	F1	P	R	F1
BiLSTM	0.878	0,904	0.8908	0.837	0.823	0.8295	0.812	0.797	0.8045
BiLSTM-CRF	0.904	0.910	0.9071	0.896	0.864	0.8800	0.863	0.863	0.8677
MT-BiLSTM	0.909	0.910	0.9092	0.870	0.837	0.8534	0.892	0.902	0.8969
MT-BiLSTM-CRF	0.916	0.919	0.9175	0.905	0.891	0.8981	0.907	0.910	0.9086
MT-BiLSTM-CCL	0.944	0.952	0.9480	0.908	0.908	0.9078	0.893	0.914	0.9033
MT-BiLSTM-CRF-CCL	0.957	0.951	0.9541	0.931	0.924	0.9274	0.916	0.926	0.9207

Table 2: Experiment results on three datasets

F1 on Shared slots VS F1 on non-shared slots

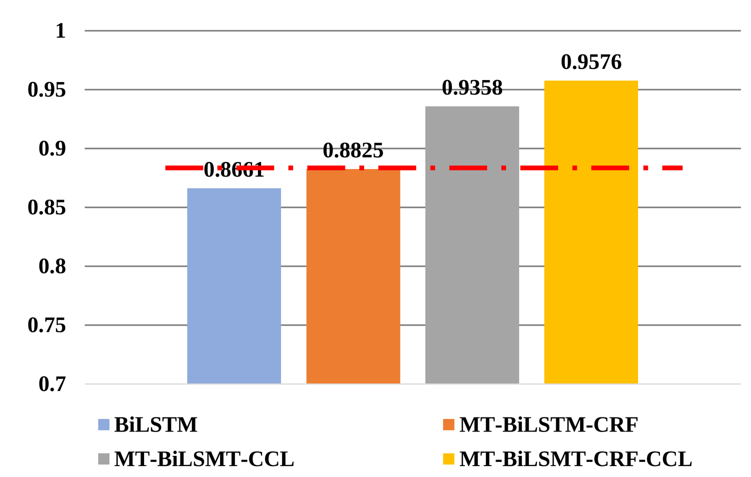

Fig. 2: F1 scores on shared slots

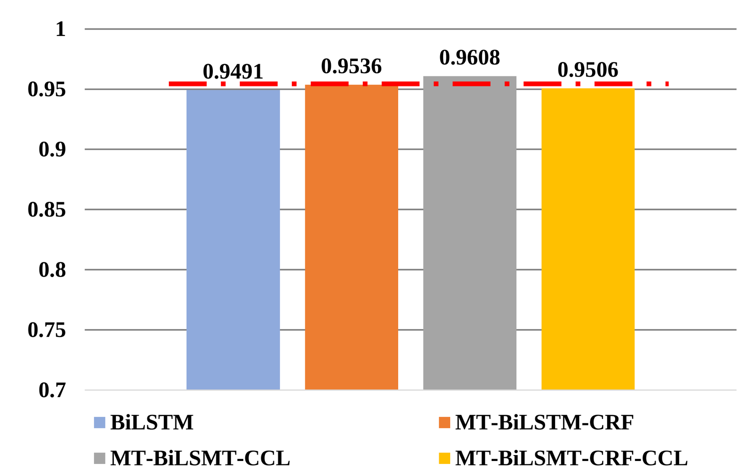

Fig. 3: F1 scores on non-shared slots

Conclusion: Compared with the other two models(BiLSTM-CRF, MT-BiLSTM-CRF), we identify that our models with CCL have a similar F1 score on non-shared slots. However, with CCL, both MT-BiLSTM-CRF-CCL and MT-BiLSTM-CCL achieves a significant improvement of F1 score on shared slots. This strongly demonstrates that CCL can effectively utilize the information of shared slots from multiple datasets.

Conclusion

1. We propose an original MTL architecture with CCL to capture the information of shared slots. The experiment results show the effectiveness of the CCL.
2. We build three datasets for slot filling tasks on three domains: computer, mobile phone, and camera. These datasets enrich the experimental data in Chinese slot filling field.