NLPCC 2019 · POSTER

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

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

## Contribution

It is common that there is semantic correspondence between slots defined in different domains. Consider these two sentences:

1. I want to buy a computer which is about $\{_{price\_middle}$ four thousand yuan$\}$ and has a ram of $\{_{ram\_size}$ 8 G$\}$.
2. I plan to get a new mobile phone of $\{_{brand}$ Huawei$\}$, which costs about $\{_{price\_middle}$ 3000 yuan$\}$.

The main contribution of this work lies on:

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.

## 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. These datasets are available online at https://github.com/JansonKong/Deep-Multi-task-Learning-with-Cross-Connected-Layer-for-Slot-Filling.

Table 1: The statistics of the 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 |

## Model

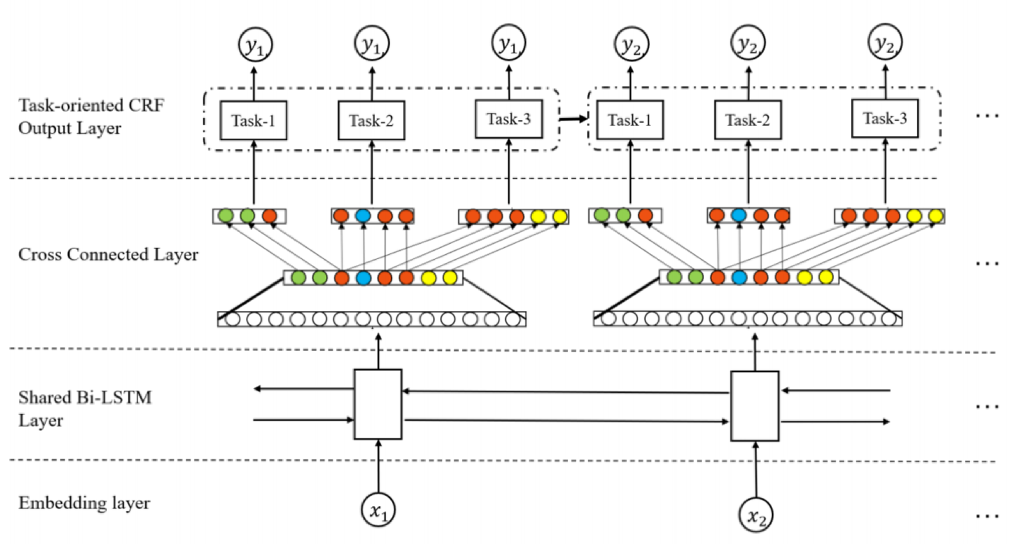


Fig. 1: The architecture of our proposed model

### Cross Connected Layer

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

$$L_u = \bigcup_t L_t. \tag{4}$$

The shared label representation $D_i$ is

$$D_i = h_i\ W_{fc}. \tag{5}$$

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 $D_i$ 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

Table 2: Experiment results on three datasets

| 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** |

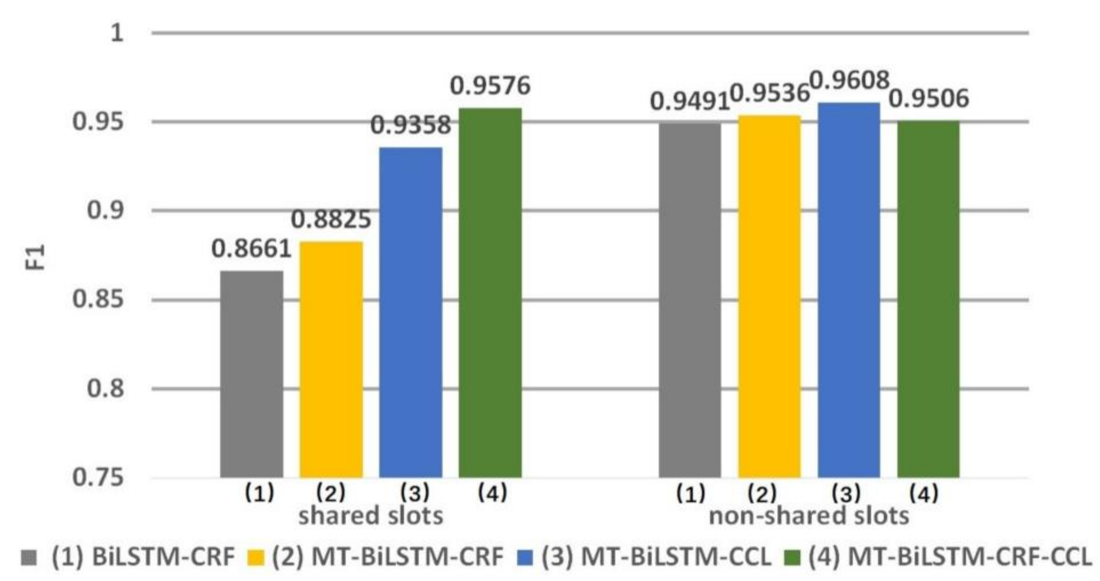


Fig. 2: F1 scores for shared slots and non-shared slots on the E-commerce camera dataset

## 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.