The Experiment of Intent Detection and Slot Filling: Based on CONDA Dataset

Yingbin Mo, Changhui He

School of Computer Science, The University of Sydney, NSW, Australia

{yimo6410, chhe9599}@uni.sydney.edu.au

+86 18022368372, +61 451353933

503/15 Garrigarrang Ave, Kogarah 2217 NSW

Abstract

Intent Detection and Slot Filling are fundamental tasks in the fields of Natural Language Understanding (NLU) and Spoken Language Understanding (SLU), and they are often found in task-based dialogue systems or log-based chat systems. In this paper, we conducted experiments on utterance-level and slot-level tasks using the CONtextual Dual-Annotated (CONDA) dataset (Weld H, et al., 2021) provided by the NLP group at the University of Sydney. We built two separate base models based on the pre-training and fine-tuning paradigm as well as Self-Attention mechanism to complete the experiment. The code and experimental result have been submitted to the AJCAI 2022 workshop.

1 Introduction

Intent Detection refers to understanding the intention of the speaker, so it is essentially a problem of classifying semantic sentences, while Slot Filling refers to the extraction of semantic components from sentences, so it is essentially a sequence annotation problem with explicit alignment information (Qin, et al., 2021).

In the traditional strategy, Intent Detection and Slot Filling would be considered as two separate tasks. A common Intent Detection scheme is RNN-based (Ravuri and Stolcke, 2015), while a common Slot Filling scheme is CRF+LSTM (Zhang, Xu and Yu, 2015). According to previous experience, we adopted a BERT pre-trained language model for the Intent Detection task and incorporated Self-Attention mechanism in BiLSTM+CRF model for the Slot Filling task.

Specifically, for the task of Intent Detection, we used a Toxic-BERT model (Hanu, 2020) as the backbone to extract utterance-level features. The model was built by Laura Hanu at Unitary (whose work is to block harmful content online by interpreting visual content in context) and pre-trained on a Wikipedia comment dataset. This pre-trained model builds a multi-headed model capable of detecting different types of toxicity, such as threat, obscenities, insults and identity-based hate. We changed the multiclassification header and fine-tuned it based on the CONDA dataset.

As for another task, Slot Filling, we used a bidirectional LSTM model incorporating a Self-Attention mechanism as the backbone to extract token-level features. And we make use of a statistical modelling approach (CRF) to make the model capable of structured prediction, i.e., context-based prediction of tokens.

2 Environment

The experiments were performed on Google Colab Platform. Training and fine-tuning processes were accelerated by a Tesla T4 GPU with 32GB RAM. CUDA version is 11.2.

3 Experiment Results

The performance of two tasks in different models, training strategies and hyperparameters is shown in Table 1. We trained and fine-tuned the models using different learning rates and training set sizes. Since the metric used in the AJCAI 2022 workshop is JSA (Joint Semantic Accuracy), that is, both utterance-level and all the token-level labels need to be correctly predicted. Therefore, accuracy is the metric for our training and tuning single-task model.

Task	Lr	Model	Training strategy	Accuracy	U_E(f1)	U_I(f1)	T_T(f1)	T_D(f1)	T_S(f1)
Intention Detection	1e-3	Pre-trained BERT: Last Hidden State	Training: 26086 Validation: 8703	0.9170	0.8666	0.7377	/	/	/
	3e-3	Pre-trained BERT: Pooler Output		0.9188	0.8714	0.7438	/	/	/
				0.9187	0.8696	0.7485	/	/	/
			Training: 31086 Validation: 3703	0.9199	0.8767	0.7576	/	/	/
Slot Filling	1e-2	BiLSTM+CRF+ Attention	Training: 26086 Validation: 8703	0.9938	/	/	0.9721	0.9554	0.9941
	1e-3		Training: 31086 Validation: 3703	0.9953	/	/	0.9822	0.9657	0.9968
	8e-3			0.9955	/	/	0.9830	0.9712	0.9960

Table 1 The performance of two tasks

For the Intention Detection, we first tried a learning rate of 0.001 and found that there was still no fit after 25 epochs of training, so we increased the learning rate to 0.003 and used pooler output instead of last hidden state as the representation of training data, the model fits faster and the model performance improves. In order to further improve the performance of the model, we took 5000 samples from the validation set to expand the training set, and the performance of the model has been further improved.

	JSA	U_E(f1)	U_I(f1)	T_T(f1)	T_D(f1)	T_S(f1)
Joint Result	0.9012	0.8661	0.7537	0.9846	0.9691	0.9952

Table 2 The final performance in the competition

For the Slot Filling, since we use a completely different model from the Intention Detection, the selection of the learning rate is also slightly different. It is worth mentioning that for the Slot Filling, we used SGD as the optimizer, while the optimizer for Intention Detection is AdamW. The reason is that compared to Intention Detection, Slot Filling can use fewer epochs to find the global minimum, which means that an optimizer such as AdamW with a faster fitting speed may not be a better choice. We train the model with a learning rate of 0.01, and then reduce the learning rate and increase the training set size to further optimize the model.

After the two tasks predict the intent and token-label respectively, the results are combined in the same file into txt format to be the final prediction result. The performance in the competition's private test set is shown in table 2.

Conclusion

In this paper, we conduct experiments on a Dual-Annotated in-game toxicity dataset (CONDA). We used pre-trained BERT and BiLSTM+CRF+Attention to perform Intention Detection and Slot Filling tasks on the CONDA dataset. We conducted several experiments to try different model structures and hyperparameters and fine-tuned the model.

References

Hanu, L. (2020). Unitary team. Detoxify. Github.

Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991.

Qin, L., Xie, T., Che, W., & Liu, T. (2021). A survey on spoken language understanding: Recent advances and new frontiers. arXiv preprint arXiv:2103.03095.

Ravuri, S., & Stolcke, A. (2015). Recurrent neural network and LSTM models for lexical utterance classification. In Sixteenth Annual Conference of the International Speech Communication Association.

Weld, H., Huang, G., Lee, J., Zhang, T., Wang, K., Guo, X., ... & Han, S. C. (2021). CONDA: a CONtextual Dual-Annotated dataset for in-game toxicity understanding and detection. arXiv preprint arXiv:2106.06213.