Method Description for CCKS 2018 Task 2: Instruction Understanding in the Field of Music

ECUST_NLP

East China University of Science and Technology, School of Information Science and Engineering, Shanghai, China 200237

Abstract. The instruction understanding in the field of music aim for intention judgment and slot filling. In this paper, we first employ Bi-LSTM model to recognize named entities for slot filling, then adopt heuristic rules for intention judgment.

Keywords: Instruction understanding, Bi-LSTM model, heuristic rule

1 Task Description

The instruction understanding in the field of music includes two tasks, namely intention judgment and slot filling. We first employ Bi-LSTM model to recognize named entities for slot filling, then adopt heuristic rules for intention judgment.

2 Methods

In this section, we will describe the methods we propose, including the Bi-LSTM model for slot filling, and heuristic rules for intention judgment.

2.1 Bi-LSTM Model for slot filling

The slot filling task aims to recognize named entities in the filed of music. The named entity recognition task can be seen as a sequence labeling problem, that is, given a sequence $X =< x_1, ..., x_n >$ from texts, the goal is to label X with tag sequence $Y =< y_1, ..., y_n >$. We adopt the $BIEO$ tag scheme, where B indicates "x_i is the beginning of a clinical named entity", I indicates "x_i is inside a clinical named entity", E indicates "x_i is the end of a clinical named entity", and O indicates "x_i is outside a clinical named entity". Take $X =<$请, 播, 放, 周, 杰, 伦, 的, 稻, 香$>$ as an example, its corresponding $Y =<$O, O, O, B-artist, I-artist, E-artist, O, B-song, E-song$>$.

We employ a Bi-LSTM Model model for this task. Note that we exploit character embedding rather than word embedding, and has no additional features used. As shown in Fig. 1, the raw natural language input sentence is processed into sequence of characters $X = [x]_1^T$. The character sequence is fed into the embedding layer, which produces dense vector representation of characters. The

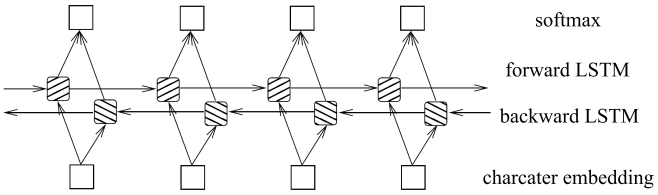

Fig. 1. Bi-LSTM model

character vectors are then fed into a bidirectional LSTM layer [1]. Specifically, for each position t, LSTM [2] computes $\boldsymbol{h}_t$ with input $\boldsymbol{e}_t$ and previous state $\boldsymbol{h}_{t-1}$, as:

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \boldsymbol{e}_t + \boldsymbol{U}_i \boldsymbol{h}_{t-1} + \boldsymbol{b}_i) \tag{1}$$
$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \boldsymbol{e}_t + \boldsymbol{U}_f \boldsymbol{h}_{t-1} + \boldsymbol{b}_f) \tag{2}$$
$$\tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{W}_c \boldsymbol{e}_t + \boldsymbol{U}_c \boldsymbol{h}_{t-1} + \boldsymbol{b}_c) \tag{3}$$
$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t \tag{4}$$
$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \boldsymbol{e}_t + \boldsymbol{U}_o \boldsymbol{h}_{t-1} + \boldsymbol{b}_o) \tag{5}$$
$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh(\boldsymbol{c}_t) \tag{6}$$

where $\boldsymbol{h}$, $\boldsymbol{i}$, $\boldsymbol{f}$, $\boldsymbol{o} \in \mathbb{R}^{d_h}$ are d_h-dimensional hidden state (also called output vector), input gate, forget gate and output gate, respectively; $\boldsymbol{W}_i$, $\boldsymbol{W}_f$, $\boldsymbol{W}_c$, $\boldsymbol{W}_o \in \mathbb{R}^{4d_h \times d_e}$, $\boldsymbol{U}_i$, $\boldsymbol{U}_f$, $\boldsymbol{U}_c$, $\boldsymbol{U}_o \in \mathbb{R}^{4d_h \times d_h}$ and $\boldsymbol{b}_i$, $\boldsymbol{b}_f$, $\boldsymbol{b}_c$, $\boldsymbol{b}_o \in \mathbb{R}^{4d_h}$ are the parameters of the LSTM; σ is the sigmoid function, and $\odot$ denotes element-wise production.

In the bidirectional LSTM, for any given sequence, the network computes both a left, $\overrightarrow{\boldsymbol{h_t}}$, and a right, $\overleftarrow{\boldsymbol{h_t}}$, representations of the sequence context at every input, x_t. The final representation is created by concatenating them as $\boldsymbol{h_t} = [\overrightarrow{\boldsymbol{h_t}}; \overleftarrow{\boldsymbol{h_t}}]$. Finally, $\boldsymbol{h_t}$ is fed into a fully connected layer with a softmax function to predict the tag y_t.

2.2 Heuristic Rules for Intention Judgment

The intention judgment task can be seen as a classification task. We simply adopt heuristic rules to deal with the task. If a sentence contains named entities recognized by Bi-LSTM model, it is music related. If not, it is music unrelated. Specifically, if the sentence contains key words such as "唱", "放", "歌" , it will be labeled as "Random" . If not, it will be labeled as "No music"

3 Experimental Settings

We first clean the utterances and only process the third sentence in it. When training, we use all the labeled data as training data. The Bi-LSTM model is trained by Adam algorithm [3], and the loss function is cross entropy function.

As for the Bi-LSTM model, each Chinese character is projected into the continuous space of 256-dimensional Euclidean space to reduce the dimensionality. The size of LSTM hidden layer is 256, and the batch size of the model is 32.

References

1. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional lstm networks. In: IEEE International Joint Conference on Neural Networks, 2005. IJCNN '05. Proceedings. (2005) 2047–2052 vol. 4
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8) (1997) 1735–1780
3. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)