

Deep Learning Coding Assignment 2

Rui Zhang

My email is zhan495@usc.edu, got 95.1% accuracy on Japanese and 95.5% on Italian, and >95th on secret language

1. Model

The model structure is one embedding layer for converting one-hot word to vector, one bi-LSTM layer for extracting the feature at each time step, and one fully connected layer to classify

During my finetuning for this model, I found bi-LSTM outperform simple RNN around 2%, and proper number of hidden units also improve accuracy around 1%. According to the paper I refer, adding a CRF layer at the end should improve the performance around 0.5%. But more RNN layers may not work well, state of the art always uses one-layer RNN not many layers. And one layer of fully connected is powerful enough for classification. Training the model at a fixed number of timesteps instead of epochs makes the performance more stable, considering different language.

2. Experiment

Optimizer used is Adam optimizer, loss is sequence loss (cross entropy loss with sequence mask)

The following parameters are chosen for training:

learning rate = $1e-2$ for 0~300 timesteps, learning rate = $1e-3$ 300~1000 timestep)

batch size: 32

timestep: 1000

word vector length: 120

hidden units: 64

3. Reference

[1] Huang, Zhiheng, Wei Xu, and Kai Yu. "Bidirectional LSTM-CRF models for sequence tagging." *arXiv preprint arXiv:1508.01991* (2015).