HW3 BoW vs CNN

Jingshu Rui

> Compare and contrast the performance of your BoW representation and CNN. Did one model demonstrate a higher dev performance than the other? What do you see as the advantages of one model over the other that might lead to this performance difference on this data?

Accuracy for BoW:

```
Epoch 0, dev accuracy: 0.844
Epoch 5, dev accuracy: 0.870
Epoch 10, dev accuracy: 0.859
Stopping training; no improvement on dev data after 10 epochs
```

Accuracy for CNN (the result of each run is different, take the result of one run as an example): in this case, the sixth epoch reached the peak, and then accuracy steadily decreased over the last few epochs.

```
Epoch 0, dev accuracy: 0.772
Epoch 1, dev accuracy: 0.801
Epoch 2, dev accuracy: 0.800
Epoch 3, dev accuracy: 0.804
Epoch 4, dev accuracy: 0.811
Epoch 5, dev accuracy: 0.814
Epoch 6, dev accuracy: 0.821
Epoch 7, dev accuracy: 0.816
Epoch 8, dev accuracy: 0.811
Epoch 9, dev accuracy: 0.806
Epoch 10, dev accuracy: 0.797
Epoch 11, dev accuracy: 0.786
Epoch 12, dev accuracy: 0.763
Epoch 13, dev accuracy: 0.757
Epoch 14, dev accuracy: 0.769

Best Performing Model achieves dev accuracy of : 0.821
```

Training Loss over Time
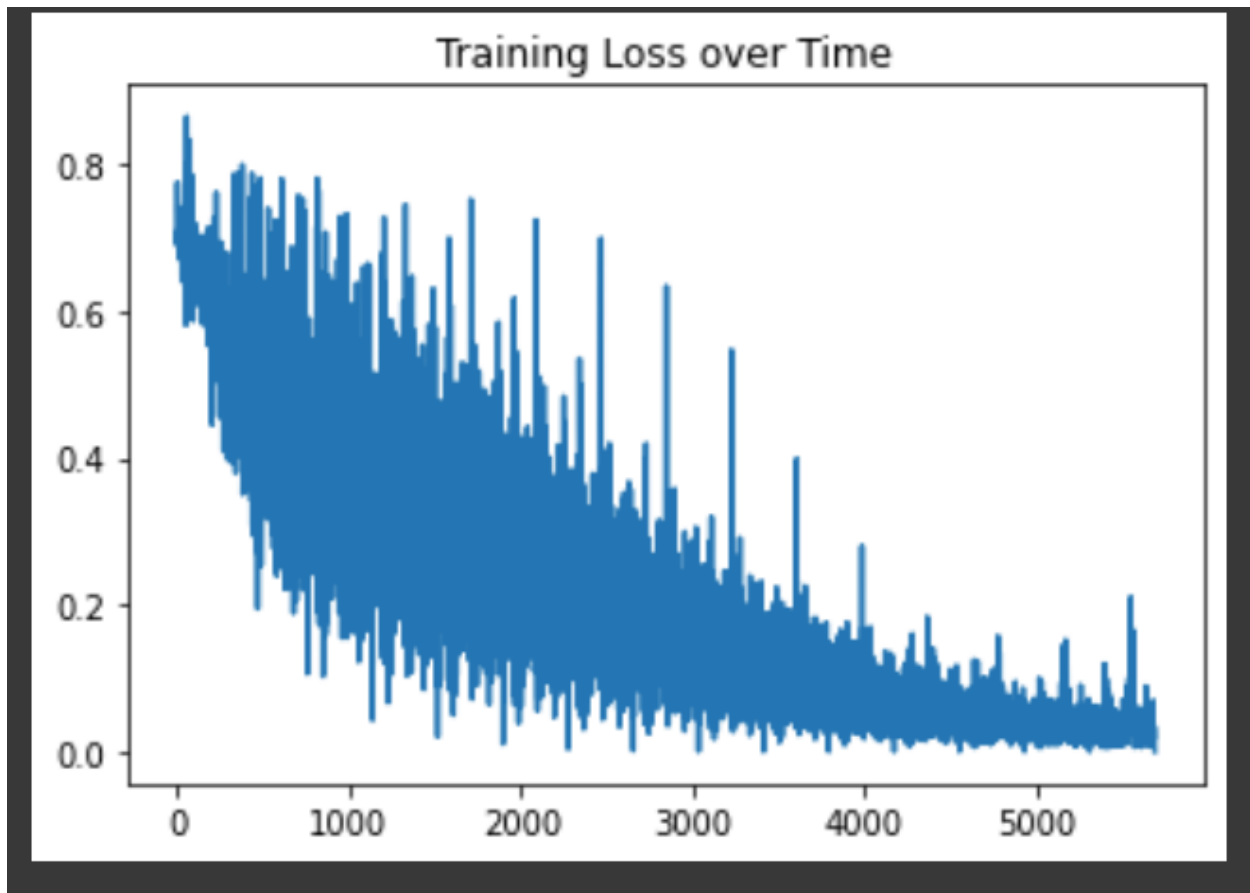
Even the best-performing model for CNN (0.821) reaches less accuracy than BoW representation (0.870). In this case, BoW is easier to implement and faster to run; it takes a shorter time to train and is more stable to converge than CNN. The visualization of the CNN loss shows that the model does not steadily decrease the loss but tends to decrease in general, yet with a large oscillation. Since the data set of this experiment is relatively small, it is not enough for training a complex network with many parameters like CNN; and with the increase of training epochs, the problem of overfitting gradually appears, so CNN is rather inferior to the simple BoW model.