

Hw3 report

Name: HSU, Chia hong

Stid: 20562937

2.1 Data Statistics:

Num_of_sent	10000
Num_of_words	592840
Vocab_size	21059
Vocab_size (freq>1)	11561
Top_10_freq_words (vocab: occurrence)	{'sound': 2962, 'ha': 3009, 'like': 3389, 'would': 3485, 'good': 3943, 'get': 3959, 'use': 4113, 'work': 4322, 'one': 5258, 'wa': 7147}
Max_sent_len	1077
Avg_sent_len	59.284
Std_sent_len	69.191
Class_Distribution	{'0': 4000, '1': 2000, '2': 4000}

2.2 Preprocessing

Cleaning methods: remove mentions, remove url, remove hashtags, remove ticks and the next character, remove punctuations, remove digits, replace the over spaces, remove stopwords, lemmatize words.

3.3 Hyperparameters

Learning Rate (dropout=0.3, kernel_num=100, kernel_sizes='3,4,5', embed_dim=100)

Learning Rate	Validation set accuracy
0.1 (Default)	best accuracy:0.564 (last accuracy:0.3472)
0.05	best accuracy:0.602 (last accuracy:0.5324)
0.03*	best accuracy:0.6036 (last accuracy:0.5428)
0.01	best accuracy:0.5304 (last accuracy:0.5204)

Dropout (lr=0.03, kernel_num=100, kernel_sizes='3,4,5', embed_dim=100)

Dropout	Validation set accuracy
0.5	best accuracy:0.5884 (last accuracy:0.5388)
0.3*	best accuracy:0.6036 (last accuracy:0.5428)
0.1	best accuracy:0.5708 (last accuracy:0.5388)
0	best accuracy:0.5836 (last accuracy:0.5504)

Kernel_num (lr=0.03, dropout=0.3, kernel_sizes='3,4,5', embed_dim=100)

Kernel_num	Validation set accuracy
50	best accuracy:0.5688 (last accuracy:0.5156)
100	best accuracy:0.6036 (last accuracy:0.5428)
150**	best accuracy:0.6148 (last accuracy:0.5612)

Kernel_sizes (lr=0.03, dropout=0.3, kernel_num=150, embed_dim=100)

Kernel_sizes	Validation set accuracy
--------------	-------------------------

'2,3,4'	best accuracy:0.5548 (last accuracy:0.5492)
'3,4,5'*	best accuracy:0.6036 (last accuracy:0.5428)
'4,5,6'	best accuracy:0.4852 (last accuracy:0.4692)

Embed_dim (lr=0.03, dropout=0.3, kernel_num=150, kernel_sizes='3,4,5')

Embed_dim	Validation set accuracy
50	best accuracy:0.5852 (last accuracy:0.5484)
100*	best accuracy:0.6036 (last accuracy:0.5428)
200	best accuracy:0.5432 (last accuracy:0.4956)

Avg pooling vs Max pooling:

(lr=0.03, dropout=0.3, kernel_num=150, kernel_sizes='3,4,5', embed_dim=100)

Parameter criteria: all the best performing parameters from above

Avg_pooling	Max_pooling*
best accuracy:0.4244 (last accuracy:0.4)	best accuracy:0.6148 (last accuracy:0.5612)

4.2 Submit your development set accuracy and test set predictions

(lr=0.02, dropout=0.2, kernel_num=150, kernel_sizes='2,3,4,5', embed_dim=100, early_stop=6, max_pooling=True)

best accuracy:0.616 (last accuracy:0.5928)

	Precision	Recall	F1-score	support
Rating 0	.701	.675	.688	1000
Rating 1	.240	.192	.213	500
Rating 2	.625	.711	.665	1000
Accuracy			.593	2500
Macro avg	.522	.526	.522	2500
Weighted avg	.579	.593	.584	2500

4.3 Analysis

Effect of kernel size:

According to the above experiment, larger kernel size [4,5,6] tends to have unstable performance (in terms of accuracy and number of run epochs). Kernel size [3,4,5] achieved highest accuracy score, while kernel size [2,3,4] also has stable performance as it converges smoothly.

However, all these might be caused by the combined effect of other factors such as dropout rate and learning rate. As I experimented larger kernel size[4,5,6] again with lower dropout rate(0.3 -> 0.2) and learning rate(0.03 -> 0.02), the accuracy score and stability improved significantly(best accuracy 0.5792, with the last three iterations, 0.5648 ,0.5564, 0.5684).

The below is another experiment comparing different kernel sizes with optimized learning rate and dropout.

Kernel_sizes (lr=0.02, dropout=0.2)	Highest validation set accuracy & Last 3 iteration accuracy
'2,3,4'	best accuracy:0.5848 (last 3 accuracy:0.574, 0.3888, 0.572)
'3,4,5'*	best accuracy:0.6004 (last 3 accuracy:0.5572, 0.5284, 0.5404)
'4,5,6'	best accuracy:0.5792 (last 3 accuracy:0.5648 ,0.5564, 0.5684)

Effect of number of kernel:

Larger number of kernels(150) performed better than that with smaller number of kernels(50).

Effect of dropout:

Among the four different dropout rates we experimented, 0.3 achieved the highest accuracy score. Although dropout rate 0.5 achieved good performance, the performance was quite unstable as it didn't seem to converge during the last three iterations. All dropout rate lower than 0.3 has showed convergence behaviour.

Effect of learning rate:

Learning rate 0.1 performance was the worst, no matter how I adjusted other parameters, it never achieved higher than 50% accuracy score, and the epoch number trained were always below 10. Lower learning rate improved significantly better.

Effect of embedding dimension:

According to experiment results, higher embedding dimension(200) didn't show better performance. Embedding dimension 50 and 100 converged to similar accuracy score, but the 100 one achieved slightly higher accuracy performance.

Max pooling vs average pooling:

Average(mean) pooling performance was notably worse than max pooling. It could be explained by having high score features neutralized by those that had lower scores. While max pooling preserved the high feature values, which could be more informative for deep learning tasks.