Forum Discussions Categorization

CS_18

Team ID	Member's names	Member's ID
CS_18	أحمد رضا قدري إبراهيم	2021170022
	أحمد أيمن محمد محمد المغربي	2021170016
	أحمد جلال حامد ابو السعادات	2021170019
	أسامة خالد محمد خيري إمام	2021170061
	إسلام رائد حسين كميل	2021170069
	عبد الرحمن أحمد ياسر محمد بيومي الوراقي	2021170274

Transformers

Transformers model Trials

Embedding size	Number of head attention	Feed forward Dimension	Number of parameters X1000	Accuracy
32	32	40	1,122	64
16	32	32	540	69
16	16	16	510	63
16	32	40	528	67
16	16	128	514	65
16	64	128	565	65
16	64	64	563	64
32	64	64	1,258	63
20	40	40	684	66
16	64	64	563	67
16	32	40	575	65
32	64	64	1,272	65
16	64	128	578	65
50	50	50	2,068	66
8	64	64	268	62

When the number of parameters decrease between 500k and 600k the accuracy increases

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 64)	0	-
embedding_4 (Embedding)	(None, 64, 32)	976,704	inputs[0][0]
add_12 (Add)	(None, 64, 32)	0	embedding_4[0][0]
multi_head_attenti… (MultiHeadAttentio…	(None, 64, 32)	134,176	add_12[0][0], add_12[0][0]
dropout_17 (Dropout)	(None, 64, 32)	0	multi_head_atten…
add_13 (Add)	(None, 64, 32)	0	add_12[0][0], dropout_17[0][0]
layer_normalizatio… (LayerNormalizatio…	(None, 64, 32)	64	add_13[0][0]
dense_24 (Dense)	(None, 64, 40)	1,320	layer_normalizat…
dense_25 (Dense)	(None, 64, 32)	1,312	dense_24[0][0]
dropout_18 (Dropout)	(None, 64, 32)	0	dense_25[0][0]
add_14 (Add)	(None, 64, 32)	0	layer_normalizat… dropout_18[0][0]
layer_normalizatio… (LayerNormalizatio…	(None, 64, 32)	64	add_14[0][0]
dense_26 (Dense)	(None, 64, 64)	2,112	layer_normalizat…
global_average_poo… (GlobalAveragePool…	(None, 64)	0	dense_26[0][0]
dropout_19 (Dropout)	(None, 64)	0	global_average_p…
dense_27 (Dense)	(None, 5)	325	dropout_19[0][0]

CNN + GRU

Optimizer Eta is gradually increases (* 1.1) in the 5 epochs then decreases (* 0.9)

```
embeding_layer_trainable = True
num_filters_in_CNN_layer = 64
kernel_size_in_CNN_layer = 3

gru_neurons_num = [128, 64, 32]

dense_neurons_num = [128]

optmizer_eta = 6e-4

batch_sz = 64
```

CNN + GRU model trials

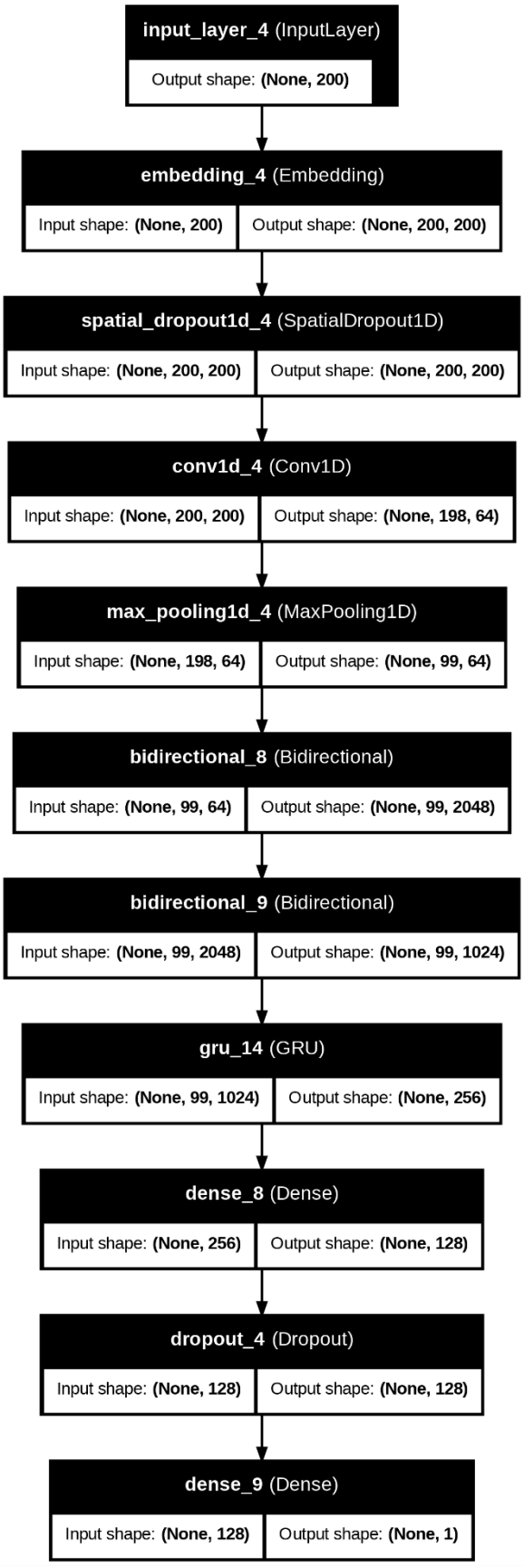
Embedding Layer weights trainable	Num Filters In CNN_layer	Kernel size in CNN_layer	Gru neurons num in layer 1	Gru neurons num in layer 2	Gru neurons num in layer 3	Dense neurons num	Optimizer Eta At the beginning	Batch size	Private Kaggle Accuracy
True	64	3	128	64	32	128	6e-4	64	0.71099
False	64	3	128	64	32	128	6e-4	64	0.70859
True	64	3	1024	512	256	128	6e-4	64	0.68290
True	64	3	64	32	16	128	6e-4	64	0.69803
True	64	3	32	16	8	128	6e-4	64	0.69131
True	64	3	16	8	4	128	6e-4	64	0.70163
True	64	3	128	64	32	128	6e-4	128	0.70619
True	64	3	128	64	32	128	6e-4	256	0.70835
True	64	3	128	64	32	128	6e-4	1024	0.68530
True	64	3	128	64	32	128	6e-4	2048	0.67882
True	128	3	128	64	32	128	6e-4	64	0.70931

Based on the trials: general conclusion: don't touch the hyperparameters of first row :,)

Batch size decrease ↠ accuracy decrease

Embedding layer weights trainable True ↠ accuracy increase

Increasing or decreasing the number of neurons in GRU layers ↠ decreases accuracy

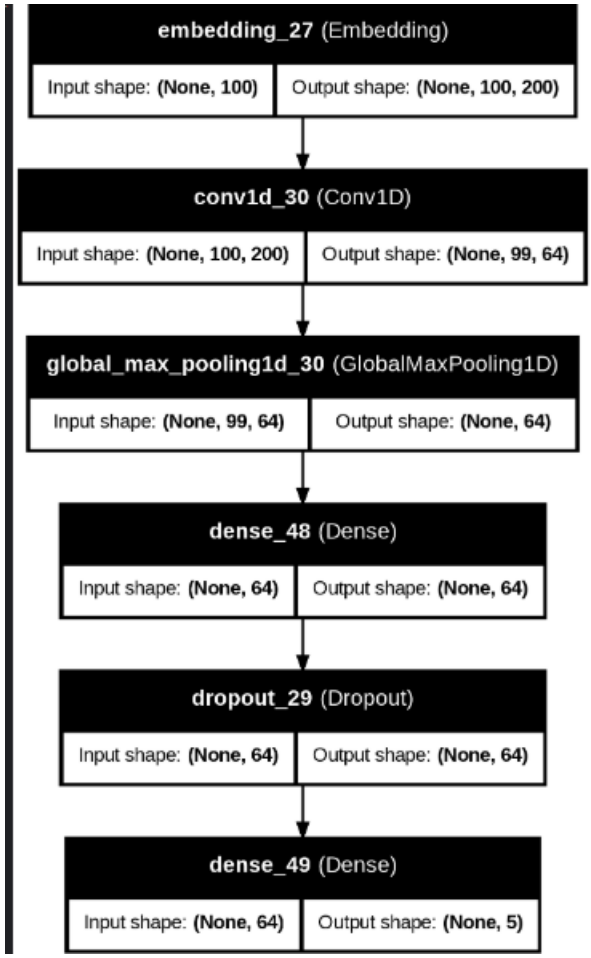

input_layer_4 (InputLayer)
Output shape: (None, 200)
embedding_4 (Embedding)
Input shape: (None, 200)
Output shape: (None, 200, 200)
spatial_dropout1d_4 (SpatialDropout1D)
Input shape: (None, 200, 200)
Output shape: (None, 200, 200)
conv1d_4 (Conv1D)
Input shape: (None, 200, 200)
Output shape: (None, 198, 64)
max_pooling1d_4 (MaxPooling1D)
Input shape: (None, 198, 64)
Output shape: (None, 99, 64)
bidirectional_8 (Bidirectional)
Input shape: (None, 99, 64)
Output shape: (None, 99, 2048)
bidirectional_9 (Bidirectional)
Input shape: (None, 99, 2048)
Output shape: (None, 99, 1024)
gru_14 (GRU)
Input shape: (None, 99, 1024)
Output shape: (None, 256)
dense_8 (Dense)
Input shape: (None, 256)
Output shape: (None, 128)
dropout_4 (Dropout)
Input shape: (None, 128)
Output shape: (None, 128)
dense_9 (Dense)
Input shape: (None, 128)
Output shape: (None, 1)

CNN

Embedding Layer weights trainable	Num of filters	Dense neurons num	Learning rate	Batch size	Num of epochs	Validation accuracy	Private Kaggle accuracy
True	128	64	0.00001	4	30	69.2	68.7
True	64	64	0.00001	8	40	69.3	68.5
False	64	64	0.00001	16	60	68.3	68.1
False	32	32	0.00005	8	100	68.1	67.7
False	64	128	0.0001	32	100	67.5	

Conclusion:

As the batch size increases the accuracy decreases. It does not matter whether the Embedding is trainable or not. I also tried with and without removing stop words but it didn't make any difference.


embedding_27 (Embedding)
Input shape: (None, 100)
Output shape: (None, 100, 200)
conv1d_30 (Conv1D)
Input shape: (None, 100, 200)
Output shape: (None, 99, 64)
global_max_pooling1d_30 (GlobalMaxPooling1D)
Input shape: (None, 99, 64)
Output shape: (None, 64)
dense_48 (Dense)
Input shape: (None, 64)
Output shape: (None, 64)
dropout_29 (Dropout)
Input shape: (None, 64)
Output shape: (None, 64)
dense_49 (Dense)
Input shape: (None, 64)
Output shape: (None, 5)

LSTM+CNN

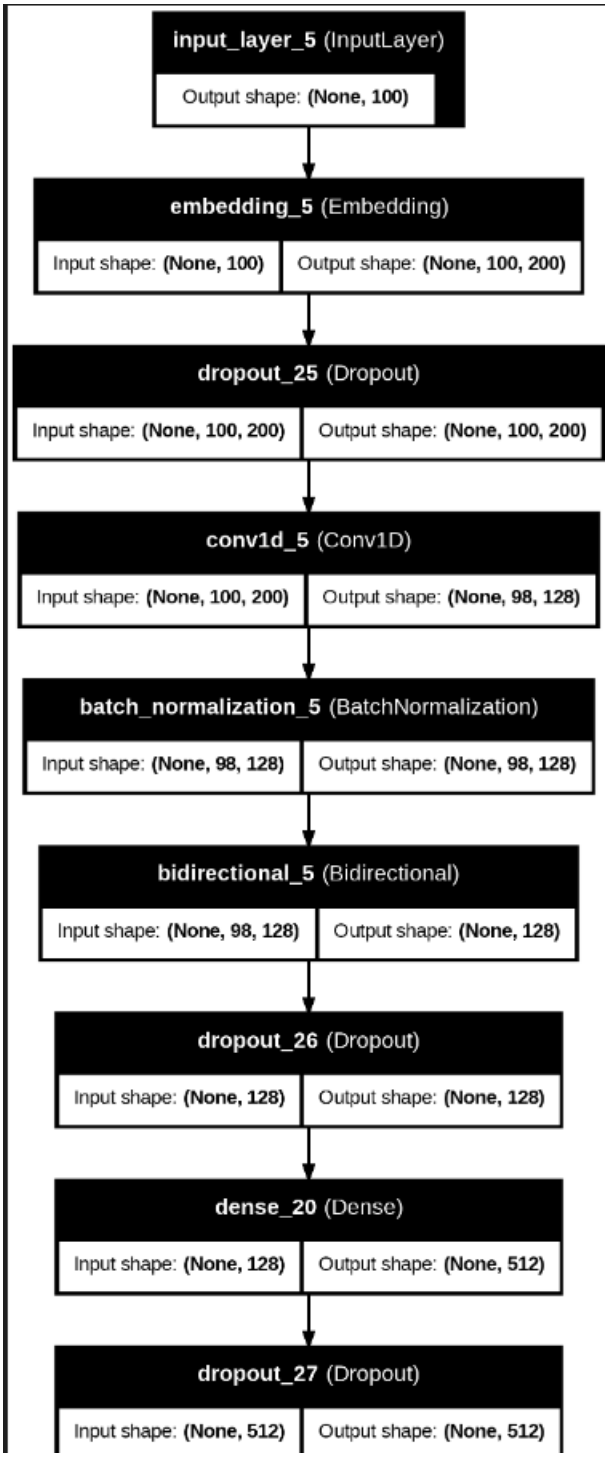
Learning rate	Regularization	Activation function	Public Accuracy	Private Accuracy
0.00002	l2 regularization in dense and LSTM 0.004, CNN 0.001	relu	69.13	67.45
0.00005	No regularization	relu	69.65	68.17
0.00005	l2 regularization in dense and LSTM 0.004, CNN 0.004	tanh	69.75	68.97
0.00005	l2 regularization in dense and LSTM 0.004, CNN 0.004	tanh	69.87	68.9

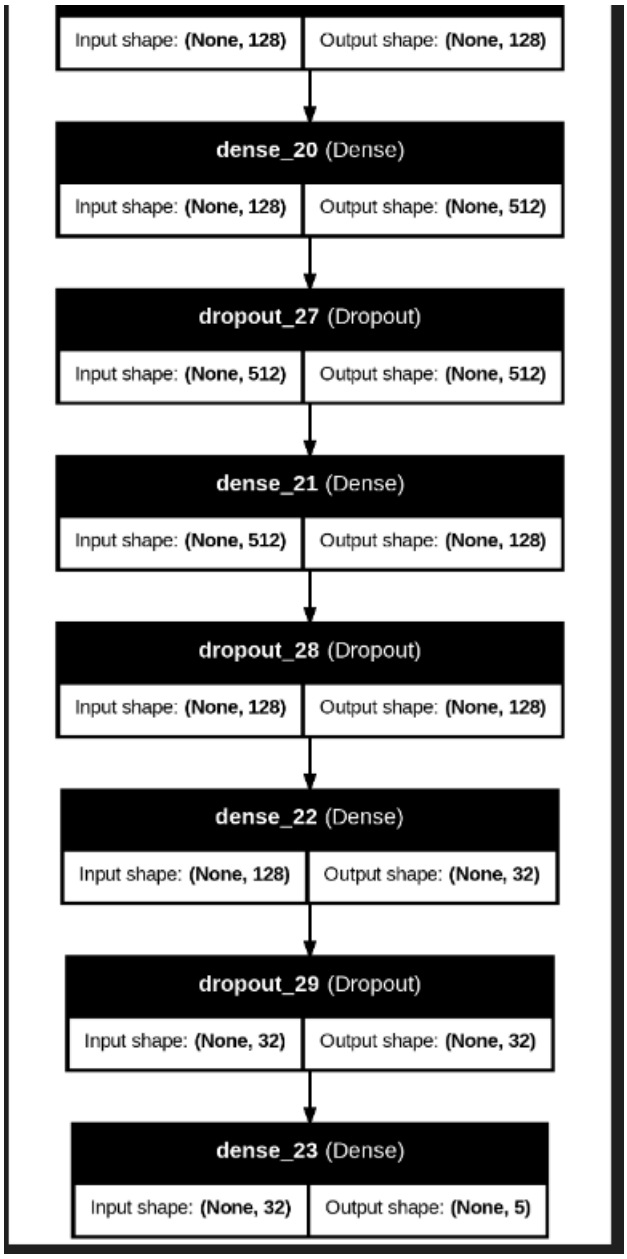
Conclusion:

Tanh is slightly better than relu.

As the learning rate decreases until 0.00005

the accuracy increases .


input_layer_5 (InputLayer)
Output shape: (None, 100)
embedding_5 (Embedding)
Input shape: (None, 100)
Output shape: (None, 100, 200)
dropout_25 (Dropout)
Input shape: (None, 100, 200)
Output shape: (None, 100, 200)
conv1d_5 (Conv1D)
Input shape: (None, 100, 200)
Output shape: (None, 98, 128)
batch_normalization_5 (BatchNormalization)
Input shape: (None, 98, 128)
Output shape: (None, 98, 128)
bidirectional_5 (Bidirectional)
Input shape: (None, 98, 128)
Output shape: (None, 128)
dropout_26 (Dropout)
Input shape: (None, 128)
Output shape: (None, 128)
dense_20 (Dense)
Input shape: (None, 128)
Output shape: (None, 512)
dropout_27 (Dropout)
Input shape: (None, 512)
Output shape: (None, 512)


Input shape: (None, 128)
Output shape: (None, 128)
dense_20 (Dense)
Input shape: (None, 128)
Output shape: (None, 512)
dropout_27 (Dropout)
Input shape: (None, 512)
Output shape: (None, 512)
dense_21 (Dense)
Input shape: (None, 512)
Output shape: (None, 128)
dropout_28 (Dropout)
Input shape: (None, 128)
Output shape: (None, 128)
dense_22 (Dense)
Input shape: (None, 128)
Output shape: (None, 32)
dropout_29 (Dropout)
Input shape: (None, 32)
Output shape: (None, 32)
dense_23 (Dense)
Input shape: (None, 32)
Output shape: (None, 5)