NLP Assignment-2
Kondipati Venkata Neeharika
SBUID 110346722

Experiment - Changing Neural Network Layers

Default Configuration :

Loss for Word Embeddings : Cross Entropy (embeddings size : 128)
Activation Function : Cubic

Configuration	UAS	LAS
#Hidden layers = 1(hidden size = 200)	75.78	68.98
#Hidden layers = 2 (hidden size =100)	72.89	65.45

Observations :
Increasing the number of hidden layers to 2, did not improve the UAS and LAS accuracy, but the loss was decreasing at a faster rate in case of 2 hidden layers.

Experiment - Capturing Interactions / Using Different Activation Functions

Default Configuration :

Loss for Word Embeddings : Cross Entropy (embeddings size : 128)
No. Hidden layers : 1

Configuration	UAS	LAS
Cubic	75.78	68.98
Tanh	64.89	55.76
ReLu	58.91	50.01
Sigmoid	63.88	55.67
Parallel Cubic	63.77	52.25

Observations :
Changing the different activation functions, I observed that highest UAS/LAS accuracy is observed for cubic activation function. Tanh and sigmoid give almost the same accuracy and ReLu do not give accuracy as good as the other activation functions.
Parallel layers implementation with cubic activation : In this case the final accuracy is not as good as normal cubic activation function as all the interaction between features are not considered i.e we have

$(E_wW_w)^3 + (E_pW_p)^3 + (E_lW_l)^3$ in case of parallel cubic activation function.

Experiment - Word embeddings using Cross Entropy / NCE Loss

Default Configuration :

Loss for Word Embeddings : Cross Entropy (embeddings size : 128)
No. Hidden layers : 1
Activation Function : Cubic

Configuration	UAS	LAS
Cross Entropy	**75.78**	**68.98**
NCE	**71.02**	**66.45**

Observations :

Word Embeddings from Cross entropy with cubic activation function gave better results than word embeddings from NCE Loss. Observed that varying the word embeddings model and trying out with different embedding sizes also effects the accuracy. Observed an improved accuracy of 78.66 when tried with CE with 256 embedding size.

Experiment – Fixing Word, POS and Dep Embeddings

Configuration	UAS	LAS
Fixing input embeddings	**55.78**	**44.98**

Observations :
By fixing the word, pos and dep embeddings by trainable = False, we observe a reduced accuracy as the model is initialized with random embeddings hence it takes lot of iterations for the loss to reduce and learn the weights in the hidden layer.

Best Configuration :

Based on the above results, the following configuration was giving the best accuracy for UAS/LAS values.
Loss for Word Embeddings : Cross Entropy (embeddings size : 128)
No. Hidden layers : 1
Activation Function : Cubic

Configuration	UAS	LAS
Best Configuration	**75.78**	**68.98**

Experiment – Without Gradient Clipping

Configuration	UAS	LAS
Without gradient clipping on best configuration	**61.34**	**49.79**

Observations :
By removing gradient clipping, we observe a very little decrease in loss in each step and hence bad accuracy.

Gradient Clipping :
It is a method useful for handling exploding or vanishing gradients.
Vanishing gradient problem : When gradients are being propagated back in time, they can vanish because they are continuously multiplied by numbers less than one.
Exploding gradient problem : This is when they get exponentially large gradients from being multiplied by numbers larger than one during back propagation.
To handle the above situations Gradient clipping will clip the gradients between two numbers to prevent them from getting too large or too small.
For our case we need gradient clipping to avoid the above situations and hence gives better results with gradient clipping.