Report- Assignment 2 - NLU- Dinesh Kumar 14428

Anonymous ACL submission

1 Dataset:-

Using the gutengerg corpus as was used in the assignment 1,datasets was divides into the train,dev adn test implemented and built the best LM in the given below setting.
S2: Train: D2-Train, Test: D2-Test

2 Taks 1:-

Built the token level LSTM-based language model using the above setting which consists of the 20 LSTM's with **1000** epochs.

2.1 Results:-

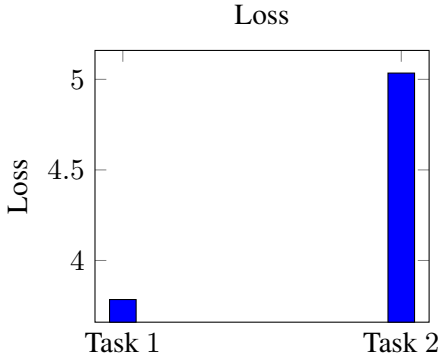
After 1000 epochs result is, loss: 3.7844 - acc: 0.2579 peplexity=exp(loss)= 44.0092

3 Task 2

Built the character level LSTM-based language model using the above setting Which consists of the 250 LSTM's with **20** epochs.

3.1 Results:

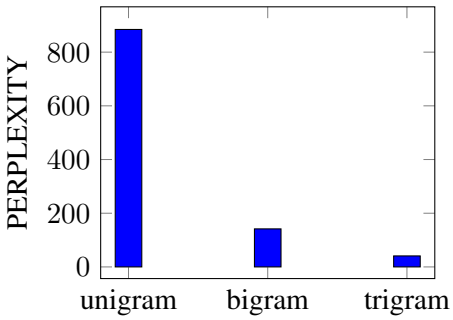
After 20 epochs result is, loss: 5.0346 peplexity=exp(loss)= 153.6381 better perplexity is expected with the larger epochs,and in that case better results will be o couldn't implement due to the lack of gpu computational power.



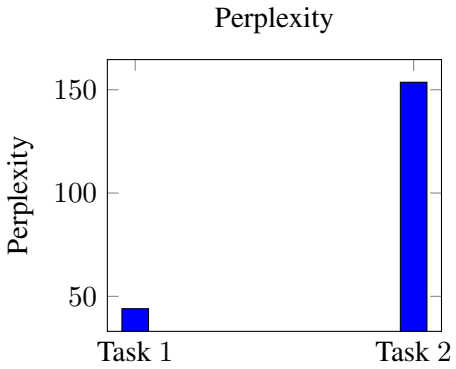
4 COMPARISON between classical and the LSTM LM:-

4.1 perplexity graph for the classical LM:-

S2: Train: D2-Train, Test: D2-Test



4.2 perplexity graph for the LSTM:-



above graphs show the perplexity comparison between the LM's.

5 Task 3:-

10 token sentence generated by task1 is:- "Sixteen minutes and the same time to be a little "

and the sentence generated by the task 2 eventhough poor performance is:- "that object which seemed to be created by entangle"