Report

Comparison and Ranking

1. Perplexity Comparison

Given below is the comparison between perplexities of neural network-based LMs (Feed Forward Neural Network (FFNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM)), i.e. the current assignment's LMs and traditional statistical models (Laplace Smoothing, Linear Interpolation, and Good-Turing),i.e. Assignment 1's LMs:

Perplexity scores of the **current LMs**: in range from 25 to 200,000

Perplexity scores of **LMs in A1**: 200 - 11,600,000

Hence, a **significant improvement** can be observed in Perplexity scores of the current LMs as compared to LMs of A1.

2. Model Ranking

1. **LSTM** - Best performing model due to its ability to capture long-term dependencies.
2. **RNN** - Performs well but struggles with longer dependencies.
3. **FFNN (n=5)** - Better than the n=3 variant due to larger context coverage.
4. **FFNN (n=3)** - Still outperforms statistical models.
5. **Linear Interpolation** - Best among traditional models.
6. **Good-Turing** - Improves over Laplace smoothing by adjusting for unseen words.
7. **Laplace Smoothing** - Worst performer, as it assigns equal probability mass to all unseen words.

3. Analysis of Results

3.1 Neural vs. Statistical Models

Neural models perform better than statistical models because they learn patterns and relationships between words instead of just using fixed probability rules. Traditional models struggle because they have limited word context and lack enough data for rare words, which makes their predictions less accurate.

3.2 Performance for Longer Sentences

LSTM is the best performer for longer sentences due to its ability to capture long-range dependencies through gating mechanisms. RNNs, while effective, struggle with vanishing gradients, making them less efficient than LSTMs for longer sequences. FFNN models, though better than statistical methods, cannot dynamically adjust their memory and struggle with very long contexts.

3.3 Effect of N-gram Size on FFNN Performance

Increasing the n-gram size improves FFNN performance since a larger context helps in better word prediction. However, beyond a certain point, increasing n leads to overfitting and computational inefficiencies. In this case, FFNN (n=5) outperforms FFNN (n=3), suggesting that five-word context is more informative for next-word prediction.

4. Conclusion

LSTM provides the best predictive performance due to its ability to model sequential dependencies efficiently. RNNs, while effective, have limitations with long-term dependencies. FFNNs benefit from larger n-grams but lack the dynamic memory capabilities of sequential models. Among statistical methods, linear interpolation is the best, though it still lags behind neural networks. This analysis confirms that deep learning-based models are significantly superior for next-word prediction tasks.

Link to the pretrained models:
https://drive.google.com/drive/folders/1YzJ805cQITGzhgmoIviBC51swIhBmcqI?usp=drive_link