Small Project 3 – Conv1D + LSTM

Jacob D. Evans

University of West Georgia

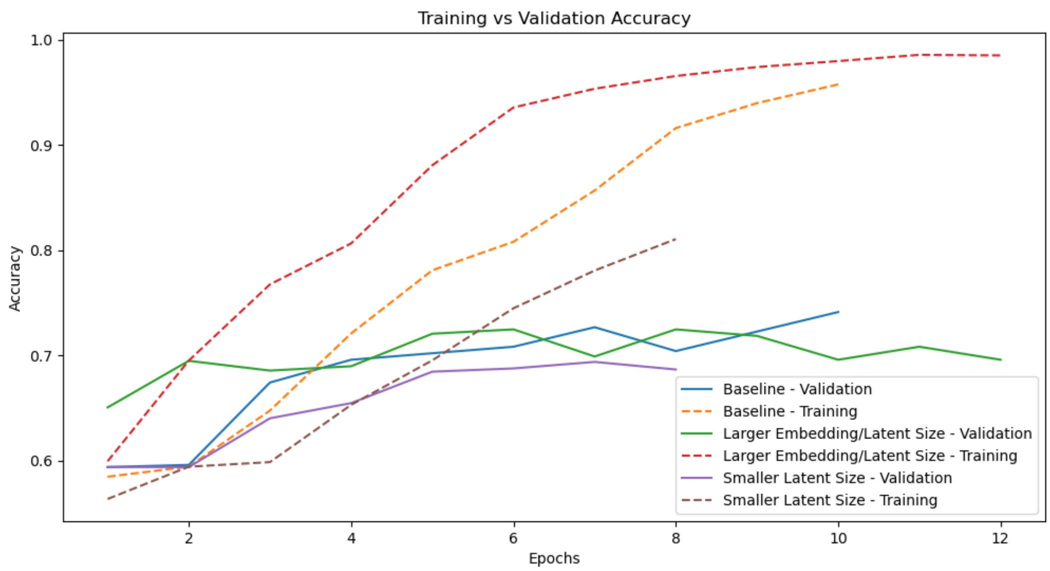
CS 4275 Machine Learning Foundations

Dr. Md Shirajum Munir

November 18, 2024

Small Project 3 – Conv1D+ LSTM

This report compares the performance of three different models that were trained with varying hyperparameters: a baseline configuration, a larger embedding/latent size, and a smaller latent size. The objective was to assess the impact of these configurations on model performance using test accuracy and loss as evaluation metrics. Additionally, the progression of training and validation accuracy across epochs was analyzed. The data remained unchanged throughout the project.



Observations

For the baseline configuration, training accuracy improved consistently, but validation accuracy plateaued and declined, indicating overfitting. Test accuracy reached 68.7%, with a relatively high test loss of 1.239, showing limited generalization. Validation accuracy improved until around epoch 4–5 before diverging, highlighting overfitting issues. The larger latent size configuration achieved the highest test accuracy of 71.2%, but its high test loss (1.563) suggests overfitting. Training accuracy increased steadily, while validation performance peaked early and then declined. The divergence between training and validation after epoch 5 further emphasized the model's inability to generalize. The smaller latent size configuration struck the best balance, with steady performance and less overfitting. It achieved a test accuracy of 68.7% and the lowest test loss (0.796), indicating better generalization. Its consistent validation accuracy after initial improvement made it the most robust option.

Recommendations

The smaller latent size configuration is the most suitable for deployment due to its balanced performance and superior generalization. To improve the larger latent size configuration, we could apply regularization techniques such as dropout, L2 regularization, or early stopping to reduce overfitting. Exploring intermediate latent sizes could reveal a better balance between complexity and performance.

To ensure the robustness of the smaller latent size, we should incorporate cross-validation to confirm consistent generalization across different data splits. This approach would validate its reliability for real-world scenarios.

Conclusion

The smaller latent size configuration emerges as the most robust option, achieving reasonable accuracy (68.7%) and the lowest test loss (0.796). However, improvements can be made to reduce overfitting in the larger latent size model, which has potential for higher accuracy if generalized effectively.