LSTM-Based Methodology for Learning Curve Prediction: Leveraging Learning Patterns of Multiple Models for New Model Performance Forecasting

Abstract

Training deep learning models requires significant time and computational resources, particularly when comparing the performance of various models such as ResNet-34, ResNet-50, Plain-34, and Plain-50. It is challenging to assess the final performance of a model based solely on insufficient data from the initial training phase. This study proposes a methodology to predict the learning curve of Plain-50 using the learning curve data of ResNet-34, ResNet-50, and Plain-34 to address this issue. Two prediction models—polynomial regression and Long Short-Term Memory (LSTM) neural networks—are employed, and their performances are compared. Notably, the LSTM model incorporates not only accuracy but also the per-epoch accuracy change (gradient) as a feature to enhance prediction accuracy. Experimental results demonstrate that the LSTM-based prediction model, trained on patterns from multiple models, captures the nonlinear patterns and convergence trends of Plain-50 more accurately than the polynomial regression model. This suggests that leveraging early training data from multiple models to predict the performance of a new model is effective in reducing unnecessary resource consumption.

1 Introduction

The training of deep learning models can require dozens or hundreds of epochs, depending on the model's complexity and the size of the dataset. This becomes particularly resource-intensive when comparing various models with different architectures, such as ResNet-34, ResNet-50, Plain-34, and Plain-50, as fully training all models demands substantial time and computational power. With only a few epochs of initial training data (e.g., 5 epochs), it is difficult to predict the final performance, making it challenging to determine which model is the most efficient and effective.

This study proposes learning curve prediction as a solution to this problem. By predicting future learning curves based on initial data accumulated during training, it becomes possible to estimate a model's final performance without completing the full training process. Specifically, this research aims to train on the learning trends of ResNet-34, ResNet-50, and Plain-34 and use this knowledge to predict the learning curve of the new model, Plain-50, thereby verifying the generalizability of the prediction model. Previous studies have primarily used parametric regression models, such as logistic functions or power-law functions, to model learning curves. However, these models assume specific curve shapes, limiting their flexibility across diverse training scenarios. To overcome this limitation, this study proposes an LSTM-based methodology that excels in sequential data analysis to learn and predict the complex nonlinear patterns of learning curves. Additionally, a traditional polynomial regression model is included for comparison to comprehensively analyze the strengths and weaknesses of both approaches.

2 Methodology

2.1 Polynomial Regression-Based Prediction

Polynomial regression is a traditional regression analysis method that identifies the polynomial function best suited to the initial learning curve data. The coefficients of the polynomial are estimated using the least squares method based on pairs of epoch and accuracy data from training. In this study, data from ResNet-34, ResNet-50, and Plain-34 were used as training data to predict the learning curve of Plain-50.

2.2 LSTM-Based Prediction

LSTM exhibits excellent performance in learning the temporal dependencies of sequential data. To maximize the predictive capability of the LSTM model in this study, features including epoch, accuracy, and gradient ($\Delta y_t = y_t - y_{t-1}$) at each time step were utilized. The LSTM model was trained on the combined learning curve data of ResNet-34, ResNet-50, and Plain-34, and subsequently used to predict the future accuracy of Plain-50 based on its initial data. Predictions are made recursively, with the model's predicted values incorporated into the input sequence for the next prediction, enabling continuous forecasting of accuracy at future time points.

2.3 Performance Evaluation Metrics

The performance of each prediction method is evaluated by comparing the error with the actual learning curve. The key evaluation metrics are as follows:

- **Root Mean Squared Error (RMSE)**: Measures the average error between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

  where y_i is the actual accuracy, $\hat{y}_i$ is the predicted accuracy, and N is the number of prediction data points.

- **Mean Absolute Error (MAE)**: Measures the average of the absolute errors, being less sensitive to outliers than RMSE.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

3 Results and Discussion

3.1 LSTM Model Performance Analysis

The LSTM model, trained on the learning patterns of ResNet-34, ResNet-50, and Plain-34, showed a significant increase in prediction accuracy as the initial training data for Plain-50 exceeded 20 epochs. Particularly at the 40-epoch mark, the predicted curve nearly perfectly converged with the actual learning curve. This indicates that the LSTM effectively learned the nonlinear convergence patterns and long-term trends of the learning curve.

3.2 Polynomial Regression Model Performance Analysis

The polynomial regression model tended to be highly sensitive to local patterns in the initial training data. Predictions were reasonably accurate up to the 10-epoch mark, but beyond 20 epochs, the model exhibited a tendency to diverge significantly from the actual learning curve. This highlights the limitation of polynomial regression in accurately approximating the convergence phase of the learning curve.

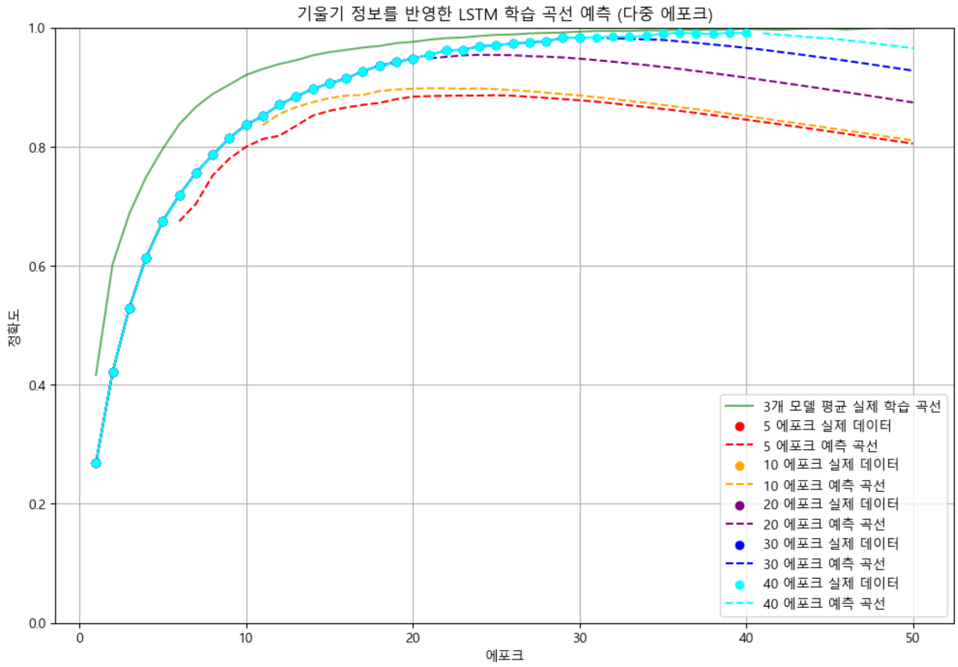

Figure 1: LSTM-based learning curve prediction results

4 4. Reference Image

4.1 Model Comparison and Conclusion

These results suggest that data-driven recurrent neural network models, such as LSTM, offer more flexible and robust prediction performance for learning curve forecasting compared to parametric regression models. In particular, the inclusion of gradient information as a time-series feature played a crucial role in enabling the model to capture the dynamic changes in the learning curve.

This study has demonstrated the effectiveness of the LSTM-based learning curve prediction methodology, and further validation across diverse datasets and models is needed to confirm its generalizability.

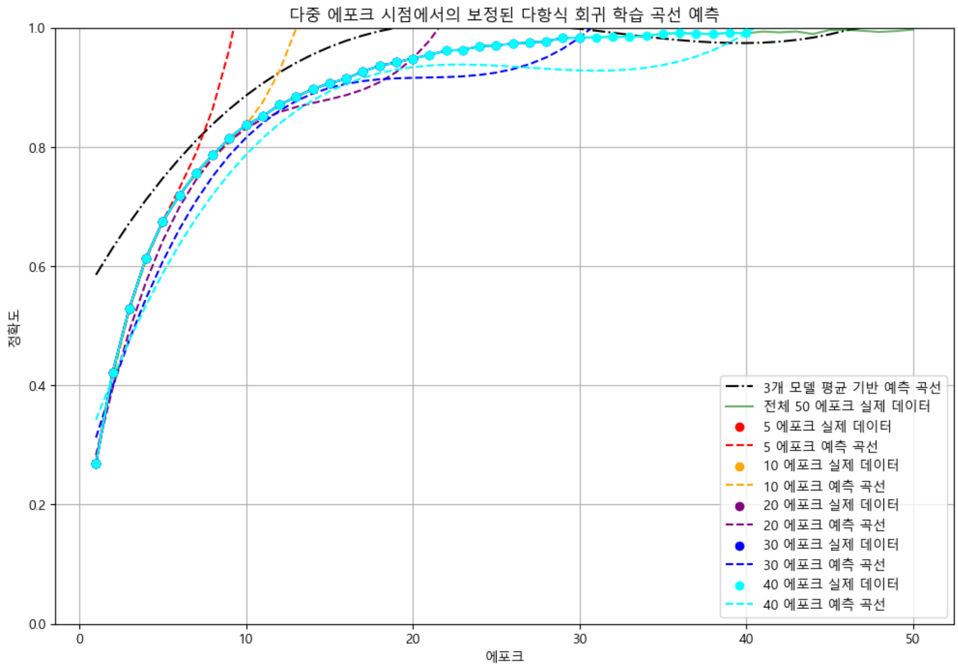

Figure 2: Polynomial regression-based learning curve prediction results

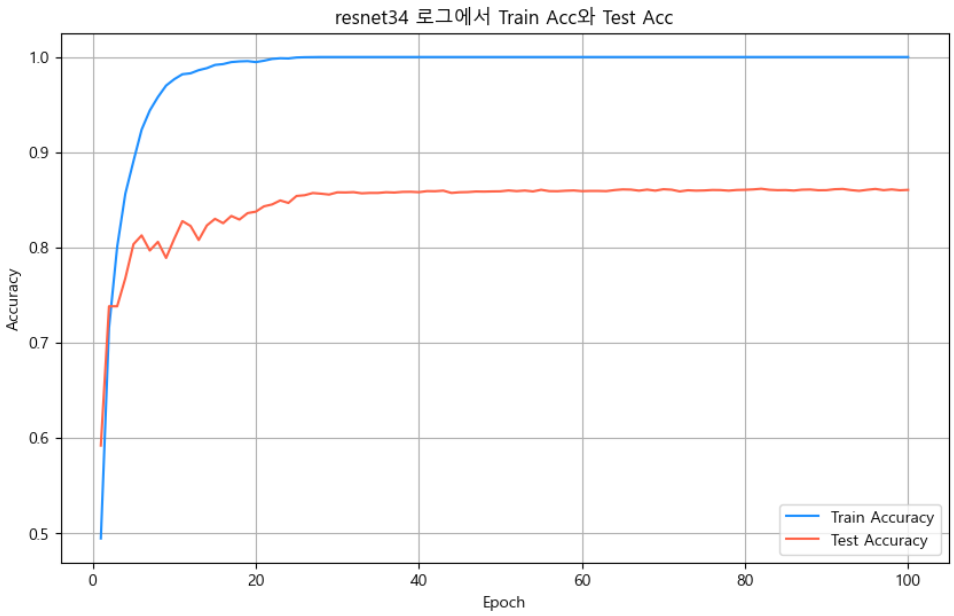

Figure 3: Comparison of Train Accuracy and Test Accuracy for ResNet-34

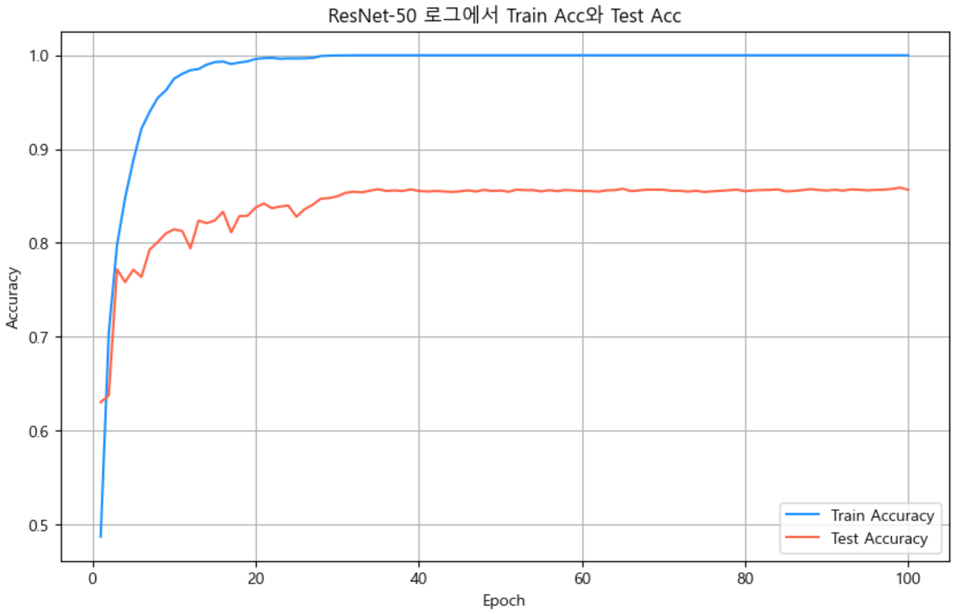

Figure 4: Comparison of Train Accuracy and Test Accuracy for ResNet-50

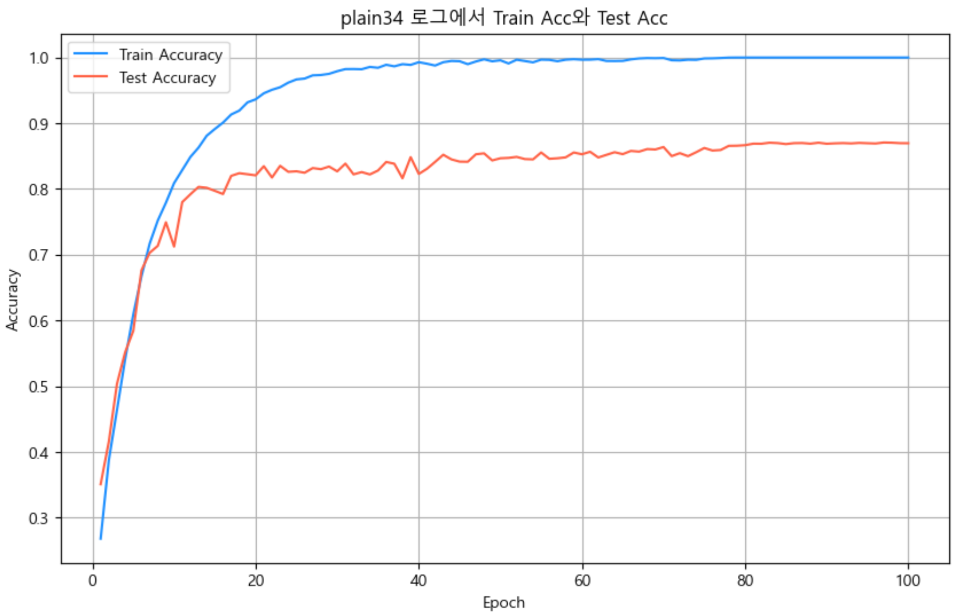

Figure 5: Comparison of Train Accuracy and Test Accuracy for Plain-34

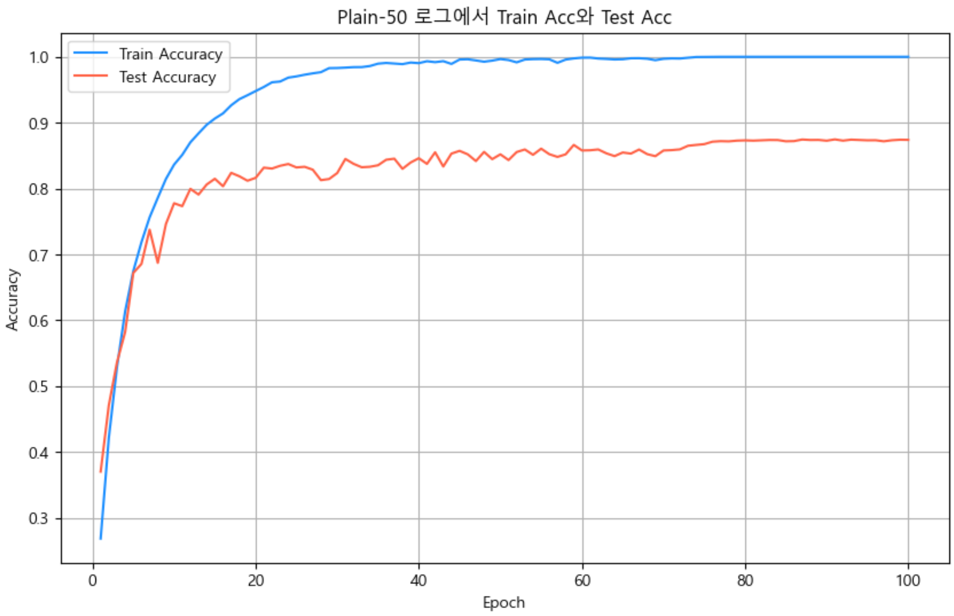

Figure 6: Comparison of Train Accuracy and Test Accuracy for Plain-50