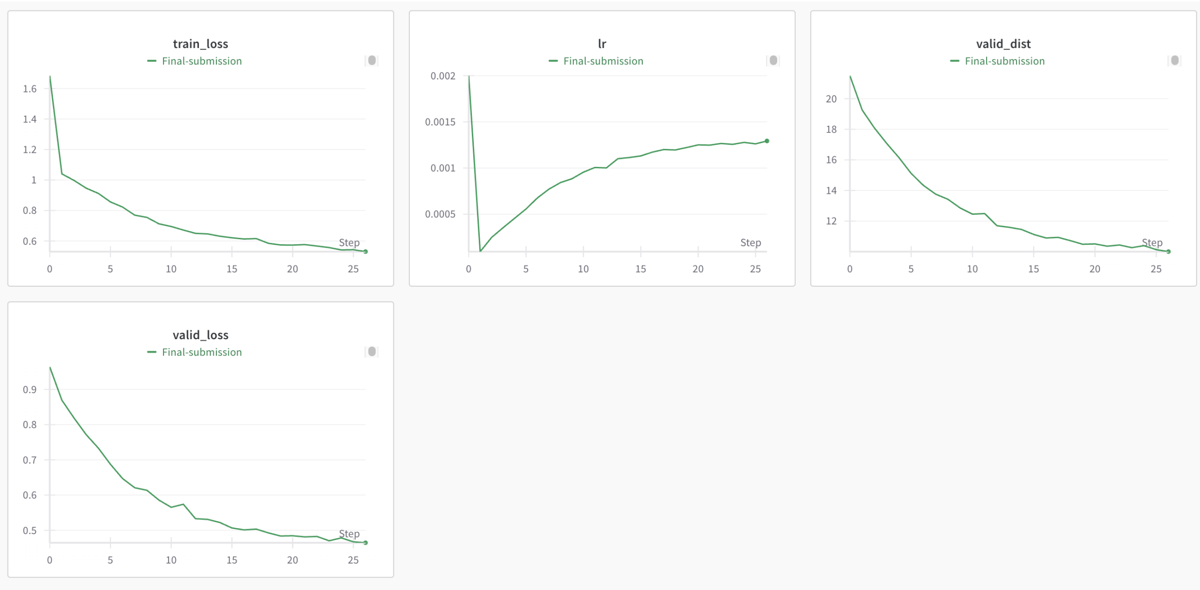
Description

This notebook implements an Automatic Speech Recognition (ASR) system using a neural network-based model. The ASR pipeline includes preprocessing, training, and evaluation steps with various configurations for loss functions, optimizers, and learning rate schedulers.

Main Features:

1. **Model Architecture**:
   - Utilizes an encoder-decoder structure.
   - Incorporates feature augmentation techniques such as Time Masking and Frequency Masking.
   - Output processed via a Connectionist Temporal Classification (CTC) loss function.
2. **Training Components**:
   - Mixed-precision training for efficiency.
   - Validation metrics include loss and Levenshtein distance calculation.
   - Optimized using the AdamW optimizer with a cosine annealing learning rate scheduler.
3. **Evaluation**:
   - Includes a beam search decoder for phoneme prediction.
   - Tracks performance using decoding and Levenshtein distance metrics.
4. **Prediction and Submission**:
   - Predicts phoneme sequences for test data using a beam decoder.
   - Prepares results for submission to a competition.

Wandb Log



Public: 7.76431

Private: 7.48600