Description

This notebook implements a Transformer-based Automatic Speech Recognition (ASR) system. The implementation includes data preparation, training, evaluation, and result generation steps with advanced configurations for neural networks, optimizers, and metrics.

Main Features

1. **Dataset Preparation:**
    - Downloads the ASR dataset from Kaggle and extracts it into the ./data directory.
    - Supports train, validation, and test partitions with configurable subsets.
2. **Model Design:**
    - Transformer-based architecture with configurable encoder and decoder layers.
    - Implements SpecAugment techniques such as frequency and time masking.
    - Embedding and down-sampling configurations for efficient processing.
3. **Training Pipeline:**
    - Optimizes the model using AdamW with a cosine annealing learning rate scheduler.
    - Supports mixed-precision training for efficiency.
    - Tracks loss and evaluation metrics (e.g., CER, WER).
4. **Evaluation:**
    - Includes metrics for Levenshtein distance, Word Error Rate (WER), and Character Error **Rate (CER).**
    - Provides attention visualization to interpret model predictions.
    - Supports beam search decoding for high-accuracy phoneme prediction.
5. **Utilities:**
    - Comprehensive tokenizer implementations (GTokenizer and CharTokenizer) to support **various tokenization strategies.**
    - Configuration files (config.yaml) for easy tuning of hyperparameters.
    - Utilities for model saving, loading, and checkpoint management.
6. **Logging and Analysis:**

- Utilizes wandb for experiment tracking and visualization.
- Provides utilities to visualize attention weights and debug outputs.

Wandb log