

Navigation

Select Function

Welcome

Start Evaluation

Dataset Analysis

About

Configuration

Upload Test Dataset

Drag and drop file here

Limit 200MB per file • JSON

Browse files

eval_data.json

17.0KB

Dataset loaded successfully with 76 samples.

Sampling Options

Temperature Settings

Model Path Configuration

Base Model Path

D:\kings\prj\Finetune_local\models

Fine-tuned Model Path

D:\kings\prj\Finetune_local\Models

Model Evaluation

Main Evaluation Results

Exact Match Rate

18.42%

vs. Base Model: +100.0%

Partial Match Rate

67.11%

vs. Base Model: +104.0%

Micro F1 Score

45.42%

vs. Base Model: +97.1%

Label-specific Performance

F1 Score by Label



Label-specific Metrics Table

Label	Base F1	Fine-tuned F1	Improvement	Improvement %	Support
6 volume	0.00%	46.15%	0.4615	+46153.85%	16
7 mid frequency	0.00%	44.44%	0.4444	+44444.44%	13
4 sound field	30.43%	53.33%	0.229	+75.24%	16
5 low frequency	12.50%	35.29%	0.2279	+182.35%	13
3 effects	0.00%	22.22%	0.2222	+22222.22%	10
2 compression	29.63%	41.67%	0.1204	+40.63%	18
1 reverb	41.38%	51.16%	0.0978	+23.64%	12
0 high frequency	50.00%	51.85%	0.0185	+3.70%	10

Prediction Examples

Success Cases Improvement Cases Failure Cases

Example 1: 整体音量太小...

Input Text: 整体音量太小

True Labels: volume

Fine-tuned Model Prediction: volume

Example 2: 回音太重...

Example 3: 回音太短...

Example 4: 立体声场太窄...

Example 5: 混响衰减调至秒...

Advanced Analysis

Confusion Matrix Analysis

Per-Label Confusion Statistics