Evaluation Metrics for Emotion Classification Project

Overview of Metrics:

This document outlines the evaluation metrics used in the CNN-RNN multimodal emotion classification project to assess model performance across audio and text modalities.

Accuracy:

The proportion of correctly classified instances among the total number of instances.

accuracy = correct / total

Usage in project:

- Primary metric for model selection during training
- Reported for both validation and test sets
- Used to compare performance across emotions and modalities

Strengths:

- Intuitive interpretation
- Provides a single summary measure of performance

Limitations:

- May be misleading for imbalanced datasets
- Doesn't provide insight into specific class performance

Loss Function

Definition: Cross-Entropy Loss measures the performance of classification models whose output is a probability value between 0 and 1.

Implementation:

criterion = nn.CrossEntropyLoss()

Usage in project:

- Optimization target during training

- Validation loss used for learning rate scheduling
- Model checkpointing based on lowest validation loss

Confusion Matrix

Definition: A table showing the counts of true positives, false positives, true negatives, and false negatives for each class.

Implementation:

cm = confusion_matrix(y_true, y_pred)

disp = ConfusionMatrixDisplay(cm, display_labels=le.classes_)

Benefits for emotion classification:

- Visualizes which emotions are commonly confused
- Identifies specific class imbalances in prediction
- Provides granular error analysis

Precision, Recall, and F1-Score

Precision: The proportion of positive identifications that were actually correct.

Precision = True Positives / (True Positives + False Positives)

Recall: The proportion of actual positives that were correctly identified.

Recall = True Positives / (True Positives + False Negatives)

F1-Score: The harmonic mean of precision and recall.

F1 = 2 * (Precision * Recall) / (Precision + Recall)

Implementation:

classification_report(y_true, y_pred, target_names=le.classes_)

Project findings:

- Performance varies across different emotions
- Some emotion pairs are more difficult to distinguish
- Balanced metrics provide comprehensive performance assessment

Class-Specific Metrics

Per-class evaluation:

- Individual precision, recall, and F1-scores for each emotion
- Particularly important for emotions with fewer samples

- Helps identify which emotions are harder to classify correctly

The combination of these metrics provides a comprehensive evaluation of the emotion classification system. While accuracy offers a high-level view of performance, the confusion matrix and class-specific metrics offer deeper insights into model behavior across different emotions. This multi-faceted evaluation is essential for refining models in emotion recognition systems where class imbalance and subtle distinctions between emotional states present unique challenges.