Analysis Report on ATIS Intent Classification Model Results

Assignment # 2

Introduction to Deep Learning

Name: Zakir Matloob

1. Dataset Overview:

- **Training Data:** The training dataset consists of intents such as atis_flight, atis_airfare, and atis_flight_time. The data includes natural language queries related to flight schedules, airfare, and flight details.
- **Testing Data:** The testing dataset contains similar intents and queries, ensuring consistency with the training dataset.
- **Data Shapes:**
    - Training Data: (4833, 46)
    - Testing Data: (799, 30)
- **Unique Intents:** 8 distinct intents are identified.

2. Model Configuration:

- **Vocabulary Size:** 872
- **Max Sequence Length:** 46
- **Number of Classes:** 8
- **Model Architecture:**
    - Embedding Layer: Converts words into 128-dimensional vectors.
    - Two Bidirectional LSTM Layers: Captures contextual information.
    - Dropout Layers: Prevents overfitting.

    - Dense Layers: Final classification into 8 intents.
- **Total Parameters:** 253,992 (All trainable)

3. Training Performance:

- **Epoch Results:** The model was trained for 10 epochs.
- **Training Accuracy:** Ranged from 96.25% (Epoch 1) to 98.77% (Epoch 10).
- **Validation Accuracy:** Peaked at 98.24%.
- **Loss Trends:** Training loss decreased from 0.1218 to 0.0457, while validation loss remained relatively stable around 0.16–0.20.
- **Observation:** Minimal overfitting, consistent accuracy improvement across epochs.

4. Testing Performance:

- **Test Accuracy:** 97%
- **Precision, Recall, and F1-Score per Intent:**
    - atis_flight: 0.99 precision, 0.98 recall
    - atis_airfare: 0.90 precision, 0.90 recall
    - atis_abbreviation: 0.94 precision, 1.00 recall
- **Weak Class Performance:** atis_quantity (F1-Score: 0.55) and atis_flight_time (F1-Score: 0.67) indicate lower performance, likely due to class imbalance.
- **Macro Avg F1-Score:** 0.86
- **Weighted Avg F1-Score:** 0.98

5. Prediction Example:

- **Input Query:** A test query was processed, and the predicted intent was correctly classified as atis_flight.

6. Key Observations:

- The model demonstrates excellent overall accuracy and consistency.
- Certain underrepresented classes (atis_quantity, atis_flight_time) require more data or handling techniques like class weighting.

- Training and validation loss curves indicate stable learning without significant overfitting.

7. Recommendations:

- Address class imbalance with techniques like oversampling, class weighting, or data augmentation.
- Fine-tune hyperparameters (e.g., learning rate, batch size) to optimize performance further.
- Explore alternative architectures (e.g., Transformer-based models) for potential improvements.

8. Conclusion: The ATIS intent classification model performs effectively with a 97% test accuracy and strong F1-scores across major intent classes. With minor adjustments for underrepresented classes, this model can achieve even better generalization and robustness.