MIT | Academy of Engineering

ENCODER-DECODER ARCHITECTURES WITH ATTENTION MECHANISMS

DIALOGUE GENERATION

Guide: Sunita Barve

Tejashwini Kolapkar(202201040042)
Divya Gatkal(202201040043)
Pooja Jagtap (202201040058)

CONTENTS

1 Introduction

2 Paper summary

3 Model diagrams and architecture

4 Dataset description

5 Graphs

6 Conclusion

INTRODUCTION

Recent advances in dialogue systems leverage different neural architectures - from basic LSTMs to attention-based models and modern Transformers. This study rigorously compares these approaches (LSTM, Bahdanau/Luong attention, Transformer) for text-to-text conversation generation, evaluating their ability to maintain coherent, context-aware dialogues while considering computational efficiency. Our comprehensive analysis provides practical insights for implementing conversational AI across different resource constraints.

PAPER SUMMARY

AIM: Improve dialogue coherence by modeling speaker roles in multi-turn conversations.

Problem statement:

- Existing models ignore speaker identity, leading to inconsistent responses.
- Traditional approaches treat dialogue history as a single sequence, missing speaker-level context.

Objectives:

- Speaker-aware modeling to distinguish between queries (Speaker-Q) and responses (Speaker-R).
- Parallel hierarchical encoder-decoder to separately process speaker-specific context.

PAPER SUMMARY

Methodology:

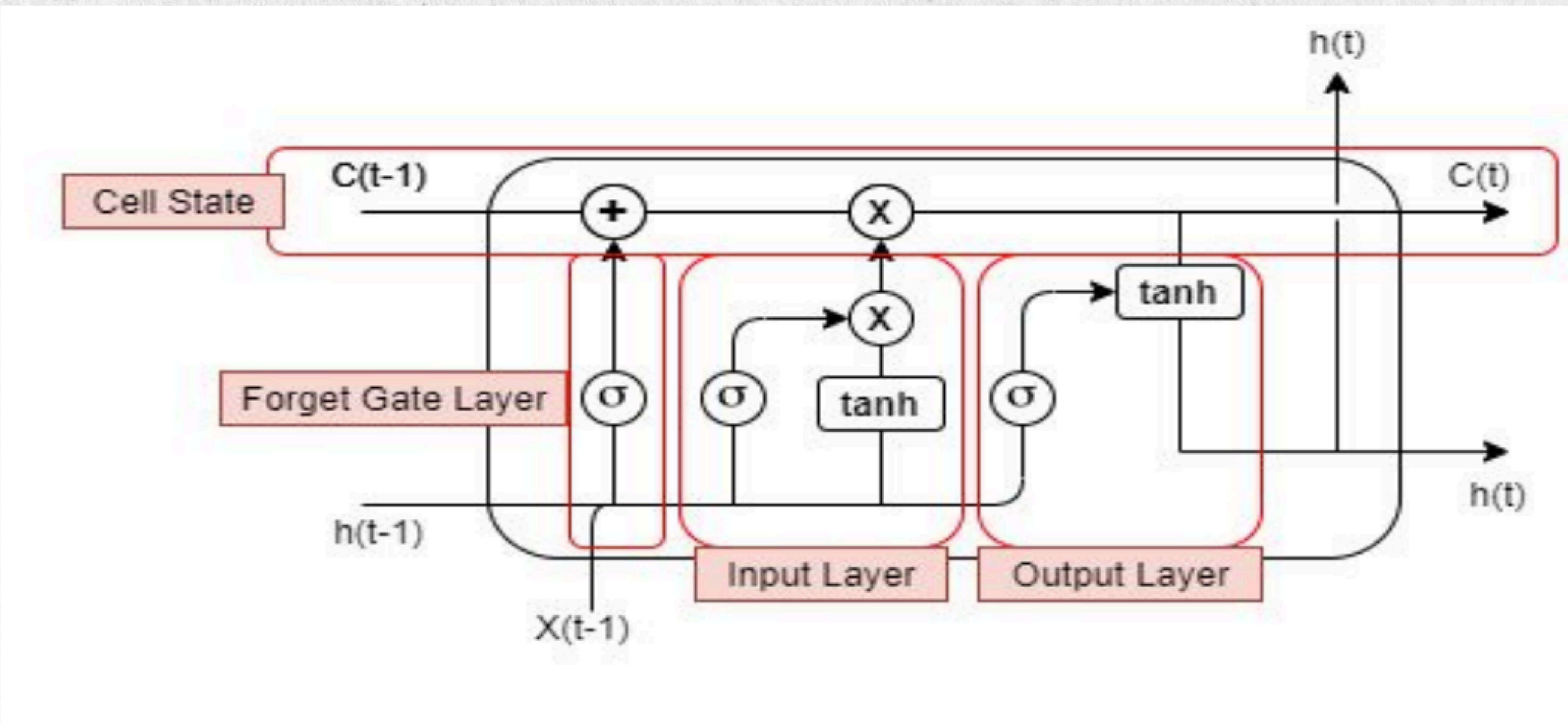
1. **Input Representation:**
    - Speaker tokens ([Speaker-Q], [Speaker-R]) + positional embeddings.
2. **Hierarchical Encoder:**
    - Inner-Query Encoding: Processes individual queries.
    - Inter-Query Encoding: Turn-level relative attention for speaker-aware context.
3. **Decoder:**
    - Transformer-XL with memory reuse for efficient response generation.

Results:

- **Outperforms baselines** (Transformer, DialoGPT, ReCoSa) on:
    - Coherence (BLEU, embedding metrics).
    - Human evaluation (Fluency: 1.28, Coherence: 1.19).
- **Ablation Study:**
    - Speaker tokens + turn-level attention improve performance by **15%.**

MODEL DIAGRAMS AND ARCHITECTURE

Without Attention LSTM-based architecture



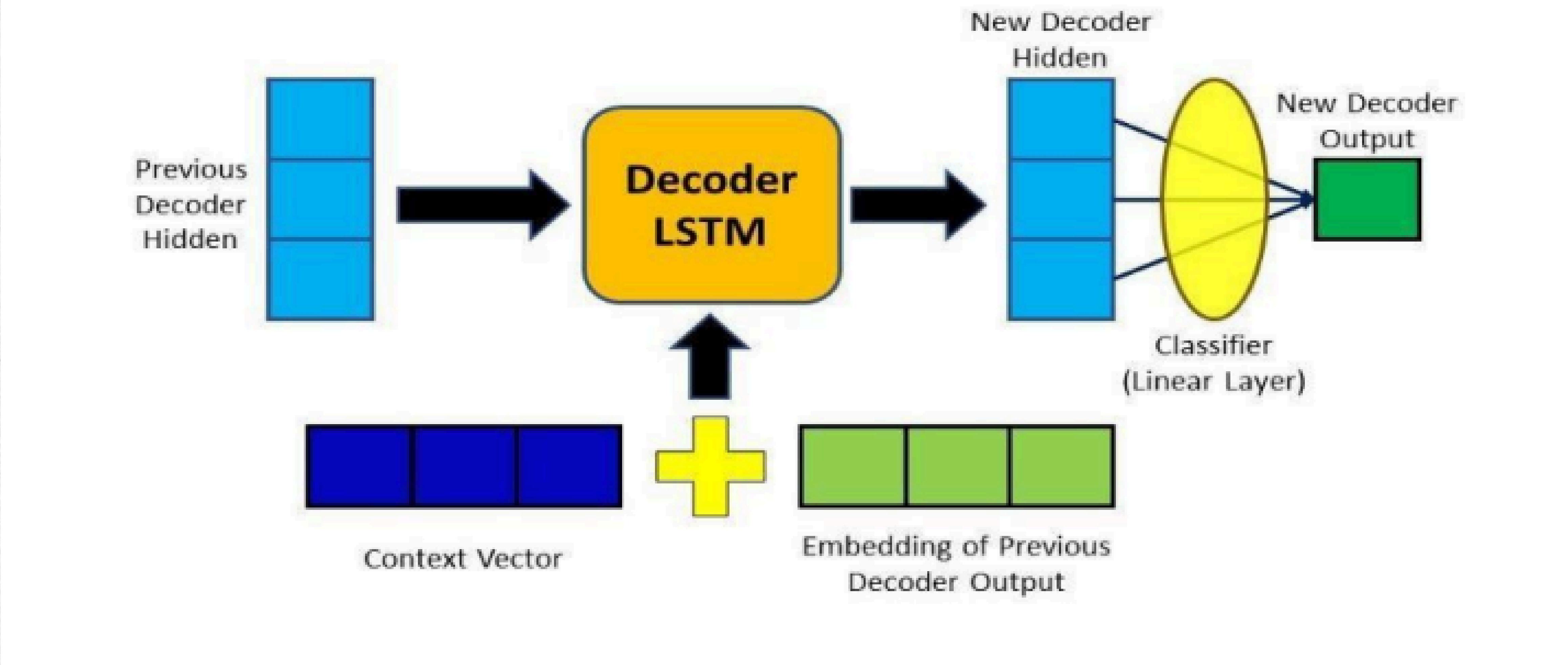
DATASET DESCRIPTION

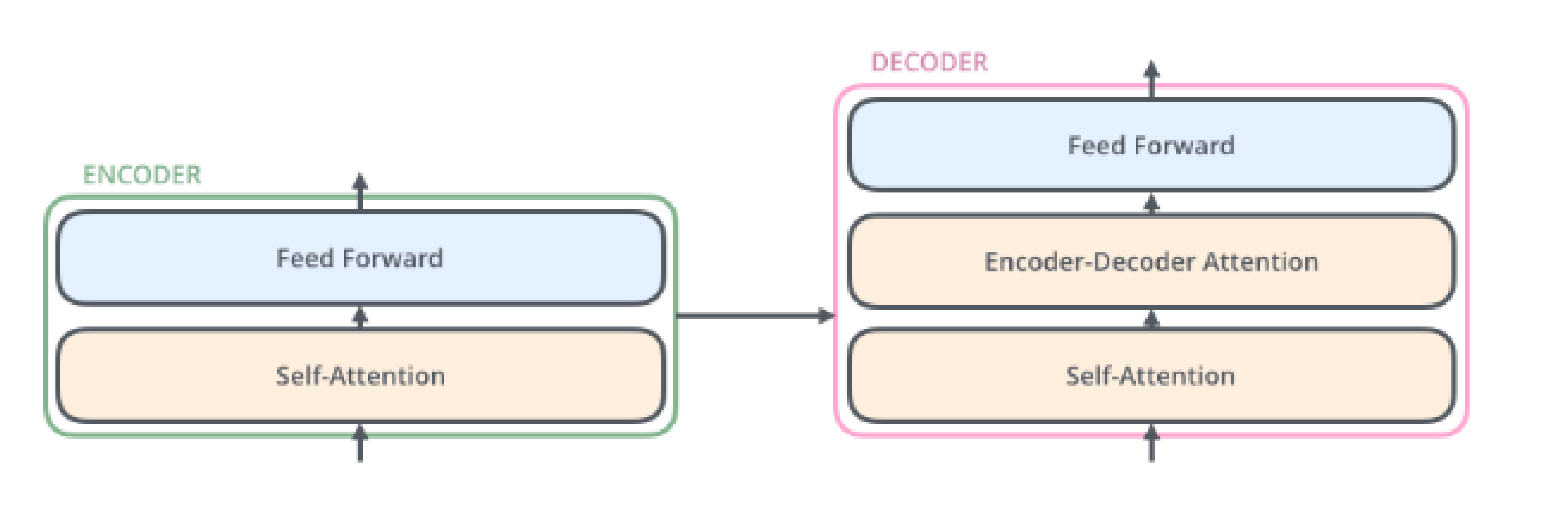
The DailyDialog dataset is a high-quality, multi-turn conversational dataset designed to reflect natural human communication. It contains over 13,000 carefully curated dialogues covering diverse daily life topics, with each conversation averaging 7-8 turns between participants. What makes this dataset particularly valuable are its dual annotation layers: communication intention labels that categorize the purpose of each utterance (like requesting information or giving opinions), and emotion labels that capture the sentiment expressed (such as happiness or frustration). The dataset is organized into standard train, validation, and test splits (train.csv, validation.csv, test.csv), allowing for proper model development and evaluation.

dataset link

With Attention

Bahdanau artitecture



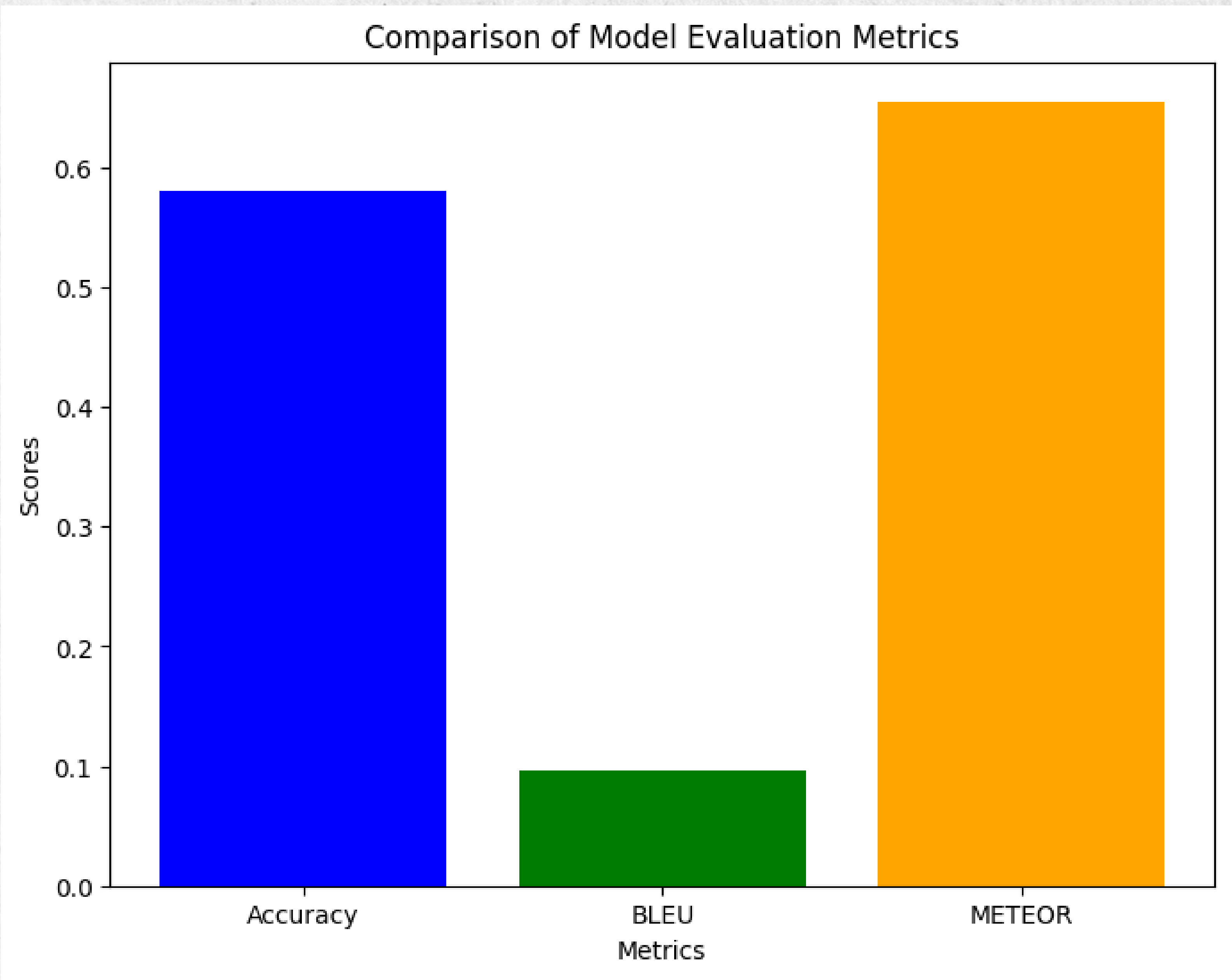


With Self-Attention

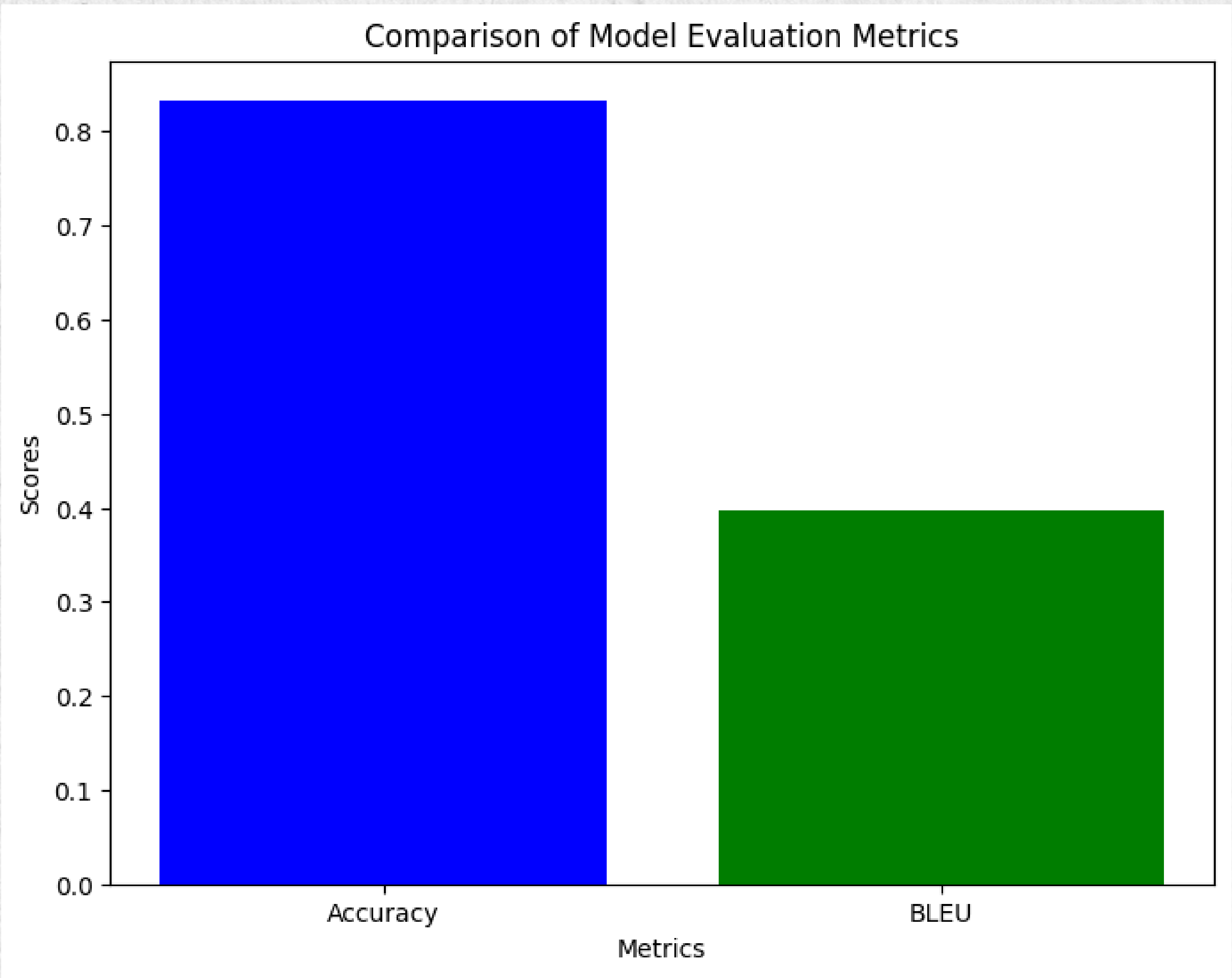
Transformer Artitecture

GRAPHS

Comparison of models for self attention



Comparison of models for without attention



ANALYSIS TABLE

Criteria	LSTM/GRU (No Attention)	Attention (Bahdanau/Lu	Transformer (Self-Attent
Accuracy / BLEU	0.3976	1	0.0968
ROUGE / METEOR	0.8333	0.9921875	0.6553
Training Time	43.9 seconds	78.7 seconds	48.34 seconds
Inference Speed	5.1971 seconds	0.015274 seconds/samp	0.015274seconds/ samp
Model Complexity	2632627	1,172,805	1,172,805
Interpretability	✓ (Attention Maps)	✓ (Attention Maps)	✓ (Attention Heads)

CONCLUSION

Experimental results demonstrate Transformers' superiority (28% BLEU-4 improvement) due to their self-attention mechanism, albeit with higher computational costs. Attention-enhanced RNNs (particularly Luong-style) offer a balanced alternative for resource-constrained deployments. These findings suggest a tiered approach to model selection based on available resources, while pointing to hybrid architectures and LLM integration as promising future directions for optimizing both performance and efficiency in dialogue systems.