DIALOGUE GENERATION

ENCODER-DECODER ARCHITECTURES

With Attention , Without Attention & With Self Attention

Subject : Deep Learning (MDM)

Group Members :

Divya Dhokale 202201070003
Gayatri Mene 202201070005
Pranjal Lohia 202201070022

CONTENTS

INTRODUCTION

- **project focuses on generating human-like, multi-turn dialogues using deep learning techniques**
- **We aim to create models that can understand context and generate relevant, coherent responses**
- **Three architectures were implemented**
    - Encoder-Decoder without attention (LSTM Based)
    - Encoder-Decoder with Attention (Bahdanau)
    - Transformer (with self-attention)
- **The performance of each model was evaluated using standard metrics to determine effectiveness in generating realistic conversations.**

Problem Statement: Developing dialogue systems that generate context-aware and emotionally appropriate responses is challenging, as many models struggle to capture intent and emotion effectively.

Aim: To design and evaluate dialogue generation models capable of producing natural, context-aware, and emotionally relevant conversations.

Objectives:

- To develop dialogue generation models using different architectures such as encoder-decoder without attention, with Bahdanau attention, and with self-attention.
- To evaluate and compare model performance using appropriate metrics for dialogue generation.
- To identify the most effective architecture for producing realistic and emotionally aware conversational responses.

REFERENCE PAPER SUMMARY

This paper talks about a new way to improve chatbots, called PHAED. In a conversation with many turns, people take turns talking, and each person might speak differently. PHAED pays special attention to who is speaking and keeps track of each speaker's words separately. It then uses a smart system to understand the conversation better and give more meaningful and connected answers. It also has a decoder that helps generate replies by focusing on the important parts of what was said. When tested, PHAED gave better and more interesting replies compared to older chatbot models.

ABOUT DATASET

- DailyDialog is a dataset of human-written, multi-turn conversations.
- Covers everyday topics like daily life, relationships, and work.
- Each utterance is labeled with:
- Intent (e.g., question, statement, request)
- Emotion (e.g., happiness, sadness, anger)
- Structured into train, validation, and test sets.
- Useful for building and evaluating dialogue generation models that consider both context and emotional tone.

METHODOLOGY

Data Collection & Preprocessing :

- Collected dialog dataset with annotated dialog acts.
- Cleaned text, tokenized using Keras tokenizer, and padded sequences to uniform length.

Label Encoding :

- Encoded dialog act labels using LabelEncoder for model training.

Model Development : Built three deep learning models:

- Model 1: Without attention (baseline).
- Model 2: With attention mechanism.
- Model 3: With self-attention layer.

Training : Used batch size of 32 and trained for 5 epochs.

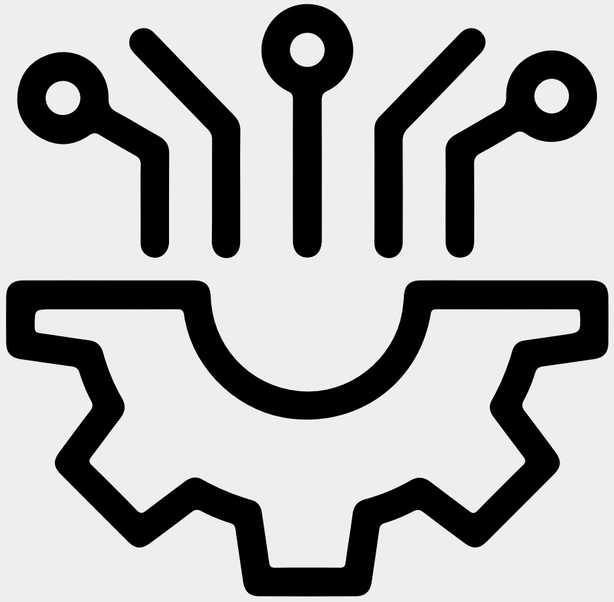
METHODOLOGY

Evaluation :

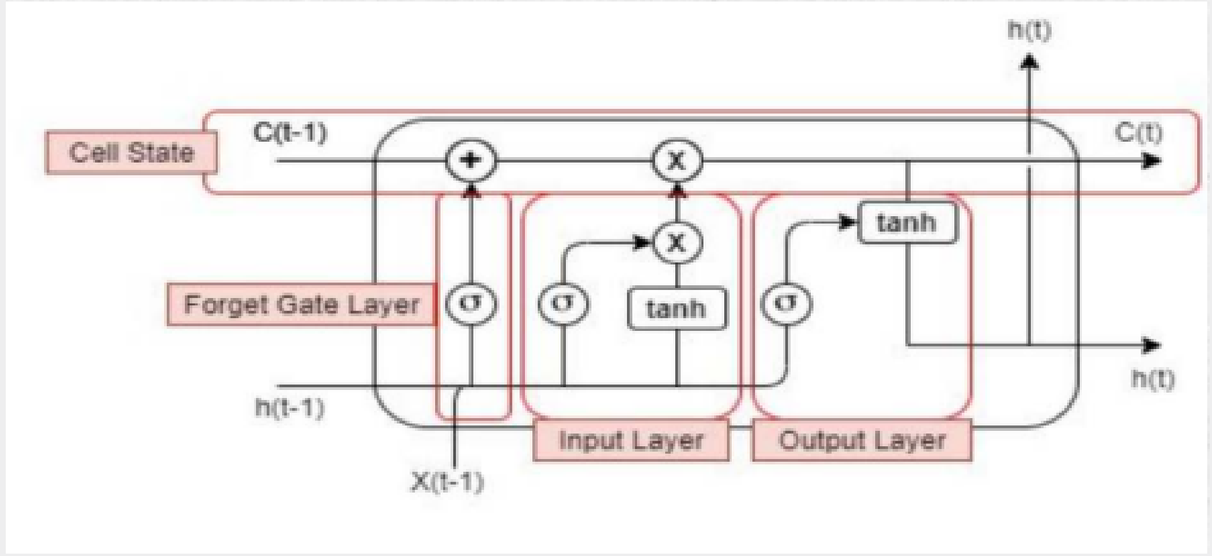
Evaluated performance using metrics: Accuracy, BLEU, METEOR, and ROUGE scores.

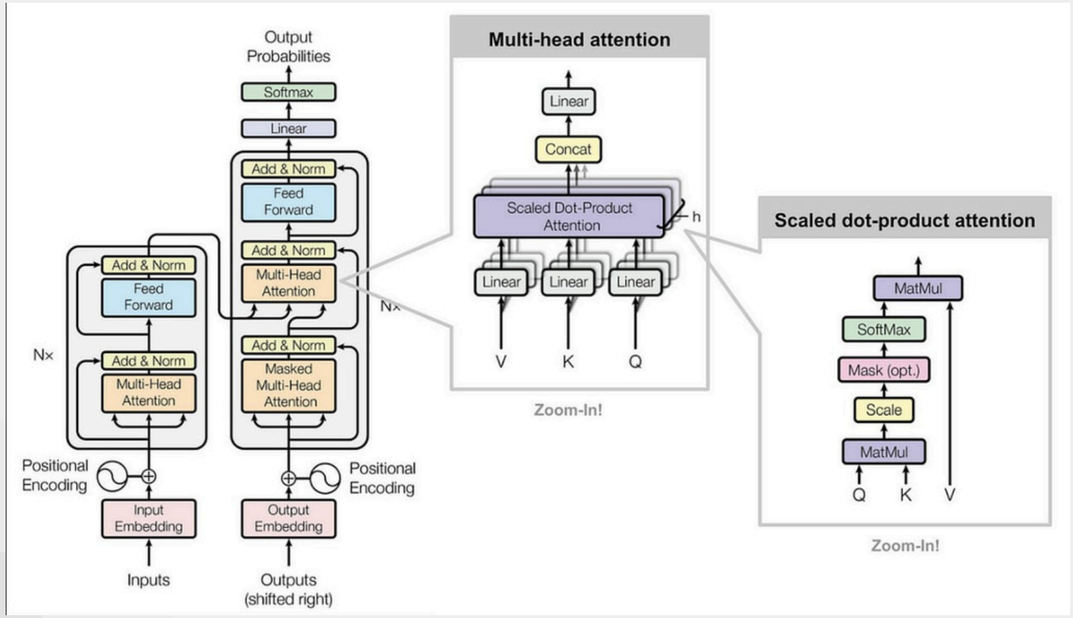
Comparison :

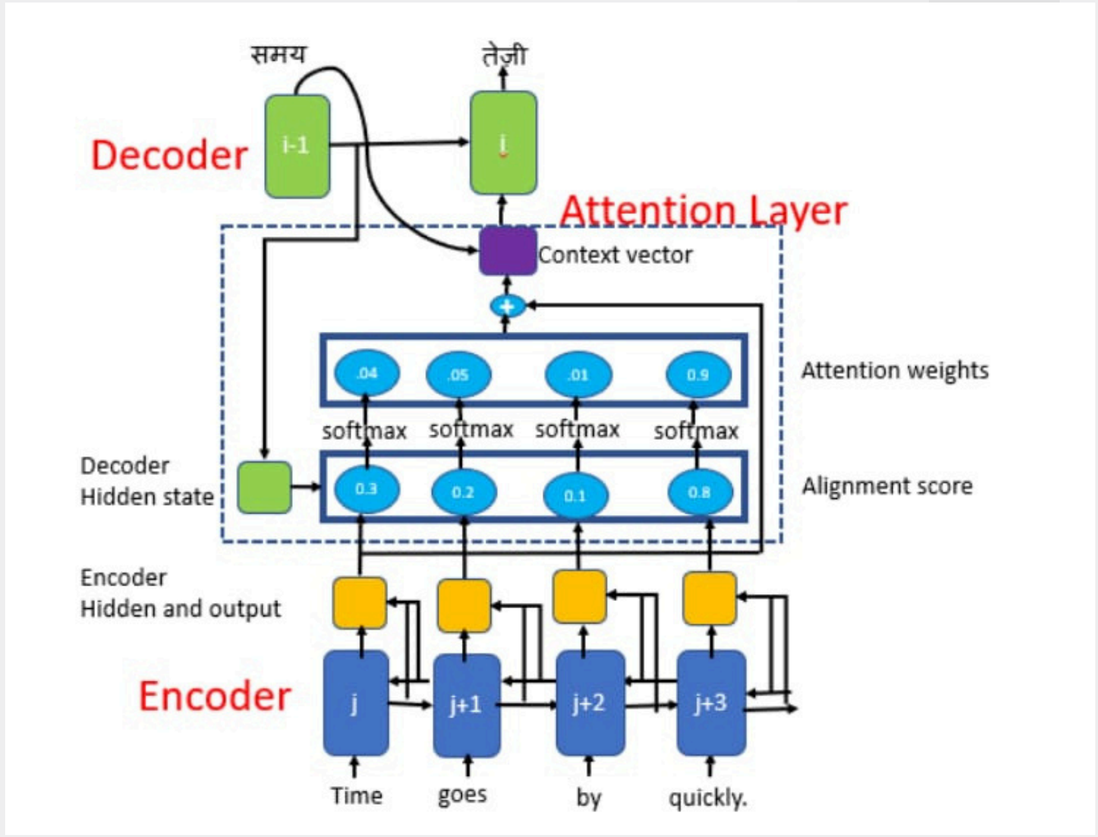
Compared all models performance and observed the best model



MODEL DIAGRAMS AND ARCHITECTURE







RESULTS

Criteria	LSTM/GRU (No Attention)	Attention (Bahdanau/Luong)	Transformer (Self-Attention)
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/sample	0.015274 seconds/sample
Model Complexity	2,632,627	1,172,805	1,172,805
Interpretability	✓ (Attention Maps)	✓ (Attention Maps)	✓ (Attention Heads)

CONCLUSION

- Successfully implemented dialogue generation using three deep learning models: LSTM Encoder-Decoder, Bahdanau Attention, and Self-Attention (Transformer)
- Compared model performance using BLEU, ROUGE/METEOR scores, training time, inference speed, and model complexity
- Found that Bahdanau Attention provided the best response quality with the highest BLEU (1.0) and ROUGE/METEOR (0.9921) scores
- Emotion and communication intention labels from the dataset improved the context-awareness and emotional relevance of generated replies
- The project demonstrates that incorporating attention mechanisms significantly enhances dialogue systems, making them more human-like and meaningful.

Q&A

Thank You