A Comparative Study of RNN with Attention and Transformer Models for English–Indonesian Neural Machine Translation

Fayshal Karan Athilla
Informatic Engineering
Universitas Darussalam Gontor
Ponorogo, Indonesia
fayshalkaranathilla12@student.cs.unida.gontor.ac.id

***Abstract*—This study discusses the implementation and evaluation of Neural Machine Translation (NMT) based on Recurrent Neural Network (RNN) with Attention and Transformer for English-Indonesian text translation. The dataset used is a bilingual sentence pair preprocessed from the raw.txt file. Experiments were conducted with training up to 50 epochs and evaluation using the SacreBLEU and chrF metrics. The results show that RNN with Attention provides more stable performance on small datasets with a SacreBLEU score of 0.16 and a chrF of 35.61, while Transformer fails to produce meaningful translations with a SacreBLEU score of 0.00 and a chrF of 1.52.**

Keywords— Keywords—Neural Machine Translation, RNN, Attention, Transformer, Low-Resource Languages, English–Indonesian Translation

I. Introduction

Machine translation (MT) is a key branch of Natural Language Processing (NLP). With the advancement of deep learning, two dominant approaches in MT are RNNs with attention and Transformers. RNNs are effective for handling sequential data but struggle to capture long-term context. Transformers with self-attention overcome these limitations through parallel processing and richer contextual representation.

This study aims to compare the performance of the two architectures on a simple English–Indonesian dataset.

II. RELATED WORK

Machine translation (MT) is a key branch of Natural Language Processing (NLP). With the advancement of deep learning, two dominant approaches in MT are RNNs with attention and Transformers. RNNs are effective for handling sequential data but struggle to capture long-term context. Transformers with self-attention overcome these limitations through parallel processing and richer contextual representation.

This study aims to compare the performance of the two architectures on a simple English–Indonesian dataset.

III. METHOD

A. Dataset

- The bilingual dataset is derived from raw.txt.
- Preprocessing included tokenization using SentencePiece, text normalization, padding, and dividing the data into train and test data.

B. Model

RNN with Attention

- GRU-based encoder–decoder.
- Attention mechanism for token alignment.

Transformer

- 256-dimensional embedding.
- 3-layer encoder & decoder.
- 512-dimensional feedforward.
- Sinusoidal position encoding.

C. Training

- Optimizer: Adam.
- Loss: CrossEntropyLoss with ignored padding.
- Epoch: 10.
- Checkpoint: Save the best model based on validation.

D. Evaluation

- Metrics: SacreBLEU and chrF.
- Evaluation was performed on 100 test sentences.

IV. EXPERIMENTS

A. Environment

- Python 3.11, PyTorch 2025.
- Hardware: Laptop Thinkpad T480 i7-8650U CPU @ 1.90GHz (8CPUs), ~2.1GHz. Memory 32 RAM.

B. Results

RNN with Attention

- SacreBLEU: 0.16
- chrF: 35.61

Table Result	
SRC	"i'll wait here."
REF	“aku akan menunggu di sini."
HYP	"a k u b erg a k an di a un gg u di s in i ."

Transformer

- SacreBLEU: 0.00
- chrF: 1.52

978-1-6654-1234-5/22/$31.00 ©2022 IEEE

Table Result	
SRC	"the night was cold."
REF	"ini milik ayahku."
HYP	"soisisis..."

V. RESULT AND DISCUSSION

Experimental results show that RNN with Attention still performs better on small datasets. Transformer fails to produce meaningful translations and only produces repetitive output. This aligns with literature that emphasizes Transformer's need for large datasets and extensive training.

Although the RNN's BLEU score is very low, the chrF metric indicates partial string matching, indicating the RNN is capable of capturing basic patterns in the target language. Technical errors such as architectural model mismatches during load checkpoints also contribute to Transformer's performance degradation.

ACKNOWLEDGMENT

This study compared RNN with Attention and Transformer for English-Indonesian translation. RNN performed better on small datasets with more stable results, while Transformer failed to learn meaningful representations. Future research can improve quality by:

1. Using larger datasets.
2. Consistently adjusting the Transformer architecture between training and evaluation.
3. Implementing better subword unit techniques to reduce out-of-vocabulary.

REFERENCES

[1] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” in *ICLR,* 2015.

[2] A. Vaswani et al., “Attention Is All You Need,” in *NeurIPS,* 2017.

[3] M. Post, “A Call for Clarity in Reporting BLEU Scores,” *WMT*, 2018.