Analisis Perbandingan Arsitektur RNN-Attention dan Transformser untuk Penerjemahan Mesin Saraf Inggris-Indonesia

Rizky Cahyono Putra  
Teknik Informatika *Universitas Darussalam Gontor*Ponorogo, Indonesia  
rizkycahyonoputra80@student.cs.unida.gontor.ac.id

*Abstract*—*Penerjemahan Mesin Saraf (Neural Machine Translation/NMT) telah menjadi pendekatan dominan untuk tugas penerjemahan otomatis, dengan arsitektur seperti RNN dan Transformer menunjukkan hasil yang canggih. Penelitian ini bertujuan untuk mengimplementasikan, membandingkan, dan menganalisis dua arsitektur NMT utama untuk tugas penerjemahan dari bahasa Inggris ke bahasa Indonesia. Dengan mengimplementasikan model baseline menggunakan Recurrent Neural Network (RNN) dengan mekanisme atensi, dan model yang lebih modern berbasis arsitektur Transformer. Kedua model dilatih pada dataset bilingual dari Anki dan dievaluasi menggunakan metrik SacreBLEU. Hasil eksperimen menunjukkan bahwa model Transformer mencapai validation loss yang lebih rendah, mengindikasikan potensi belajar yang lebih baik. Namun, dengan waktu pelatihan yang terbatas, kedua model menunjukkan skor BLEU yang sangat rendah, menandakan bahwa model belum sepenuhnya konvergen. Studi ablasi pada kedalaman model Transformer juga mengonfirmasi pentingnya arsitektur yang dalam untuk kapasitas belajar model. Penelitian ini menyoroti pentingnya waktu pelatihan yang cukup untuk mencapai kualitas terjemahan yang memadai.*

Keywords—Neural Machine Translation, Transformer, RNN, Attention Mechanism, BLEU, PyTorch

# PENDAHULUAN

Penerjemahan mesin otomatis adalah salah satu sub-bidang kecerdasan buatan yang paling fundamental, dengan tujuan untuk menerjemahkan teks dari satu bahasa (sumber) ke bahasa lain (target) secara otomatis. Dalam beberapa tahun terakhir, paradigma telah bergeser dari metode statistik (Statistical Machine Translation/SMT) ke Penerjemahan Mesin Saraf (Neural Machine Translation/NMT) [1]. NMT menggunakan model jaringan saraf tiruan untuk memprediksi probabilitas sebuah urutan kata, yang umumnya menghasilkan terjemahan yang lebih fasih dan akurat.

Dua arsitektur yang menjadi tonggak dalam pengembangan NMT adalah model berbasis Recurrent Neural Network (RNN) dengan mekanisme atensi [2] dan arsitektur Transformer [3]. Model RNN memproses teks secara sekuensial, sementara Transformer menggunakan mekanisme *self-attention* untuk memproses seluruh kalimat secara paralel, memungkinkannya menangkap dependensi kontekstual jarak jauh dengan lebih efektif.

Penelitian ini berfokus pada implementasi dan perbandingan kedua arsitektur tersebut untuk pasangan bahasa Inggris-Indonesia, sebuah pasangan bahasa yang memiliki tantangan tersendiri karena perbedaan struktur gramatikal. Tujuan dari penelitian ini adalah:

1. Mengimplementasikan model NMT baseline menggunakan arsitektur Encoder-Decoder RNN dengan atensi.
2. Mengimplementasikan model NMT menggunakan arsitektur Transformer.
3. Mengevaluasi dan membandingkan performa kedua model secara kuantitatif menggunakan metrik SacreBLEU dan secara kualitatif melalui analisis hasil terjemahan.
4. Melakukan studi ablasi untuk memahami dampak komponen arsitektur terhadap performa model.

# METODOLOGI

## Dataset dan Preprocessing

Penelitian ini menggunakan dataset bilingual Inggris-Indonesia yang bersumber dari [manythings.org](https://www.manythings.org/anki/), yang berisi 14.881 pasang kalimat. Proses preprocessing data meliputi beberapa tahap. Pertama, data dibersihkan dari informasi atribusi. Selanjutnya, dataset dibagi menjadi tiga bagian: 80% untuk data latih (11.904 kalimat), 10% untuk data validasi (1.488 kalimat), dan 10% untuk data uji (1.489 kalimat).

Untuk tokenisasi, digunakan metode Byte-Pair Encoding (BPE) [4] dengan ukuran kosakata (vocabulary) sebesar 16.000 untuk masing-masing bahasa. Tokenisasi BPE efektif dalam menangani kata-kata langka atau yang tidak ada dalam kamus (*out-of-vocabulary*) dengan memecahnya menjadi unit sub-kata.

## Arsitektur Baseline (RNN + Attention)

Model baseline yang dipakai adalah arsitektur Encoder-Decoder berbasis GRU (Gated Recurrent Unit) [5] dengan mekanisme atensi.

* **Encoder**: Terdiri dari GRU dua arah (*bidirectional*) yang membaca kalimat sumber dan menghasilkan serangkaian status tersembunyi (*hidden states*). Status tersembunyi terakhir dari kedua arah digabungkan untuk membentuk *context vector* awal bagi Decoder.
* **Decoder**: Terdiri dari GRU satu arah yang menghasilkan kalimat target kata per kata. Pada setiap langkah waktu, mekanisme atensi [2] menghitung skor relevansi antara status tersembunyi Decoder saat ini dengan semua status tersembunyi dari Encoder, memungkinkannya untuk fokus pada bagian kalimat sumber yang paling relevan.

## Arsitektur Transformer

Sebagai perbandingan, diimplementasikan arsitektur Transformer [3] yang sepenuhnya berbasis atensi.

* **Encoder**: Terdiri dari tumpukan 3 *encoder layer* identik. Setiap layer memiliki mekanisme *multi-head self-attention* yang memungkinkan setiap kata dalam kalimat sumber untuk "melihat" kata lain, diikuti oleh *feed-forward network*.
* **Decoder**: Terdiri dari tumpukan 3 *decoder layer*. Setiap layer memiliki dua sub-layer atensi: satu *masked multi-head self-attention* untuk token target, dan satu *multi-head attention* yang menghubungkan output Encoder dengan input Decoder.
* **Positional Encoding**: Karena model tidak memiliki sifat rekuren, informasi posisi token ditambahkan ke *embedding* input untuk mempertahankan urutan kalimat.

# EKSPERIMEN

## Pengaturan Pelatihan

Semua model diimplementasikan menggunakan framework PyTorch. Pelatihan dilakukan menggunakan optimizer Adam dengan fungsi kerugian (*loss function*) Cross-Entropy. Model baseline dilatih selama 10 epoch, sedangkan model Transformer dilatih selama 15 epoch. Untuk studi ablasi, model Transformer yang dimodifikasi dilatih selama 10 epoch.

## Metrik Evaluasi

Performa model dievaluasi secara kuantitatif menggunakan metrik SacreBLEU [6]. BLEU (Bilingual Evaluation Understudy) mengukur kesamaan antara terjemahan yang dihasilkan mesin dengan terjemahan referensi berkualitas tinggi dengan menghitung presisi n-gram. Skor yang lebih tinggi menunjukkan kualitas terjemahan yang lebih baik.

# HASIL DAN DISKUSI

Hasil kuantitatif dari semua eksperimen dirangkum dalam Tabel 1.

1. Hasil Perbandingan Kinerja Model

| **Model** | ***Konfigurasi*** | Val. Loss | Val. PPL | BLEU Score |
| --- | --- | --- | --- | --- |
| RNN + Attention | Baseline | 4.482 | 88.44 | 0.07 |
| Transformer | 3 Layers | 4.213 | 67.59 | 0.01 |
| Transformer | 1 Layer (Ablasi) | 4.458 | 86.28 | 0.01 |

Dari hasil di atas, beberapa poin penting dapat didiskusikan. Pertama, semua model menghasilkan skor BLEU yang sangat rendah (mendekati nol). Analisis kualitatif pada hasil terjemahan (seperti "<s> Kamu m el ihat b el ihat b erapa ? </s>") mengonfirmasi bahwa model menghasilkan kalimat yang tidak koheren dan seringkali repetitif. Hal ini secara kuat mengindikasikan bahwa jumlah epoch pelatihan (10-15 epoch) tidak cukup bagi model untuk belajar dan konvergen pada tugas yang kompleks ini.

Meskipun skor BLEU tidak informatif, metrik *validation loss* dan *perplexity* (PPL) memberikan wawasan yang lebih baik. Model Transformer dengan 3 layer berhasil mencapai *validation loss* dan PPL terendah, menunjukkan bahwa arsitektur ini memiliki kapasitas belajar yang lebih baik dan lebih cepat dalam memahami distribusi data dibandingkan model RNN.

Studi ablasi, di mana jumlah layer Transformer dikurangi menjadi satu, menghasilkan *validation loss* yang lebih tinggi dibandingkan model 3 layer. Ini membuktikan bahwa kedalaman arsitektur merupakan faktor penting yang berkontribusi pada kapasitas model untuk mempelajari representasi yang lebih kompleks, meskipun dampaknya belum terlihat pada skor BLEU karena pelatihan yang tidak memadai.

# HASIL DAN DISKUSI

Dalam penelitian ini, telah berhasil diimplementasikan dan dibandingkan model NMT berbasis RNN-Attention dan Transformer untuk penerjemahan Inggris-Indonesia. Hasil eksperimen menunjukkan bahwa arsitektur Transformer memiliki potensi belajar yang lebih unggul, yang terlihat dari nilai *validation loss* yang lebih rendah. Namun, kesimpulan utama dari penelitian ini adalah bahwa waktu pelatihan yang singkat menjadi kendala utama dalam mencapai kualitas terjemahan yang berarti. Kedua model gagal menghasilkan terjemahan yang dapat dipahami, yang tercermin dalam skor BLEU yang sangat rendah.

Untuk penelitian di masa depan, langkah yang paling krusial adalah memperpanjang durasi pelatihan secara signifikan. Selain itu, eksplorasi lebih lanjut pada penyesuaian hyperparameter, penggunaan dataset yang lebih besar, dan teknik seperti *learning rate scheduling* dapat dilakukan untuk meningkatkan performa model.

Referensi

1. [1] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, 27.
2. [2] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv preprint arXiv:1409.0473*.
3. [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30.
4. [4] Sennrich, R., Haddow, B., & Birch, A. (2015). Neural Machine Translation of Rare Words with Subword Units. *arXiv preprint arXiv:1508.07909*.
5. [5] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv preprint arXiv:1406.1078*.
6. [6] Post, M. (2018). A Call for Clarity in Reporting BLEU Scores. *Proceedings of the Third Conference on Machine Translation: Research Papers*.