

Analisis Model Encoder-Decoder untuk Generasi Judul Berita

Nama : Stevani Dwi Utomo
NIM : 24/546969/PPA/06865

1. Ringkasan Eksperimen

Pada eksperimen ini, kita melatih dan mengevaluasi tiga jenis arsitektur model berbeda untuk tugas *text summarization* (khususnya, membuat judul dari artikel berita), yaitu:

- **Basic LSTM Encoder-Decoder**
- **Attention-based LSTM Encoder-Decoder**
- **Transformer Model**

Setiap model dilatih menggunakan dataset pasangan *artikel-judul* dari berita teknologi yang telah di-scrape. Evaluasi dilakukan dengan melihat:

- **Kualitas prediksi** secara kualitatif.
- **Efisiensi komputasi** berdasarkan waktu training.
- **Metode evaluasi BLEU Score** untuk mengukur kualitas prediksi.

2. Hasil Training

Model	Waktu Training	Rata-Rata BLEU Score	Observasi Kualitatif
Basic LSTM	13.5 detik	0.0000	Prediksi tidak akurat, hanya mengulang "the the of"
Attention LSTM	71.2 detik	0.0000	Prediksi mengulang kata "of" dan "to" secara berlebihan
Transformer	8.5 detik	0.0010	Prediksi jauh lebih beragam, meskipun tetap tidak akurat

3. Evaluasi Kualitatif

Basic LSTM Encoder-Decoder

- Model hanya mampu menghasilkan frasa "the the of" berulang-ulang.
- Tidak menangkap konteks isi artikel sama sekali.
- BLEU score 0 menunjukkan tidak ada *overlap* signifikan dengan ground truth.
- **Kemungkinan penyebab:** Data terlalu sedikit, atau model terlalu sederhana untuk menangkap relasi kompleks dalam teks panjang.

Attention LSTM Encoder-Decoder

- Walaupun menggunakan *attention*, prediksi malah menjadi lebih buruk.
- Output berulang seperti "of of of of" menandakan *overfitting* pada token yang sering muncul.
- Attention belum mampu memperbaiki performa karena kemungkinan model terlalu dalam atau kurang data yang cukup.

Transformer

- Prediksi mengandung variasi kata seperti "this", "in", "a", "the", meskipun masih belum membentuk kalimat yang berarti.
- BLEU score meningkat sedikit menjadi 0.0010.
- **Transformer lebih cepat dan relatif lebih mampu** memahami pola dibandingkan LSTM, bahkan dengan data terbatas.

4. Analisis Efisiensi Komputasi

- **Transformer** melatih model tercepat (8.5 detik), dibandingkan LSTM (13.5 detik) dan LSTM+Attention (71 detik).
- Attention LSTM paling lambat karena kompleksitas tambahan dari mekanisme perhatian di tiap langkah decoding.
- Dari sisi waktu, **Transformer jauh lebih unggul** dan lebih scalable untuk dataset besar.

5. Kesimpulan

- Dengan dataset kecil ini, **semua model gagal** menghasilkan judul berita yang masuk akal.
- **Transformer** menunjukkan tanda-tanda pengertian konteks dibandingkan model berbasis LSTM.
- **BLEU score mendekati nol** untuk semua model mengindikasikan ketidakcocokan antara hasil prediksi dan ground truth.
- Untuk perbaikan:
 - Tambahkan jumlah data latih secara signifikan.

- Lakukan *hyperparameter tuning*.
- Implementasikan teknik *teacher forcing* yang lebih baik untuk melatih decoder.
- Gunakan pre-trained embeddings seperti GloVe atau BERT embeddings untuk input artikel.