

Báo Cáo Tóm Tắt Kết Quả Mô Hình Dịch Máy Anh-Việt

1. Đề bài

Xây dựng và đánh giá hai loại mô hình dịch máy (Transformer và RNN) cho bài toán dịch tiếng Anh sang tiếng Việt sử dụng tập dữ liệu Opus100.

2. Mô hình

Hai loại kiến trúc chính được triển khai:

2.1. Transformer

- Kiến trúc dựa trên mô hình Transformer gốc
- Sử dụng cơ chế Multi-head attention
- Bao gồm position encoding và feed-forward networks
- Số lớp encoder/decoder: 2-3

2.2. BasicRNNSeq2Seq

- Mô hình encoder-decoder sử dụng GRU
- Cơ chế attention giữa encoder và decoder

3. Siêu tham số

3.1. Transformer

Ba cấu hình được thử nghiệm:

Cấu hình	hidden_size	n_layers	learning_rate	batch_size	Số tham số
T1	128	2	0.001	128	3,121,802
T2	96	2	0.001	256	2,219,818
T3	192	3	0.001	64	5,914,186

3.2. RNN

Ba cấu hình được thử nghiệm:

Cấu hình	embed_size	hidden_size	n_layers	learning_rate	batch_size	Số tham số
R1	128	256	1	0.001	128	8,010,122
R2	96	192	2	0.001	64	6,608,362
R3	64	128	1	0.002	256	3,622,730

4. Kết quả

4.1. Transformer

Quá trình huấn luyện (10 epochs):

Cấu hình	Epoch 1	Epoch 5	Epoch 10	Val Loss Cuối	Test Loss
T1	7.733	4.688	2.278	2.459	1.956
T2	8.409	5.582	4.480	4.888	4.466
T3	6.646	3.414	1.014	1.320	1.056

4.2. RNN

Quá trình huấn luyện (3 epochs):

Cấu hình	Epoch 1	Epoch 2	Epoch 3	Val Loss Cuối	Test Loss
R1	6.326	5.731	5.575	6.148	5.748
R2	6.061	5.554	5.338	6.077	5.665
R3	6.619	5.800	5.692	6.241	5.886

5. Nhận xét

5.1. So sánh mô hình

- **Transformer** đạt hiệu suất vượt trội hơn **RNN** với test loss thấp hơn đáng kể (1.056 so với 5.665)
- Cấu hình Transformer T3 có hiệu quả tốt nhất trong tất cả các mô hình được thử nghiệm

5.2. Ảnh hưởng của siêu tham số

- **Transformer:**
 - Tăng hidden_size (từ 96 lên 192) và số lớp (từ 2 lên 3) cải thiện đáng kể hiệu suất
 - Batch size nhỏ hơn (64) cho kết quả tốt hơn so với batch size lớn (256)
 - Mô hình với nhiều tham số hơn (T3: 5.9M) hiệu quả hơn mô hình nhỏ (T2: 2.2M)
- **RNN:**
 - Mô hình 2 lớp (R2) hiệu quả hơn một chút so với mô hình 1 lớp
 - Cấu hình với hidden_size lớn hơn thường cho kết quả tốt hơn
 - RNN cần thêm epochs để đạt hiệu suất tốt hơn (3 epochs không đủ để hội tụ hoàn toàn)

5.3. Tốc độ hội tụ

- **Transformer** hội tụ nhanh hơn, với loss giảm mạnh sau mỗi epoch
- **RNN** hội tụ chậm hơn, với sự cải thiện nhỏ hơn giữa các epoch

5.4. Kết luận

- Mô hình Transformer với cấu hình T3 (hidden_size=192, n_layers=3, batch_size=64) đạt hiệu suất tốt nhất trong bài toán dịch Anh-Việt
- Mô hình RNN, mặc dù có số lượng tham số lớn hơn, không đạt được hiệu suất tương đương với Transformer
- Dựa trên kết quả, nên ưu tiên sử dụng kiến trúc Transformer với hidden_size lớn và nhiều lớp hơn cho nhiệm vụ dịch máy Anh-Việt