

2024年3月11日

## 卒業論文要旨

# 機械翻訳における Transformer の系列最適化に関する研究

森岡 拓（愛媛大学 工学部 工学科 コンピュータ科学コース）

指導教員：二宮 崇 教授，梶原 智之 講師

近年ニューラルネットワークの導入によって機械翻訳の性能は向上し、Google 翻訳<sup>1</sup>や DeepL 翻訳<sup>2</sup>などの Web サービスをはじめとして、実用的な国際コミュニケーション手段として利用されている。現在最も多く用いられている機械翻訳モデルは Transformer をベースとしたものである。このモデルは、推論時に時系列に沿って、文頭から文末に向けて、1 単語ずつ順番に出力を生成する。そのため、Transformer は将来出力されるであろう文末側の単語列を参照して推論を行うことができない。

そこで、本論文では、将来生成される文末側の単語列も含めて出力系列全体を最適化することを目指し、出力系列全体にわたって最適化を行う SO モジュール（系列最適化モジュール, Sequence Optimizer Module）を Transformer の Decoder 出力に対して適用する SeqOT（Sequence Optimized Transformer）を提案する。また、SO モジュールとして Transformer-default, Transformer-CRF, Transformer-DCRF の 3 つのバリエーションを提案する。SeqOT のモデル全体図を図 1 に示す。

---

1 <https://translate.google.co.jp/>

2 <https://www.deepl.com/>

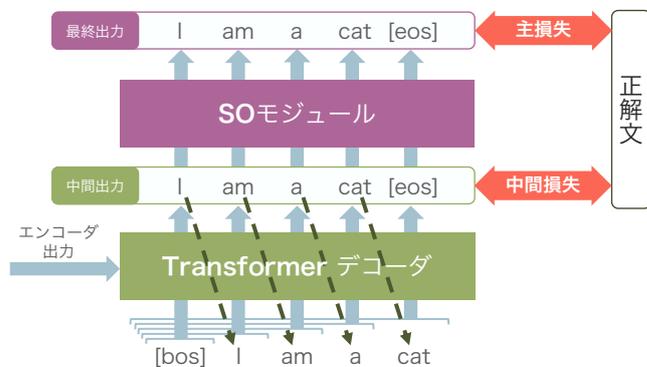


図 1: SeqOT の全体図

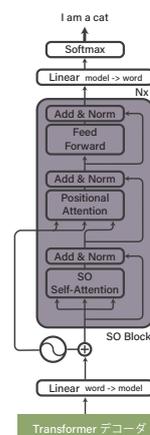


図 2: SeqOT-default

各 SO モジュールには、NAT（非自己回帰 Transformer, Non-Autoregressive Transformer）の分野の技術を活用する。Transformer が推論時に、単語列「I am」を入力して次の単語「a」を得るような自己回帰的な出力を利用するのに対し、NAT は一括生成のアプローチをとっている。SeqOT-default のモデル構造を図 2 に示す。SeqOT-default では、Transformer の Decoder 構造を SO モジュールのベースとし、Cross Attention 構造を NAT で利用される Positional Attention 構造に変更している。SeqOT-CRF のモデル構造を図 3 に示す。SeqOT-CRF は、系列全体の最適化を行う CRF モデルを SO モジュールとして用いた SeqOT であり、ラベル候補のすべての組み合わせをグラフでモデル化して最適なラベルを探索する。ラベル候補数が数十程度である一般の系列解析タスクでは CRF を用いても特に問題とはならないが、機械翻訳タスクにおいて CRF を用いる場合は、ラベル候補数が語彙サイズ (数万) と等しくなるため、計算量の観点からそのままでは CRF の学習を行うことは難しい。そのため、NAT の分野では、NART-CRF で提案された、遷移行列の分割と、Beam 近似によって解決した。SeqOT-DCRF のモデル構造を図 4 に示す。SeqOT-DCRF では SeqOT-CRF に NART-CRF で提案された Dynamic Matrix を導入している。

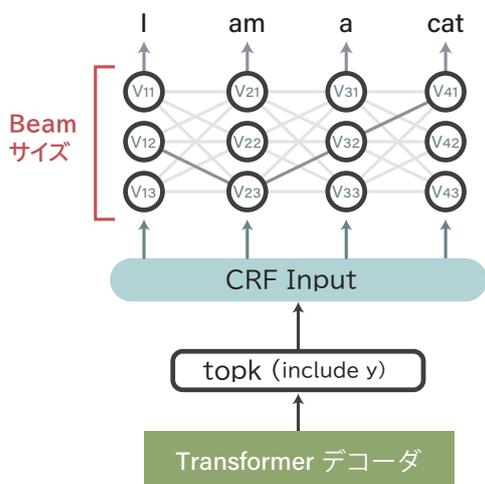


図 3: SeqOT-CRF

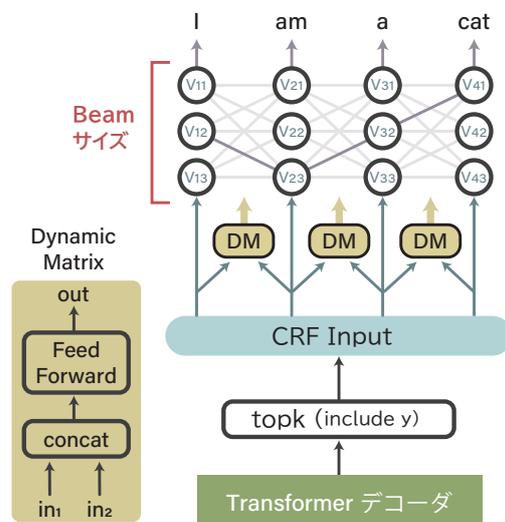


図 4: SeqOT-DCRF

実験では、アジア学術論文抜粋コーパス (ASPEC) を用いて、英語から日本語への機械翻訳の性能評価を行った。2 万文対の訓練データで各モデルを学習し、1,812 文対の評価用データに対し、翻訳精度の評価を行った。翻訳精度の評価にはモデルの出力と人手で作成した正解文の間の一致度を測る指標である BLEU を使用した。

表 1: 実験結果

| モデル         | 最終出力 BLEU           | 中間出力 BLEU           |
|-------------|---------------------|---------------------|
| Transformer | -                   | 16.78(+0.00)        |
| SeqOT       | 17.98(+1.20)        | 18.02(+1.24)        |
| SeqOT-CRF   | <b>18.32(+1.54)</b> | <b>18.48(+1.70)</b> |
| SeqOT-DCRF  | 17.91(+1.13)        | 18.12(+1.34)        |

実験結果は表 1 のようになった。すべての SeqOT モデルがベースラインの Transformer を上回る性能を発揮した。また、SeqOT-CRF の中間出力が最大の BLEU 値 18.48 を達成し、ベースラインに対して 1.7 ポイントの向上が見られた。

実験結果の分析から、SeqOT の中間出力の品質向上が最終出力の性能向上につながっていると考えられることと、SeqOT の中間出力から最終出力に変換される際の品質低下の原因が自己回帰的な中間出力生成によるノイズによるものだと考えられることが分かった。出力例の分析では、SeqOT は中間出力を最終出力にコピーすることが多い傾向があること、SeqOT-CRF は文の一部に限り修正していること、SeqOT-DCRF は入力系列が誤りを含む場合過剰な修正をして性能を下げていると考えられることが分かった。また、SO モジュールは出力系列長を変化させるほどの修正能力は持たないとみられることが分かった。