

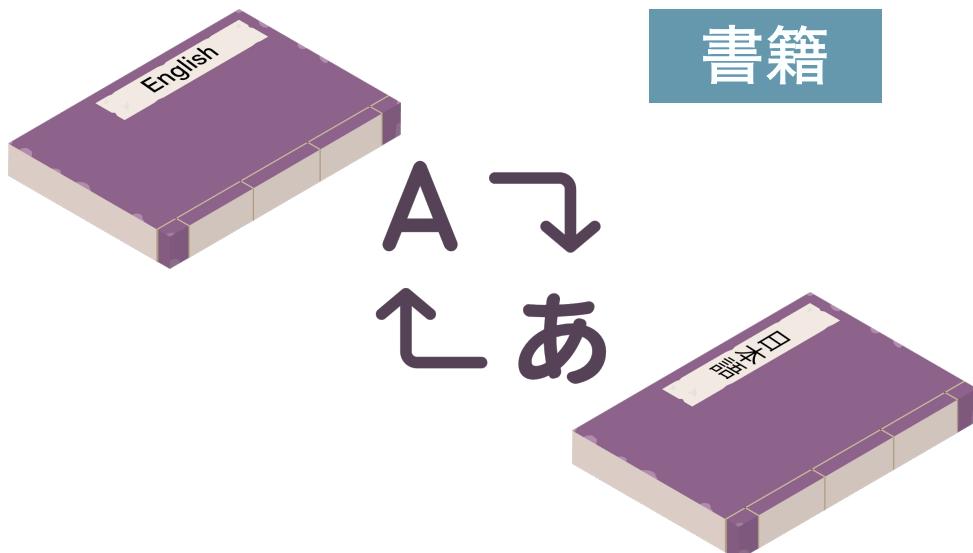
機械翻訳におけるTransformerの 系列最適化に関する研究

人工知能研究室 学部4年 森岡拓
(指導教員： 二宮崇 梶原智之)

卒業論文資料
2024年3月11日

背景：機械翻訳

- ・機械翻訳はある言語を他の言語に自動的に翻訳するタスク
- ・人手の翻訳と比較して、読めない言語の文書を低成本で読める文書に変換できる
- ・実用化例：Google翻訳、DeepL翻訳など

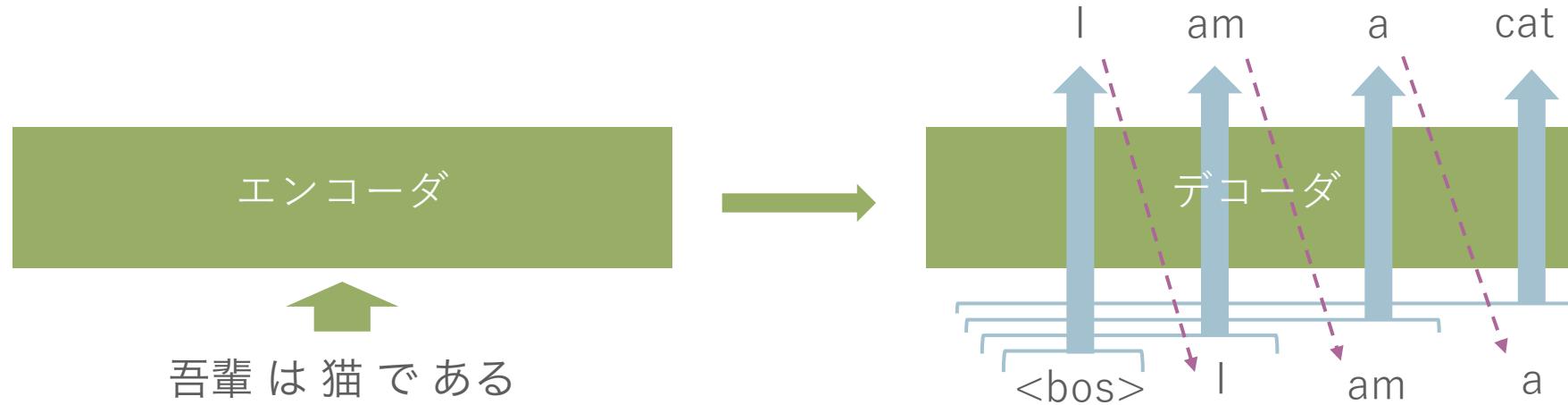


書籍



背景：Transformer [Vaswani+ 2017]

- ・機械翻訳で主流のモデル
- ・推論時はデコーダで翻訳文を前から順に生成する



Transformerは過去に出力された単語列のみを考慮して生成する
= 未来(文末側)に出力される単語列を考慮できない



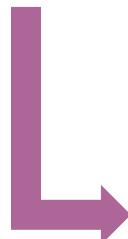
系列全体を見て最適な翻訳文を生成したい

研究の目的

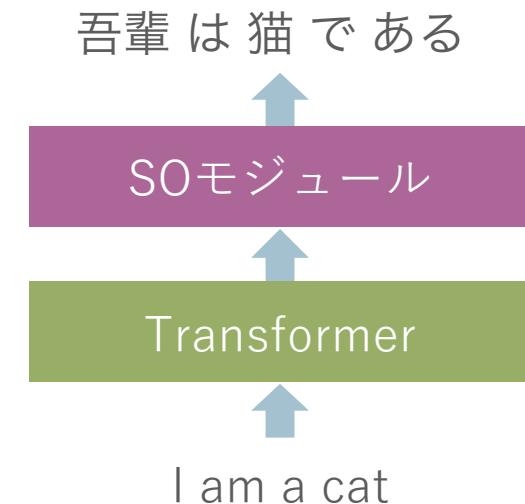
系列全体について最適化することで、より高品質な翻訳を可能にすることを目指す



Transformerに**SOモジュール**(系列最適化モジュール)を付加した
SeqOT(Sequence Optimized Transformer)を提案する

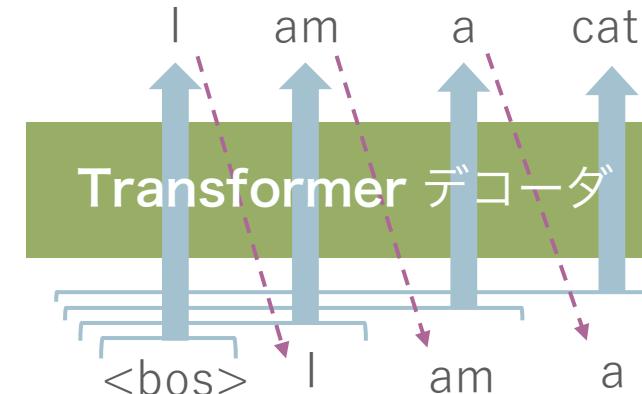
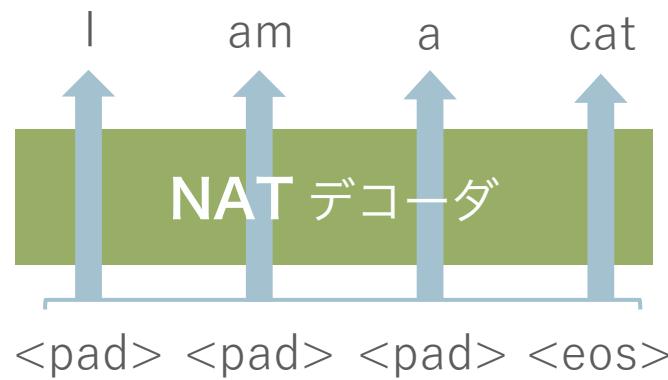


- SeqOT-default
- SeqOT-CRF
- SeqOT-DCRF



関連研究：NAT

- 系列を一括で最適化する事例
- NAT(Non-Autoregressive Transformer)[Gu+ 2018]
 - Transformerの推論パスの自己回帰をなくして高速化を目指すアプローチ
 - 出力単語同士の独立性により同一単語の連續出力などが課題
 - Positional Attentionによる位置情報の強い導入

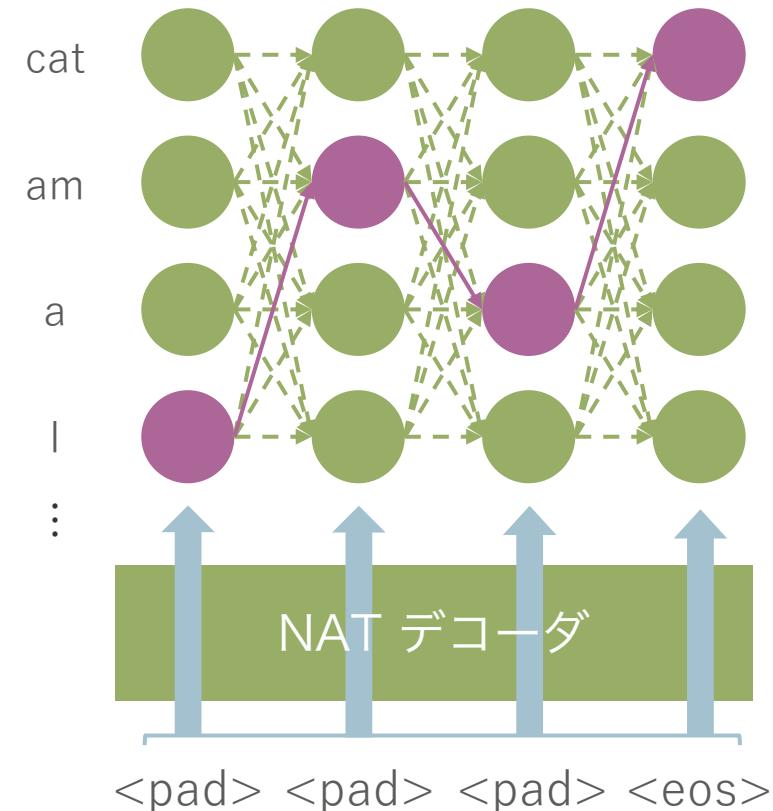


関連研究：NART-CRF [Sun+ 2019]

NATにおいて、出力単語間の独立性を緩和するためにCRFを導入

CRF(Conditional Random Fields) [Lafferty+ 2001]

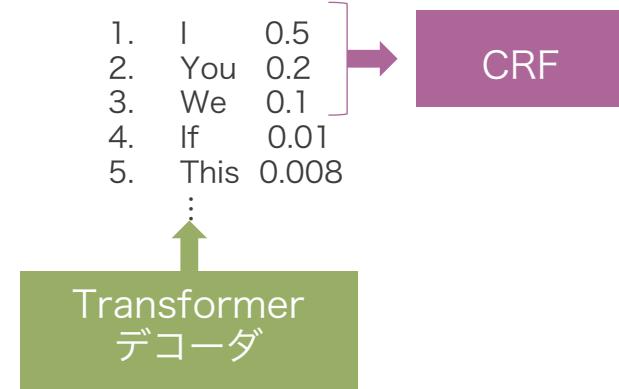
- 系列ラベリングタスクで利用される
(品詞推定、固有表現認識など)
- 全てのラベル列候補のうち、
最もスコアが大きいものを出力する
 - 観測素性：ある時点で各ラベルが
選ばれるスコア
 - 遷移素性：次の時点のあるラベルへ
遷移する際のスコア
- 出力パスの前後関係を遷移素性によって
明確にモデル化できる



NART-CRF : CRFの軽量化

Beam近似

- 全ての語彙に対してlossを計算するのは困難
→デコーダ出力のうち、数値が大きい
上位の単語k個だけをCRFに渡して学習する



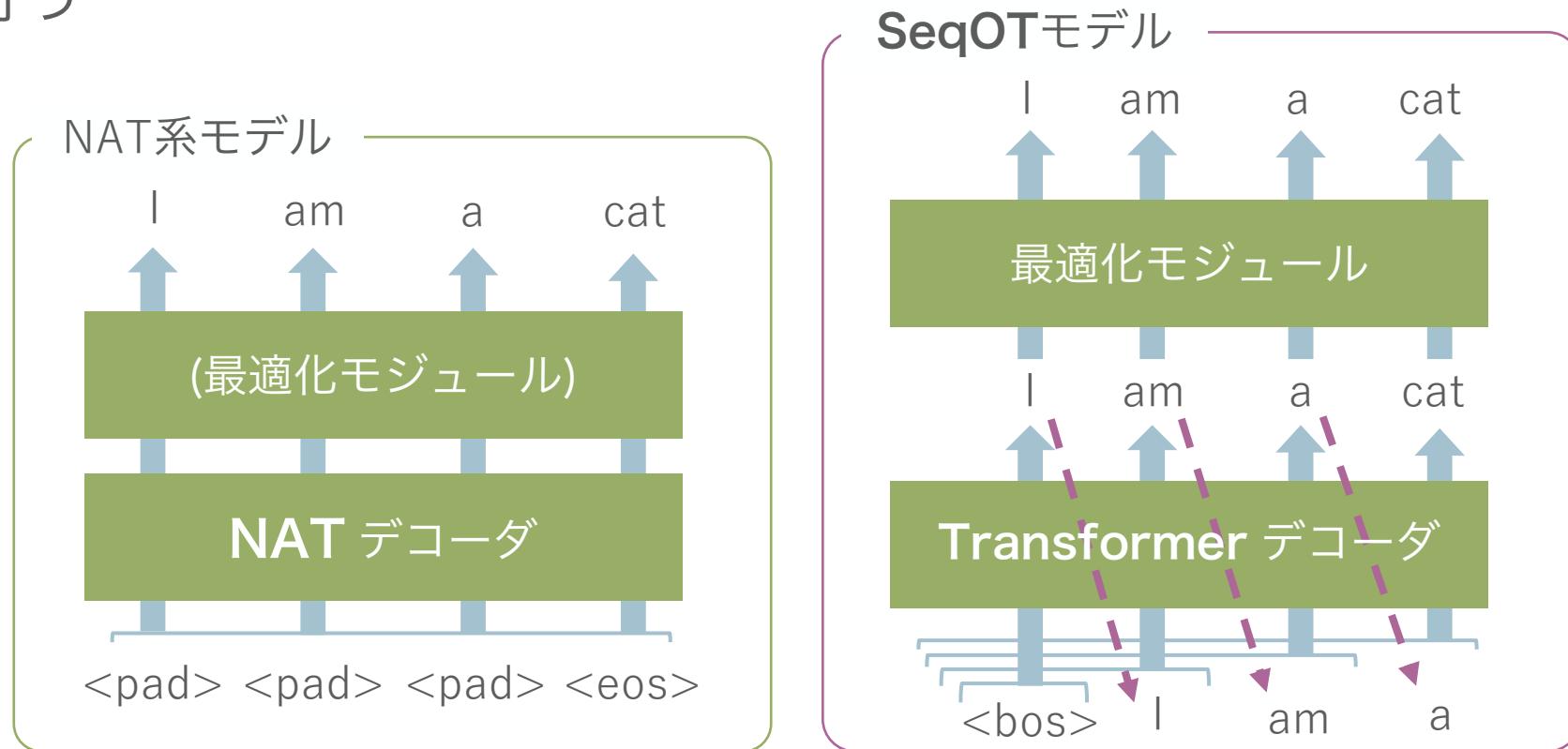
遷移行列の分割

- 遷移素性は語彙数の2乗のサイズが必要である
→遷移素性を2つのベクトルの積で表現
メモリ空間を節約

$$\begin{matrix} \text{語彙数} \\ \text{遷移行列} \end{matrix} = \begin{matrix} \text{語彙数} \\ \text{E1} \end{matrix} \times \begin{matrix} \text{圧縮} \\ \text{E2} \end{matrix}$$

提案手法との関係

- NATでは**一括出力**に対してモジュール処理を行う
- 提案手法ではTransformerの**自己回帰的な出力**に対してモジュール処理を行う



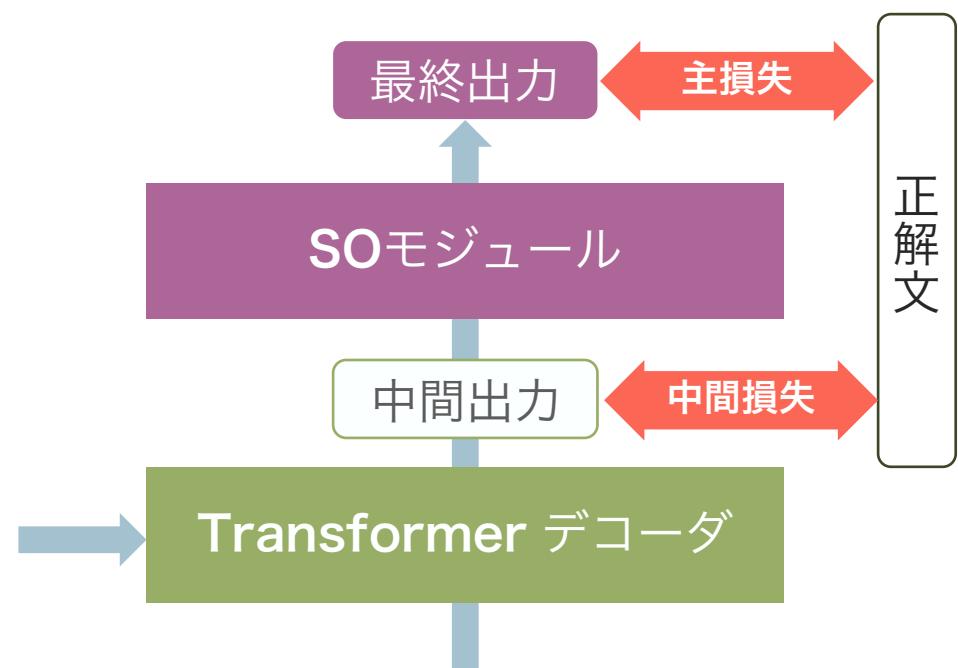
提案手法：SeqOT

SeqOT(Sequence Optimized Transformer)

- SOモジュール(Sequence Optimizer Module)
 - Transformerデコーダの系列全体に渡った出力に対して最適化を行う
 - 3つのモジュールを提案する
- 学習損失 L
$$L = \text{主損失} \times \mu + \text{中間損失} \times (1 - \mu)$$

主損失：SOモジュールで計算する損失

中間損失：Decoder Blockの出力を線形変換したものに対する損失（共通）



提案手法：SeqOT-default

- SO Block
 - Transformer Decoder Blockをベースとする
 - Cross Attention削除
 - Positional Attention追加

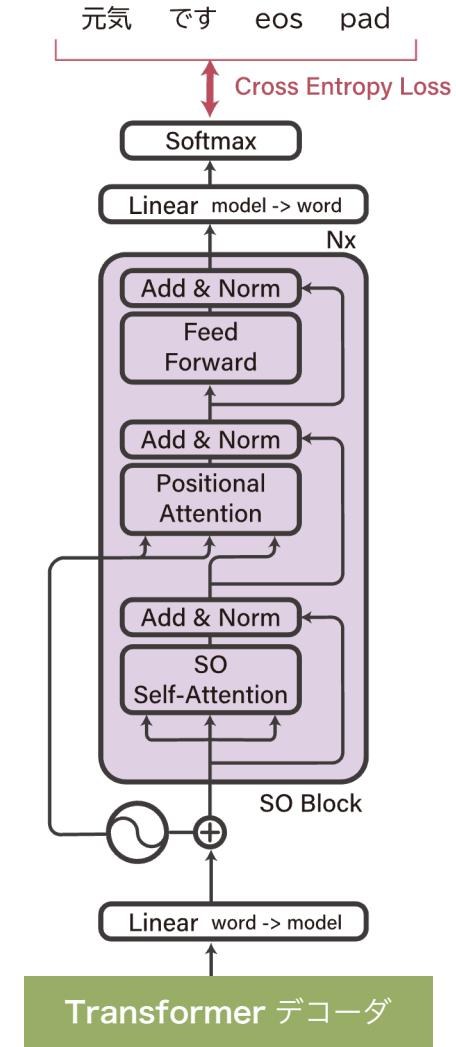
- 学習損失

$$\text{主損失} = \text{Cross Entropy}(W_{out} \cdot \text{SO Block}_{\times N}(W_{in}m, PE), y)$$

y : 正解文

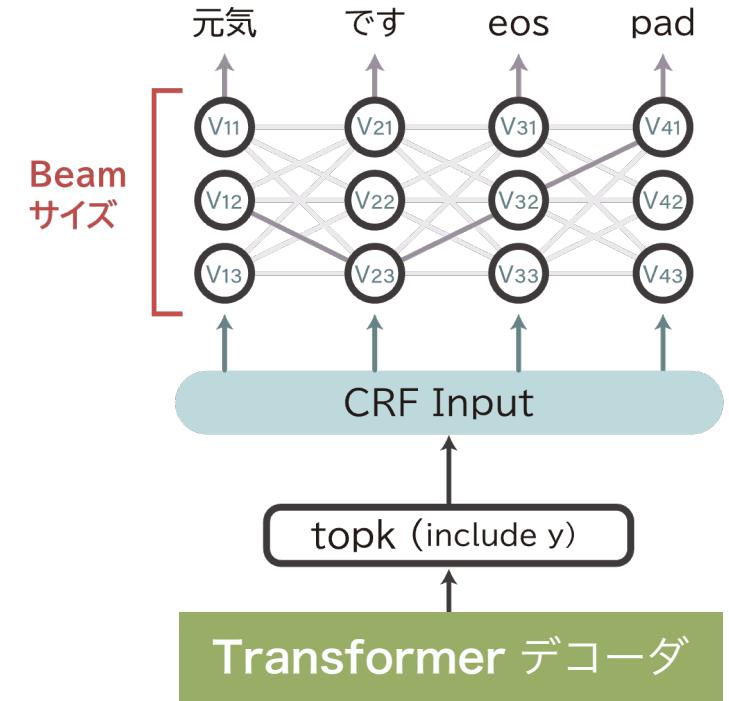
m : デコーダ出力行列

PE : 位置埋め込み行列



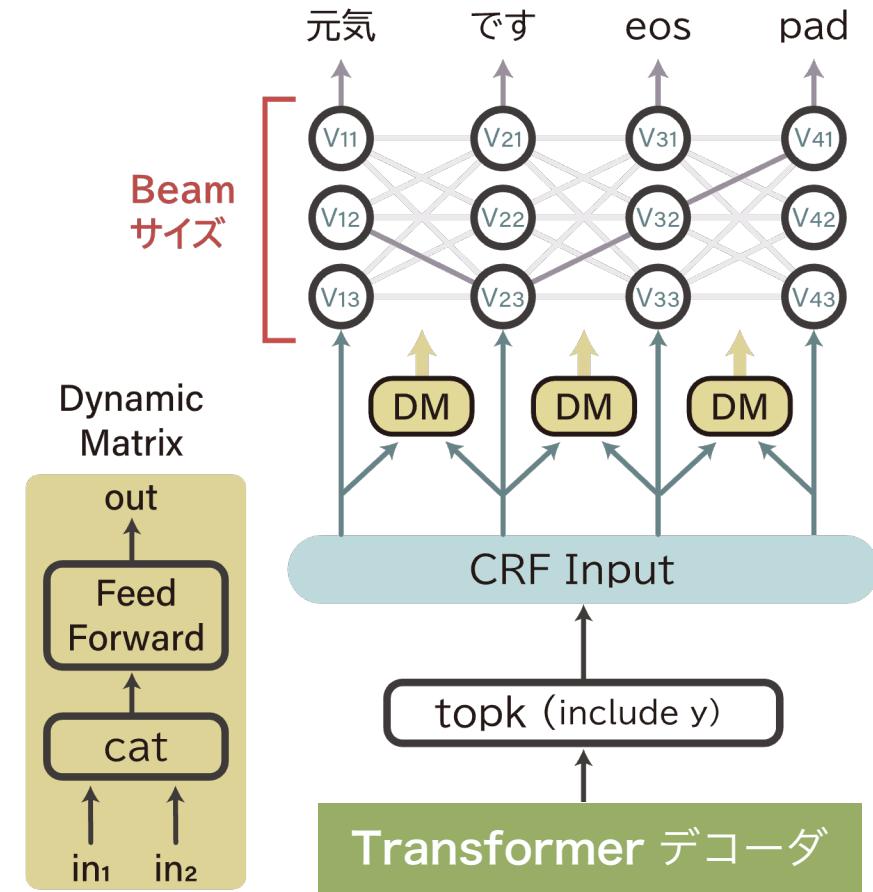
提案手法：SeqOT-CRF

- 遷移行列分割
 - 遷移行列を2つの行列 $E_1 E_2$ の積として表現
 $TM = E_1 E_2$
- Beam近似
 - デコーダ出力の行列 m をbeam sizeに圧縮
 - $\text{topk}_{\text{include } y}()$: 正解系列 y を含む上位beam size個の要素を抽出
- 学習損失
主損失=CRF loss($TM, \text{topk}_{\text{include } y}(m, y, \text{beam size}), y$)



提案手法：SeqOT-DCRF

- Dynamic Matrix
 - CRFの遷移行列にDecoder出力を埋め込む
 - i 回目の遷移行列は m_i と m_{i+1} から構成される
 $d_i = \text{FeedForward}(\text{concat}(m_i, m_{i+1}))$
 - $\text{TM}_i = E_1 d_i E_2$
- 学習損失 $\text{TM} = \{\text{TM}_1, \text{TM}_2, \dots, \text{TM}_{T-1}\}$ $m' = \text{topk}_{\text{include } y}(m, y, \text{beam size})$ 主損失 = CRF loss(TM, m' , y)



評価実験：設定

- 評価タスク：翻訳タスク(英語→日本語)
 - 3つのシード値で実験し、その平均値を報告
- データセット：ASPEC(アジア学術論文抜粋コーパス)
 - 学習用：20,000、開発用：1,790、評価用：1,812
 - 低リソース設定(Transformerの出力を修正する能力を測る)
- 評価指標：BLEU
 - 生成文と正解文の一一致度を測る指標
- モデル設定
 - 学習エポック数：90
 - 開発用データに対して最大BLEUを記録したエポックのモデルを利用

評価実験：結果

- すべてのSeqOTモデルがベースラインを上回るBLEU値を達成
- SeqOT-CRFの中間出力が**最大**のBLEU値
- SeqOT-CRFの最終出力が**2番目**に高いBLEU値

※カッコ内はベースラインとの差分

モデル	最終出力BLEU	中間出力BLEU
Transformer(ベースライン)	-	16.78 (+0)
SeqOT-default	17.98 (+1.2)	18.02 (+1.24)
SeqOT-CRF	18.32 (+1.54)	18.48 (+1.7)
SeqOT-DCRF	17.91 (+1.13)	18.12 (+1.34)

分析

- SeqOTの中間出力がベースラインより高品質
→ SOモジュールの損失がTransformer部分の学習に有効であった
- SeqOTでは最終出力が中間出力より低い品質
→ 推論時の中間出力にはノイズが混ざるためSOモジュールが学習時と大きく異なる入力を受け取り最終出力の品質を悪化させた

※カッコ内はベースラインとの差分

モデル	最終出力BLEU	中間出力BLEU
Transformer(ベースライン)	-	16.78 (+0)
SeqOT-default	17.98 (+1.2)	18.02 (+1.24)
SeqOT-CRF	18.32 (+1.54)	18.48 (+1.7)
SeqOT-DCRF	17.91 (+1.13)	18.12 (+1.34)

出力例

SeqOT：中間出力と最終出力が一致している

SeqOT-CRF：「の」を「その」に修正している

SeqOT-DCRF：「，面圏」という中間出力の誤りにつられて最終出力で大きな誤りを含んでいる

モデル	出力例
正解文	ファセットに非吸収領域を形成することで光学損傷を無くし，最大180mWの光出力を得た。
Transformer	光損傷は，を形成した領域を除去することで可能で，最大光学出力を得た。
SeqOT-default 中間	高感度の形成により光学損傷を形成し，光出力の最大光学出力を得た。
SeqOT-default 最終	高感度の形成により光学損傷を形成し，光出力の最大光学出力を得た。
SeqOT-CRF 中間	三面体領域を形成することにより可能な光学損傷を除去し， の 最大光学出力を得た。
SeqOT-CRF 最終	三面体領域を形成することにより可能な光学損傷を除去し， その 最大光学出力を得た。
SeqOT-DCRF 中間	面性について は，面圏 の高光学損傷を軽減し，光出力を得た。
SeqOT-DCRF 最終	面性について inhibition, cfb圏 のメトロ光学損傷を軽減し，フレンド出力を得た。

まとめ

背景：機械翻訳ではTransformerが広く使われている

課題：Transformerは未来(文末側)に生成される単語列を考慮して生成できない

手法：Transformerに系列全体を最適化するSOモジュールを加えた**SeqOT**を提案した

結果：**SeqOT**の各モデルはTransformerベースラインを超える性能を示した

