

ACL2018読み会@名大

Context-Aware Neural Machine Translation Learns Anaphora Resolution

Authors: **Elena Voita**^{1,2}, **Pavel Serdyukov**¹, **Rico Sennrich**^{3,4}, and **Ivan Titov**^{3,2}

¹Yandex, Russia

²University of Amsterdam, Netherlands

³University of Edinburgh, Scotland

⁴University of Zurich, Switzerland

発表者: 綱川 隆司 (静岡大学)

概要

- 文脈を考慮したニューラル機械翻訳(Context-aware NMT)の有効性を検証
 - Transformer (Vaswani et al., 2017) を拡張
 - BLEU: 29.46 → 30.14 (+0.68)
- 英露翻訳において、代名詞の曖昧性解決に貢献することを示した
- アテンションの分布と共参照関係の対応を分析し、Context-aware NMTが照応関係を暗黙的に捉えていることを実験的に示した

背景

- 従来の機械翻訳は文単位で行われていたが、談話(discourse)あるいは文脈を考慮することで訳文は改善されるはず
- 統計的機械翻訳(SMT)時代には、代名詞、接続詞、語彙の一貫性、話題適合といった観点の特徴量として加えた研究が行われていた
- ニューラル機械翻訳(NMT)では、文脈情報を直接入力する手法が加わった → **Context-aware NMT**

本研究の目的

- Context-aware NMTの有用性を示す。
- 文脈情報がどう用いられているかを分析する。特に、代名詞の翻訳に着目する。
- Context-aware NMTが暗黙的に照応関係を捉えていることを、人手による実験で示す。

ベースモデル：Transformer

(Vaswani et al., 2017)

- 現在最先端のニューラル機械翻訳 (NMT) モデル
 - RNNと異なり反復構造を持たず、自己アテンションと呼ばれるメカニズムにより系列入力に対応
 - 自己アテンションをN層重ねることで、各階層で文内の単語間の依存関係を捉えられる

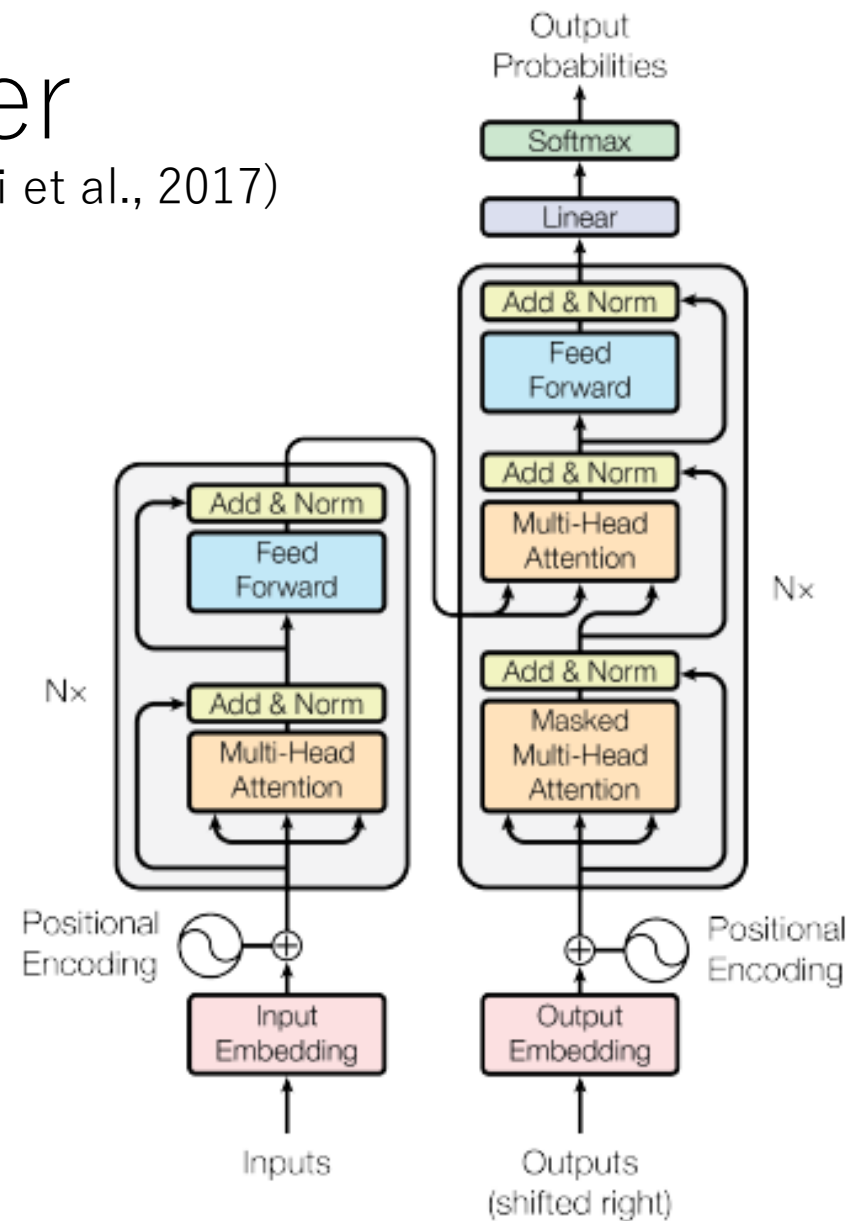
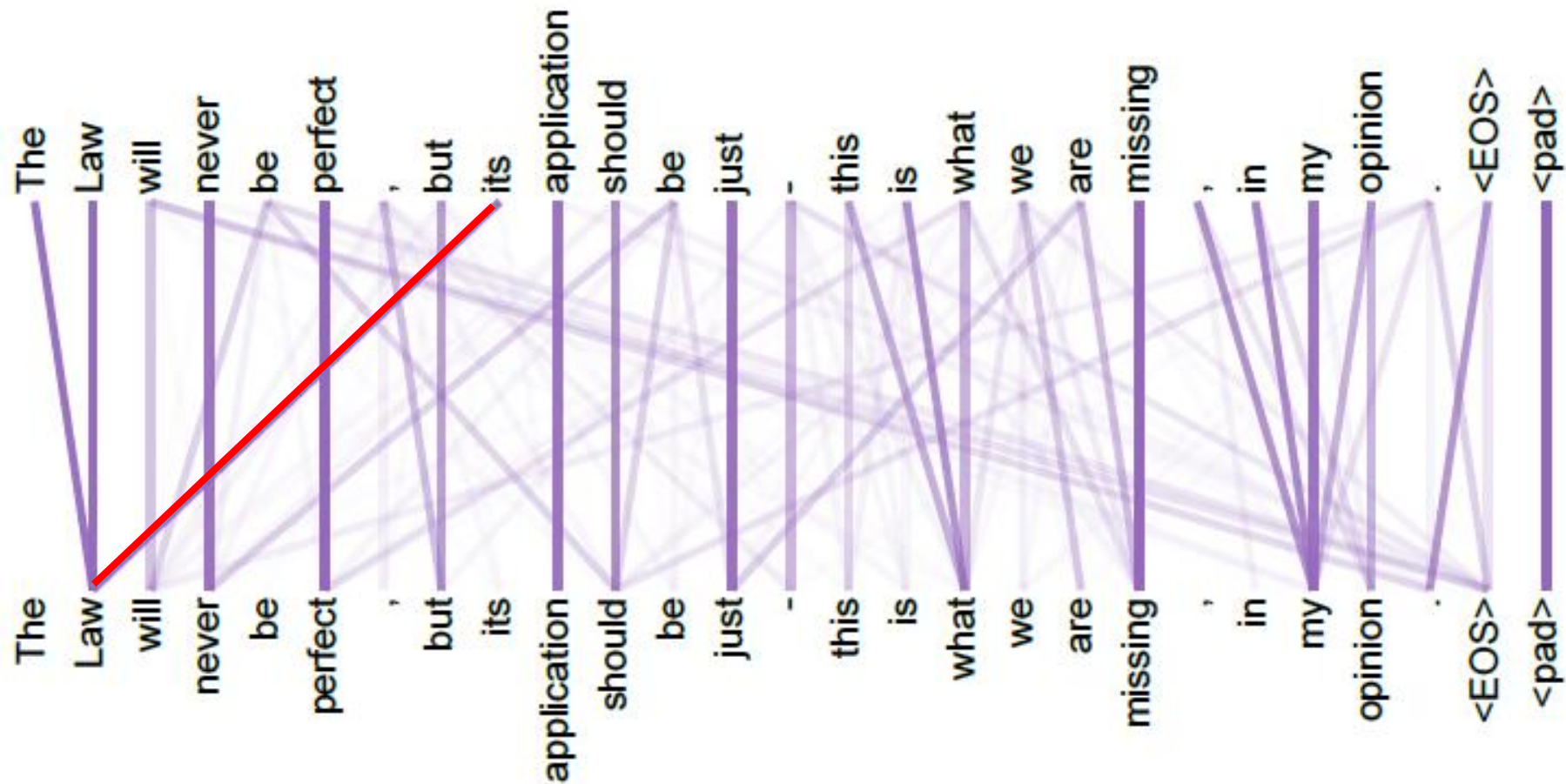


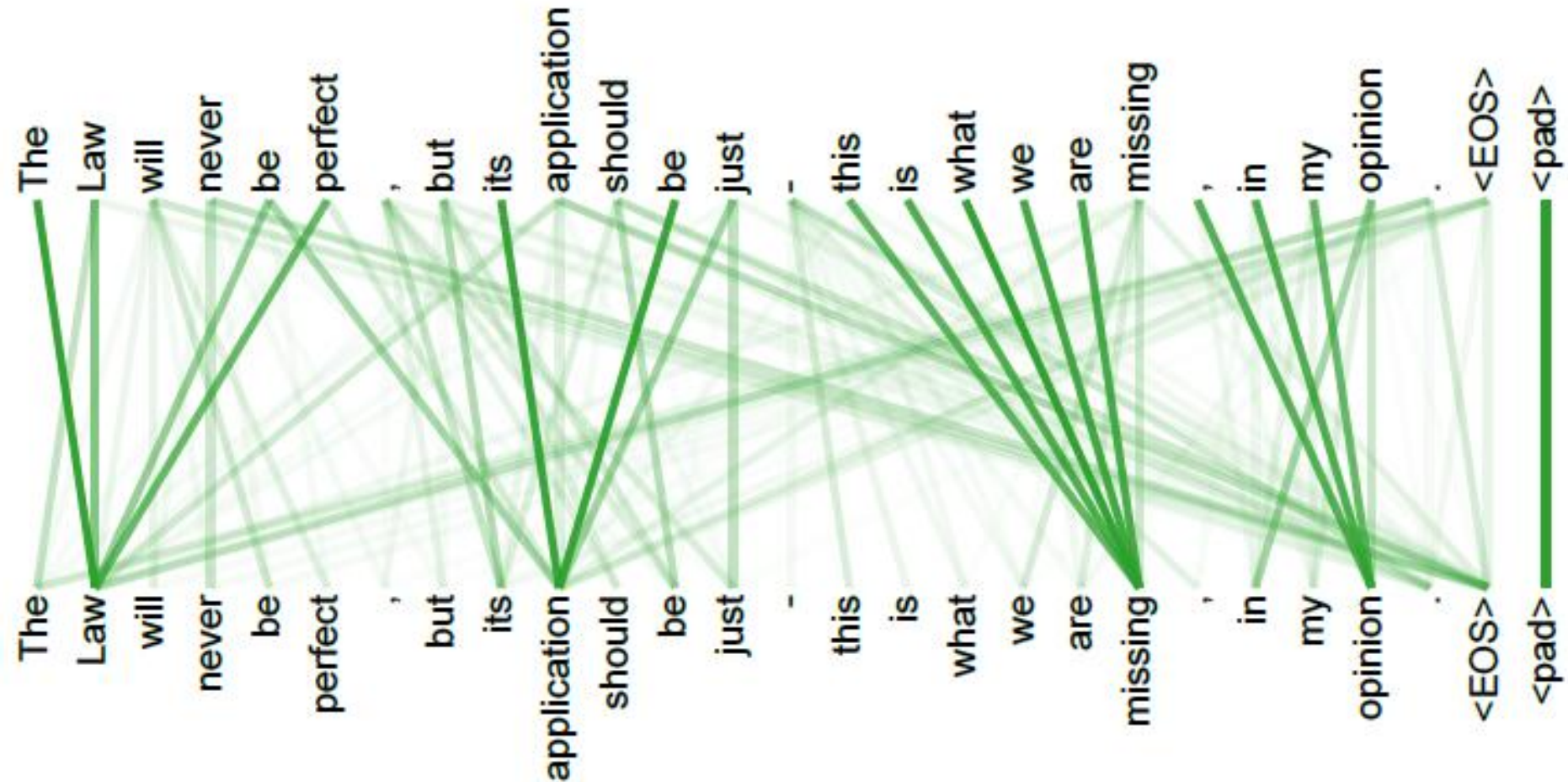
Figure 1: The Transformer - model architecture.

自己アテンションの例 (Vaswani et al., 2017)



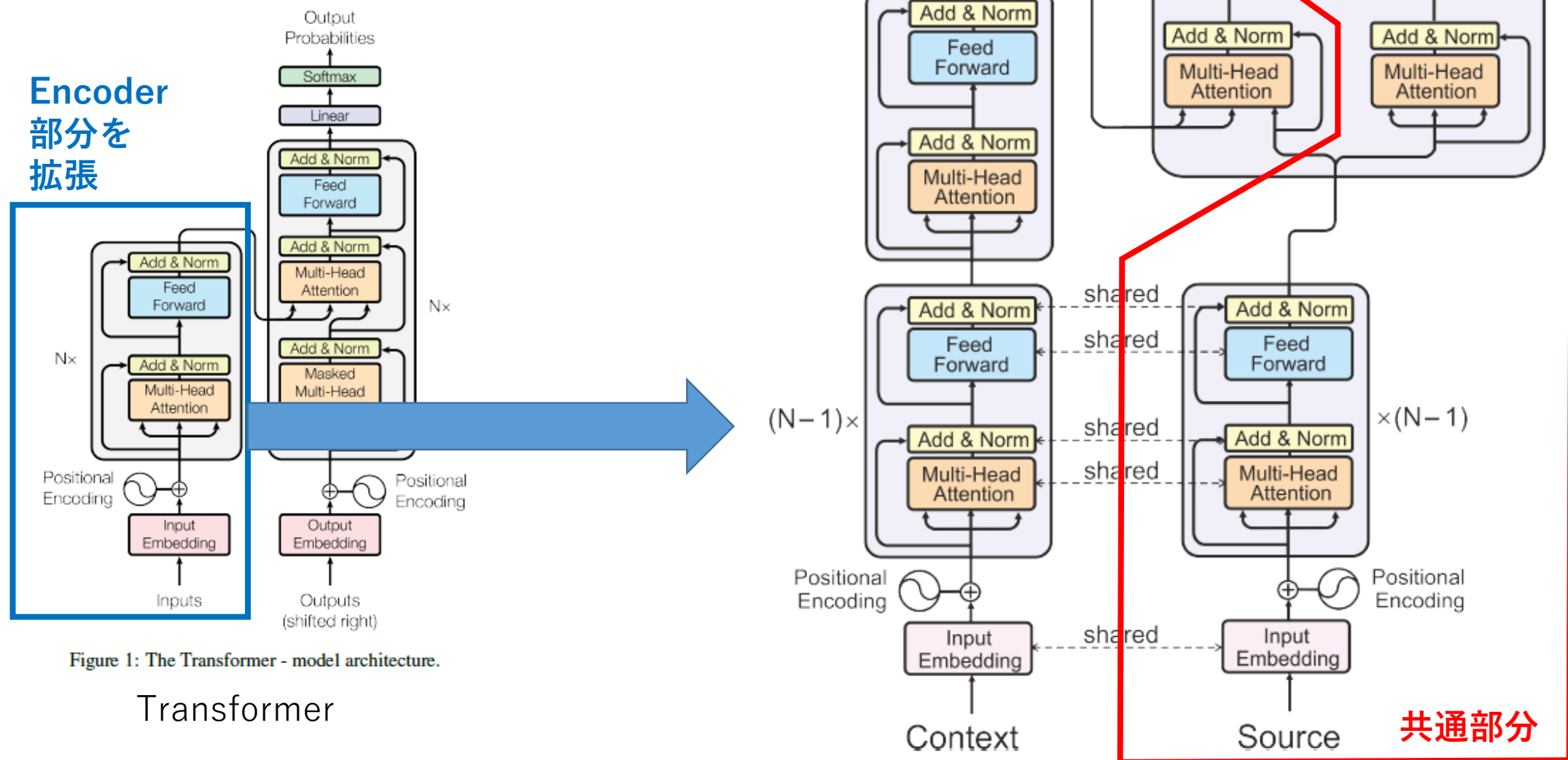
The Law will never be perfect, but **its** application should be just - this is what we are missing, in my opinion.

自己アテンションの例 (Vaswani et al., 2017)



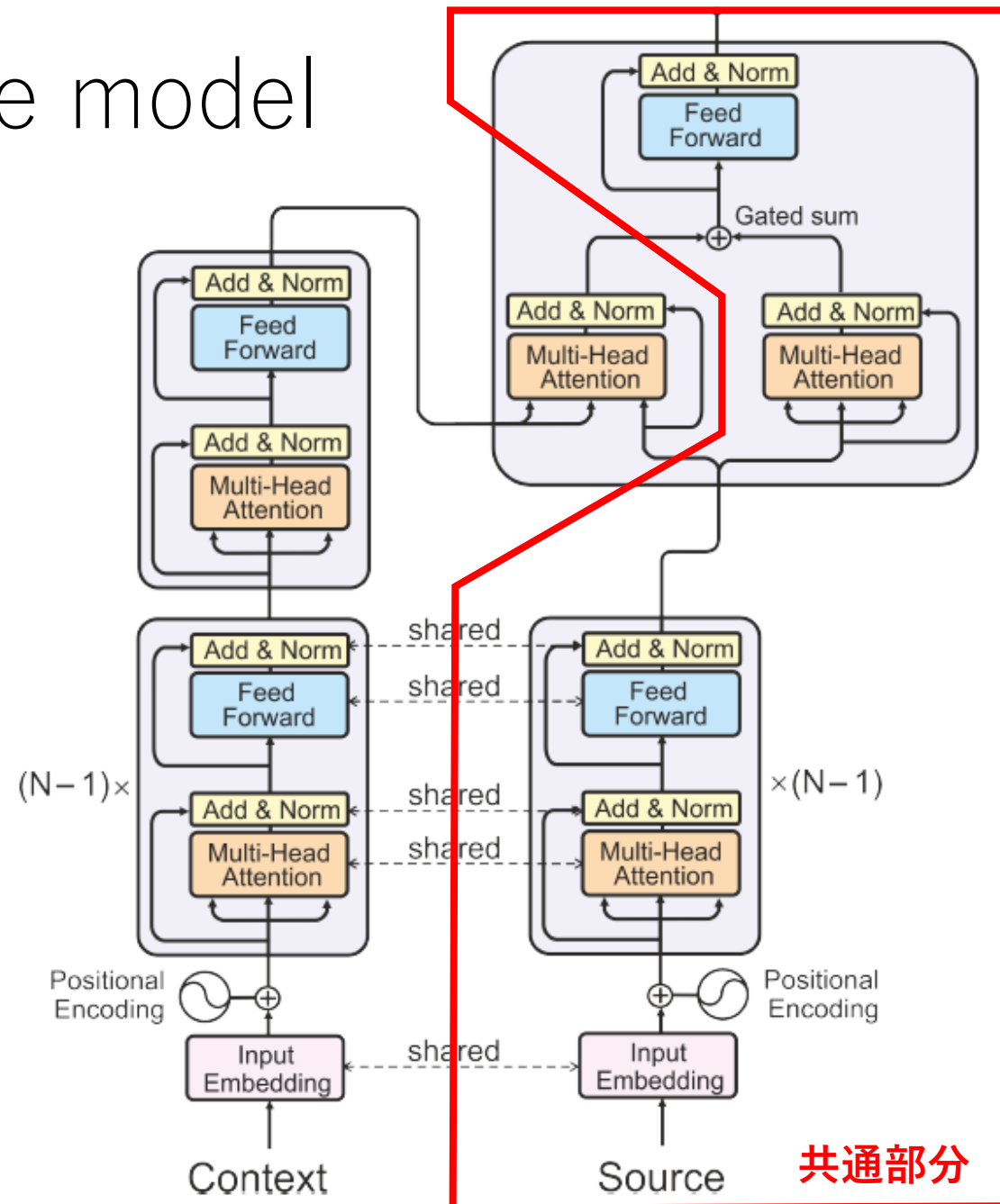
The **Law** will never be perfect, but its **application** should be just - this is what we are **missing**, in my **opinion**.

提案手法：Context-aware model



提案手法：Context-aware model

- TransformerのEncoder部分のうち、 $(N - 1)$ 層までは文脈と入力文でパラメータを共有する
- Source encoderの N 層目で、文脈と入力文のアテンションを求め、入力文のみの自己アテンションとのGated sumをとる



実験

- データ
 - OpenSubtitles2018（映画字幕のコーパス）（Lison et al., 2018）
 - 英語・ロシア語のペアを利用
 - 訓練：200万文対
 - 開発・テスト：各1万文対
- 実験設定
 - Byte-pair encoding (Sennrich et al., 2016)を適用
 - 英語・ロシア語で約32000種類のトークンに変換
 - パラメータ等はTransformerのものを使用

実験結果：翻訳性能

model	BLEU
baseline	29.46
concatenation (previous sentence)	29.53
context encoder (previous sentence)	30.14
context encoder (next sentence)	29.31
context encoder (random context)	29.69

Table 1: Automatic evaluation: BLEU. Significant differences at $p < 0.01$ are in bold.

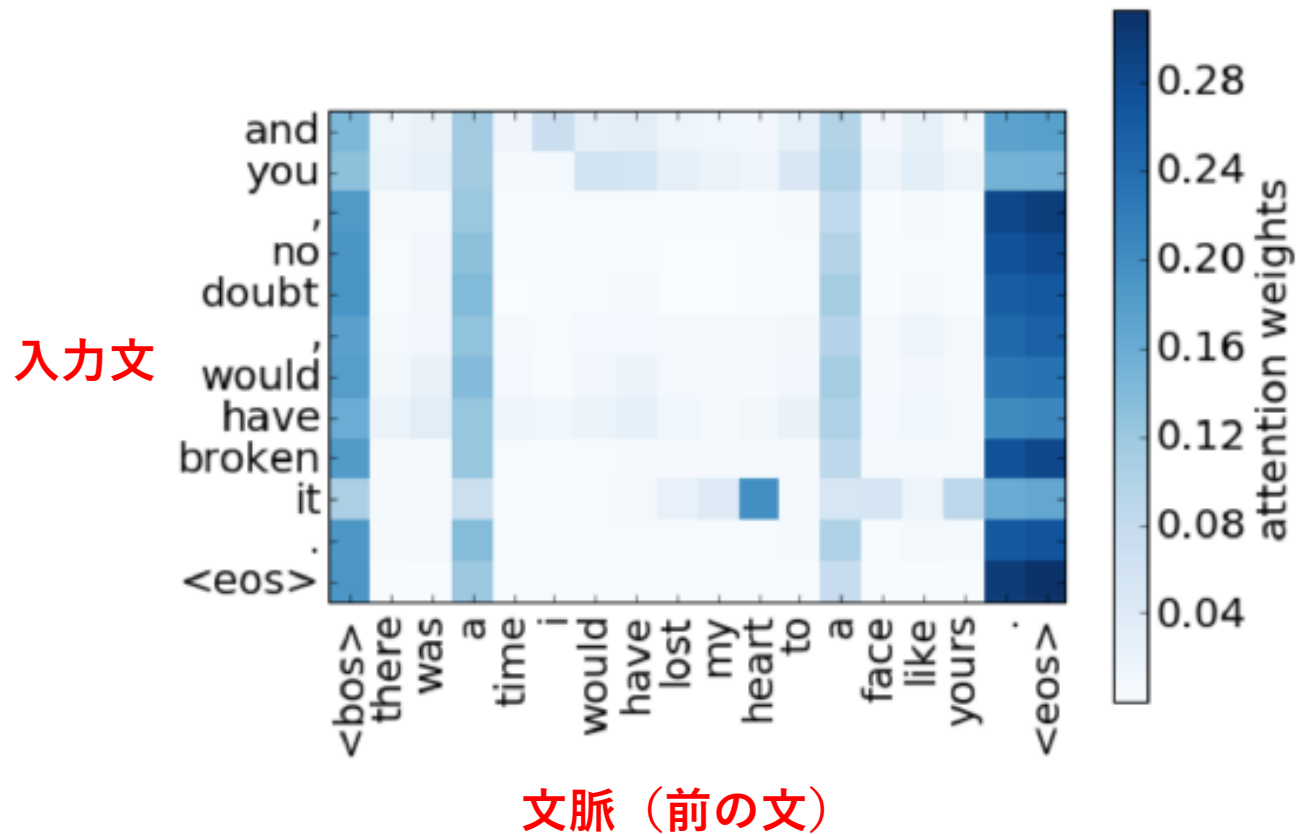
- **concatenation**: 入力文の前の文を結合して入力するモデル (Tiedemann and Scherrer, 2017)
- **(random context)**: ランダムに選択した文を文脈として入力

- Context-aware modelによりBLEUで約0.7ポイント向上
- 入力文に文脈を結合するモデルよりも有意に向上

分析

- 本研究における課題
 - **入力文の中で、文脈に頼って翻訳される語はどれか？**
 - 文長や単語の位置は文脈へのアテンションや性能に影響を与えるか？
 - Context-aware NMTは素性エンジニアリングをすることなく暗黙的に共参照を捉えられるか？

文脈の寄与度の求め方



- 入力文・文脈間のアテンションに着目
- 入力文の語のほとんどは、文脈中の無意味な語 (<bos>, <eos>, “a” など) と対応付く
- それ以外の文脈中の語に対応付く部分が、文脈の寄与がある箇所とみなせる
 - 重みの最大値を寄与度とする

文脈に依存しやすい語

寄与度平均 単語位置

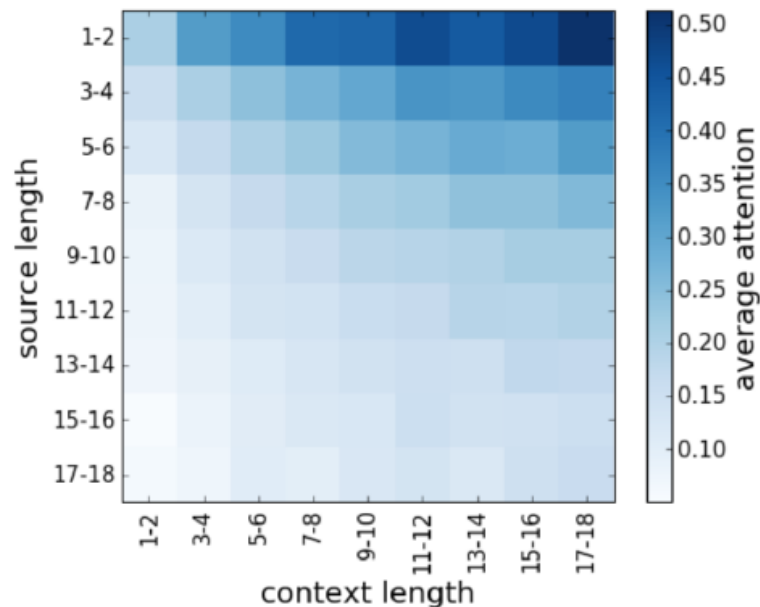
word	attn	pos	word	attn	pos
it	0.376	5.5	it	0.342	6.8
yours	0.338	8.4	yours	0.341	8.3
yes	0.332	2.5	ones	0.318	7.5
i	0.328	3.3	'm	0.301	4.8
yeah	0.314	1.4	you	0.287	5.6
you	0.311	4.8	am	0.274	4.4
ones	0.309	8.3	i	0.262	5.2
'm	0.298	5.1	's	0.260	5.6
wait	0.281	3.8	one	0.259	6.5
well	0.273	2.1	won	0.258	4.6

- 文脈からの寄与度が高い上位10単語をリストアップ
- ロシア語への翻訳時に非常に曖昧な“it”, “yours”といった代名詞がランクイン
- “I”は、過去時制において動詞との性の一致が必要
- “yes”, “yeah”などは非常に短い文で文頭に出現するためにランクインし、既存研究と同じ傾向

分析

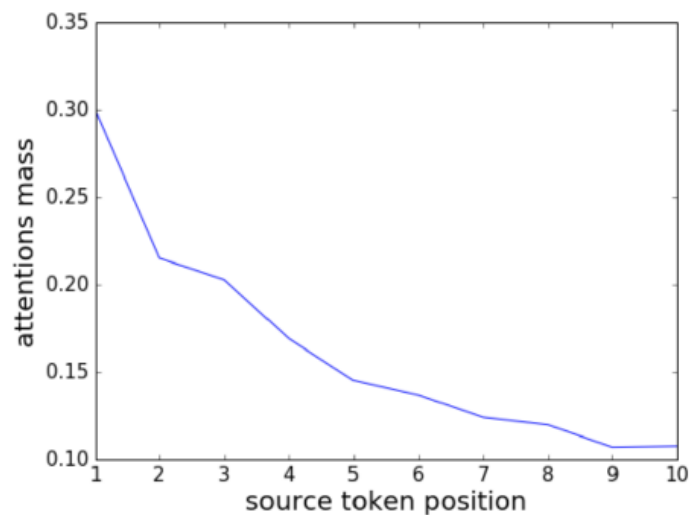
- 本研究における課題
 - 入力文の中で、文脈に頼って翻訳される語はどれか？
 - **文長や単語の位置は文脈へのアテンションや性能に影響を与えるか？**
 - Context-aware NMTは素性エンジニアリングをすることなく暗黙的に共参照を捉えられるか？

文長、単語の位置の影響



- 入力文長・文脈長と文脈寄与度
の関係

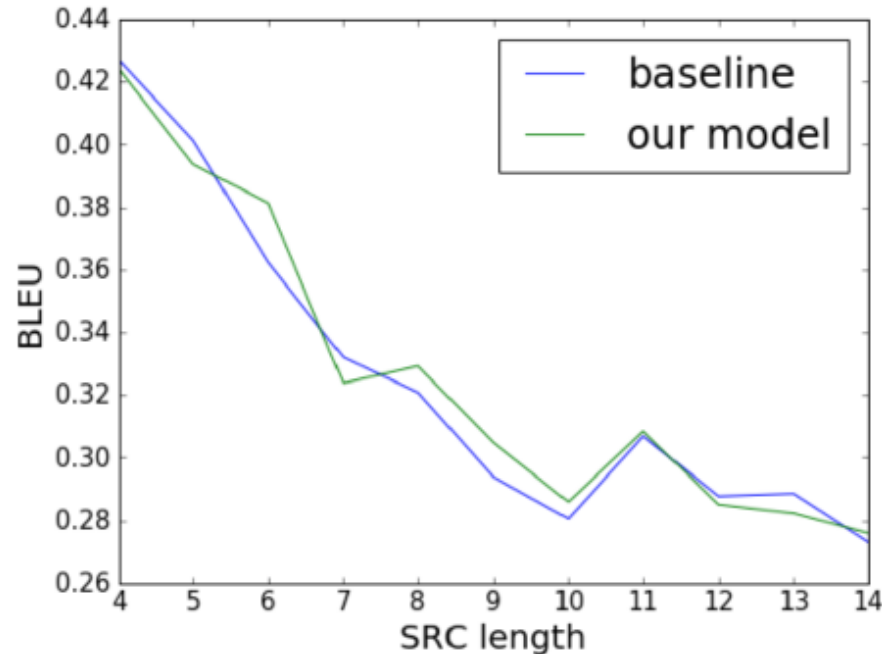
- 入力文が短いほど寄与度は高い
- 文脈が長いほど寄与度は高い



- 単語の位置と文脈寄与度の関係

- 入力文の先頭に近い単語ほど寄与度は高い
- 英語では、聞き手が知っている情報を知らない情報より前に持ってくる傾向がある

文長、単語の位置の影響（続き）



- 入力文長別のBLEUスコアの比較
 - ベースラインモデルとの性能差は入力分の文長には依存しない
 - 文脈は、短い入力文だけに有効なわけではない

分析

- 本研究における課題
 - 入力文の中で、文脈に頼って翻訳される語はどれか？
 - 文長や単語の位置は文脈へのアテンションや性能に影響を与えるか？
 - **Context-aware NMTは素性エンジニアリングをすることなく暗黙的に共参照を捉えられるか？**

代名詞の翻訳における性能改善

pronoun	N	#pronominal antecedent	baseline	our model	difference
it	11128	6604	25.4	26.6	+1.2
you	6398	5795	29.7	30.8	+1.1
yours	2181	2092	24.1	25.2	+1.1
I	8205	7496	30.1	30.0	-0.1

N: その代名詞を含む入力文の数

#pronominal antecedent: 代名詞の先行詞も代名詞である入力文の数

- 各代名詞を含む文のみを抽出し、BLEUスコアの性能差を比較
 - 照応関係はCoreNLPのオープンソースシステム (Manning et al., 2014)により得られたもの
 - 形式主語の“it”などは除外
- 多くの代名詞について、全体の改善(+0.7)を超える性能差がみられる
 - 先行詞も代名詞である場合、先行詞自体からは有用な情報が得られないが、それでも有効性がみられる

代名詞の翻訳における性能改善（続き）

word	N	baseline	our model	diff.
it	4524	23.9	26.1	+2.2
you	693	29.9	31.7	+1.8
I	709	29.1	29.7	+0.6

type	N	baseline	our model	diff.
masc.	2509	26.9	27.2	+0.3
fem.	2403	21.8	26.6	+4.8
neuter	862	22.1	24.0	+1.9
plural	1141	18.2	22.5	+4.3

- 先行詞が名詞の場合に限定した場合の性能差
 - 改善幅がより大きく
 - “I”でも有意な改善
- “it”のロシア語訳の性・数ごとの性能差
 - “it”が女性形または複数形に訳される場合に顕著な改善
 - 男性形のほうがよく出現し、またベースラインでは男性形に訳される傾向があるため

本当に照応解析ができているのか？

- Context-aware NMTシステムは、本当に照応解析ができているのか、それとも単に何らかの単純なヒューリスティックルールを発見しているだけなのか？
 - 文脈として文末の単語を選ぶ、など
- 先行詞が1つの名詞である文対を選び、システムが選んだ（アテンションで最も値が高い）先行詞との一致率を測る
 - 比較対象
 - random: 文脈からランダムに名詞を選択
 - first: 文脈の最初の名詞を選択
 - last: 文脈の最後の名詞を選択

Context-aware NMTシステムが選んだ先行詞の一致率 (正解率)

pronoun	agreement (in %)			
	random	first	last	attention
it	69	66	72	69
you	76	85	71	80
I	74	81	73	78

- “it”では最後の名詞、“you”, “I”では最初の名詞を先行詞に選ぶと最も一致率が高い
 - ほとんどの文脈が名詞を1つしか含まず、必ず一致することになる影響が大きい

pronoun	agreement (in %)			
	random	first	last	attention
it	40	36	52	58
you	42	63	29	67
I	39	56	35	62

- 名詞を2つ以上含む文脈に限定した場合、システムの一一致率が最も高い
 - 先行詞をうまく選択できた
→ **照応解析ができています**

人手で選択した先行詞による評価

- ここまでの実験ではCoreNLPで求めた先行詞を用いていた
- “it”を含む500文対に対して人手で先行詞を特定し、CoreNLPが出力した先行詞との一致率と本システムによる先行詞の一致率を比較

	agreement (in %)
CoreNLP	77
attention	72
last noun	54

	CoreNLP	
	right	wrong
attn right	53	19
attn wrong	24	4

- CoreNLPと比較して遜色ない性能を示す
- CoreNLPとの混合行列から、誤りの傾向が異なることが示された

まとめ

- 文脈を考慮したニューラル機械翻訳(Context-aware NMT)の有効性を示した
 - En-Ru BLEU: 29.46 → 30.14 (+0.68)
- 代名詞の曖昧性解決に貢献することを示した
- Context-aware NMTが照応関係を暗黙的に捉えていることを実験的に示した

