

P3-20 BERTを用いた機械翻訳の自動評価

嶋中宏希†

梶原智之‡

小町守†

shimanaka-hiroki@ed.tmu.ac.jp

†: 首都大学東京

‡: 大阪大学

1. 研究の背景と概要

- WMT-2017で最高性能を達成した RUSE [Shimanaka et al., 2018] は、事前学習された文の分散表現が機械翻訳の自動評価に有効であることを示している。
- 近年、文単位の表現学習モデルであるBERT [Devlin et al., 2018] が多くの応用タスクで最高性能を更新し、注目を集めている。
- 本研究では、この BERT を用いて機械翻訳の自動評価を行い、詳細な分析を行った。
- 実験の結果、BERT は文単位の全ての to-English 言語対で、**最高性能を大幅に更新し、機械翻訳自動評価への応用も可能であることを示した。**
- 詳細な分析の結果、RUSE との主な**3つの相違点**が、それぞれ BERT の性能改善に貢献していることが明らかになった。

2. 先行研究 RUSE

- ◆ 大規模コーパスで事前学習された文の分散表現から素性を抽出し、人手評価値データで訓練を行う。
- 特に、**隣接文推定**で事前学習された文の分散表現により高い性能を示している。

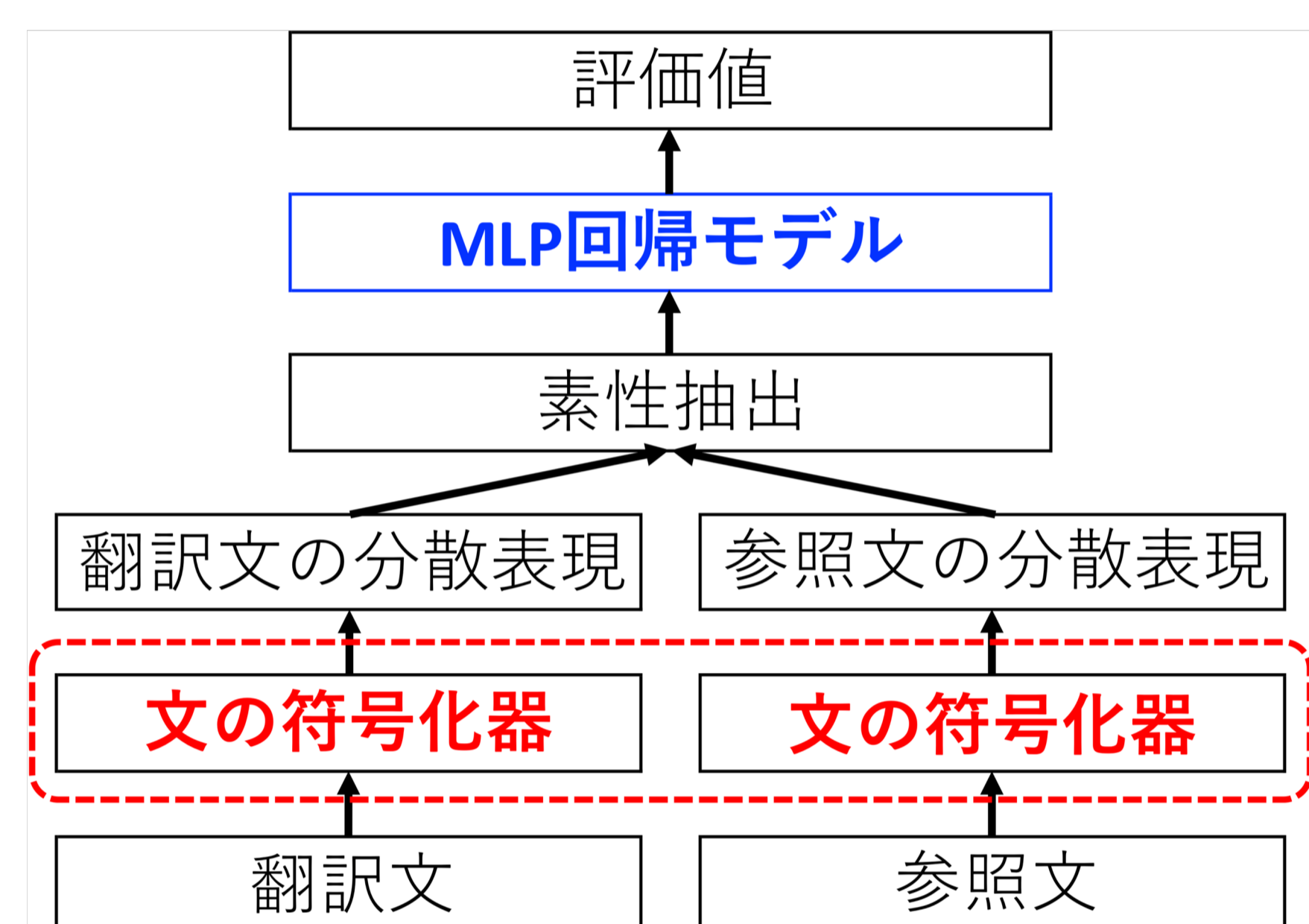


図1: RUSEによる機械翻訳の自動評価

3. 提案手法 BERTを用いた機械翻訳自動評価

- ◆ BERT も大規模コーパスで事前学習を行った後に、人手評価値データで訓練を行う。RUSE との主な相違点は以下の3つである。
- **事前学習の方法** ... 事前学習の際、隣接文推定に加え、双方向言語モデルの学習も行う。
- **文対モデリング** ... 図3のように、文対を同時に符号化する。
- **符号化器の再訓練** ... 回帰モデルの訓練だけでなく符号化器の再訓練も行う。

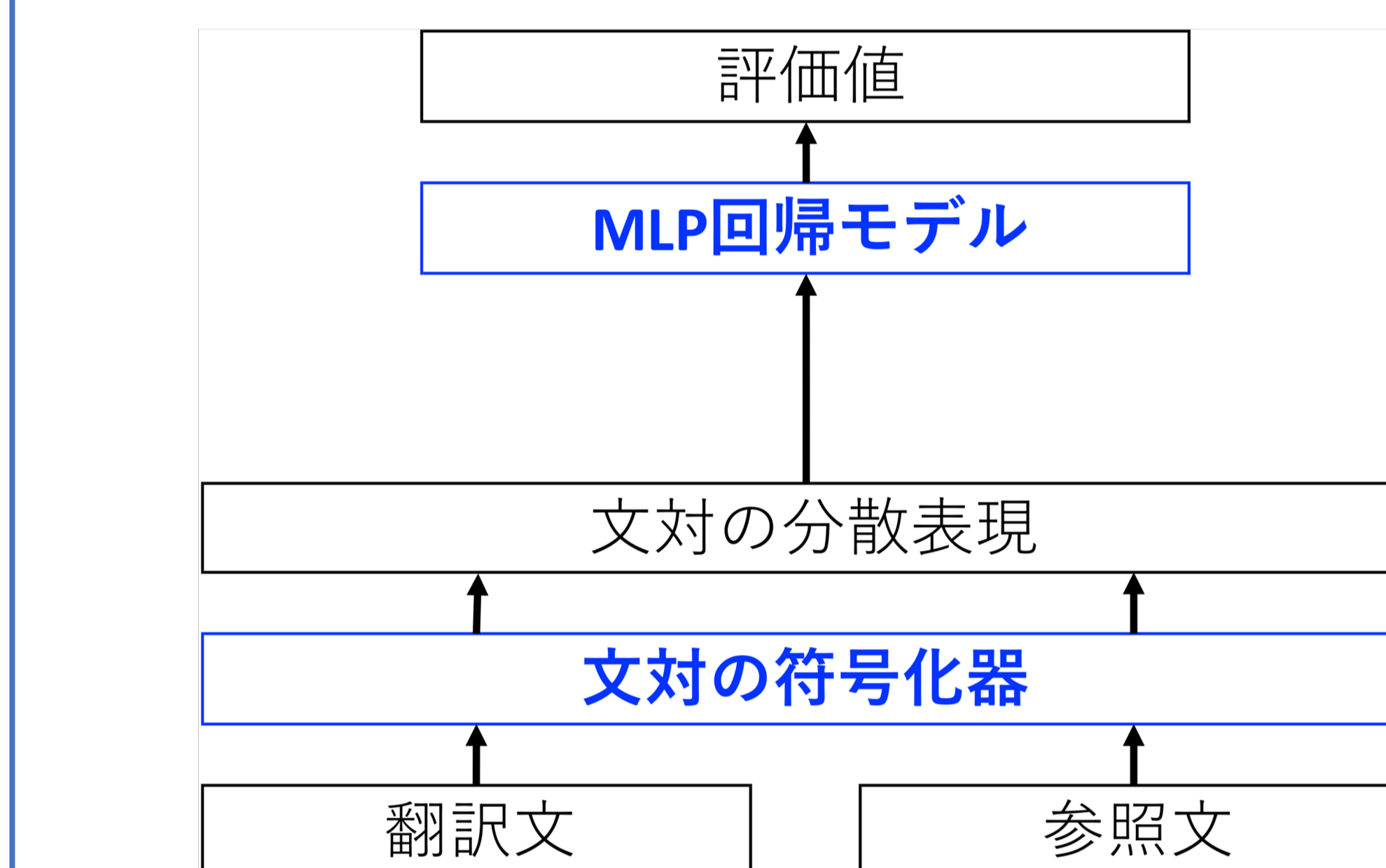


図2: BERTによる機械翻訳の自動評価

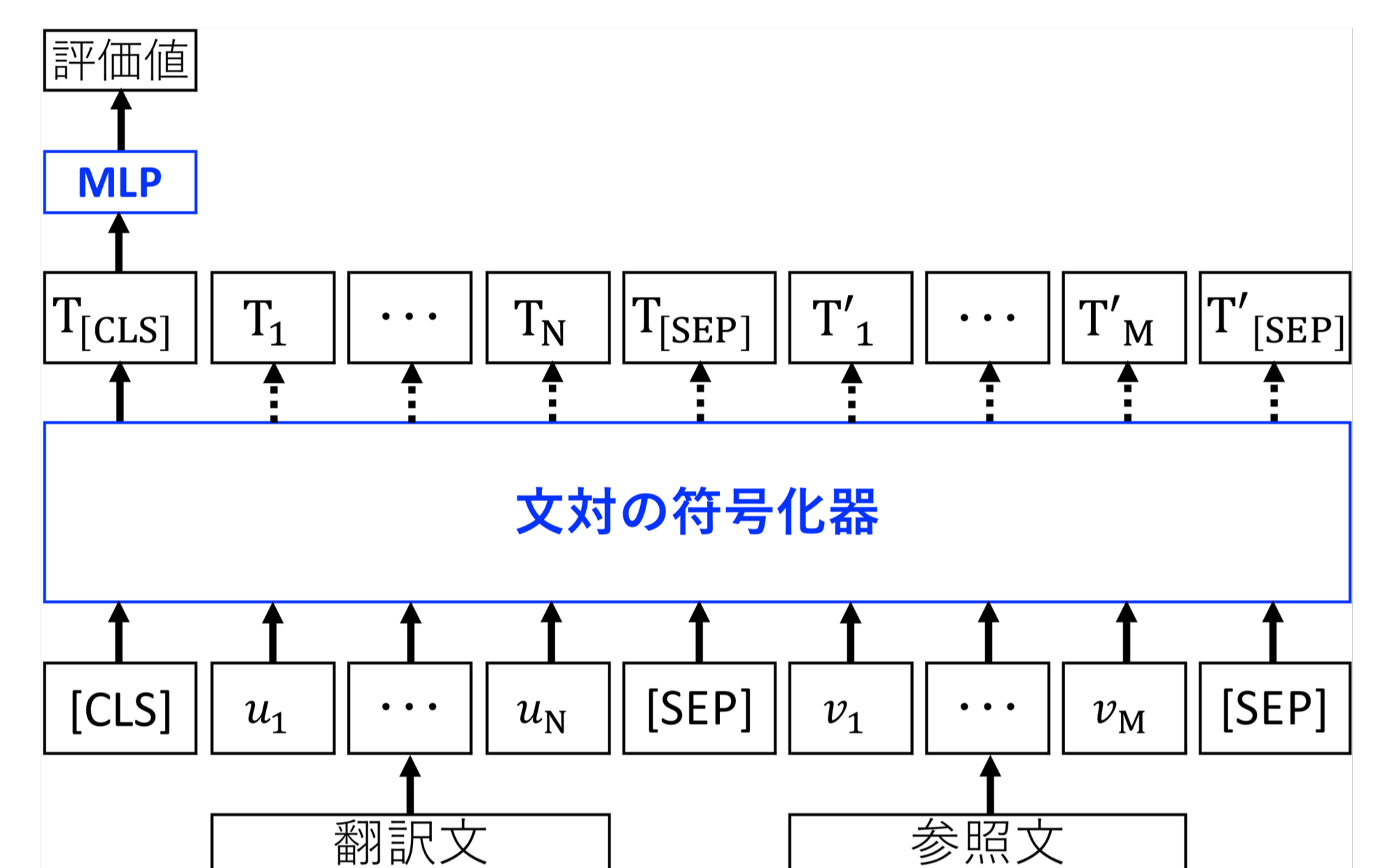


図3: BERTの文対モデリング

3. 実験設定

◆ 人手評価値の学習データセット

- 出力文と参照分に対して人で評価値が付与
- WMT15, 16を5,360文の9割を訓練用, 1割を開発用に分割し, WMT17の各言語対 560文に対して評価。

表1: WMTの人手評価値付きの文対数 (英語方向)

	cs	de	fi	lv	ro	ru	tr	zh
WMT15	500	500	500	-	-	500	-	-
WMT16	560	560	560	-	560	560	560	-
WMT17	560	560	560	560	-	560	560	560

◆ BERT_{BASE}

- BookCorpus (800M)と English Wikipedia (2,500M) で事前学習

4. 実験結果

- BERTを用いた機械翻訳の自動評価を行った結果、WMT18で最高性能を示したRUSEを大幅に上回り、**SoTAを更新した。**
- ➔ BERTの機械翻訳自動評価への応用も可能である

表2: WMT17における文単位での人手評価とのピアソンの相関係数

	cs-en	de-en	fi-en	lv-en	ru-en	tr-en	zh-en	Avg.
SentBLEU	0.435	0.432	0.571	0.393	0.484	0.538	0.512	0.481
Blend (WMT17で最高性能)	0.594	0.571	0.733	0.577	0.622	0.671	0.661	0.633
RUSE (WMT18で最高性能)	0.614	0.637	0.756	0.705	0.680	0.704	0.677	0.682
BERT _{BASE}	0.732	0.751	0.856	0.829	0.795	0.811	0.763	0.791

5. 分析

- ◆ 異なる素性を用いたRUSEとBERTを比較
- **RUSE with GloVe-BoW** ... 単語分散表現 GloVe の平均ベクトルを用いる。
- **RUSE with Quick-Thought** ... 隣接文推定によって事前学習されたQuick-Thoughtを用いる。
- **RUSE with BERT_{BASE} (文)** ... 双方向言語モデルと隣接文推定によって事前学習された単一文入力のBERTを用いる。
- **RUSE with BERT_{BASE} (文対)** ... 文対を入力とするBERTの出力を用いる。
- BERTの性能改善には、**事前学習の方法**、**文対モデリング**、**符号化器の再訓練**の3点が大きく貢献していると考えられる。

- 赤→青: 素性を単語の分散表現から文の分散表現に変更
- 青→橙: 事前学習に双方向言語モデルを追加
- 橙→黄: 文の符号化器から文対の符号化器へ変更
- 黄→緑: 符号化器の再訓練も行うように変更

