多言語文符号化器の言語表現と意味表現の分離に基づく機械翻訳の品質推定

黒田勇斗*　梶原智之**　荒瀬由紀***　二宮崇**

*愛媛大学工学部　**愛媛大学大学院理工学研究科
***大阪大学大学院情報科学研究科

背景

- 研究・開発の場では、BLEU[Papineni+ 2002]など参照訳を用いた自動評価が行われている

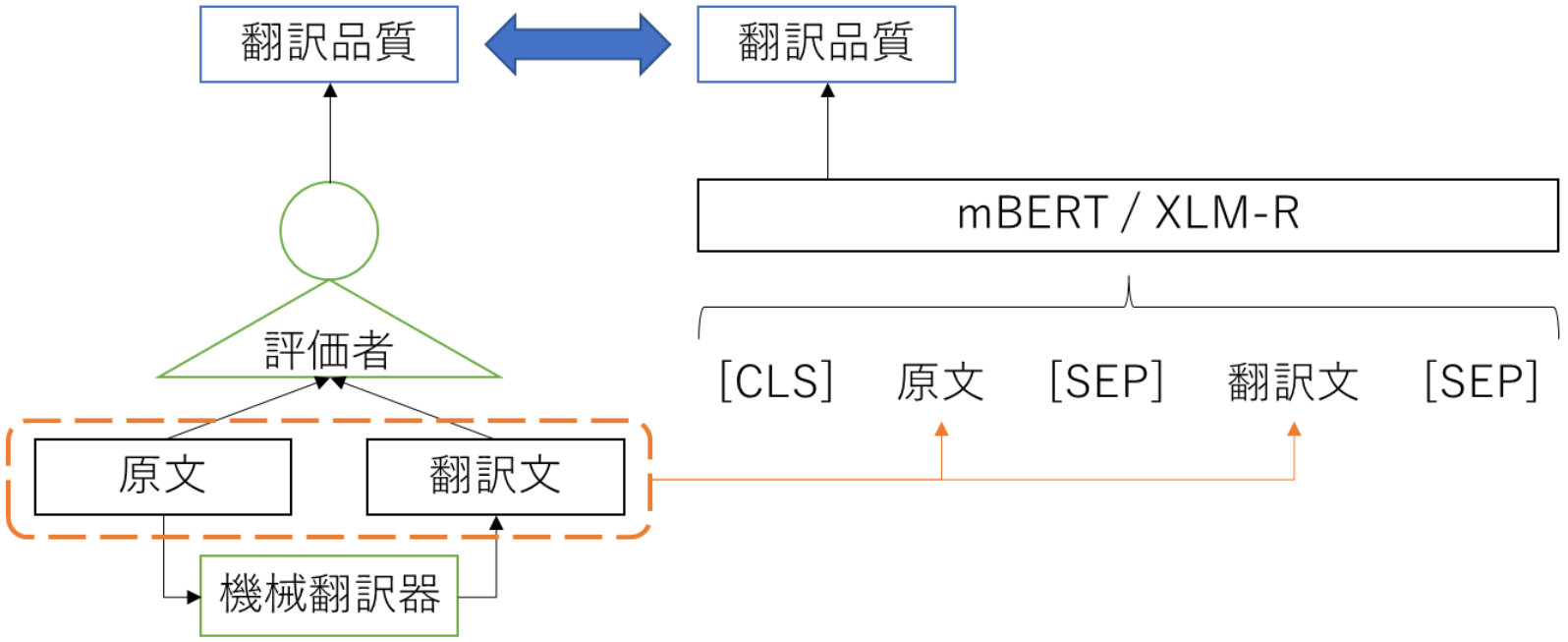
→機械翻訳システムの使用者は参照訳を用意できない

- 機械翻訳の利用促進のために参照訳を用いない品質推定を行いたい

[Papineni+ 2002] Papineni et al. BLEU: a method for automatic evaluation of machine translation, In Proc. of ACL, 2002

教師あり品質推定の課題

- 多言語文符号化器[Devlin+ 2019, Conneau+ 2020]を再訓練した品質推定モデルが多く提案されてきた

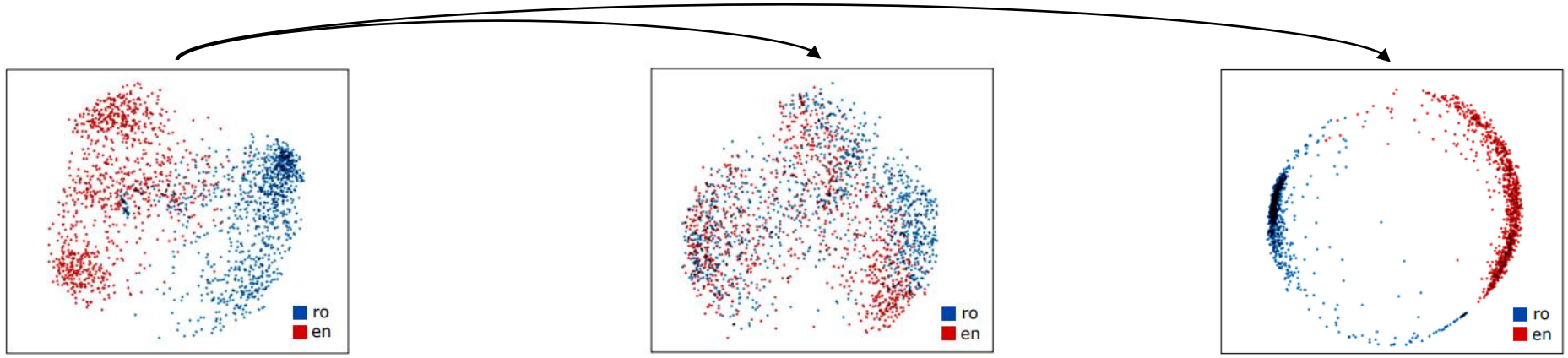


[Devlin+ 2019] Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proc. of NAACL, 2019

[Conneau+ 2020] Conneau et al. Unsupervised Cross-lingual Representation Learning at Scale. In Proc. of ACL, 2020

教師なし品質推定の利点

- 対訳コーパスのみで訓練できるため100以上の言語で利用できる
  - LaBSE [Feng+ 2020]
    - 対訳文間の余弦類似度が高くなるように再訓練
  - DREAM [Tiyajamorn+ 2021]
    - LaBSEの文表現を意味表現と言語表現に分離



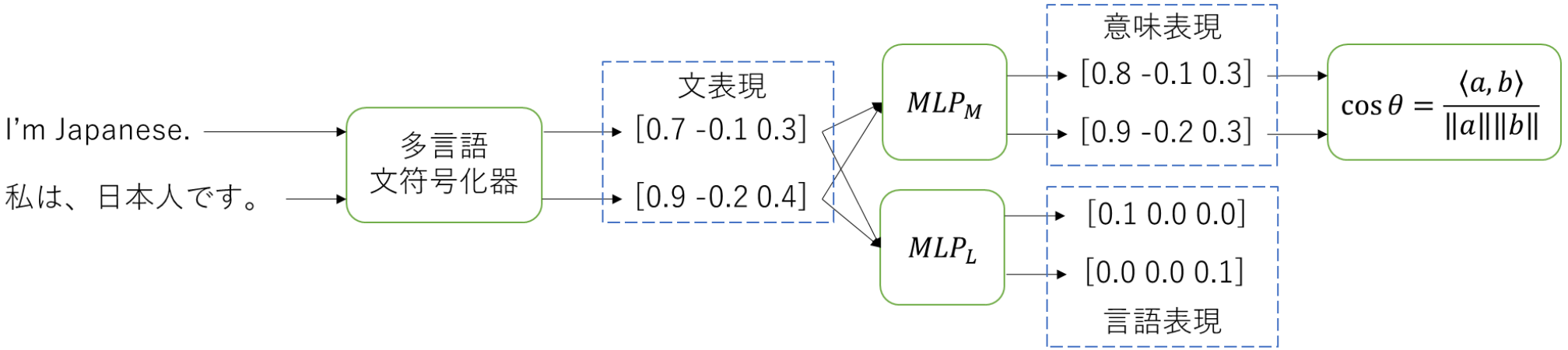
文表現　　意味表現　　言語表現

[Feng+ 2020] Feng et al. Language-agnostic BERT Sentence Embedding, In arXiv:2007.01852. 2020.

[Tiyajamorn+ 2021] Tiyajamorn et al. Language-Agnostic Representation from Multilingual Sentence Encoders for Cross-Lingual Similarity Estimation, In Proc. of EMNLP, 2021

提案手法の概要

- 多言語文符号化器の文表現から意味表現を抽出
  - 文表現を言語表現と意味表現に分離する
  - 対訳文のみを用いたマルチタスク学習
- DREAMとの相違点
  - 敵対的訓練により意味表現に言語固有の情報がないことを保証する



提案手法の詳細

(4) 敵対的損失：L_A

意味

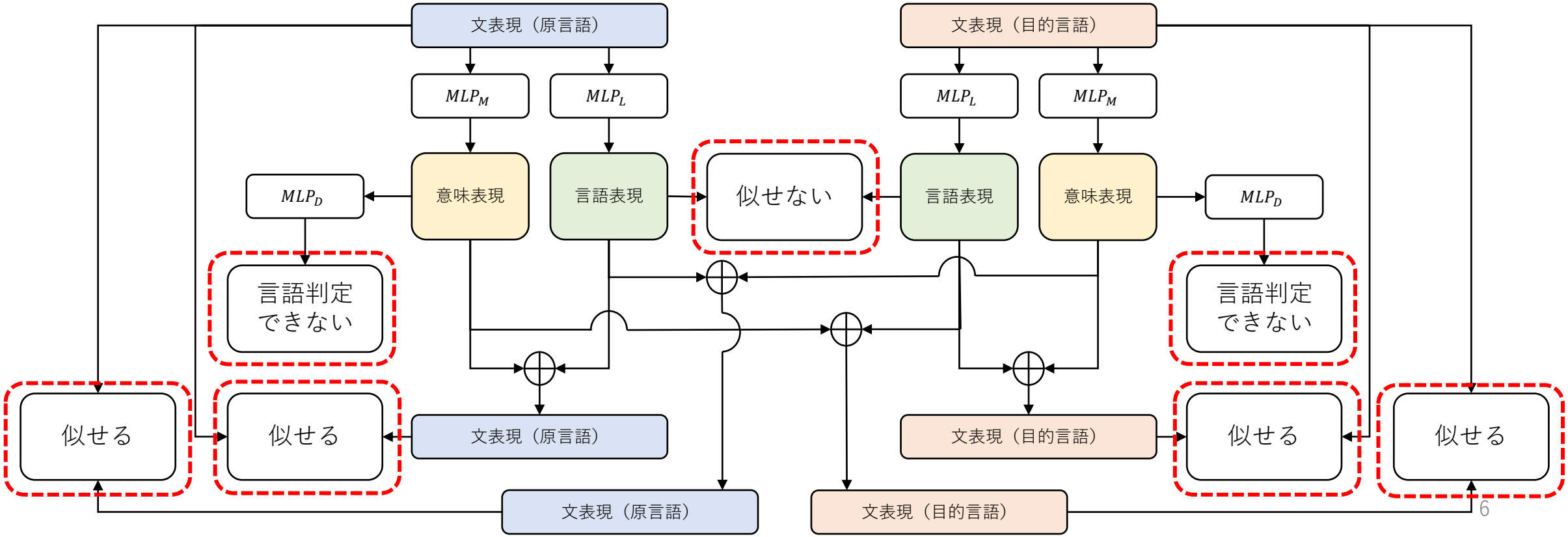
(1) 復元損失：L_R

言語表現と意味表現から元の文表現を復元できる

$L = L_R + L_C + L_A$

$\hat{e_M}, \hat{e_L} = MLP_M(e), MLP_L(e)$

$L_R = 1 - cos(e, (\hat{e_M} + \hat{e_L}))$



実験

- データ：WMT2020品質推定タスク [Specia+ 2020]
- 方法：意味表現を用いて余弦類似度を計算
- 評価指標：人手評価値とモデル評価値のPearson相関

学習データ

	言語対	対訳文数
多資源言語対	en-de	1,000,000
	en-zh	1,000,000
中資源言語対	ro-en	200,000
	et-en	200,000
少資源言語対	ne-en	50,000
	si-en	50,000

原言語文	It inhabits the Atlantic, Indian, and Pacific Oceans and the Mediterranean Sea.
目的言語文	它居住在大西洋、印度洋、太平洋和地中海。
人手評価値	70

[Specia+ 2020] Specia et al. Findings of the WMT 2020 Shared Task on Quality Estimation, In Proc. of EMNLP, 2020

比較手法

- 多言語文符号化器による教師なし品質推定
  - ベースライン：LaBSE [Feng+ 2020]
  - DREAM[Tiyajamorn+ 2021]：先行研究で抽出した意味表現
  - 提案手法：提案手法で抽出した意味表現
- Enc-Decモデルによる教師なし品質推定
  - D-TP[Fomicheva+ 2020]：評価対象の機械翻訳器のパラメータが必要
  - Prism[Thompson+ 2020] ：大量の対訳コーパスが必要

[Feng+ 2020] Feng et al. Language-agnostic BERT Sentence Embedding, In arXiv:2007.01852. 2020.

[Tiyajamorn+ 2021] Tiyajamorn et al. Language-Agnostic Representation from Multilingual Sentence Encoders for Cross-Lingual Similarity Estimation, In Proc. of EMNLP, 2021

[Fomicheva+ 2020] Fomicheva et al. Unsupervised Quality Estimation for Neural Machine Translation. In TACL, 2020

[Thompson+ 2020] Thompson et al. Automatic Machine Translation Evaluation in Many Languages via Zero-Shot Paraphrasing. In Proc. of EMNLP, 2020

実験結果

- LaBSEの文表現をそのまま用いるよりも、
  意味表現を用いることで結果が向上した
- DREAMよりも良い意味表現を抽出できている

	多資源言語対		中資源言語対		少資源言語対		
	en-de	en-zh	ro-en	et-en	ne-en	si-en	Avg.
ベースライン	0.084	0.036	0.705	0.550	0.545	0.455	0.396
DREAM	0.151	0.156	0.711	0.549	0.627	0.552	0.458
提案手法	**0.216**	**0.222**	**0.718**	**0.587**	**0.634**	**0.571**	**0.491**
D-TP	0.259	0.321	0.693	0.642	0.558	0.460	0.489
Prism	0.464	0.303	0.829	0.694	-	-	0.573

分析

- (g)および(h)より、L_RおよびL_Cの影響は小さい
- (d)より、$L_R + L_A$でも有効
- (f)より、L_Lの影響は大きい

	L_R 復元損失	L_c 交差復元損失	L_L 言語表現損失	L_A 敵対的損失	Avg.
LaBSE					0.396
提案手法	✓	✓	✓	✓	**0.491**
(a)	✓				0.390
(b)	✓	✓			0.072
(c)	✓		✓		0.082
(d)	✓			✓	0.434
(e)	✓	✓	✓		0.439
(f)	✓	✓		✓	0.327
(g)	✓		✓	✓	0.483
(h)		✓	✓	✓	0.488

まとめ：多言語文符号化器の言語表現と意味表現の分離に基づく機械翻訳の品質推定

背景：参照訳を用いない品質推定により機械翻訳を利用促進したい

課題：意味表現に言語固有の情報が含まれうる

手法：対訳文のみを用いた、意味表現と言語表現を分離するための4つの損失関数を提案した

- 復元損失
- 交差復元損失
- 言語表現損失
- 敵対的損失

結果：人手評価との相関が向上

少資源言語対においては教師なし品質推定の最高性能を達成