機械翻訳のための言い換え生成による前編集の検討

惟高日向・岩本裕司・梶原智之・二宮崇（愛媛大学）・藤田篤（NICT）

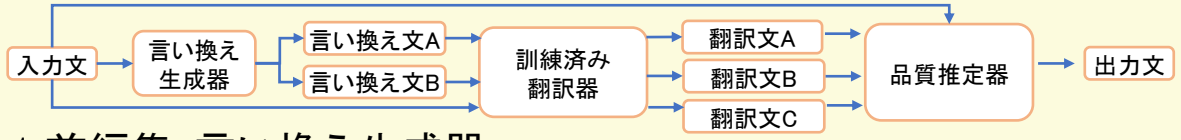
1. 背景

訓練済み翻訳器には得意な表現と不得意な表現がある

入力文	Google翻訳
私は愛媛大学を受けます。	I will attend Ehime University.
私は愛媛大学を受験します。	I will take the Ehime University entrance exam.

2. 提案手法

- 提案手法：訓練済み翻訳器を用いる際の外側の工夫
- 前編集：入力文を翻訳器が得意な表現に言い換えれば、より良い出力が得られる
- リランキング：出力文の品質推定を行い、より良い出力文を選択する



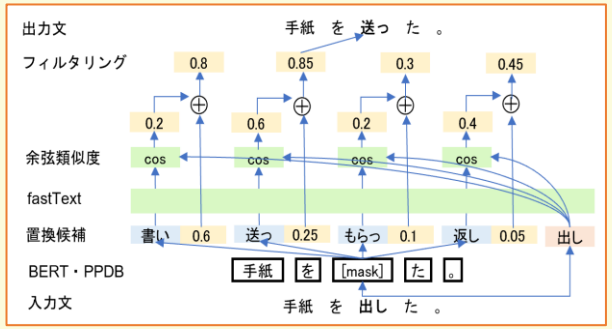
2.1 前編集：言い換え生成器

- 東北大BERTとPPDBによる候補生成
  - 入力文の単語数×10種類の候補生成

PPDB：ParaPhrase DataBase（https://ahcweb01.naist.jp/old/resource/jppdb/）

- 言い換え生成の流れ

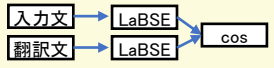
1. 入力文のMASKした部分をBERTまたはPPDBを用いて置換
2. 各置換候補単語を、MASKした元の単語との余弦類似度（fastText）とBERTの生成確率（PPDBは言い換え確率）との和でスコアリングする
3. スコアの上位k個の置換単語を元の単語と置換した文を言い換え文として使用する（k=10）



2.2 リランキング：品質推定器

- 品質推定器：(k+1)種類の翻訳文をリランキング

①LaBSE：cos(LaBSE(入力文), LaBSE(翻訳文))の高い順
②mBART：mBARTによる forced decoding 確率の高い順※
（品質推定器としては訓練済み言語モデルをそのまま使用）



LaBSEによるスコアリング

mBART

$$Q(y|x) = \frac{1}{|y|}\sum_{t=1}^{|y|} logp(y_t|y_{i<t}, x)$$

mBARTによるスコアリング

- 品質推定器の性能について
  - 中～多資源言語対では、mBARTの方が高性能（本実験はこの設定）

※ 西原+ (2022) 多言語雑音除去自己符号化器による教師なし品質推定 https://doi.org/10.5715/jnlp.29.669

	多資源言語対		中資源言語対		小資源言語対		
	En-De	En-Zh	Ro-En	Et-En	Ne-En	Si-En	Avg.
LaBSE	0.08	0.04	0.71	0.55	**0.55**	0.46	0.40
mBART	**0.28**	**0.32**	**0.82**	**0.70**	0.54	**0.47**	**0.52**

WMT20における機械翻訳の品質推定（人手評価とのピアソン相関）

3. 実験結果：日→英の機械翻訳におけるBLEU

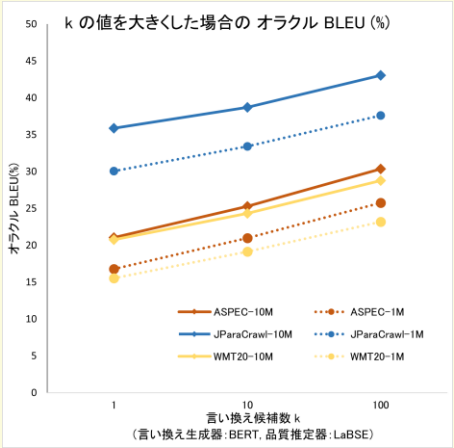
- データ

訓練用：JParaCrawl (ver.3.0)		1,000 万文対 または 100万文対
評価用：ASPEC	（学術論文）	1,812 文対
評価用：WMT20	（ニュース）	993 文対
評価用：JParaCrawl		2,000 文対

- 比較手法
  - リランキングのみ
    言い換えなしで翻訳器からビームサーチを用いて10個の翻訳文を出力し、リランキング後のトップ1を評価したもの
  - オラクル
    (k+1)種類の翻訳文の中からsentence BLEUが最も高いものを選ぶことができた場合の結果

(1,000万文対)	ASPEC				WMT20				JParaCrawl			
言い換え生成器	BERT	PPDB	BERT	PPDB	BERT	PPDB	BERT	PPDB	BERT	PPDB	BERT	PPDB
品質推定器	LaBSE		mBART		LaBSE		mBART		LaBSE		mBART	
前編集なし	19.78				19.49				35.05			
リランキングのみ	19.88		19.85		19.31		19.61		32.56*		33.86*	
前編集+リランキング	19.86	19.79	20.04	20.15*	19.11*	19.32	19.76	19.68	32.90*	32.86*	33.85*	33.72*
オラクル（k=10）	25.29	25.85	25.09	25.84	24.34	24.53	24.44	24.53	38.72	39.25	38.72	39.27

(100万文対)	ASPEC				WMT20				JParaCrawl			
言い換え生成器	BERT	PPDB	BERT	PPDB	BERT	PPDB	BERT	PPDB	BERT	PPDB	BERT	PPDB
品質推定器	LaBSE		mBART		LaBSE		mBART		LaBSE		mBART	
前編集なし	15.76				13.99				29.12			
リランキングのみ	16.17*		16.48*		14.21		14.68*		27.92*		28.93	
前編集+リランキング	15.56	15.65	16.42*	16.26*	14.60*	14.27	15.02*	14.77*	28.29*	27.73*	29.17	28.88
オラクル（k=10）	20.99	21.29	20.76	21.07	19.16	19.16	19.16	18.89	33.42	33.62	33.42	33.58

※青は改善、赤は改悪、*は有意差検定を行いp値が0.05未満のもの

- 分かったこと
  - 言い換え生成器による前編集は効果的
  - 良いリランキング手法で一貫してBLEUが向上
  - オラクルの結果から良い文を選択することができれば、BLEUは向上
- 今後の予定
  - 言い換え生成器を訓練済み翻訳器に特化させる
  - 他のリランキング手法を検討（教師ありQEの利用など）
  - 効果のある言い換えの分析