

# A Deeper Exploration of the Standard PB-SMT Approach to Text Simplification and its Evaluation

Sanja Stajner and Hannah Bechara and Horacio Saggion

# 紹介文献

- title: A Deeper Exploration of the Standard PB-SMT Approach to Text Simplification and its Evaluation
- author: Stajner, Sanja and B'echara, Hannah and Saggion, Horacio
- journal: Volume 2: Short Papers
- pages: 823
- month: July
- year: 2015
- publisher: Association for Computational Linguistics
- pages: 594--598

# 概要

- 句に基づく統計的機械翻訳をつかったテキスト・シンプリフィケーション
- 学習データサイズによる影響
- 学習データの質の違いによる効果比較

# 機械翻訳による平易化の研究

- Specia, 2010
  - Translating from complex to simplified sentences.
- Coster and Kauchak, 2011
  - Learning to Simplify Sentences Using Wikipedia
- Stajner, 2014
  - Translating sentences from 'original' to 'simplified' spanish.

# Translating from complex to simplified sentences

- ブラジリアン・ポルトガル語の簡易化翻訳
- 学習コーパスの簡易文の特徴
  - 語彙言い換え、文分割
- 結果:
  - 60.75

# Learning to Simplify Sentences Using Wikipedia

- English Wikipedia と Simple English Wikipedia
- 翻訳モデルに”削減”を追加した新モデル
- このモデルを使っただけの結果は60.46
- 削減なしモデルで59.87
- Simplificationしないモデルで59.37
- 上記3つの結果より多くの文は不変のまま出力されている

# Translating Sentences from ‘original’ to ‘simplified’ Spanish

- スペイン語 Text Simplification コーパス
  - 異なるレベルの平易化が行われている
- 1,000文未満のサイズで構築
- データセットサイズより質に大きく依存

# 学習データ構築

- David Kauchuk による Wikipedia アライメント・データ
- 文対を質、量にもとづいて学習データの分類
- 質分類:
  - オリジナル文と参照訳文との分単位BLEU値(S-BLEU)による分類
  - 区間範囲: (0 - 0.3], (0.3 - 0.4], (0.4 - 0.5]... (0.9, 1]
- 量分類
  - 2,000, 4,000, 6,000, 8,000 & 10,000 文対
- これらの組み合わせ40の組み合わせができる



# S-BLEU

- 0.87

- Austen was not well known in Russia and the first Russian translation of an Austen novel did not appear until 1967.
- Austen was not well known in Russia. The first Russian translation of an Austen novel did not appear until 1967.

- 0.08

- In women, the larger mammary glands within the breast produce the milk.
- The breast contains mammary glands.

# テストセット

- テストセット
- WikiTest
  - 各区間から30文よりなる240文対のデータセット
- EncBritTest
  - Encyclopedia Britannica からの引用文と子供向けに書かれた文
  - 601文対のデータセット

テストセットのBLEU値

Test set	Size	BLEU
WikiTest	240	62.27
EncBritTest	601	12.40

# WikiTestによる翻訳結果

Table 3: BLEU scores on the WikiTest dataset

S-BLEU	Size of the training set				
	2,000	4,000	6,000	8,000	10,000
[0, 0.3]	56.38	56.38	56.15	57.75	57.89
(0.3, 0.4]	60.89	61.35	61.76	61.52	61.37
(0.4, 0.5]	61.27	61.36	61.74	61.55	<b>62.11</b>
(0.5, 0.6]	60.96	61.30	61.52	61.77	61.98
(0.6, 0.7]	60.96	61.30	61.60	61.69	61.80
(0.7, 0.8]	61.56	61.38	61.67	61.77	61.89
(0.8, 0.9]	61.54	61.49	61.51	61.57	61.61
(0.9, 1]	61.57	61.57	61.59	61.55	61.55

# EncBritTestによる翻訳結果

Table 4: BLEU scores on the EncBritTest dataset

S-BLEU	Size of the training set				
	2,000	4,000	6,000	8,000	10,000
[0, 0.3]	13.84	13.84	13.87	13.68	13.59
(0.3, 0.4]	14.05	13.95	14.08	14.06	14.01
(0.4, 0.5]	14.02	14.09	14.17	14.15	14.12
(0.5, 0.6]	14.09	14.22	14.27	14.16	14.13
(0.6, 0.7]	14.25	14.30	14.35	14.35	14.32
(0.7, 0.8]	14.30	14.29	14.30	14.30	14.28
(0.8, 0.9]	14.38	14.40	14.40	14.40	<b>14.41</b>
(0.9, 1]	12.71	12.52	12.46	12.39	12.54

- 学習データサイズによる大きな変化見られない
- 統計的有意性もみられない(t検定)
- データサイズの質区間端では低値をだしている

# 主観評価

Table 7: Results of the human evaluation

System	G	M	S
Original	4.85	/	2.60
S-03-200	4.03	3.95	2.57
S-03-1000	4.20	4.03	<b>2.85</b>
S-06-200	<b>4.50</b>	4.45	2.68
S-06-1000	4.43	<b>4.48</b>	2.72
S-10-200	3.25	2.92	2.45
S-10-1000	2.92	2.95	2.53

- 選定した6個を以下の3つク  
ライテリアを5段階評価
  - G: 文法
  - M: 意味保持
  - S: 簡易性
- 評価者: 3人
- 質的にはG、Mでは0.6の  
BLEU値付近
- 量的違いはない

# まとめ

- 今までの機械翻訳によるテキスト平易化の研究では成功・失敗の理由について言及されてこなかった
- 学習データを質と量に分類し、40種類のシステムを構築した
- 結果、量より質ということがわかった