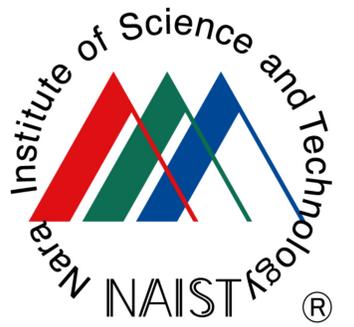


単語分散表現に基づいた誤差によるニューラル機械翻訳の学習

帖佐 克己, 須藤 克仁, 中村 哲 (NAIST)

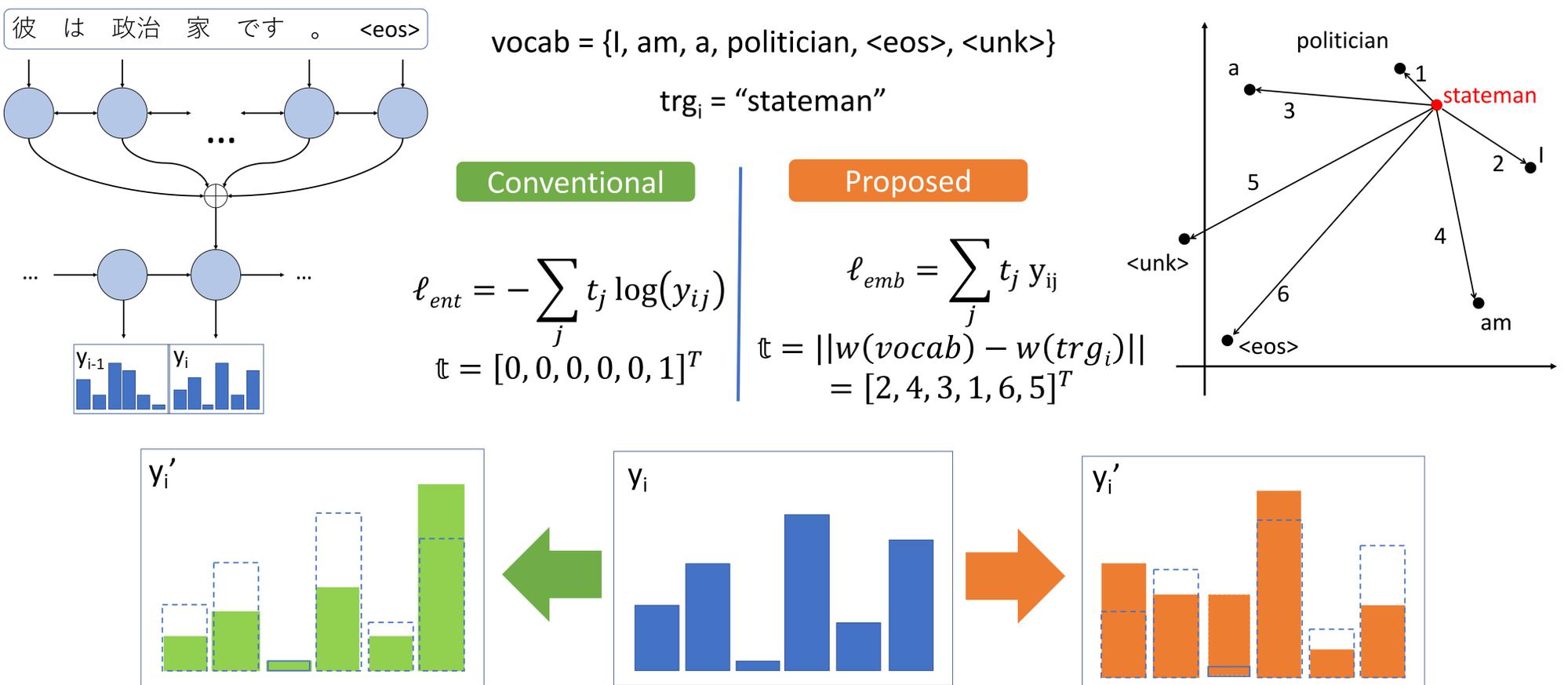
E-mail: k-chousa@is.naist.jp



Abstract

- 語彙サイズを小さくすると Out-of-Vocabulary (OOV) になる単語が増えて翻訳精度が著しく減少してしまう問題
 - Softmax Cross-Entropy が **クラスごとに独立に**ペナルティを与えていることが原因
 - 特に正解単語がOOVの場合 **すべての単語に一律に**ペナルティがかかってしまう
- 正解単語との **意味空間上での距離に応じて**各単語への **誤差を平滑化**する手法を提案
 - 正解単語と意味が近いものには小さなペナルティを 遠いものには大きなペナルティを与える
 - 語彙内の各単語と正解単語との間の距離を計算し **生成確率による加重平均**を誤差として使用する
 - 正解単語がOOVでも **語彙中の似た意味の単語への誤差が小さくなる**ため、それらの単語で **言い換えられやすくなる**
- Ja-En (ASPEC)とEn-Fr (IWSLT17)の2つのコーパスにおける実験で提案手法が有効であることを確認
 - 特に **語彙サイズが小さい**場合には **同義語などを考慮した評価尺度で精度が大きく向上**

Proposed Method – Word Embedding Loss



Experiments

実験設定 コーパス: ASPEC(Ja-EN) / IWSLT17(En-Fr)
 モデル: Attention EncDec (Luong et al., 2015)
 ソース側 語彙: 20000 words
 評価尺度: BLEU / METEOR
 単語分散表現: word2vec
 English: Google News dataset (pre-trained)
 French: Wikipedia (train with gensim)

実験結果

- 誤差関数の組み合わせ方 (on ASPEC 10000 words)
 ℓ_{emb} を使用したケース全てにおいて精度が向上し、語彙サイズを倍にした場合の精度と同等のBLEUスコアが得られた。特に ℓ_{ent} で事前学習をして ℓ_{emb} で最適化する方法が良い結果を示した。この結果から提案手法は、**比較的小さい語彙の状況において有効**であることが考えられる。
- 少ない語彙での有効性 (on ASPEC 1000 words)
 語彙が限られた状況でも精度の向上が確認でき、**METEORが +1.72 points**と大幅に向上したケースも見られた。この結果より提案手法は、**非常に語彙サイズが限られた状況においても有効**であり、語彙に無い表現を **同義語などに言い換えられている**ことが分かる。
- 言語対による影響 (on ASPEC&IWSLT 10000 words)
 Ja-Enと同じ設定でEn-Frにおける実験を行ったところ、同様な傾向での精度の改善が見られた。この結果より提案手法は**特定の言語対に依存したものでない**と考えられる。

Results about ASPEC by BLEU/METEOR

target vocab.	loss	pre-train	BLEU	METEOR
20,000	ℓ_{ent}	None	24.91	30.71
10,000	ℓ_{ent}	None	23.78	29.39
	$\ell_{ent} + \ell_{emb}$	None	24.75 (+0.97)	29.93 (+0.54)
	$\ell_{ent} + \ell_{emb}$	ℓ_{ent}	24.60 (+0.82)	29.52 (+0.13)
	ℓ_{emb}	ℓ_{ent}	24.85 (+1.07)	29.81 (+0.41)
1,000	ℓ_{ent}	None	14.21	18.43
	$\ell_{ent} + \ell_{emb}$	None	14.35 (+0.14)	18.66 (+0.23)
	$\ell_{ent} + \ell_{emb}$	ℓ_{ent}	14.72 (+0.51)	18.88 (+0.45)
	ℓ_{emb}	ℓ_{ent}	14.74 (+0.53)	20.15 (+1.72)

Results about IWSLT17 by BLEU/METEOR

target vocab.	loss	pre-train	BLEU	METEOR
10,000	ℓ_{ent}	None	33.89	56.37
	$\ell_{ent} + \ell_{emb}$	None	33.94 (+0.05)	57.20 (+0.83)
	$\ell_{ent} + \ell_{emb}$	ℓ_{ent}	35.46 (+1.57)	58.35 (+1.98)
	ℓ_{emb}	ℓ_{ent}	35.60 (+1.72)	58.35 (+1.99)

Example in ASPEC

source (Ja):	リザーバ内流動パターンと押出物形態。
target (En):	The flow pattern in the reservoir and the shape of <unk:moldings>.
baseline:	Flow patterns in reservoir and <unk> forms.
proposed:	The flow pattern in the reservoir and the extrusion form.