

階層的な部分単語を入力としたニューラル機械翻訳



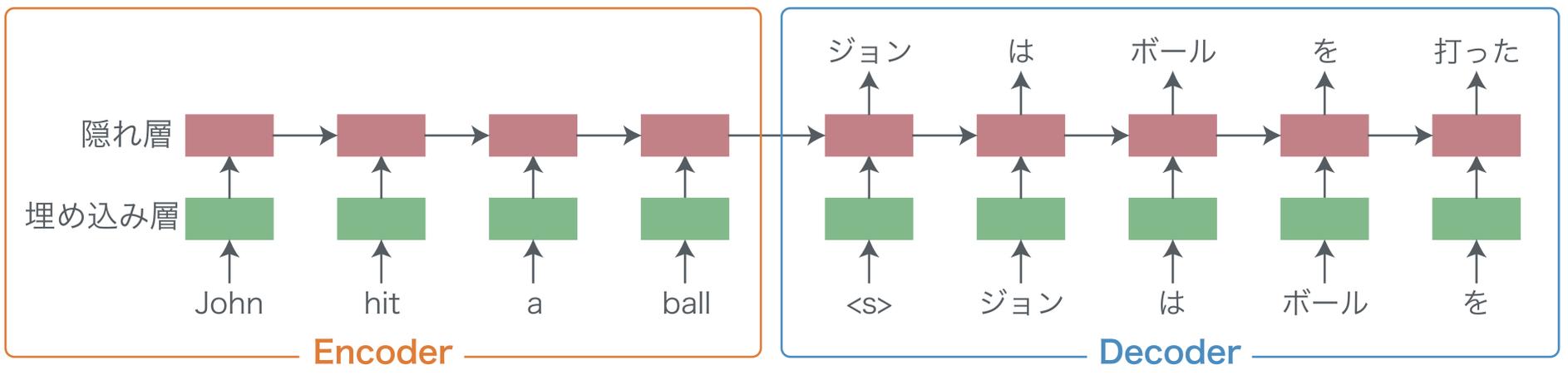
Innovative R&D by NTT

森下 睦, 鈴木 潤, 永田 昌明
NTT コミュニケーション科学基礎研究所

概要

- BPE には**階層的特性**が存在する
e.g. 1 万回マージした語彙には 100 回マージした際の語彙を包含している
- BPE の階層的特性を利用して**ニューラル機械翻訳の埋め込み層を改良**
- 提案法により, **翻訳精度の向上**が確認できた

ニューラル機械翻訳 (Neural Machine Translation, NMT)

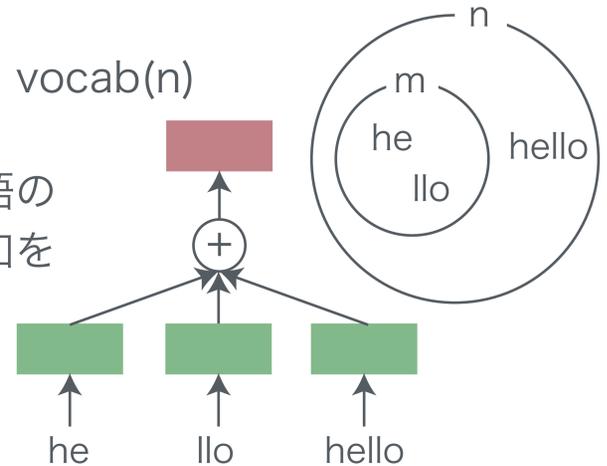


部分単語単位

- NMT では**低頻出語や未知語を扱いにくい**
→ 埋め込み層が学習しにくい
↓ **改良**
- **部分単語単位** (Byte Pair Encoding, BPE)
 - (1) 全ての単語を文字単位に分割
hello → h e l l o
 - (2) 頻出する 2 つの部分単語のペアを 1 つの部分単語とする
h e l l o → he l l o
 - (3) (2) を n 回繰り返す
 - (4) 語彙サイズ ≒ n になる

提案法：階層的部分単語の活用

- **BPE の階層的特性**
n 回マージしたときの語彙を vocab(n) とする
m < n のとき
vocab(m) ⊂ vocab(n)
- 複数の部分単語の埋め込み層の和を隠れ層に入力
- **語彙サイズを増やさずに精度向上**が見込める



実験

- TED トークコーパスを使用
(独→英, 英→独)

	単語数	文数
Train	3.2M	189.3K
tst2012	30.9K	1.7K
tst2013	21.0K	1.0K
tst2014	25.0K	1.3K

Enc		Dec		独→英			英→独		
単位	素性	単位	素性	tst2012	tst2013	tst2014	tst2012	tst2013	tst2014
16k	16k	16k	16k	31.70	34.38	28.81	26.00	28.67	24.05
16k	16k, 1k	16k	16k	31.78 (+0.08)	34.29 (-0.09)	29.17 (+0.36)	26.26 (+0.26)	28.85 (+0.18)	24.59 (+0.54)
16k	16k, 300	16k	16k	33.04 (+1.34)	34.90 (+0.52)	30.12 (+1.31)	26.72 (+0.72)	28.81 (+0.14)	24.29 (+0.24)
16k	16k, 1k, 300	16k	16k	33.22 (+1.52)	35.04 (+0.66)	30.81 (+2.00)	27.23 (+1.23)	29.47 (+0.80)	25.35 (+1.30)
16k	16k	16k	16k, 1k	31.59 (-0.11)	33.37 (-1.01)	28.87 (+0.06)	26.19 (+0.19)	28.81 (+0.14)	24.20 (+0.15)
16k	16k	16k	16k, 300	32.46 (+0.76)	34.43 (+0.05)	30.12 (+1.31)	26.13 (+0.13)	28.96 (+0.29)	24.21 (+0.16)
16k	16k	16k	16k, 1k, 300	31.47 (-0.23)	33.34 (-1.04)	29.10 (+0.29)	26.79 (+0.79)	28.78 (+0.11)	24.38 (+0.33)
16k	16k, 1k, 300	16k	16k, 1k, 300	32.86 (+1.16)	34.78 (+0.40)	30.38 (+1.57)	26.42 (+0.42)	29.19 (+0.52)	24.48 (+0.43)

- **Encoder 側**にのみ階層的素性を適用する手法が最も良い
- **Decoder 側との併用はあまりうまくいかない**