

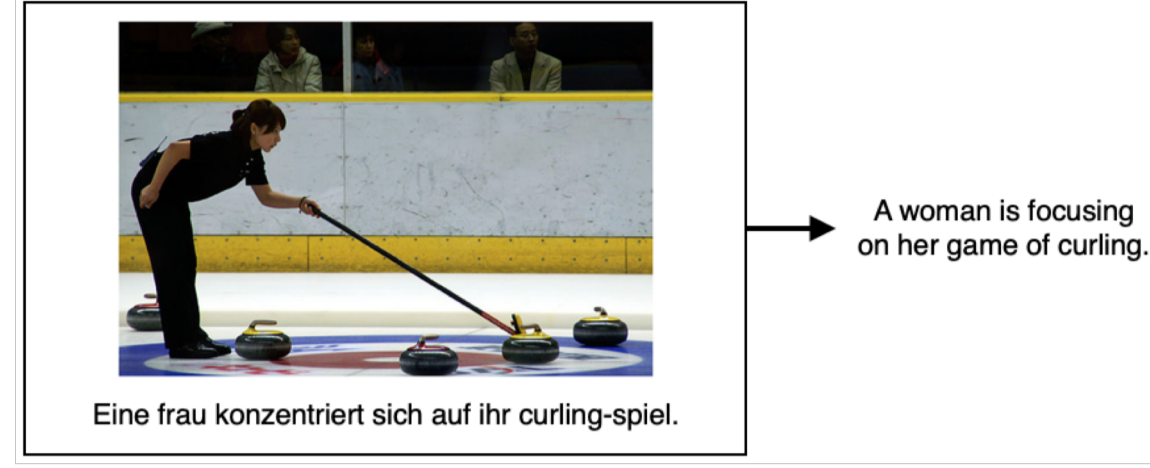
事前学習した単語分散表現を利用したマルチモーダル機械翻訳

首都大学東京 平澤 寅庄 松村 雪桜 山岸 駿秀 小町 守

1 背景：マルチモーダルニューラル機械翻訳

マルチモーダルニューラル機械翻訳 (Multimodal NMT, MNMT)

- 非言語情報を利用する機械翻訳
- WMT Multimodal Shared Task
 - 入力：原言語のキャプション+画像
 - 出力：目的言語のキャプション



問題点

- 低頻度語の翻訳精度が低く、出力の Adequacy が低下する。
- 低頻度語が上位概念の単語に置き換わる (例: corn => food)

本研究

- マルチモーダル NMT のモデルに、単語分散表現を予測する Decoder を導入することにより、低頻度語の出力を安定化させ、翻訳モデルの Adequacy 向上を目指した。

2 先行研究

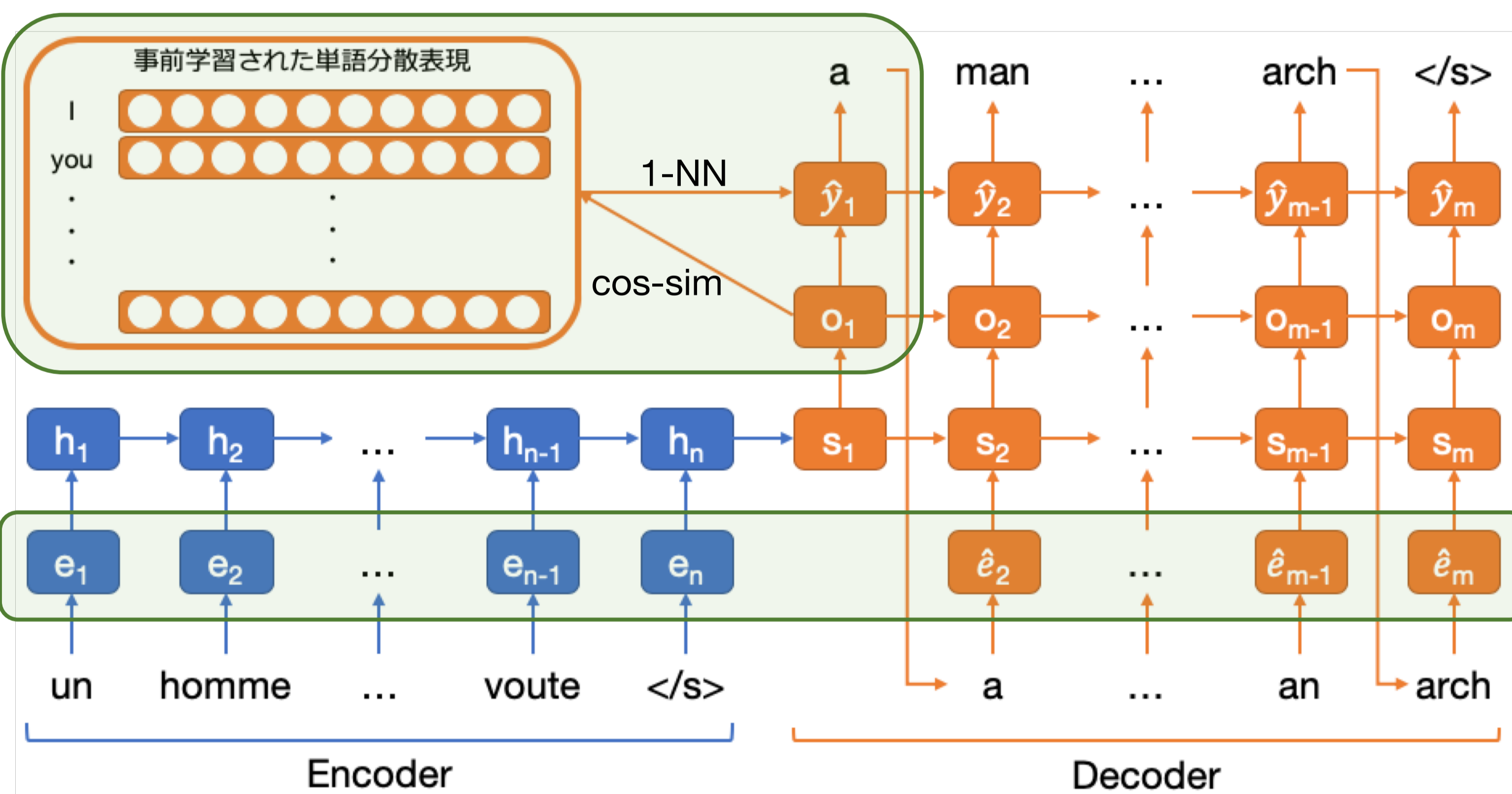
Continuous Outputs [Kumar et al., 2019]

- Decoder の出力層で、単語分散表現を使用した手法を提案した。
 - FastText で単語分散表現を事前学習する。
 - 出力層では次の単語分散表現を予測し、単語類似度を計算し出力を得る。
- 従来の単語生成確率を予測するモデルと同等以上の性能を達成した。
- 特に低頻度語の翻訳精度を顕著に改善した (IWSLT16)。
- softmax 関数を使用しないため、大きな語彙を使用することができる。

Imagination [Elliot et al., 2017]

- 機械翻訳と潜在共有空間構成の2つのタスクを学習する。
 - 潜在共有空間構成：文と画像を同一の潜在共有空間にマッピングするとき、ある文とそれに対応する画像の距離が近くなるように学習する。
- 損失関数はタスクごとの損失関数の線形補間として与える。
- 潜在共有空間構成タスクの損失関数には max-margin 損失を使用する。

3 提案手法：事前学習した単語分散表現を利用したマルチモーダル機械翻訳



1. 単語分散表現を予測する機械翻訳

- 単語分散表現を基準として最近傍の単語を検索し出力する。
 - 単語の生成確率ではなく、分散表現を予測する。
 - コサイン類似度を使用して、1. と出力候補の単語の距離を計算する。
 2. のうち、最も距離が近い単語をシステム出力とする。
- 出力候補の単語分散表現には事前学習した FastText を使用。

2. 単語分散表現を使った初期化

- 事前学習した単語分散表現を機械翻訳システムに組み込む。
 - Encoder の埋込み層を初期化する (追加学習あり)
 - Decoder の埋込み層を初期化する (追加学習なし)
- 未知語の分散表現には、FastText の学習に使用したコーパスに含まれるが、モデルの訓練データに含まれない単語分散表現の平均を使用する。
- 事前学習した FastText を使用。

4 実験設定

Multi30K データセット

- Flickr30K データセット (1つの画像に5文の英語の説明がついている) にフランス語の翻訳を追加したコーパス。
- 原言語文、目的言語文、画像の三つ組のデータを持つ。

モデル

- 言語対: フランス語 (FR) -> 英語 (EN)
- 語彙数: 10,000
- 埋め込み層: 300 次元
- Encoder 隠れ層: bi-GRU, 256 次元*2
- Decoder 隠れ層: 256 次元
- Attention 隠れ層: 256 次元
- ビーム幅: 1
- FastText: 公開されている事前学習モデルを使用 (CBOW、窓幅: 5)

訓練

- 損失関数の補間係数: 0.01
- Optimizer: Adam
- 初期学習率: 0.0004
- バッチサイズ: 64
- Dropout: 0.3

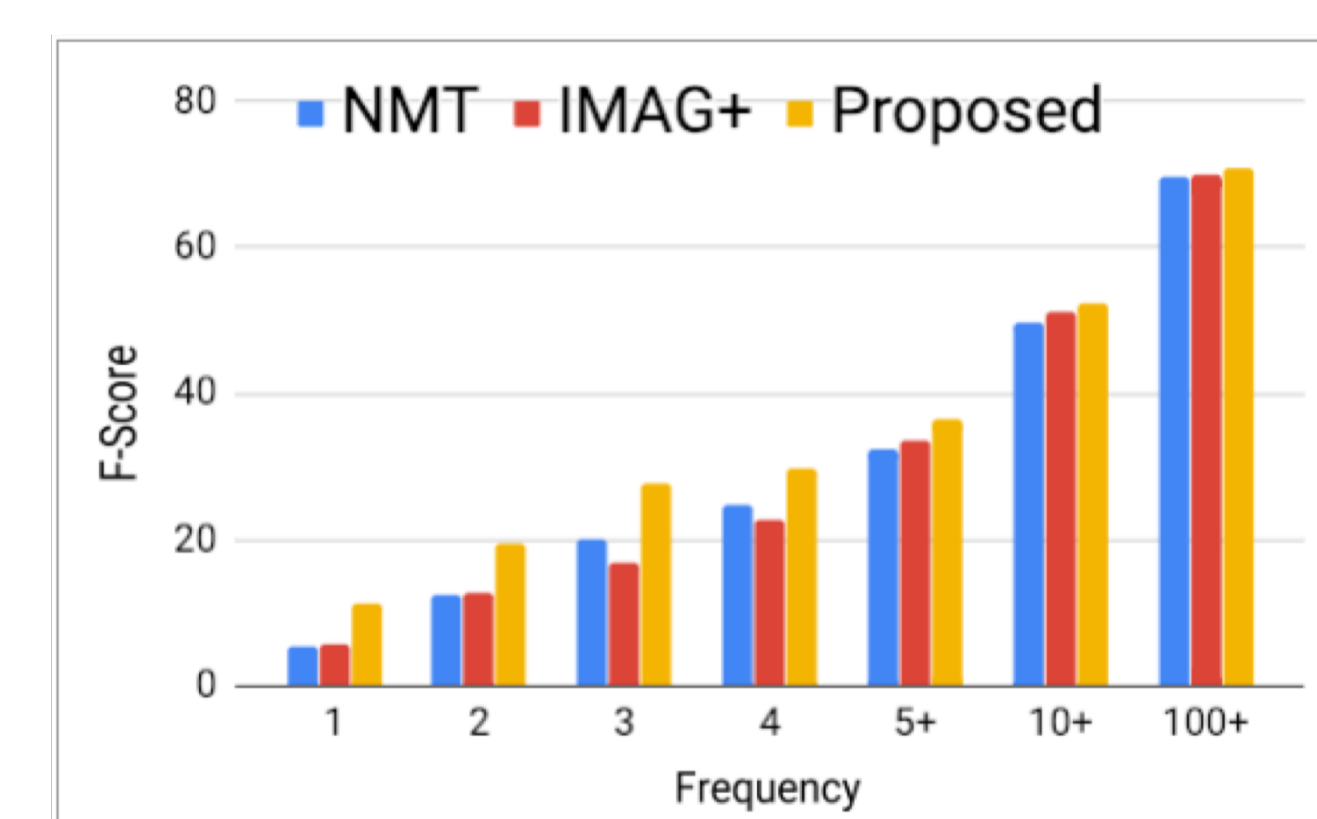
5 結果と考察

システム全体の性能

model	dev		test	
	BLEU	BLEU	BLEU	METEOR
NMT	50.83	51.00 ± .37	42.65 ± .12	
IMAG+	51.06	51.18 ± .16	42.80 ± .19	
Proposed	53.14	53.49 ± .20	43.89 ± .14	
- Images	52.97	53.25 ± .21	43.91 ± .11	

- 検証・評価ともにベースラインを上回っている。
- 文中に含まれる低頻度語や未知語がより適切な単語に置き換えられており、Adequacy の向上に貢献している。
- 長い文になると、繰り返しなどが発生し、出力の Fluency が低下する (6. 出力例・下)。

単語頻度別の翻訳精度



- 単語の出現する文をラベルとして F 値を計算する。
 - 正解: 参照データ中の出現文
 - 予測: システム出力中の出現文
- 訓練データ中の出現頻度ごとに区分けし、平均値を求める。
- 低頻度語で顕著な改善が見られた。また、高頻度語でも限定的ではあるが改善が見られた。

6 出力例

Source	un homme en vélo pédale devant une voûte .
Reference	a man on a bicycle pedals through an archway .
NMT	a man on a bicycle pedal past an arch .
IMAG+	a man on a bicycle pedals outside a monument .
Proposed	a man on a bicycle pedals in front of a archway .
Source	quatre hommes , dont trois portent des kippas , sont assis sur un tapis à motifs bleu et vert olive .
Reference	four men , three of whom are wearing prayer caps , are sitting on a blue and olive green patterned mat .
NMT	four men , three of whom are wearing aprons , are sitting on a blue and green speedo carpet .
IMAG+	four men , three of them are wearing alaska , are sitting on a blue patterned carpet and green green seating .
Proposed	four men , three are wearing these are wearing these are sitting on a blue and green patterned mat .

事前学習した単語分散表現の効果 (test セット、補間係数: 0.5)

Encoder	Decoder	Fixed	BLEU	METEOR
FastText	FastText	Yes	52.90	43.70
random	FastText	Yes	52.06	43.23
FastText	random	No	50.77	42.44
random	random	No	50.22	41.97
FastText	FastText	No	50.29	42.25
random	FastText	No	49.69	41.68

- Decoder の埋込み層を固定する必要がある。
 - Decoder 埋め込み層の追加学習を行うことにより、低頻度語の代わりに頻度の高い類似語が出力されやすくなる。
- 事前学習した単語分散表現を Encoder 埋込み層の初期化に使用するのに比べ、Decoder 埋込み層の初期化に使用の方が有用である。