Nearest Neighbor Machine Translation

(Khandelwal et al., 2020)

出口 祥之

■ deguchi@ai.cs.ehime-u.ac.jp

2020/11/05 二宮研論文輪読会

ව Links

Paper

https://arxiv.org/abs/2010.00710

Introduction

k 近傍法を利用した生成により翻訳性能の大幅改善

- 翻訳時に訓練データを参照して k 近傍法を適用
 - 翻訳時の処理のみが変わるため,追加の訓練無しで 従来の訓練済み NMT モデルをそのまま利用可能

■ 翻訳性能の大幅改善

- SOTA な独英翻訳システムと比較して +1.5 BLEU %
- 多言語翻訳システムにおいて英独・中英・英中翻訳 などで +3 BLEU %
- ドメインに特化したデータ (翻訳のドメイン適応) に 対して平均で +9.2 BLEU %

k-Nearest Neighbor (kNN)

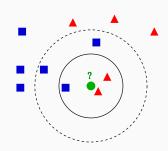
最も近い例に基づいた分類法

距離が近いサンプル k 個を参照し,最も一般的なクラスを割り当てる

例: 赤三角と青四角の2クラス分類

k = 3 **のとき** 緑丸 → 赤三角クラス

k = 5 **のとき** 緑丸 → 青四角クラス



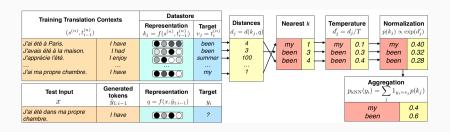
https:

//en.wikipedia.org/wiki/
K-nearest_neighbors_algorithm
(CC-BY-SA 3.0; by Antti Ajanki)

Proposed Method

Nearest Neighbor Machine Translation

翻訳時,ステップ毎に k 近傍法を適用して探索



データストア: 訓練データ (対訳コーパス) の目的言語文 の各単語位置における中間表現を事前に作成

 $q = f(x, \hat{y}_{1:i-1})$: 翻訳時の各ステップにおける中間表現

Nearest Neighbor Machine Translation

Training Translation	Contexts	Datastore	,] [1			ı		
$(s^{(n)},t^{(n)}_{i-1})$		Representation $k_j = f(s^{(n)}, t_{i-1}^{(n)})$	Target $v_j = t_i^{(n)}$		Distances $d_j = d(k_j, q)$		Neares	st k		Temper $d'_j = d$			Norma $p(k_j) \propto$	$exp(d'_j)$
J'ai été à Paris. J'avais été à la maison. J'apprécie l'été. J'ai ma propre chambre.	I have I had I enjoy I have		been been summer my	 	4 3 100 1	X	my been been	1 3 4	→	my been been	0.1 0.3 0.4	\exists	my been been	0.40 0.32 0.28
Test Input	Generated tokens $\hat{y}_{1:i-1}$	Representation $q = f(x, \hat{y}_{1:i-1})$	Target y_i							p	knn(y	_i) =	$\sum_{j} 1_{y_i}=$	$v_j p(k_j)$
J'ai été dans ma propre chambre.	I have	•••	?									my een		0.4 0.6

翻訳時の動作

- **(1)** q とデータストア内の中間表現との距離から k 近傍を獲得
- (2) k 近傍内の各距離に対して温度付き softmax 関数を適用
- (3) 2. の計算結果を集約し、確率に変換
- (4) モデルの出力確率と 3. の確率を線形補間

Datastore creation

全翻訳過程の中間表現を事前に作成

$$(\mathcal{K}, \mathcal{V}) = \{ (f(s, t_{1:i-1}), t_i), \forall t_i \in t \mid (s, t) \in (\mathcal{S}, \mathcal{T}) \}$$

- % f: NMT モデル (デコーダの中間表現を返す)
- ※ (S,T):対訳コーパス
 - lacktriangleright lack
 - 原言語文の情報はキーの中に暗黙的に含まれる

Generation

温度付き softmax により距離に基づいた確率を算出

$$p_{kNN}(y_i|x, \hat{y}_{1:i-1}) \propto \sum_{(k_j, v_j) \in \mathcal{N}} \mathbb{1}_{y_i = v_j} \exp\left(\frac{-d(k_j, f(x, \hat{y}_{1:i-1}))}{T}\right)$$

- ※ ŷ:翻訳時の生成済みトークン

NMT モデルの出力確率 $p_{MT}(y_i|x,\hat{y}_{1:i-1})$ と線形補間

$$p(y_i|x,\hat{y}_{1:i-1}) = \lambda$$
 $p_{kNN}(y_i|x,\hat{y}_{1:i-1})$ $+(1-\lambda)$ $p_{MT}(y_i|x,\hat{y}_{1:i-1})$ 距離に基づく確率 モデル出力確率

Experimental Setup

データセット

- (1) 通常の翻訳 (WMT'19 De-En)
- (2) 多言語翻訳 (訓練: CCMatrix, 評価: newstest2018, newstest2019, TED Talks)
- (3) ドメイン適応 (multi-domains; Medical, Law, IT, Koran, Subtitles)

モデル

NMT システム Transformer (Fairseq)

kNN-MT, Computational Cost

k NN-MT Faiss (高速なk 近傍探索ライブラリ) を使用

- キー: Transformer 最終層 FFN への入力
 - 多言語翻訳: キー 5M, 131K クラスタ
 - ドメイン適応: キー 1M, 4K クラスタ
- 64 近傍,32 クラスタを探索(k = 64)

計算時間と引き換えに追加訓練無しで翻訳性能を改善

データストア作成 事前に全てのサンプルに対して一度順方向 に計算すればよい (訓練1エポック分と同等のコスト)

翻訳時 データストア全体 (10 億のオーダー) から 64 近傍 を得るため,生成速度が 2 桁程遅くなる

Experiments

WMT'19 De-En

モデル	BLEU (%)
baseline	37.59
+kNN-MT	39.08 (+1.5)

■ 追加の訓練無しで 1.5 BLEU % の性能改善

Multilingual Machine Translation

評価する言語対と同じ言語対のデータストアを構築

Test set sizes	de-en 2,000	ru-en 2,000	zh-en 2,000	ja-en 993	fi-en 1,996	lt-en 1,000	de-fr 1,701	de-cs 1,997	en-cs 2,000
Base MT +kNN-MT	34.45 35.74	36.42 37.83	24.23 27.51	12.79 13.14	25.92 26.55	29.59 29.98	32.75 33.68	21.15 21.62	22.78 23.76
Datastore Size	5.56B	3.80B	1.19B	360M	318M	168M	4.21B	696M	533M
Test set sizes	en-de 1,997	en-ru 1,997	en-zh 1,997	en-ja 1,000	en-fi 1,997	en-lt 998	fr-de 1,701	cs-de 1,997	Avg.
Test set sizes Base MT +kNN-MT									Avg. - 26.00 27.40

Multilingual Machine Translation

原言語側を英語にしてデータストアを構築

	Ted Talks					Ne	Avg.		
	de-ja	ru-ja	uk-ja	de-ru	de-zh	fr-de	cs-de	de-cs	
Test set sizes	4,442	5,090	3,560	4,288	4,349	1,701	1,997	1,997	-
Base MT	10.11	9.69	8.36	17.24	20.48	26.04	22.78	21.15	16.98
+kNN-MT (en-*)	11.08	10.42	9.64	18.02	21.22	27.85	23.71	21.74	17.96
Datastore Size	433M	433M	433M	4.23B	1.13B	6.50B	6.50B	533M	-

Domain Adaptation

5 つの異なるドメインに対し zero-shot 翻訳

	Newstest 2019	Medical	Law	IT	Koran	Subtitles	Avg.
Test set sizes	2,000	2,000	2,000	2,000	2,000	2,000	-
Aharoni & Goldberg (2020):							
one model per domain	-	56.5	59.0	43.0	15.9	27.3	40.34
one model for all domains	-	53.3	57.2	42.1	20.9	27.6	40.22
best data selection method	-	54.8	58.8	43.5	21.8	27.4	41.26
Base MT	37.59	39.91	45.71	37.98	16.30	29.21	33.82
+kNN-MT:							
in-domain datastore	39.08	54.35	61.78	45.82	19.45	31.73	42.63
WMT'19 datastore	39.08	40.22	46.74	40.27	17.99	29.23	34.89
all-domains datastore	38.88	54.54	61.11	48.63	19.22	31.70	43.04
Datastore Size (in-domain)	770M	5.70M	18.3M	3.10M	450K	159M	-

Tuning kNN-MT (on validation set)

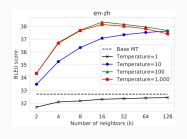
探索幅 k

- 実験では 64 で固定
- 大きくしても性能は変わらず、下がることも (ノイズ?)

Softmax の温度パラメータ T

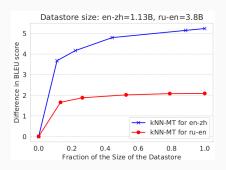
■ *T* が大きくなるほど:

- 分布が平坦になる
- コンテクストへの過剰適合 が抑えられる
- 多様性が高まる



Tuning kNN-MT (on validation set)

データストアの大きさと翻訳性能



Qualitative Analysis

kNN 分布のみで生成される場合 ($\lambda=1$)

Test Input: Dabei schien es, als habe Erdogan das Militär gezähmt.
Generated tokens: In doing so it seems as if Erdogan has tamed the

Training Set Translation Context (s	Training Set Target	Context Probability	
Dem charismatischen Minis- terpräsidenten Recep Tayvip Erdoğan, der drei aufeinanderfol- gende Wahlen für sich entscheiden komnte, ist es gelungen seine Autorität gegenüber dem Militär geltend zu machen.	The charismatic prime minister, Re- cep Tayyip Erdoğan, having won three consecutive elections, has been able to exert his authority over the	military	0.132
Ein bemerkenswerter Fall war die Ermordung des gemäßigten Pre- mierministers Inukai Tsuyoshi im Jahre 1932, die das Ende jeder wirklichen zivilen Kontrolle des Militärs markiert.	One notable case was the assas- sination of moderate Prime Minis- ter Inukai Tsuyoshi in 1932, which marked the end of any real civilian control of the	military	0.130
Sie sind Teil eines Normal- isierungsprozesses und der Her- stellung der absoluten zivilen Kontrolle über das Militär und bestätigen das Prinzip, dass niemand über dem Gesetz steht.	They are part of a process of nor- malization, of the establishment of absolute civilian control of the	military	0.129
Diese hart formulierte Erklärung wurde als verschleierte, jedoch un- missverständliche Warnung ange- sehen, dass das Militär bereit wäre einzuschreiten	That toughly worded statement was seen as a veiled but unmistakable warning that the	military	0.123

Final Translation: In doing so, Erdogan seemed to have tamed the military.

Reference: In doing so, it seems as if Erdogan has tamed the military.

Related Work (Retrieval in Translation)

NMT による翻訳と検索の組み合わせ

- 対訳コーパスから *n*-gram 翻訳句を検索 ¹
- キャッシュに保存した翻訳履歴から検索²

用例に基づく機械翻訳

- アナロジーに基づく機械翻訳³
- 編集距離に基づく似た原言語文から翻訳例を検索 4
- NMT と翻訳メモリ検索の組み合わせ 5 6

^{1.} Guiding Neural Machine Translation with Retrieved Translation Pieces", Zhang et al., 2018.

^{2.} Learning to Remember Translation History with a Continuous Cache", Tu et al., 2018.

^{3.} A Framework of a Mechanical Translation between Japanese and English by Analogy Principle", Nagao, 1984.

^{4.} Search Engine Guided Neural Machine Translation", Gu et al., 2018.

^{5...} Neural Fuzzy Repair: Integrating Fuzzy Matches into Neural Machine Translation", Bulte et al., 2019.

^{6.} Boosting Neural Machine Translation with Similar Translations", Xu et al., 2020.

Conclusion

追加訓練無しに従来の NMT モデルに適用できる 新たな翻訳法 kNN-MT を提案

- 翻訳時、NMT モデルの中間表現と近いコンテクスト ベクタを k 近傍法により探索
- kNN-MT による確率とモデルの出力確率との線形補間から出力単語を得ることで翻訳性能を大幅改善

今後の課題

- 手法の効率化
 - 頻出する単語をダウンサンプリングするなど