

NMT における Word vector size と 1 文単位の fine tuning の候補文数と翻訳文選択方法の考察

村上仁一

鳥取大学工学部

murakami@tottori-u.ac.jp

概要

NMT は、単語を Word vector に変換して、計算をしている。従来 Word vector size は約 500 である。この数は、意味を考慮した品詞の数と、ほぼ同程度である。そしてこの数を削減することで、計算量やメモリ量を削減することが可能である。

本研究では、Word vector size を削減したとき、どの程度の翻訳精度の劣化が見られるか調査した。その結果、大量の対訳データを学習したモデルに、1 文ごとに fine tuning を行う条件では、翻訳精度の劣化が少ないことを示した。

この結果は、従来と比較して、計算量やメモリ量を大幅に削減できる、ことを示している。

1 はじめに

1.1 NMT と AR モデル

ニューラルネットワーク機械翻訳 (以後 NMT) は、現在の機械翻訳において主流である。NMT は、encoder-decoder モデルに基づいている。この encoder-decoder モデルは、自己回帰モデル (以後 AR モデル) に極めて類似している。この図を図 1 に示す。

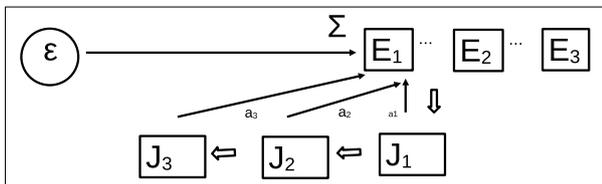


図 1: 自己回帰モデルに基づく機械翻訳の概念図 (ARMT)

NMT は、AR モデルの近似として解釈できる。そこから、NMT は、以下のように解釈できる。

1. Word Vector を利用して、離散モデルから連続モデルに変換する。

2. 次に、AR モデルで、回帰係数を求める。
3. 最後に Soft Max 関数で、連続モデルから離散モデルに変換する。

1.2 Word vector size

以上の考察から Word vector size と翻訳精度には、以下のように推定できる。

1. Word Vector size は、翻訳精度に大きな影響を与えない。
2. 翻訳精度に大きな影響を与えるのは、回帰係数を決定する学習データ量である。

以上の考察から、本研究では、Word Vector size における、翻訳精度の影響を調査する。なお、Word vector size の削減した場合、NMT のモデルの計算量やメモリー量を削減できる。

2 翻訳システム [1]

NLP は、大量の対訳文を利用して学習をおこなう。しかし、入力文には、分野依存性がある。そのため、分野依存性を考慮して、大量の対訳文を学習したモデルに、分野依存した少量のデータを利用して、fine tuning を行うことが一般的である。この fine tuning は、ドメイン適応 (Domain Adaptation) とも呼べるもので、In-Context Learning や Embedding や RAG や転移学習 (Transfer Learning) などの多くの手法が提案されている。そして、ドメインの特徴を表現するため、ある程度のデータ量が必要になる。

本研究で使用した翻訳システムは、1 文ごとに、fine tuning をおこなう。tuning の手法として、大量の学習データを利用して得られた学習済みモデルに、追加学習をおこなう。具体的には、1 文ごとに、大量の学習データから、類似文検索を行う。そして、この複数の類似文を利用して、学習済みモデルに、追加学習をおこなう。本研究で使用した翻訳システムの概要を以下に示す [1]。

2.1 学習

学習方法を以下に示す。

1. 大規模データを利用したモデル作成
大量の対訳データを利用して、NMT のモデルを作成する。
2. 入力文ごとに類似文を抽出
入力文 1 ごとに、大量の対訳データから、類似文を抽出する。
3. 入力文ごとに類似を利用した fine tuning
入力 1 文ごとに、大量のモデルで作成した NMT に対して、上記の類似文を利用して、追加学習を行う。つまり、fine tuning を追加学習により行う。

2.2 翻訳 (多数決)

NMT は AR モデルの変形と考えることができる。そして学習や翻訳には、常にホワイトノイズが加算されていると仮定できる。この性質を利用して、本研究では、翻訳は複数のモデルを学習して、複数の翻訳候補文を出力した後、多数決もしくは信頼度最大で選択をおこなう。基本的な翻訳選択方法を以下に示す。

1. 入出力の順序
入力が順逆および出力が順逆の計 4 方向で、モデルを学習する。
2. 初期値
初期値を 4 回変更して、それぞれでモデルを作成する。
3. N -best
出力を 4best とする。

以上の工程で得られた文を、信頼度最大で選択し [1]、最終的な翻訳文とする。

2.3 学習データ

実験に利用した学習データを表 1 にまとめる。表中“単文”や“複文”は、電子辞書などから抽出した日英対訳例文 [2] である。“追加学習”は Jparacrawl から抽出した日英対訳例文 [3] である。

表 1: 学習データ

単文	163,188 文対
複文	92,427 文対
追加学習	130,974 文対

日本語の総語彙数は 83629 単語で、英語の総語彙数は 92584 単語である。

2.4 類似文検索

類似文検索には TFIDF を利用する。TFIDF の計算には sklearn を利用する。類似文の数は任意とする。

3 実験結果

Word vector size を変えた実験結果を表 2 にしめす。テスト文数は、日本語複文 100 文である。

表 2: 翻訳精度と Word Vector size

Word Vector size	500	64	32	16	8
BLEU	0.263	0.259	0.270	0.279	0.248
meteor	0.558	0.558	0.552	0.564	0.539
TER	0.528	0.536	0.535	0.521	0.537
RIBES	0.831	0.821	0.810	0.828	0.812
STR [4]	0.04	0.04	0.05	0.06	0.05
COMET	0.802	0.789	0.783	0.794	0.782

Word Vector size は OpenNMT では 500 が default である。この結果から、Word vector size 64 でも、かなり高い翻訳精度が維持されていることがわかる。さらに、Word vector size 8 にしても、かなり高い翻訳精度が維持できている。

NMT は、単語を Word vector に変換して、計算をしている。Word vector size を小さくすると、NMT における計算量やメモリー量は、大きく削減できる。したがって、実験結果から、ある程度、翻訳精度が低下を許容した場合、NMT における計算量やメモリー量は、大きく削減できる。

出力例を表 6 に示す。人手評価では、Word vector size が 500 の実験結果が最も良く、Word vector size が小さくなるに従って、緩やかに翻訳性能が低下していた。また、Word vector size 500 における翻訳誤りを分析すると、テストデータには、学習データには存在しない単語の存在（いわゆる未知語）が誤りの原因であった。

4 考察

今回採用した翻訳システムにおいて、多くの任意の数がある。その数を変化させたときの実験結果を以下に示す。

4.1 類似文の数

入力文の類似文の数を変化させたときの翻訳精度の変化を表 3 にしめす。

表 3: 翻訳精度と類似文の数

類似文の数	8	64	256	512	1024
BLEU	0.272	0.295	0.274	0.274	0.263
meteor	0.540	0.580	0.576	0.560	0.558
TER	0.549	0.516	0.507	0.527	0.528
RIBES	0.807	0.833	0.828	0.826	0.831
STR [4]	0.06	0.06	0.06	0.05	0.04
COMET	0.779	0.802	0.800	0.796	0.802

この結果から、類似文の数が多いほど、翻訳精度が高いことがわかる。しかし、類似文の数が少なくても（例えば 8）、かなり高品質な翻訳文が得られている。この原因として、未知語が挙げられる。類似文が少ない場合、fine tuning において未知語が増加する。そのため、翻訳精度の低下につながったと考えられる。ただし、この低下は緩やかである。

なお、入力文から類似文を選択する方法として、今回は TFIDF を利用した。他にも BERT を利用した方法など、様々な方法がある。今後調査したい。

4.2 翻訳文の選択方法を変えた実験

NMT は AR モデルの近似としてみなせる。そのとき、候補文は、常にホワイトノイズが加わっているとみなせる。このホワイトノイズの影響を軽減するために、複数のモデルを利用して、各モデルから得られた複数の候補文から、最終的に翻訳文を選択する。その選択方法として、以下の 4 つがある [1]。

1. 尤度最大
尤度最大で選択する。
2. 尤度合計
同一の候補文の尤度は、加算して、選択する。
3. 多数決
最も多く出現した候補文を選択する。
4. 信頼度最大 [1]
入出力順逆、計 4 方向において、最大の候補を選択する。

これらの方法における翻訳精度の変化を調査した。TFIDF において類似文 64 と 256 と 1024 のときの実験結果を表 4 に示す。

表 4: 評価方法

類似文数	64	256	1024
comet			
尤度最大	0.791	0.784	0.771
尤度合計	0.798	0.789	0.791
多数決	0.803	0.797	0.799
信頼度最大	0.802	0.800	0.802
BLEU			
尤度最大	0.312	0.274	0.235
尤度合計	0.305	0.264	0.261
多数決	0.297	0.278	0.266
信頼度最大	0.295	0.274	0.263

これらの結果から、“多数決”もしくは“信頼度最大”が良い推定方法と言える。そして、NMT においてよく利用されている“尤度最大”は、最も低いことがわかる。

4.3 google との比較

google 翻訳における同一テスト文 100 文の実験をおこなった。結果を表 5 にしめす。

表 5: google の結果

BLEU	0.260
meteor	0.554
TER	0.577
RIBES	0.820
STR [4]	0.05
COMET	0.831

この結果から、提案手法は、google 翻訳には及ばない。google に及ばない翻訳例を調査すると、テストデータには、学習データには存在しない単語の存在（いわゆる未知語）がある。google 翻訳には、学習データ量に大きな差があることや、closed data が含まれている可能性があることから、提案手法は、高い翻訳精度が得られていると推定している。

4.4 自動評価と人手評価

今回の実験では、自動評価として主に BLEU と comet を利用した。また、各実験において、絶対値を用いた人手評価を、様々な場面で利用した。そこからの見解を以下に示す。

1. 人手評価（絶対値 5 段階評価）と comet は、かなり相関が高い。
2. 人手評価（絶対値 5 段階評価）と BLEU は、相

関が高いとはいえない。

3. 自動評価で、ある程度信頼性が高いのは、comet である。
4. ただし、comet も、信頼ができない。BLEU より、ましであるレベルである。
5. 翻訳文と参照文において、使われている単語が大きく異なっているが、意味があっている場合がある。このような場合、comet BLEU とともに、自動評価と人手評価に大きな差がでる。

5 おわりに

本研究では、Word vector size と翻訳精度の関係を調査した。その結果 Word vector size が低くても、翻訳精度があまり低下しないことがしめされた。つまり、ある程度、翻訳精度の低下をある程度許容した場合、計算量やメモリー量は、大きく削減できる。

また、考察として fine tuning に利用する文数の調査や、評価方法を調査した。その結果 fine tuning に利用する文数が多いほど、翻訳精度が高いこと。また、翻訳文の決定には、尤度よりも、多数決や信頼度が有効であることを示した。

参考文献

- [1] 村上仁一. Nmt における多様性最大基準による翻訳文の選択. 第 264 回自然言語処理研究会, 2025.
- [2] 村上仁一, 藤波進. 日本語と英語の対訳文対の収集と著作権の考察. 第一回コーパス日本語学ワークショップ, pp. 119-130, 2012.
- [3] 森下他. Jparacrawl v3.0: 大規模日英対訳コーパス. 言語処理学会第 28 回年次大会, 2022.
- [4] 村上仁一. 機械翻訳における文一致率による評価. 人工知能学会全国大会論文集 第 27 回, 2013.

表 6: 出力例

<p>2 この花が咲くともう春です。 When this flower blooms , you will know it is spring . (参照文) When this flower blooms, it's already spring. (google) When this flower blooms, it's already spring. (提案手法 Vord Vector 500, fine tuning1024, 信頼度最大) 3 - 2 (多数決 - 信頼度) This flower is already spring . (提案手法 Vord Vector 64, fine tuning1024, 信頼度最大)</p>
<p>27 わたしはその金を受け取ることを断った。9 - 4 I refused to accept the money . I refused to accept the money. I refused to take the money . I refused to receive the money .</p>
<p>36 きみのあいまいな説明ではあの娘がだれだかわからなかった。 I could not recognize the girl from the vague description of her you gave me . I couldn't figure out who that girl was from your vague description. I couldn't recognize the girl from the vague description of her you gave me . 16 - 4 Your vague explanation gave me no idea who she was .</p>
<p>39 メーカーは「ヒースーモ」絶縁材を最低の価格で貴社に提供する用意があります。 The manufacturer is prepared to supply you with "Heethemo" insulating material at the lowest possible prices . The manufacturer is ready to offer you the lowest price for the "Heathymo" insulation. The manufacturer is prepared to supply you with the "Heethemo" material at the lowest prices . 4 - 2 The manufacturer is prepared to supply you with "Heethemo" insulation at the lowest price .</p>
<p>50 クロスワードに字を当てはめるのは面白い。 It is fun to fit a word into a crossword puzzle . It's fun to fit letters into crosswords. It is interesting to apply words to crosswords . 14 - 4 It is interesting to apply characters to crosswords .</p>
<p>62 一息吐いてから次の仕事にかかろう。 Let us have a short break before we set about our next job . Let's take a deep breath and then move on to the next task. Let 's get down and get down to the next job . 1 - 1 Let 's take a break to the next job .</p>
<p>63 君の言うことの逆も同程度に当たっているようだ。 The converse of your statement seems equally true . The opposite of what you say seems to be equally true. The reverse of what you say is true to the same extent . 3 - 3 The reverse of what you say seems to be in the same way .</p>
<p>65 真を尽くして説明したら、相手も分かってくれました。 I explained it to him in all sincerity , and he understood . I tried my best to explain, and the other person understood. When I explained the truth , the reverse was true . 1 - 1 When I explained the truth , I understood the other person .</p>
<p>75 最期まで奮闘したのは敵ながら天晴であった。 To his honour be it said that the enemy fell fighting to the last . The one who struggled until the end was the enemy, Appare. It was the enemy to fight to the last . 2 - 2 It was the enemy who fought to the last .</p>
<p>77 送り状第 2929 号に従って船積みした商品の損傷のため、貴殿に迷惑をおかけし申し訳ございません。 We apologize for any inconvenience you have experienced due to damage of our merchandise shipped under Invoice No. 2929 . We apologize for the inconvenience caused to you due to the damage to the goods shipped according to invoice number 2929. We apologize to you for the inconvenience caused according to the products shipped by shipping No. 333 . 1 - 1 We apologize for any inconvenience this may cause you .</p>
<p>78 彼がその絵に見入るがままにさせておいた。 We left him gazing at the picture . I let him admire the painting. I let him stare at the painting . 2 - 2 He let the picture remain .</p>
<p>82 この事件は近所で知らない者はない。 There is not anybody in the neighborhood who does not know about this incident . There is no one in the neighborhood who does not know about this incident. This case is not known in the neighborhood . 9 - 4 This case is not known in the neighborhood .</p>
<p>84 らっぱが鳴り響いて国王を迎えた。 This case is not known in the neighborhood . A blare of trumpets uprose to salute the king . A trumpet sounded to welcome the king. 8 - 3 The bugle sounded the King .</p>
<p>87 私は針のむしろに座る心地がした。 I felt as if sitting on thorns . I felt comfortable sitting on the mat of the needle. I felt as if I sat on thorns . 4 - 2 I felt comfortable to sit on the needle .</p>
<p>94 そう考えるのはまったく大違いだ。 It is a gross mistake for you to think so . It's a big mistake to think so . It is quite a big difference to think so . 5 - 3</p>