

2020年度（令和2年度） 卒業論文

ニューラル機械翻訳に  
乱数を与える影響

電気情報系学科 卒業論文検印	
学科長	

指導教員

村上仁一

鳥取大学工学部 電気情報系学科

自然言語処理研究室

B17T2114B 矢野 貴大

## 概要

機械翻訳の方式の一つに，Neural Machine Translation(以下，NMT)[1]がある．NMTは学習を行う際に乱数を用いている．そのため，同じデータで学習を行ったとしても出力が異なる．しかし，精度に大きな差はないとされ，今まであまり考慮されてこなかった．

本研究では，NMTにおいて同じデータで学習を行った際の，翻訳結果の変動について調査を行う．調査結果より，同一の学習データで学習させたとしても出力は異なることが分かった．また，各翻訳評価においても差が生じるということが分かった．考察より，PRED値を用いることでNMTの出力を改善出来ることが分かった．

# 目次

第1章	はじめに	1
第2章	NMTについて	2
	2.0.1 NMT(Transformer)	2
第3章	翻訳評価の手法	3
	3.1 翻訳確率 (PRED)	3
	3.2 BLEU,METEOR,TER	3
	3.3 人手評価	3
第4章	調査方法	4
	4.1 学習	4
	4.2 翻訳	4
第5章	実験環境	6
	5.1 実験条件	6
	5.2 実験データ	6
第6章	実験結果	7
	6.1 出力結果	7
	6.2 自動評価結果	8
	6.3 人手評価結果	9
第7章	考察	10
	7.1 PRED 値平均	10
	7.2 出力結果 (PRED と人手の一致した例)	11
	7.3 出力結果 (PRED と人手の一致しなかった例)	14
	7.4 各文に対して最良の PRED 値を選択した際の人手評価結果	17

7.5 ステップ数の妥当性について . . . . .	18
第8章 おわりに	19

# 目 次

4.1 調査手順 1,2,3 . . . . .	4
4.2 調査手順 4,5 . . . . .	5

# 表 目 次

3.3.1 人手評価の基準 . . . . .	3
5.2.1 実験データ . . . . .	6
6.1.1 出力結果 . . . . .	7
6.2.1 自動評価結果 (16000 文) . . . . .	8
6.3.1 人手評価結果 (100 文) . . . . .	9
7.1.1 人手評価と対応する PRED 値の平均 . . . . .	10
7.2.1 PRED と人手評価の一致した例 1 . . . . .	11
7.2.2 PRED と人手評価の一致した例 2 (繰り返し文) . . . . .	12
7.2.3 PRED と人手評価の一致した例 3 (He , She 入れ替わり) . . . . .	13
7.3.1 PRED と人手評価が一致しなかった例 1 (肯定否定が逆でも PRED に大差なし) . . . . .	14
7.3.2 PRED と人手評価が一致しなかった例 2 (意味不明でも PRED はそこまで悪くない) . . . . .	15
7.3.3 PRED と人手評価が一致しなかった例 3(PRED 低いけど人手評価は ) . . . . .	16
7.4.1 人手評価結果 (100 文) . . . . .	17
7.5.1 ステップ数と各評価値の関係 . . . . .	18

# 第1章 はじめに

機械翻訳の方式の一つに，Neural Machine Translation(以下，NMT)[1]がある．NMTは学習を行う際に乱数を用いている．そのため，同じデータで学習を行ったとしても出力が異なる．しかし，精度に大きな差はないとされ，今まであまり考慮されてこなかった．

本研究では，NMTにおいて同じデータで学習を行った際の，翻訳結果の変動について調査を行う．

本論文の構成は以下の通りである．第2章では，翻訳評価の手法について述べる．第3章では，調査方法について述べる．第4章では，実験環境について述べる．第5章では，実験結果について述べる．第6章では，考察について述べる．第7章では，本実験の簡単なまとめについて述べる．

## 第2章 NMTについて

### 2.0.1 NMT(Transformer)

NMTはニューラルネットワークを用いた機械翻訳の方式であり、現在機械翻訳において主流の方式となっている。今回の調査ではOpenNMTを用いる。OpenNMTは翻訳の際にアテンション機構を用いる。以下に翻訳の流れを示す。

- 1 入力文をベクトル表現に符号化(変換)する
- 2 1を基に入力文の中の注目すべき単語をアテンション情報としてまとめる
- 3 符号化された入力文とアテンション情報を基に翻訳文を生成する



## 第3章 翻訳評価の手法

本章では、翻訳評価の手法について記述する。

### 3.1 翻訳確率 (PRED)

翻訳確率とは、入力が  $a$  であったときに  $b$  が出力される尤もらしさを数値化したものである。この翻訳確率の対数をとった値が、PRED である。PRED の値は 0 に近いほど、良い翻訳である。

### 3.2 BLEU, METEOR, TER

BLEU[2]、METEOR[3]、TER[4] は機械翻訳の自動評価の手法である。参照文と翻訳結果をもとに評価を行う。

### 3.3 人手評価

著者自身が各出力文に対し、 $\checkmark$ 、 $\circ$ 、 $\times$  の 3 つの評価基準をもって正確さに基づいて評価を行う。それぞれの基準を表 3.3.1 に示す。

表 3.3.1: 人手評価の基準

	参照文と同じである 大体の意味を捉えている google 検索で相当量ヒットする
	意味が部分的に合っている 入力文が推定可能である
$\times$	意味がほとんど合っていない 入力文の推定が困難である

## 第4章 調査方法

本章では, 本研究における調査方法についての説明を行う.

### 4.1 学習

調査方法における学習の過程では, 同じ学習データを異なる乱数を用いた上で NMT に対して学習を行うことで, 8 個の NMT を生成する.

### 4.2 翻訳

テスト文を 8 個の NMT それぞれに翻訳させ, 8 個の翻訳結果を生成する. ここで生成された翻訳結果を用いることで, 翻訳結果の変化についての調査を行う.

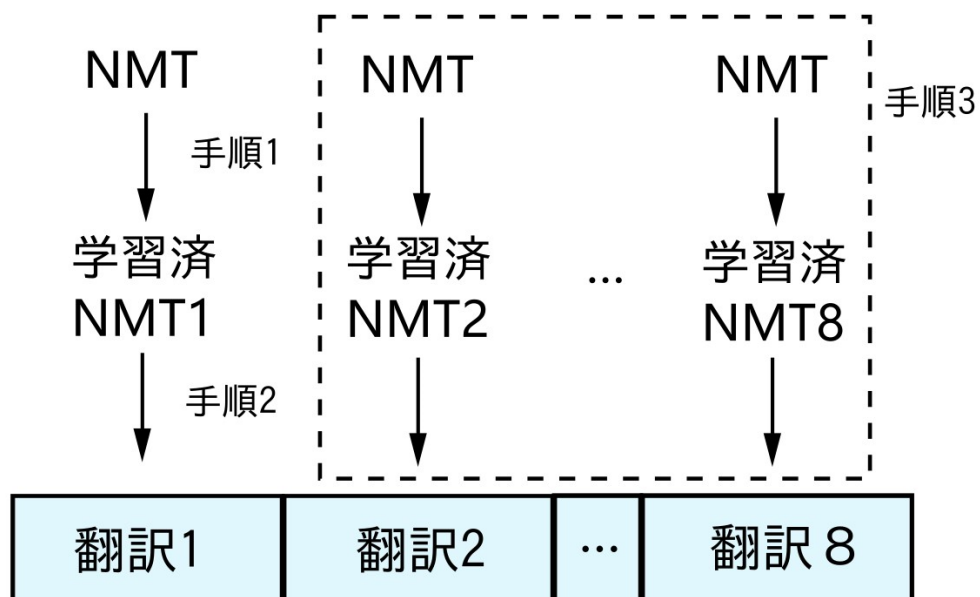


図 4.1: 調査手順 1,2,3

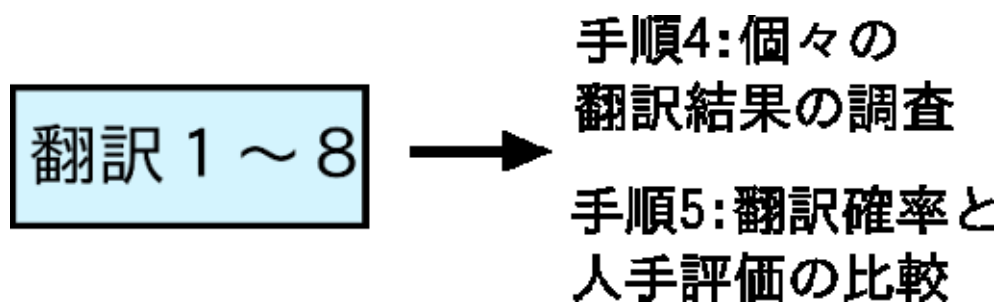


図 4.2: 調査手順 4,5

# 第5章 実験環境

## 5.1 実験条件

本研究における調査では，OpenNMT[5] を用いて日英ニューラル機械翻訳を行う．実験データ

## 5.2 実験データ

調査に用いるデータの内訳を表 5.2.1 に示す．なお，学習文は全て単文である．

表 5.2.1: 実験データ

対訳学習文 (単文)	160,000 文
ディベロップメント文	1,000 文
テスト文	16,000 文
ステップ数	100,000

## 第6章 実験結果

### 6.1 出力結果

出力結果を表 6.1.1 に示す。なお，表 6.1.1 における評価の列は，人手評価の結果を示す。

表 6.1.1: 出力結果

入力文	彼はまだ歩けない赤ん坊だ。	評価
参照文	He is still a baby in arms .	
NMT1	He is still a short baby .	
NMT2	He is still a baby .	
NMT3	He is still a little baby .	
NMT4	He is still a baby boy .	
NMT5	He is still a walking boy .	×
NMT6	He is still an baby's baby .	
NMT7	He is still a little baby .	
NMT8	He is still hard .	×

表 6.1.1 の結果より，乱数によって出力結果が異なってくることを確認できた。

## 6.2 自動評価結果

自動評価結果を表 6.2.1 に示す。なお，表 6.2.1 に示す PRED 値は，テスト文 16,000 文各文に対して導出された PRED 値の平均値である。

表 6.2.1: 自動評価結果 (16000 文)

	PRED	BLEU	METEOR	TER
NMT1	-0.4061	0.1844	0.4569	0.6228
NMT2	-0.3980	0.1838	0.4574	0.6218
NMT3	-0.414	0.1827	0.4528	0.6248
NMT4	-0.4076	0.1838	0.4582	<b>0.6172</b>
NMT5	-0.4066	<b>0.1855</b>	0.4574	0.6190
NMT6	-0.4014	0.1850	0.4547	0.6241
NMT7	<b>-0.3962</b>	0.1853	<b>0.4587</b>	0.6222
NMT8	-0.4006	0.1849	0.4575	0.6182

表 6.2.1 の結果より，自動評価の値も異なってくることを確認できた。この表の PRED で最良の値を示したのは NMT 7 であった。また，BLEU で最良の値を示したのは NMT 5 であった。

## 6.3 人手評価結果

翻訳を行ったテスト文中のランダムな 100 文に対し, 評価を行った。結果を表 6.3.1 に示す。

表 6.3.1: 人手評価結果 (100 文)

			×
NMT1	47	25	28
NMT2	43	25	32
NMT3	39	35	26
NMT4	43	29	28
NMT5	39	30	31
NMT6	44	22	34
NMT7	41	23	36
NMT8	41	30	29

表 6.3.1 の結果を見ると, モデルによって人手評価の結果が大きく異なることが確認できる。つまり, NMT が乱数の影響を強く受けることが分かる。

また, ここで表 6.3.1 の結果と表 6.2.1 の結果を比較する。PRED の最も良かった NMT 7 と, BLEU の最も良かった NMT 5 は, どちらも人手評価では振るわなかった。

つまり, 自動評価と人手評価の結果が一致するとは限らないと言える。

## 第7章 考察

### 7.1 PRED 値平均

人手評価に対応する PRED 値の平均を表 7.1.1 に示す .

表 7.1.1: 人手評価と対応する PRED 値の平均

			×
NMT1	-2.1564	-3.4963	-8.1842
NMT2	-1.7890	-3.4370	-7.7541
NMT3	-1.7239	-3.4888	-8.1038
NMT4	-2.0195	-3.1141	-7.5009
NMT5	-1.8041	-2.9416	-8.4624
NMT6	-2.1846	-4.6117	-7.7291
NMT7	-1.9401	-3.8695	-7.1248
NMT8	-1.9080	-3.2454	-7.5257
平均	-1.9407	-3.5256	-7.7981

表 7.1.1 より , PRED 値が高いほど良い人手評価になる傾向にあることが確認できた .



## 7.2 出力結果 (PRED と人手の一致した例)

表 7.2.1 に PRED が人手評価と対応がとれている例を示す。

表 7.2.1: PRED と人手評価の一致した例 1

入力文	再び こういう こと の ない よう に 注意 し なさい。	評価	PRED
参照文	Look to it that this does not happen again .		
NMT1	Be careful what you do .		-3.2388
NMT2	Be careful of such a thing .		-3.2991
NMT3	See that there is no occasion .	×	-4.8158
NMT4	Look like this again .	×	-2.9777
NMT5	Take care of such a thing again .		-4.0140
NMT6	Be careful not to do such a thing again .		-3.6025
NMT7	Be careful not to do anything like this again .		-2.3996
NMT8	Be careful not to do such a thing again .		-1.7900

表 7.2.1 については，人手評価 の文に対しては高めの PRED 値が，×の文に対しては低めの PRED 値が出ていることが確認できる。

表 7.2.2: PRED と人手評価の一致した例 2 (繰り返し文)

入力文	大統領の媒介によって争議は解決した。	評価	PRED
参照文	The strike was settled by the intervention of the President .		
NMT1	The dispute was settled by the President of the President .		-1.8957
NMT2	The dispute was settled by the president's invasion .		-1.8790
NMT3	The dispute caused by the president caused a solution .	×	-3.3359
NMT4	The dispute was settled by the dispatch of the president .		-2.6528
NMT5	The dispute was settled by the mediation of the president .		-3.5173
NMT6	The dispute was settled by the President .		-2.2755
NMT7	The President's formation was resolved by the President .		-2.8033
NMT8	The dispute was settled by the president of the president .		-1.5124

表 7.2.2 について, NMT 1, 8 の出力では the president が繰り返し発生している。しかし意味は取れている文であるため, 人手評価では とした。そして PRED についても良い値を示している。このことより, この例において PRED は繰り返し文に対応出来ていた。

表 7.2.3: PRED と人手評価の一致した例 3 (He, She 入れ替わり)

入力文	イギリス 紳士 の まさに 典型 で あった 。	評価	PRED
参照文	He was the very ideal of an English gentleman .		
NMT1	This was a very typical example of a gentleman .		-3.0452
NMT2	It was a sensible example of a British gentleman .		-3.2549
NMT3	That was the worst example of a gentleman .		-3.7659
NMT4	He was a typical example of a British gentleman .		-2.2104
NMT5	She was a typical classic of a British gentleman .		-3.7124
NMT6	He was a classic example of a gentleman .		-2.9573
NMT7	He was the epitome of a gentleman of English .		-1.9124
NMT8	He was the very epitome of a gentleman .		-1.9405

表 7.2.3 において NMT 5 の出力は He であるべき部分が She となっている。そしてこの出力の PRED 値は下から 2 番目に悪い。今回 She の出力で PRED が悪かったのは、類似の例文に He で始まるものが多かったからではないかと考える。

### 7.3 出力結果 (PRED と人手の一致しなかった例)

表 7.3.1: PRED と人手評価が一致しなかった例 1 (肯定否定が逆でも PRED に大差なし)

入力文	自分の将来に不安を感じていない。	評価	PRED
参照文	He feels secure about his future .		
NMT1	I don't feel any apprehension about my future .		-1.3998
NMT2	I feel uneasy about my future .	×	-2.5883
NMT3	I don't feel feeling uneasy about my future .		-2.4133
NMT4	I feel unsettled about my future .	×	-1.2790
NMT5	I feel not worried about my future .		-2.6342
NMT6	I don't feel anxiety about my future .		-2.6342
NMT7	I don't feel anxiety about my future .		-1.4979
NMT8	I feel uneasy about my future .	×	-1.8851

表 7.3.1 について説明する。NMT2, 4, 8 の出力は翻訳すると「自分の将来に不安を感じている」といったニュアンスになる。しかし入力文は「自分の将来に不安を感じていない」であり、これらの出力文は意味が逆になって出力されている。しかし PRED 値においては正解の出力文と大して変わらない値となっている。

表 7.3.2: PRED と人手評価が一致しなかった例 2 (意味不明でも PRED はそこまで悪くない)

入力文	彼は閑職にまわされた。	評価	PRED
参照文	He has been transferred to an easy post .		
NMT1	He was deeply engrossed .	×	-2.5850
NMT2	He was carried into a small section .		-2.8353
NMT3	He was relegated to a minor post .		-0.6828
NMT4	He was stabbed by a sudden death .	×	-0.9691
NMT5	He was attacked by a fraction .	×	-1.0670
NMT1	He was put into a post .	×	-2.7056
NMT2	He was carried into a strange telephone .	×	-3.1215
NMT3	He was driven to the gallows .	×	-1.0629

表 7.3.2 において, NMT4, 5, 8 の出力は入力文と大きく異なる訳であるにもかかわらず, PRED の値は良い。

この原因については, 類似する例文が少なかったからではないかと推測する。

表 7.3.3: PRED と人手評価が一致しなかった例 3(PRED 低いけど人手評価は )

入力文	老後 は 楽 に 暮らしたい。	評価	PRED
参照文	I want to live comfortably when I become old .		
NMT1	I want to live in my old age .		-1.6963
NMT2	I would like to live at an easy age .	×	-2.8527
NMT3	I want to get along with old age .		-3.1904
NMT4	I want to live comfortably in old age .		-3.6880
NMT5	I want to live in old age .		-1.0251
NMT6	I want to live at ease in my old age .		-3.7577
NMT7	I want to live in old age .		-1.6900
NMT8	I want to live in a comfortable age .		-2.8026

表 7.3.3 において，NMT4, 6 の出力は人手評価では とした．しかし PRED 値で見ると，ワーストの 2 文である．PRED は短い文ほど良い値が出る傾向にあるが，この出力例はその傾向が顕著に現れている．

## 7.4 各文に対して最良の PRED 値を選択した際の人手評価結果

表 7.4.1 に示す。なお、ここでの PRED は、文単位で出力される PRED のことを指す。表の「PRED 最良」の行では各入力文に対する出力文に最も低い PRED 値を選択した際の結果を示す。また、「理論値」の行では各入力文に対しての出力文に人手評価が最も良いものを選択した際の結果を示す。

表 7.4.1: 人手評価結果 (100 文)

			×
PRED 最良	49	24	27
理論値	68	22	10

表 6.3.1 と表 7.4.1 を比較すると、PRED 値が最良の出力を選択することで人手評価結果が改善することが確認できた。このことより、PRED によってある程度は翻訳の良し悪しを判定できることが分かった。しかし理論値と比較すると、改善の余地があることが分かる。

## 7.5 ステップ数の妥当性について

ステップ数の値が 100,000 で適切であるかどうかを検証した．ステップ数別の PRED と BLEU の値を表 7.5.1 に示す．

表 7.5.1: ステップ数と各評価値の関係

ステップ数	PRED	BLEU
60000	-0.5393	0.1793
65000	-0.4524	0.1821
70000	-0.4615	<b>0.1875</b>
75000	-0.4428	0.1863
80000	-0.4353	0.1868
85000	-0.4408	0.1859
90000	-0.4324	0.1862
95000	-0.4328	0.1865
100000	<b>-0.4228</b>	0.1867
200000	-0.4245	0.1873

表 7.5.1 を見ると，PRED の値は 100,000 以降で，BLEU の値は 70,000 以降頭打ちになっている．そのため，ステップ数の値は 100,000 が適切であると判断した．



## 第8章 おわりに

NMTの問題点の一つとして、学習時に乱数が用いられているという問題点がある。しかし、性能に大きな差はないとされ、今まであまり考慮されてこなかった。本研究では、NMTにおいて同じデータで学習を行った際の、翻訳結果の変動について調査を行った。調査結果より、同一の学習データで学習させたとしても出力は異なることが分かった。また、各翻訳評価においても差が生じるということが分かった。考察より、PRED値を用いることでNMTの出力を改善出来ることが分かった。

## 謝辞

最後に、一年間に渡り、本研究のご指導をいただきました鳥取大学工学部知能情報工学科自然言語処理研究室の村上仁一准教授、村田真樹教授に深く感謝すると共に、厚く御礼申し上げます。そして、日常の議論を通じて多くの知識や示唆を頂いた同研究室の皆様に深謝いたします。

## 参考文献

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. CoRR, abs/1409.0473, 2014.
- [2] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu. BLEU: a method for automatic evaluation of machine translation. In ACL, 2002.
- [3] S. Banerjee and A. Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. pages 6572, 2005.
- [4] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. Proceedings of Association for Machine Translation in the Americas, 2006.
- [5] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. OpenNMT: Open-source toolkit for neural machine translation. ArXiv e-prints, 2017.