

相対的意味論に基づく統計機械翻訳における翻訳確率

中村勇太
鳥取大学 工学部
m19j4033b@edu.tottori-u.ac.jp

村上仁一
鳥取大学 工学部
murakami@tottori-u.ac.jp

1 はじめに

機械翻訳において様々な手法が提案されている。しかし、それらの手法で得られる出力が異なる場合が多い。

そこで、異なる機械翻訳手法によって得られた複数の出力(候補文)から、最尤の出力を選択する手法である“相対的意味論に基づく統計機械翻訳(RSMT)”が提案されている。この手法を用いることで、各手法の出力から最尤の出力を得られる。

RSMTでは出力を選択するために、学習文対の日本語側(A)、学習文対の英語側(B)、入力文(C)、候補文(D)について類似度及び翻訳確率を計算する。類似度は同一言語の文章(AとCまたは、BとD)がどれだけ類似しているかを示す。翻訳確率はCがDに翻訳される確率を示す。なお、類似度及び翻訳確率には様々な計算方法が存在する。

本稿の目的は、RSMTで用いる翻訳確率の計算方法として最適な方法を調査することである。5つの翻訳確率の計算方法を用いてRSMTを動作させ結果を比較する。この実験によって、最適な翻訳確率の計算方法を調査する。

2 Neural Machine Translation(NMT)

本論では、候補文を得るために“Neural Machine Translation(NMT)”を用いる。本章でNMTの概要を述べる。

2.1 NMTの概要

NMTとはニューラルネットワークを用いて機械翻訳を行う手法である。2言語コーパス等を用いてAttentionモデルと呼ばれる異言語間の単語の対応を学習し、それを元に入力文の翻訳を行う。

2.2 NMTの学習における乱数の影響

今回の実験では、NMTの一種であるOpen-NMTを用いる。Open-NMTは、学習で乱数を用いている。よって同一のコーパスで学習を行っても、翻訳結果が異なるモデルが生成される。本論ではOpen-NMTのモデルを複数生成し、同一の入力文を翻訳することで、RSMTの候補文を獲得する。

2.3 Open-NMTのN-best出力

Open-NMTは第n候補まで、複数の出力文を出力できる。また、一般的に第1候補が最も優れている。しかし、第2候補以降に優れた出力が存在することも考えられる。本論ではRSMTを用いることで、第1候補より優れた出力文が選択可能か調査する。

3 相対的意味論に基づく統計機械翻訳(RSMT)^[6]

“相対的意味論に基づく統計機械翻訳”では異なる機械翻訳手法によって得られた複数の出力(候補文)から、最尤の出力を選択する。そのために、学習文対の日本語側(A)、学習文対の英語側(B)、入力文(C)、候補文(D)について類似度及び翻訳確率を計算する。出力を選択する手順を示す。

3.1 選択の手順

手順1 学習文対の日本語文と入力文の類似度を計算
学習文対の日本語文と入力文の類似度を計算する。計算された値を sim_{AC} とする。

sim_{AC} の計算式を式(1)に示す。

$$sim_{AC} = \frac{count(Match)}{count(A_{all})} \times \frac{count(Match)}{count(C_{all})} \quad (1)$$

$count(Match)$: 学習文対の日本語文と入力文において共通している単語の総数

$count(A_{all})$: 学習文対の日本語文の単語の総数

$count(C_{all})$: 入力文の単語の総数

手順2 学習文対の英語文と各候補文の類似度を計算
学習文対の英語文と各候補文の類似度を計算する。計算された値を sim_{BD} とする。

sim_{BD} の計算式を式(2)に示す。

$$sim_{BD} = \frac{count(Match)}{count(B_{all})} \times \frac{count(Match)}{count(D_{all})} \quad (2)$$

$count(Match)$: 学習文対の英語文と候補文において共通している単語の総数

$count(B_{all})$: 学習文対の英語文の単語の総数

$count(D_{all})$: 候補文の単語の総数

手順3 入力文と候補文に類似した学習文対を抜粋
入力文と各候補文に類似した学習文対を一定数抜粋する。抜粋された学習文対の集まりをDBとする。
各候補文について式(3)の数値が高い順に抜粋する。

$$sim_{AC} \times sim_{BD} \quad (3)$$

手順4 DB内の学習文対の翻訳確率を計算
DB内の学習文対の翻訳確率を計算する。計算された値を $trans_{AB}$ とする。

ここで用いる式は第4章で説明する。

手順5 入力文から各候補文への翻訳確率を計算
入力文から各候補文への翻訳確率を計算する。計算された値を $trans_{CD}$ とする。

ここで用いる式は手順4と同様である。

手順6 出力文の決定
計算された値を用いて出力文を決定する。
式(4)の値が最も高い候補文を、出力文とする。

$$sim_{AC} + sim_{BD} + trans_{AB} + trans_{CD} \quad (4)$$

RSMT の流れ図を図 1 に示す。

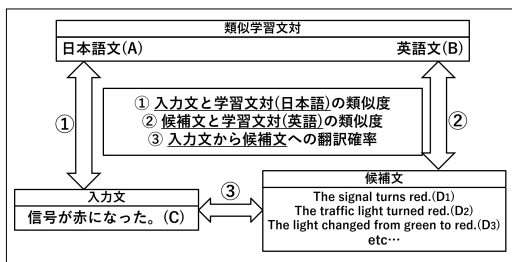


図 1 RSMT の流れ図

4 翻訳確率の計算方法

実験で用いる翻訳確率の計算方法を説明する。なお実際の計算では、2つの単語の連なりを1つの単語と見なす。

4.1 自己相互情報量 (PMI)^[2]

自己相互情報量とは、2つの事象の間の関連度合いを測る尺度である。日英単語の共起をこの尺度に当てはめることで翻訳確率として用いる。計算式を式(5)に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\text{count}(J_j) \text{count}(E_i)} \quad (5)$$

$\text{count}()$: 学習文対中での登場回数

J_j : 入力文中 j 番目の単語 E_i : 候補文中 i 番目の単語

M : 入力文の単語数 N : 候補文の単語数

4.2 Joint Probability(Joint)^[5]

Joint Probability とは、交差エントロピーを拡張した翻訳確率の計算方法である。計算式を式(6)に示す。

$$\sum_{j=1}^M \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\text{count}(J_{all}) \text{count}(E_{all})} \times \log_2 \frac{\text{count}(J_j, E_i)}{\text{count}(J_j) \text{count}(E_i)} \quad (6)$$

$\text{count}(J_{all})$: 全学習文対における日本語単語の総数

$\text{count}(E_{all})$: 全学習文対における英語単語の総数

4.3 Jaccard 係数 (Jaccard)^[3]

Jaccard 係数とは2つの集合がどれだけ類似しているかを表す係数である。入力文と候補文を「文を構成する単語の集合」とみなすことで、Jaccard 係数を翻訳確率として用いる。計算式を式(7)に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\text{count}(J_j) + \text{count}(E_i) - \text{count}(J_j, E_i)} \quad (7)$$

4.4 Dice 係数 (Dice)^[4]

Dice 係数とは Jaccard 係数と同じく、2つの集合がどれだけ類似しているかを表す係数である。計算式を式(8)に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{2 \text{count}(J_j, E_i)}{\text{count}(J_j) + \text{count}(E_i)} \quad (8)$$

4.5 Simpson 係数 (Simpson)^[1]

Simpson 係数とは Jaccard 係数と同じく、2つの集合がどれだけ類似しているかを表す係数である。Simpson 係数

は、「片方の集合がもう片方の集合の真部分集合である時、類似度が最大になる」という特徴がある。計算式を式(9)に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\min(\text{count}(J_j), \text{count}(E_i))} \quad (9)$$

5 Open-NMT の第 1 候補を用いる実験

5.1 実験目的と方法

本章では、Open-NMT の第 1 候補を、RSMT の候補文として実験を行う。5つの翻訳確率計算方法で RSMT を動作させ、翻訳精度を調査する。翻訳精度の調査は自動評価及び人手評価で行う。

5.2 実験に使用するデータ

本研究では、実験に電子辞書などの例文より抽出した単文コーパス^[7]を使用する。

RSMT の候補文は Open-NMT の第 1 候補を用いる。また、使用する Open-NMT のモデル数は 32 個である。よって候補文は、入力文 1 文につき 32 文存在する。使用するデータの内訳を表 1 に示す。

表 1 実験データの内訳

学習文対	160,000 文
入力文 (自動評価)	1,000 文
入力文 (人手評価)	100 文

5.3 実験結果

5.3.1 自動評価の結果

自動評価の結果を表 2 に示す。

表 2 自動評価の調査結果 (1000 文)

	BLEU	METEOR	RIBES	TER
PMI	0.183	0.484	0.776	0.583
Joint	0.164	0.468	0.772	0.652
Jaccard	0.173	0.474	0.789	0.605
Dice	0.176	0.477	0.788	0.606
Simpson	0.174	0.471	0.787	0.615

表 2 によると、多くの評価指標において PMI が最も優れた結果を示す。

5.3.2 人手評価の結果

100 文を対象に、人手評価を行った結果を表 3 に示す。また、人手評価の評価基準を示す。

- : 入力文の意味と合致かつ文法の誤りがない
- △: 入力文の意味と部分的に合致
- ×: 入力文の意味と合致していない

表 3 人手評価結果 (100 文)

	○	△	×
PMI	32	22	46
Joint	25	22	53
Jaccard	32	22	46
Dice	32	23	45
Simpson	29	24	47

表 3 によると、Joint が他の手法と比較して性能が劣る。また、Joint 以外の手法間に大きな差はない。表 4 から表 6 に人手評価の例を示す。

表 4 人手評価の例 1

入力文	彼は晩年をベニスで過ごした。
Ref	He spent the afternoon of his life in Venice .
PMI Jaccard Dice	He spent his late years in Venice .
Joint Simpson	He spent his late years to Venice .
人手評価の結果	PMI Jaccard Dice ○, それ以外は△

表 5 人手評価の例 2

入力文	私はその仕事に慣れていない。
Ref	I am not used to the task .
Jaccard Dice	I am not used to the work .
Joint	I am not yet used to the work .
PMI Simpson	I am unused to the work .
人手評価の結果	PMI Jaccard Dice ○, Joint △, PMI Simpson ×

表 6 人手評価の例 3

入力文	あの女の人は美しい顔をしている。
Ref	That woman has a beautiful face .
Jaccard Dice	That woman has a beautiful face .
PMI Joint Simpson	That woman's face has a beautiful face .
人手評価の結果	Jaccard Dice ○, それ以外は△

6 Open-NMT の第 8 候補までを用いる実験

6.1 実験目的と方法

ここでは、Open-NMT の第 8 候補までを RSMT の候補文として用いて実験を行う。候補文となる基準を引き下げることで、Open-NMT のモデルを複数生成するよりも容易に、多くの候補文を得ることができる。ここでは、5 章の実験との比較のために、候補文の数を揃える。

6.2 実験に使用するデータ

実験に使用するデータは表 1 と同様である。使用する Open-NMT のモデル数は 4 つである。よって候補文は、入力文 1 文につき 32 文存在する。

6.3 実験結果

6.3.1 自動評価の結果

自動評価の結果を表 7 に示す。

表 7 自動評価の調査結果 (1000 文)

	BLEU	METEOR	RIBES	TER
PMI	0.189	0.465	0.777	0.613
Joint	0.173	0.451	0.733	0.702
Jaccard	0.193	0.465	0.769	0.630
Dice	0.194	0.466	0.765	0.635
Simpson	0.189	0.465	0.759	0.655

表 7 によると、PMI と Dice が優れた結果を示す。

6.3.2 人手評価の結果

100 文を対象に、人手評価を行った結果を表 8 に示す。

表 8 人手評価結果 (100 文)

	○	△	×
PMI	36	16	48
Joint	20	11	64
Jaccard	35	19	46
Dice	32	18	50
Simpson	28	17	55

表 3 と表 8 を比較すると、PMI, Jaccard, Dice 以外の手法は○が減少している。特に、Joint の減少数は顕著であ

る。一方で、PMI, Jaccard は○が増加している。また、○の数は PMI が最多である。このことから、PMI を用いることで、第 1 候補以外の優れた出力文を選択できることが分かる。

表 9 から表 11 に人手評価の例を示す。

表 9 人手評価の例 1

入力文	彼の話には深みがある。
Ref	There is something deepin what he said .
PMI	His story has a depth of meaning .
Jaccard	There is a depth in his story .
Simpson	His story has a depth of depth .
Joint Dice	His story has a depth of depth .
人手評価の結果	PMI ○, Jaccard △, それ以外×

表 10 人手評価の例 2

入力文	この部屋での喫煙はご遠慮ください。
Ref	Please refrain from smoking in this room .
Joint Jaccard Dice	Please refrain from smoking in this room .
PMI Simpson	Kindly refrain from smoking in this room .
人手評価の結果	Joint Jaccard Dice ○, それ以外は×

表 11 人手評価の例 3

入力文	わが国の刑務所制度はとんでもなく時代遅れなものとなっている。
Ref	Our prison system has become an absurd anachronism .
Joint Simpson	The prison system of this country is out of date .
それ以外	Our prison system is in a sad tone .
人手評価の結果	Joint, Simpson ○, それ以外は×

7 実験のまとめ

本章では第 5 章と第 6 章の結果から判明した事項を述べる。最後に総評として、本論で述べた 5 つの翻訳確率の計算方法から、RSMT で用いる最適な方法を決定する。

7.1 5 章のまとめ

5 章の実験では、翻訳確率の各計算方法の性能を比較するための実験を行った。その結果、Joint は他 4 つの計算方法より性能が若干劣ることが分かった。また、Joint 以外の計算方法間に性能差はほぼなかった。

7.2 6 章のまとめ

6 章の実験では、候補文に用いる Open-NMT の出力の順位を引き下げることで、第 1 候補以外の優れた候補文を選択できるか調査した。その結果、PMI 以外の計算方法は一様に性能が低下した。しかし、PMI は他 4 つと異なり性能が向上した。

7.3 総評

総評すると、5 章の実験で Joint 以外の計算方法にはほぼ差がなく、6 章の実験で PMI が第 1 候補以外の優れた候補文を選択できることが判明した。よって、RSMT で用いる翻訳確率の計算方法は、PMI が最も適している。

8 考察

8.1 Open-NMT の第 1 候補と RSMT の性能比較

RSMT の性能を評価するために Open-NMT の第 1 候補と性能を比較した。また、人手評価の結果から RSMT が誤った出力をする原因を考察した。実験に使用するデータは第 6 章と同様とする。比較のために用いる RSMT の結果は、人手評価で最も○が多かった第 6 章の PMI の結果とする。

8.1.1 比較結果

Open-NMT の 1best と RSMT の、自動評価と人手評価の結果をそれぞれ表 12 と表 13 に示す。

表 12 自動評価結果 (1000 文)

	BLEU	METEOR	RIBES	TER
Open-NMT	0.191	0.453	0.765	0.628
RSMT	0.189	0.458	0.777	0.613

表 13 人手評価結果 (100 文)

	○	△	×
RSMT	36	16	48
Open-NMT	34	21	45

表 12 によると、RSMT の方が BLEU 以外の指標において優れた結果を示した。また、人手評価も RSMT の方が優れた結果を示した。人手評価の例を表 14 から表 16 に示す。また、RSMT が誤った候補文を選んだ原因を考察する。

表 14 人手評価の例 1

入力文	洞穴の中でぼくたちの声が反響した。
Ref	Our voices echoed in the cave .
Open-NMT	The inside of the cave echoed .
RSMT	Our voices echoed in the cave .
人手評価の結果	RSMT ○

表 15 人手評価の例 2

入力文	良心が彼女を苦しめた。
Ref	Her conscience stung her .
Open-NMT	Her conscience worried her .
RSMT	My conscience worried her .
類似学習文 (日本語)	良心がとがめる。
類似学習文 (英語)	My conscience pricksme .
人手評価の結果	Open-NMT ○

表 16 人手評価の例 3

入力文	この手形は 8 月 30 日で満期となる。
Ref	The bill is due on August 30 .
Open-NMT	The bill matures on August 30 .
RSMT	The bill matures on June 30 .
類似学習文	日英共に長文 (後述)
人手評価の結果	Open-NMT ○

表 14 では、前置詞の使い分けが RSMT によって改善された。しかし、表 15 では、RSMT は所有格が誤っている。これは日本語と英語の所有格の明記の差異が原因で、誤っている候補文の類似度が正しい候補文より高くなっている。

表 16 では入力文及び RSMT の出力文が、学習文対「6 月 30 日号付け財務諸表の写しをお送り申し上げます。」“It is a pleasure to send you the enclosed copy of our June 30 Financial Statements.” との類似度が高かったことが原因で出力された。

RSMT はこの例のように、ある学習文対と極度に類似した候補文を選択する場合が多い。また計算方法の都合、学習文対及び候補文の単語数が多いほど、翻訳確率が高くなるため、単語数が多い候補文が選択されやすい。これらが誤りの原因となる可能性がある。

8.1.2 単語数の多さが原因である誤りの解決法

この節では、8.1.1 の例における誤りの原因を解決する方法を提案する。

解決方法の 1 つは、文の単語数によって正規化を行うことである。正規化を行うことで、単語数が多いほど選択されやすい問題を解決できる

8.2 過剰生成を含む候補文

NMT の出力文の特徴の 1 つとして「過剰生成」がある。これは表 17 に示すように、同じ単語が複数回、出力に現れることを指す。過剰生成は NMT の出力数を増やすほど、現れる傾向がある。多くの場合、RSMT は過剰生成を含む候補文を選択しないことが望ましい。

表 17 過剰生成を含む出力の例

入力文	左舷に塔が見えてきた。
Ref	We opened a tower to the port .
出力	The tower was sighted in sight in sight .

しかし 6 章の実験において、過剰生成を含む候補文が選択される入力文が増加した。また、それが人手評価で × とする理由になった数を表 18 に示す。

表 18 過剰生成が理由で人手評価 × とした数

PMI	5
Joint	20
Jaccard	6
Dice	9
Simpson	17

それぞれの手法で過剰生成が精度低下の原因となった入力が存在する。その中でも、Joint と Simpson は過剰生成が精度低下の原因として多かったことを表 18 が示している。

8.3 類似度の計算方法

本論では、翻訳確率の計算方法の調査を行った。しかし 7.1 節で示したように、RSMT の誤りの理由の 1 つとして類似度の計算方法が挙げられる。また、RSMT において類似度の計算方法も翻訳確率と同様に、様々な方法が考えられる。今後の研究では、類似度の計算方法についても調査が必要である。

8.4 出力文の決定に用いる計算式

本論では、出力文の決定の時に式 (4) を用いた。しかし、この時に用いる式によって結果が変化する。今回用いたパラメータは 4 種類あるので、それぞれのパラメータの使用または不使用の組み合わせだけでも 16 通りの計算式が考えられる。また、7.1 節の問題を解決するならば、翻訳確率に重みを付与することも考えられる。

今後の研究では、式 (4) を用いた時以上の性能を得られる式を調査する。

9 おわりに

相対的意味論に基づく統計機械翻訳 (RSMT) に最適な翻訳確率の計算方法の調査を目的として、5 つの計算方法を用いた翻訳結果を比較した。実験により、自己相互情報量 (PMI) が他の計算方法とほぼ同等の性能であると分かった。さらに PMI は、Open-NMT の第 1 候補以外の優れた候補文を選択できることが分かった。

よって、今回述べた計算方法の中では、PMI が RSMT で用いるのに最も適している。

参考文献

- [1] Bollegala, D., Matsuo, Y., Ishizuka, M. A web search engine-based approach to measure semantic similarity between words. *IEEE Trans. on Knowl. and Data Eng.*, Vol. 23, No. 7, pp. 977–990, 2011.
- [2] Church, K., Hanks, P. Word association norms, mutual information, and lexicography. *Computational Linguistics*, Vol. 16, No. 1, pp. 22–29, 1990.

- [3] Salton, G. , McGill, M. Introduction to modern information retrieval. 自然言語処理学会第 25 回年次大会, 1983.
- [4] Smadja, F. , McKeown, K. R. , Hatzivassiloglou, V. Translating collocations for bilingual lexicons: a statistical approach. *Comput. Linguist.* , Vol. 22, No. 1, pp. 1–38, 1996.
- [5] 松本大輝, 村上仁一. 翻訳における分野依存性を軽減する言語モデルの調査. 自然言語処理学会第 25 回年次大会, 2019.
- [6] 村上仁一. 相対的意味論と機械翻訳の応用. 自然言語処理学会第 27 回年次大会, 2021.
- [7] 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, 2012.