

英日機械翻訳システムに対する自動評価法 RIBES の補正

浦田 明日美 田添 丈博

鈴鹿工業高等専門学校 専攻科 電子機械工学専攻

1. はじめに

英日機械翻訳結果に関する正当性を人手で評価するには膨大な時間と手間がかかってしまう。その負担を軽減、効率よく正当性の高い機械翻訳システムを構築するために、自動評価法を用いることで翻訳結果の正当性を人手よりも早く評価することができる。

このような自動評価法において最も重要な要素は「人間の評価との相関の高さ」であり、英語と日本語の間には文法間で大きな違いがあるため、機械翻訳文が日本語である場合に人間の評価との相関において正しい文章であるにも関わらず自動評価法の値が低くなってしまおうという問題点がある。

本研究では自動評価法 RIBES を英日機械翻訳で用いた際に生じる、人間の評価との相関のずれを補正する手法を提案する。Web 上にある英日機械翻訳システムで多様な英文を翻訳し、翻訳された日本語文の結果が正しいものなのか、誤ったものなのかをその補正手法を用いて評価した実験の結果、より人間の評価に近いものになった。

本論文の構成を以下に示す。2 章では自動評価法 RIBES について概説し、3 章では RIBES を用いて人間の評価とどれ程の相関があるのかを調べるための実験について説明する。4 章で語順のゆらぎに基づく補正手法を提案し、5 章で補正手法を用いた相関を調べる実験について説明する。6 章でまとめ、今後の課題について述べる。

2. 自動評価法 RIBES

RIBES[1] とは、機械翻訳文と参照文の間における共通単語の位置関係、順序に注目して作り出された自動評価法である。

$$RIBES(H, R) = \frac{\sum_{h_i \in H} \max_{r_j \in R_i} \{NKT(h_i, r_j) \cdot P(h_i, r_j)^\alpha \cdot BP_s(h_i, r_j)^\beta\}}{|機械翻訳文集}| \quad (1)$$

$$P(h_i, r_i) = \frac{機械翻訳文と参照文で一致した単語数}{機械翻訳文の単語数} \quad (2)$$

$$BP_s(h_i, r_i) = \min \left(1, \exp \left(1 - \frac{参照文の単語数}{機械翻訳文の単語数} \right) \right) \quad (3)$$

$$NKT = \frac{\tau + 1}{2} \quad (4)$$

$$\tau = \frac{\sum_{i=1}^{n-1} T_i - \sum_{i=1}^{n-1} U_i}{n(n-1)/2} \quad (5)$$

α は単語適合率の重み、 β は BP の重みであり、 P は単語正解率である。 BP_s は機械翻訳文が参照文より短かった場合のペナルティである。

NKT はケンドール順位相関係数 τ を正規化したものであり、 T_i は機械翻訳文から得た単語出現位置 ID リストについて i 番目の要素よりも大きな要素が $i+1$ 番目から n 番目までに出現する数、 U_i は逆に小さな要素が出現する数である。図 1 に RIBES 値による翻訳文の比較を示す。

(英文) He caught a cold because he got soaked in the rain.
(参照文) 雨に濡れたので、彼は風邪をひいた。
(機械翻訳文 1) 彼は雨に濡れたので、風邪をひいた。 RIBES 値: 0.85
(機械翻訳文 2) 彼は風邪をひいたので、雨に濡れた。 RIBES 値: 0.41

図 1. RIBES 値の比較

3. RIBES と人間との相関

3.1 実験方法

RIBES が英日機械翻訳において人間の判断とどの程度の相関を持つのかを比較するにあたり、読売新聞と The Daily Yomiuri から自動生成された『日英対応付けコーパス、1対1の日英文 [2]』を用いて比較実験を行う。この翻訳課題は合計 100 文あり、実験では英文を機械翻訳への入力文とし、日本語を参照文とする。これらの英文を Microsoft が提供している機械翻訳 Bing-Translator [3]を用いて翻訳したものを機械翻訳文とする。RIBES 値の計算には、オープンソフトウェアとして公開されているものを使用する。我々が評価した結果と RIBES 値とを比較する。

3.2 人手評価結果

まず、機械翻訳文だけで文法的に我々が判断したものを「日本語らしさ」として○、△、×の3段階で評価を行った。日本語らしさの分布を表 1 に示す。さらに日本語らしさが○であった機械翻訳文のみを参照文と比較し、意味の一致度合いを「正確さ」として、同様に○、△、×の3段階で評価を行った。正確さの分布を表 2 に示す。

表 1. 日本語らしさ (100 文)

日本語らしさ	文の数
○	87
△	10
×	3

表 2. 正確さ (日本語らしさ○、87 文)

正確さ	文の数
○	49
△	34
×	4

3.3 実験結果

表 2 の正確さの結果と RIBES 値を比較した。その結果を図 2 に示す。

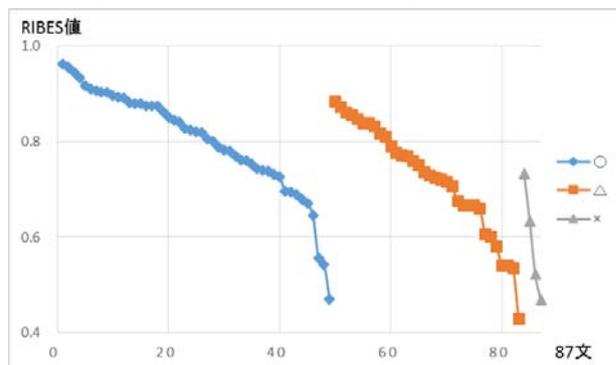


図 2. 正確さと RIBES 値との比較

正確さと RIBES 値は比較的相関があるように思われるが、図 2 より正確さが○であっても RIBES 値が低い (0.6 未満) ものが確認できた。

3.4 考察

正確さが○、RIBES 値が 0.6 未満であった文は 3 文であり、その 1 文を図 3 に示す。

(英文) The Federal Constitutional Court decides on the question of unconstitutionality.
 (参照文) 違憲の問題については、連邦憲法裁判所が決定する。
 (機械翻訳文) 連邦憲法裁判所は違憲の問題を決定します。 RIBES 値: 0.54

図 3. 正確さが○、RIBES 値が 0.6 未満の文

これらの共通点は参照文と機械翻訳文において語句の順が逆になっていることである。これが原因で評価のずれが起こっていると考えられる。日本語は語順に関して寛容であり、因果関係さえ正しければ、どのように語句を並べても意味が通じてしまう。

語順を重視した評価法である RIBES においてこれらの問題点は正確な評価をしていくにあたり、障害になると考えられる。

4. RIBES 値の補正

4.1 並列関係の並び替えによる補正

機械翻訳文に対して構文解析を行い、文中の並列関係にある文節を並び替えて、評価し直すことにより、より人間の評価と相関の高い RIBES 値を得ることができる。図 3 で示した機械翻訳文における並列関係を図 4 に示す。

「連邦憲法裁判所は」と「違憲の問題を」の並列関係を考慮して文節の並び替えを行った際の RIBES 値の変化を図 5 に示す。

連邦憲法裁判所は
違憲の
問題を
決定します。

図 4. 並列関係

(参照文) 違憲の問題については、連邦憲法裁判所が決定する。
(機械翻訳文) 連邦憲法裁判所は違憲の問題を決定します。 RIBES 値: 0.54
(並び替え文) 違憲の問題を連邦憲法裁判所は決定します。 RIBES 値: 0.85

図 5. RIBES 値の補正

4.2 構文解析を利用した補正アルゴリズム

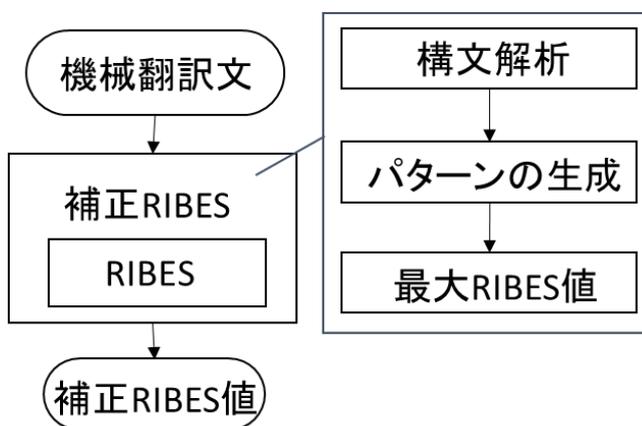


図 6. 補正アルゴリズムの手順

語順のゆらぎによる評価のずれを考慮した RIBES 値の補正アルゴリズムを提案する。手順を図 6 に示す。

1: 構文解析

構文解析を行い、並列関係を持つ文節の有無を調べる。

2: パターンの生成

並列関係にある文節同士を入れ替え、全てのパターンを生成する。例えば、文節が 2 つならば 2 パターン、3 つならば 6 パターンとなる。

3: 最大 RIBES 値

並列要素の入れ替えにより作成された各パターンの RIBES 値を計算し、その中で最も大きかった値をその機械翻訳文の RIBES の評価とする。

5. 補正 RIBES と人間との相関

5.1 実験方法

3 章と同様に日英対応付けコーパスを、4 章で提案した RIBES の補正プログラムを用いて再度評価を行い、正確さを比較する。構文解析には CaboCha[4]を用いた。

5.2 実験結果

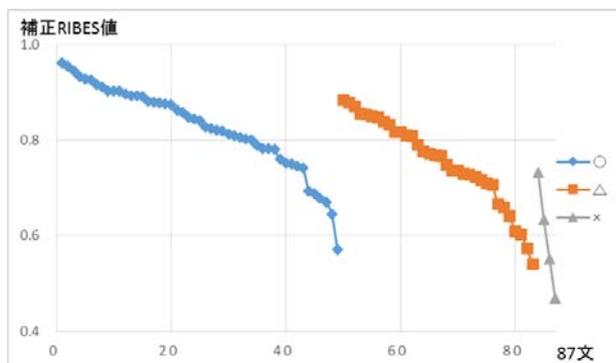


図 7. 正確さと補正 RIBES 値との比較

今回の結果を図7に示す。図2の結果と比較すると全体的に RIBES 値が高くなり、人間の評価により近いものになっていることが確認された。

5.3 考察

改善後も正確さが○にも関わらずRIBES 値が0.6未満の文は1文のみで、あまり大きく数値が改善されなかった。その文を図8に示す。

(英文)

2. The Government of Japan and the Government of the Russian Federation share the perception that Japan's experiences of its postwar economic development can be useful in supporting the efforts of the Russian Federation for its transition toward a market economy.

(参照文)

2 日本国政府及びロシア連邦政府は、市場経済への移行を目指すロシア連邦の努力に対する支援に当たって戦後の日本国の経済発展の経験が有効となり得ることにつき、共通の理解を有する。

(機械翻訳文)

2.、日本政府とロシア政府は戦後の経済発展の経験は、ロシア連邦の市場経済への移行を支援に有用であること意識を共有します。(RIBES 値 : 0.472)

(並び替え文)

2.、日本政府とロシア政府はロシア連邦の市場経済への移行を支援に有用であること意識を経済発展の戦後の経験は、共有します。(RIBES 値 : 0.554)

文の長さが機械翻訳文の方が短いことと参照文中の文節と共通する文節が機械翻訳文は少ないことが原因であると考えられる。このため、参照文にはある表現が翻訳文にはない場合、表現のゆらぎが大きい場合における対処法を確立させる必要がある。

6. おわりに

本研究では英日機械翻訳に対する自動評価法 RIBES を人間の評価と相関が高くなるようにするため、語句の位置関係や順序に着目した補正アルゴリズムを用いて実験を行った。日本語の順序における寛容さから補正アルゴリズムを提案し、人間の評価に近づけることができた。

今後は表現の統一や省略の補完などを考えることにより、人間の評価との相関を更に高めることを目指す。

参考文献

- [1] 平尾努, 磯崎秀樹, 須藤克仁, Kevin Duh, 塚田元, 永田昌明: 語順の相関に基づく機械翻訳の自動評価法, 自然言語処理 Vol. 21, No. 3, pp.421-444, Jun. 2014.
- [2] Masao Utiyama and Hitoshi Isahara: Reliable Measures for Aligning Japanese-English News Articles and Sentences, ACL-2003, pp.72-79, 2003.
- [3] Bing-Translator
<http://www.bing.com/translator/>
- [4] CaboCha
<https://code.google.com/p/cabochoa/>

図8. あまり改善されなかった文