

参照訳を必要としない単語分散表現による 異言語間類似度を用いた訳文の自動評価

藤川 寛基[†] 越前谷 博[‡] 荒木 健治[†]

[†] 北海道大学大学院情報科学研究科 〒060-0814 北海道札幌市北区北 14 条西 9 丁目

[‡] 北海学園大学 〒064-0526 北海道札幌市中央区南 26 条西 11 丁目 1-1

E-mail: [†] {f_hiroki2@eis,araki@ist}.hokudai.ac.jp, [‡] echi@lst.hokkai-s-u.ac.jp

あらまし 従来より参照訳との表層的な一致に基づき機械翻訳が出力する訳文を評価する自動評価法が提案されている。しかし、実際ユーザが機械翻訳を利用する際には参照訳がない場合が多いことから、参照訳を用いずに訳文を評価することのニーズが高まっている。そこで、大規模な対訳コーパスから参照訳を用いずに訳文の品質推定を行う研究もある。しかし、英日のように大規模な対訳コーパスが乏しい言語間では用いることが困難である。よって、本稿では大規模対訳コーパスを用いずに翻訳対象文と訳文の異言語間類似度を用いて訳文を評価する手法を提案する。提案手法では、単語分散表現を翻訳行列によって異言語空間へマッピングした後 Earth Mover's Distance により異言語間類似度を計算することで対訳コーパスを用いることなく訳文を評価する。

キーワード 機械翻訳, 自動評価, Word2Vec, Earth Mover's Distance,

Automatic Evaluation Method for Machine Translation based on Cross-Lingual Similarity using Word Embeddings without Reference Translation

Hiroki FUJIKAWA[†] Hiroshi ECHIZEN[‡]YA[‡] and Kenji ARAKI[†]

[†] Graduate School of Information Science and Technolgy, Hokkaido University, Kita 14, Nishi 9, Kita-ku, Sapporo-shi, Hokkaido 060-0814 Japan

[‡] Hokkai-Gakuen University 1-1 Minami 26 Zyo Nishi 11 Chome, Toyohira-ku, Sapporo-shi Hokkaido, 062-8605 Japan

E-mail: [†] {f_hiroki2@eis,araki@ist}.hokudai.ac.jp, [‡] echi@lst.hokkai-s-u.ac.jp

Abstract We propose a new method for automatic evaluation method for machine translation without reference. Automatic evaluation of machine translation is generally based on the use of reference. However, when the user wants to know the quality of the translation, there is no reference in most cases. Though there is a research to estimate the quality of translation without reference, it is difficult to use it because there is no large-scale bilingual corpus between English and Japanese. In this paper, our proposed method based on Cross-Lingual Similarity evaluates translation without bilingual corpus using word embeddings and translation matrix.

Keywords Machine Translation, Automatic Evaluation, Word2Vec, Earth Mover's Distance

1. はじめに

近年ニューラル機械翻訳^[1]によって Google 翻訳などを代表に機械翻訳の研究は大幅な進化を遂げている。だが、現状の機械翻訳システムが必ずしも正しい訳を出力するわけではなく、特に小説のように抽象性が高い文書に関しては正しい結果が得られないことも多い。機械翻訳システムを利用するユーザは訳文のよし悪しを判定できない場合が多いことから、訳文を自動評価する研究のニーズが高まっている。

しかし、従来の自動評価手法は参照訳と表層的な単

語の一致を基に訳文を評価しているため参照訳の質や量に評価の結果が左右されてしまうという問題がある。このような問題を解決するために、参照訳を用いずに訳文を評価する研究が近年活発化しているが、大規模な対訳コーパスの利用を前提としているため全ての言語に利用することは困難である。

そこで我々は、単語の分散表現を用いて異言語間類似度から最適な訳を選択する手法を提案した^[2]。その中で、異言語間類似度を計算する際の単語アライメントが不完全で改良の余地があり、また単語アライメントの有無を判定するしきい値についての考察

が不十分であった。

そこで本研究では新たに被験者を増やし、また単語アライメントに用いるアライメントスコアの一部に重みをつけることで手法の改良を試みた。その上で実験を行い、アライメントスコアに用いるしきい値と、一致数、順位相関係数の推移について調査することで提案手法の有効性について更なる検討を行った。その結果、従来の提案手法によって得られていた結果を上回りつつ手法の有効性を確認した。

2. 関連研究

従来の自動評価の研究は、BLEU^[3]に代表されるようにNグラム的一致率など表層的な情報を用いて訳文を評価していた。しかし、表層的な情報だけでは単語の表記ゆれや単語の意味を考慮できず、また英日間のように翻訳した後に語順が大きく変わってしまう言語では精度が下がってしまうという問題がある。これに対し、これまで様々な手法が提案されてきているが、英日間においては参照訳と訳文の語順の近さを測るRIBES^[4]がその一つである。RIBESは局所的に注目するNグラムではなく順位相関係数から単語の出現順序を基にして訳文を評価することで英日間において高い精度で訳文を評価することができる。しかし、BLEUなどと同様にRIBESも訳文の評価には参照訳が必要であり、参照訳が存在しない状況下では評価を行うことができない。

それに対して、WMTでは訳文の品質を推定するQuality Estimation Task^[5]の研究が盛んに行われている。この研究では訳文の評価に参照訳を用いずに翻訳対象文を基にすることで訳文の品質を推定している。しかし、訳文の品質推定には大規模な対訳コーパスの利用が前提になっており、対訳コーパスなどの言語資源が豊富な欧州言語間では利用できても、英日間など言語資源の少ない言語間に対する利用は適当ではなく汎用性に欠ける。

文書の類似度を計算する研究としては柳本の研究^[6]がある。柳本は類似画像検索の分野で用いられている2つの分布間の距離を定義できるEarth Mover's Distance(EMD)を用いて、単語分散表現を特徴量に単語の意味を反映させた文書間の類似度を計算している。これらのことより、単語分散表現を特徴量に用いてEMDを自動評価に用いることで表記ゆれや意味の近さを考慮しながら訳文を評価できると考えられる。しかし、EMDの類似度計算は全ての単語特徴量間の距離を用いて行われるため、文間の類似度を計算するには必要のない情報を多く含むという問題がある。

本手法では、参照訳を用いずに翻訳対象文と訳文を直接比較するために、柳本の手法と同様に単語の分散表現を用いて単語の意味を反映させながらEMDに

より異言語間類似度を計算することで訳文を評価する。その際、単語アライメントを用いることで対応係のある単語間でのみ類似度を計算している点で柳本の手法とは異なっている。また、本手法では訳文を評価するために大規模な対訳コーパスを必要としないという点でQuality Estimation Taskとも異なっている。

3. 提案手法概要

ここでは提案手法の概要について示す。英語対象文と分かち書きされた日本語訳文をWord2Vec^[7]を用いてそれぞれ分散表現に変換する。ただし、異言語間の単語分散表現は直接比較することができないため、異言語間での類似度を計算する前に翻訳行列によって異言語空間へ片方の単語ベクトルをマッピングする。マッピングを行った後に、異言語間で単語アライメントを行い、対応する単語を決定する。そして、EMDにより特徴量をWord2Vecによる単語ベクトルを用い、重みにはtf-idfによる単語重要度を用いて類似度計算を行うことで訳文を評価する。図1に提案手法の概要を示す。

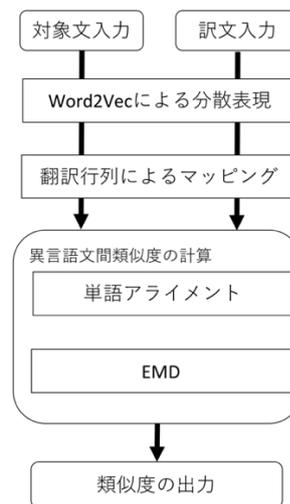


図1 提案手法概要

3.1. 翻訳行列によるマッピング

ここではWord2Vecによって得られた単語ベクトルを異言語間において比較するために翻訳行列^[8]を用いる。Word2Vecによって得られた単語ベクトルは異言語間においても単語間の関係が類似する。しかし、単語ベクトルは単言語コーパスから学習されるため異言語間において直接類似度を計算することはできない。この問題を解決するために、ある単語ベクトルを他の言語の同意味のベクトルへ線形変換をなす翻訳行列Wを用いる。このことにより翻訳行列を片方の単語ベクトルに乗算することで異言語間でのベクトルを比較することが可能となる。翻訳行列Wは対訳語ペア $\{x_i, z_i\}_{i=1}^n$ を用いて最小二乗法により近似することで学習できるため、以下の式(1)により表すことができる。

$$W = \operatorname{argmin}_W \sum_{i=1}^n \|Wx_i - z_i\|^2 \quad (1)$$

3.2. 異言語文間類似度の計算

3.2.1. Earth Mover's Distance

本研究では EMD を用いて対象文と訳文間の類似度を計算している。EMD は 2 つの分布間の距離を最適化する問題の 1 つである輸送問題を基に定義し、それに基づき文書間の類似度を計算する手法である。それぞれの分布は特徴量と重みからなるシグネチャの集合から構成されており、ある分布 P が m 個の集合からなる分布 P は $P = \{(p_1, w_{p_1}) \cdot \cdot \cdot (p_m, w_{p_m})\}$ と表すことができる。ここで p_i は特徴量であり w_{p_i} はその特徴量に対する重みである。同様に分布 Q についても $Q = \{(q_1, w_{q_1}) \cdot \cdot \cdot (q_n, w_{q_n})\}$ と表現することができる。この分布 P, Q 間において EMD により最小化するコスト関数は以下の式(2)で表せる。

$$WORK = \sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij} \quad (2)$$

式(3)において d_{ij} は特徴量間の距離であり、 f_{ij} は p_i から q_j への輸送量である。 d_{ij} は特徴量により一様に決定できることから、式(2)は f_{ij} を最小にしたときに最小となる。また式(2)のコスト関数を最小にするとき、以下の 4 つの制約条件を満たす必要がある。

$$f_{ij} \geq 0 \quad 1 \leq i \leq m, 1 \leq j \leq n \quad (3)$$

$$\sum_{j=1}^n f_{ij} \leq w_{p_i} \quad 1 \leq i \leq m \quad (4)$$

$$\sum_{i=1}^m f_{ij} \leq w_{q_j} \quad 1 \leq j \leq n \quad (5)$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j} \right) \quad (6)$$

式(3)は輸送量 f_{ij} が正の値を取るという条件であり、式(4)と式(5)は輸送元や輸送先の重みを超えて輸送できないという条件である。また、式(6)は輸送量はどちらかの分布の重みの総和の少ない方に制限されるという条件である。以上の制約条件を満たしコスト関数を最小にする最適な f_{ij} を f^*_{ij} とすると、EMD が定義する分布 P, Q 間の距離は以下の式(7)で表すことができる。

$$EMD(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n d_{ij} f^*_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f^*_{ij}} \quad (7)$$

本手法では、EMD に用いる特徴量を Word2Vec から得られる単語ベクトルを用いた。また、 d_{ij} は単語ベクトル間のコサイン類似度となり、重みには tf-idf によ

る単語重要度を用いた。また、得られたコサイン類似度を 0 から 1 の範囲に正規化し、分布毎に単語重要度の総和を 1.0 となるように正規化することで式(7)で得られる EMD による距離を 0 から 1 の範囲に正規化している。

3.2.2. 単語アライメント

本手法では、異言語文間の類似度を計算する際に単語アライメントを行う。EMD によって定義される距離は特徴量間の距離の総和を用いて文間の距離を計算しているが、文間の類似度を計算するうえで不必要な単語間の計算も多く、それらは類似度を計算する際のノイズになってしまう。そこで単語アライメントを行い対応関係のある単語間のみを類似度計算に用いることでノイズを減らし、EMD の計算の精度を向上させることが可能となった^[2]。

単語アライメントには、単語ベクトル間のコサイン類似度と単語間の共起情報から計算される DICE 係数を用いてアライメントスコアを定義し使用した。アライメントスコアに DICE 係数を用いたのはマッピングを行った単語ベクトル間では類似度計算の精度が十分ではないと考えたからである。DICE 係数を用いることによってより強固なアライメントを行うことが可能となる。本手法では、北村らが提案した重み付き DICE 係数^[9]を基にした値を用いている。これは出現頻度が高い単語に重みをつけることによって高頻出語と低頻出語を区別する目的がある。全ての単語間でアライメントスコアを計算し、アライメントスコアがしきい値 t を上回った場合は単語間に対応関係があるとみなしアライメントを行い、アライメントスコアがしきい値 t を下回った場合は単語間に対応関係がないとして単語間の距離 d_{ij} をコサイン類似度が 0 のときの値 1.0 とする。アライメントスコアの計算式を以下に示す。

$$\text{alignment score}(x, y) = \frac{\cos(x, y) + wDICE(x, y)}{2.0} \quad (8)$$

$$wDICE(x, y) = \log_2(1 + f_{xy}) \times \frac{2 \times f_{xy}}{f_x + f_y} \quad (9)$$

$$d_{ij} = \begin{cases} 1.0 & t < \text{alignment score}(x, y) \\ 1.0 - \cos(x, y) & \text{alignment score}(x, y) < t \end{cases} \quad (10)$$

ここで f_x は単語 x が対象文全文で出現した回数であり、 f_y は単語 y が訳文全文で出現した回数である。また、 f_{xy} は対応する対象文と候補文のペアに単語 x と単語 y が同時に出現した回数である。

アライメントによって得られた式(10)の d_{ij} を式(7)に代入することで二文間の EMD の距離を計算できる。式(7)で得られた距離を用いて EMD による文間の類似度を以下の式(12)によって定義する。(0 < EMD < 1)

$$\text{類似度} = 1 - EMD \quad (11)$$

4. 性能評価実験

4.1. 実験データ

翻訳対象文には川端康成の「古都」を原文とした「The Old Capital」^[10]を用いた。対象文のデータに小説を用いたのは、Word2Vecによる分散表現を評価するために意識が含まれる小説が適当であると考えたためである。また、訳文は「Google 翻訳」^[11]「Microsoft Translator」^[12]「Excite 翻訳」^[13]の3つの機械翻訳サイトから獲得したものを使用した。高い精度を持つとされるニューラル機械翻訳を使用している「Google 翻訳」と「Microsoft translator」に統計的機械翻訳を使用している「Excite 翻訳」を加えることによって、異なる3つの訳文が得られる。

さらに、Word2Vecの学習コーパスはWikipediaのダンプデータ^[14]を用いた。本来学習コーパスには小説が考えられるが、英語と日本語において一貫性のある十分な量の小説のデータを集めることができなかったため、Wikipediaのダンプデータを学習コーパスに用いた。学習コーパスは、英語版Wikipediaが1.3GB、日本語版Wikipediaが850MBであった。ここで、学習コーパスのサイズが言語毎に異なるのは、Wikipediaの記事は言語毎に内容や記事数が異なるページがあり、英語版Wikipediaの方が記事数は多いためである。翻訳行列の学習に用いた単語ペアは英語版Wikipediaの頻出単語上位8,000語とGoogle翻訳より得た対訳語のペアとした。そのうち、Word2Vecにおいて未知語となった単語は2,433語含まれていた。その際未知語となった単語は学習に用いなかった。

また類似度を計算する際の特徴量にはWord2Vecから得られた単語ベクトルを使用し、特徴量の重みには対象文全文と3つの訳文全文それぞれに対してtf-idfを文単位で対応させて獲得した単語重要度を使用した。またアンケートによって8人の被験者(20代理系大学院生8名(男性6名、女性2名))に人手で3つの訳文を評価してもらうことで正解データを作成した。評価は5段階で点数をつけてもらい、評価点の平均点を訳文の点数とし、点数を基に訳文を順位付けしたものを正解順位とした。その際、翻訳対象文、訳文、参照訳を確認してもらいながら被験者に評価してもらった。

4.2. 実験方法

翻訳対象文とした小説の冒頭100文を用いて、1文ずつ異なる機械翻訳サイトによって翻訳を行い3つの訳文を獲得した。得られた3つの訳文と翻訳対象文間で、それぞれ異言語文間類似度を計算し訳文と翻訳対象文の類似度を基に訳文を順位付けした。システムから得られた順位と人手によって得られた正解の順位との相関を順位が完全に一致しているかの完全一致数、また訳文の順位の大小関係の相関をケンドールの順位

相関係数を用いて求めた。また、しきい値を0から2の範囲で変化させながら同様に実験を行い最適なしきい値について考察した。ここで、しきい値を0から2としたのはDICE係数に重みをかけたことによって、アライメントスコアが1を超えてしまう場合があったためである。

比較手法として、提案手法における単語アライメントの有効性を示すために、単語アライメントを行わず異言語文間類似度を求める手法、単語アライメントを行うが重みを付与しないDICE係数を単語アライメントスコアに用いる手法、提案手法の有効性を示すために既存の文単位の自動評価法であるMETEOR^[15]と英日において高い精度を持つとされるRIBESの4つの比較手法でも同様に訳文を順位付けし正解との相関を評価した。

4.3. 実験結果

表1に提案手法と比較手法を用いて実験をして得られた一致数と順位相関係数を示す。提案手法の結果はアライメントスコアのしきい値を0.73としたときの値であり、重みなし提案手法で用いたアライメントスコアのしきい値は0.62とした。これらの値は両手法の一致数と順位相関が最も良かった場合のしきい値である。

表1 実験結果

	一致数	順位相関
提案手法	22	0.090
重みなし提案手法	20	0.08080
アライメントなし提案手法	13	-0.17006
METEOR	20	0.06206
RIBES	24	0.10204

表1に示すように、提案手法はアライメントなしの手法と比べると一致数、及び順位相関に関して高い結果であることが確認された。また、既存の自動評価法についても、METEORと比較すると提案手法は一致数、及び順位相関係数に関して上回る結果となった。これに対し、RIBESと比較するとRIBESには一致数、及び順位相関双方に関して提案手法が追従する結果となった。参照訳を用いる自動評価法に対し参照訳を用いない提案手法は同等の結果となり提案手法の優位性を確認した。また、提案手法におけるアライメントスコアのしきい値を変化させたときの一致数と順位相関係数の推移についてのグラフを図2、図3に示す。

4.4. 考察

しきい値を変化させながら実験を行った結果、提案手法では0.73で最大の一致数と相関係数となった。図2よりしきい値と一致数の推移を見ると0.72で最大になった後徐々に減少していく。これはしきい値を上げていくことで、全く対応関係のない単語間の類似度

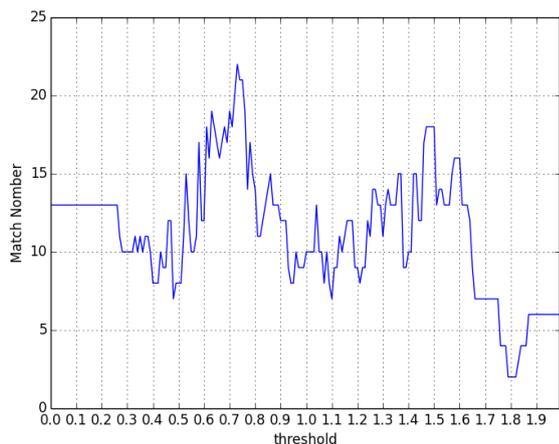


図 2 しきい値と一致数の推移

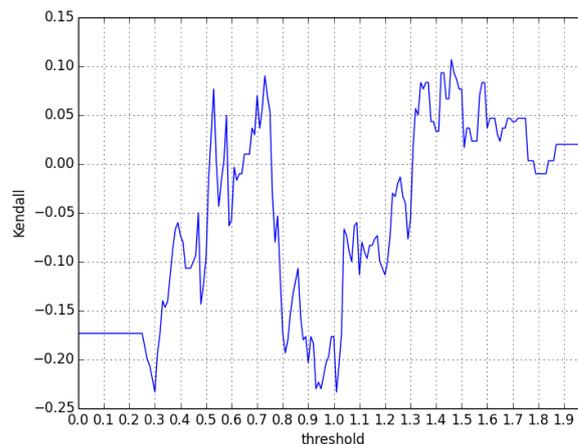


図 3 しきい値と順位相関係数の推移

表 2 訳文と類似度の例

翻訳対象文	"Ah. They've bloomed again this year,"she said as she encountered the gentleness of spring.
参照訳	「ああ、今年も咲いた。」と、千重子は春のやさしさに出会った。
訳 1	"ああ、今年もう一度咲きました"と彼女は春の優しさに遭遇した
訳 2	"ああ、彼らは今年も咲いている"と彼女は春の優しさに遭遇したと述べた
訳 3	「ああ。それらは今年再び咲き、彼女が春の親切に遭遇すると、"彼女は言った。

提案手法		RIBES		正解順位
順位	類似度	順位	類似度	
訳 3	0.396651	訳 1	0.801306	訳 1
訳 1	0.278574	訳 2	0.782542	訳 2
訳 2	0.261096	訳 3	0.590952	訳 3

が文間類似度計算に用いられなくなったためノイズが減ったことにより一致数が上がったと考えられる。しかし、それ以上しきい値を上げると類似度計算に必要な情報まで排除されてしまうため一致数は減少していると考えられる。実際、アライメントスコアにおいて出現回数が 1 回しかない単語間では重み付き DICE 係数は 1.0 となってしまう。Word2Vec によるコサイン類似度は 0.5 近傍なので、このことから最適なしきい値を超えると適当な単語アライメントまで排除されてしまうということが確認できる。

次に、しきい値と相関係数の推移をみるとしきい値 0.73 のときの順位相関係数は 0.090 となっているが、しきい値 1.46 ではそれよりも高い順位相関係数 0.10666 となっている。一見有効な値に見えるが、完全一致数は 17 件と少なくなっている。これはしきい値を上げていくことによりアライメントされる単語が極端に少なくなり、内容語をほとんどアライメントで

きておらず DICE 係数が高くなりやすい機能語や記号などだけで評価を行っているためであると考えられる。実際に全くアライメントがされず、異言語文間類似度が 0 となった文を訳文に含んでいたものが 100 件中 19 件あった。このことから、しきい値 1.46 よりも 0.73 のほうが適していることがわかる。

次に、提案手法と自動評価法から得られた結果について考察する。提案手法は METEOR の結果を上回り RIBES の結果を超えられなかったが追従することができた。まず自動評価法によって得られた結果がそもそも低い理由として、実験で参照訳として用いた小説の文自体に意識や難しい言い回しが多く、参照訳との表層的な一致を基にした自動評価法では判定が難しい文が多く含まれていたためと考えられる。つまり、こういった参照訳の質が十分担保されていないような文に対しては参照訳を用いずに対象文と直接比較する提案手法の方が有効な場合があることが確認できた。提案手

法の結果より RIBES が勝っていた原因として RIBES は統計的機械翻訳である Excite 翻訳を評価するのに有効であったためと考えられる。表 2 に用いた翻訳対象文と訳文、またそれに伴う提案手法と RIBES の類似度と正解順位を示す。表 2 の例から RIBES は人手で見て評価が低い訳 3 を低く評価しているのに対して提案手法は訳 3 の評価を高くしてしまっている。これは、提案手法は単語間の距離を基に訳文を評価していることから直訳調の訳文の評価を高くしてしまうという弱点によるものであると考えられる。ここで、訳 3 に用いられている Excite 翻訳は統計的な機械翻訳システムであるため翻訳が難しい文については直訳調の訳文が出力されてしまう。RIBES は統計的機械翻訳の文に対して高い精度で評価を行えるが、提案手法はむしろ直訳調の文のスコアを高く出してしまうため RIBES のほうが相関は高くなった。よって、直訳調の訳文にペナルティを課すことができれば RIBES の結果を超えることが期待される。

次に、提案手法における単語アライメントの精度について考察する。単語アライメントを改良することによって実験結果が従来の我々の手法より一致数で 2 件、順位相関係数で 0.0092 ポイント上回った。これは文間類似度の計算に必要な内容語のアライメントをより多く獲得することができたためである。だが、従来の我々の提案手法と比べると機能語のアライメントも増えてしまっているという問題があった。これは DICE 係数の高頻出語に重みをかけた結果、文書で高い頻度で出現する機能語に強い重みがかかってしまいアライメントスコアが高くなってしまったためであると考えられる。例えば、表 2 の翻訳対象文と訳 1 では”the”と”は”はアライメントスコアが 1.6336 となり”the”と”の”はアライメントスコアが 1.2026 となりアライメントされている。しかし、機能語はアライメントされてしまっても tf-idf 法の重み付けにより内容語と比べて非常に低い重みとなっているので EMD の計算を行う際には結果に悪影響を与えていないことから、大きな問題とならず提案手法は従来の手法と比べて結果が改善されたと考えられる。

5. まとめと今後の課題

参照訳を用いずに異言語間の類似度を計算することで訳文を評価する手法を改良した。その結果、従来の我々の手法よりも一致数では 2 件、順位相関係数では 0.0092 ポイント良い精度になった。また、参照訳を用いている自動評価手法の一つ METEOR を順位相関係数では 0.028 ポイント、一致数では 2 件上回り、RIBES に追従する結果を確認できた。提案手法は参照訳を用いずに、参照訳を用いている自動評価法の結果より良いか、同等の性能であり提案手法は有効である

ことが確認された。

今後は、単語アライメントをより精度良く改善していくことが考えられる。単語アライメントの際にアライメントスコアがしきい値を超えた単語全てを対応付けるのではなく、品詞等の情報からさらに単語の対応関係を絞っていくことで、手法を改善することが考えられる。他にも、統計的機械翻訳の評価に対しては RIBES が有効的に働いていたことから語順情報を考慮して類似度計算を行うことによって提案手法の精度をより上げていくことが可能である。

参考文献

- [1] Y. Wu, M. Schuster, Z. Chen, QV. Le, M. Norouzi, W. Macherey, ... and J. Klingner, "Google's neural machine translation system: Bridging the gap between human and machine translation.", arXiv preprint arXiv:1609.08144, 2016.
- [2] 藤川寛基, 越前谷博, 荒木健治, "単語の分散表現を用いた異言語文間類似度に基づく最適訳選択", 第 16 回情報科学技術フォーラム(FIT2017)講演論文集, 第 2 分冊, pp.185-186, 2017.
- [3] K Papineni, S Roukos, T Ward, W Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation", In Proceedings of the 40th Annual Meeting of the Association. for Computational Linguistics, 311-318, 2002.
- [4] 平尾努, 磯崎秀樹, Duh, K., 須藤克仁, 塚田元, 永田昌明, "RIBES: 順位相関に基づく翻訳の自動評価法", 言語処理学会第 17 年次大会発表論文集, pp.1111-1114, 2011.
- [5] A.J.Yepes, A.Névél, M.Neves, K.Verspoor, O.Bojar, A.Boyer... P.Pecina, "Findings of the WMT 2017 biomedical translation shared task." In Proceedings of the Second Conference on Machine Translation, pp. 234-247, 2017.
- [6] 柳本 豪一, "単語の分散表現を利用した文書類似度", The 29th Annual Conference of the Japanese Society for Artificial Intelligence, 2015.
- [7] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, "Distributed representations of words and phrases and their compositionality", In Advances in neural information processing systems ,pp.3111-3119, 2013.
- [8] T Mikolov , QV Le , I Sutskever , "Exploiting similarities among languages for machine translation", arXiv preprint arXiv:1309.4168,2013.
- [9] 北村美穂子, 松本裕治, "対訳コーパスを利用した対訳表現の自動抽出", 情報処理学会論文誌, Vol.38, No.4, pp.727-736, 1997.
- [10] Yasunari Kawabata, "The Old Capital", J Martin Holman, pp.3-8, Tuttle Publishing, New York, 1987.
- [11] Google 翻訳 <https://translate.google.co.jp>
- [12] MicrosoftTranslator <https://translator.microsoft.com/neural>
- [13] エキサイト翻訳 <http://www.excite.co.jp/world/>
- [14] WikipediaDownloads <https://dumps.wikimedia.org>
- [15] MDA Lavie "Meteor universal: Language specific translation evaluation for any target language", ACL2014, 2014.