

## 小論文採点支援のための関連文書取得法の考察

泉仁 宏太<sup>†</sup> 竹内 孔一<sup>†</sup> 大野 雅幸<sup>†</sup> 田口 雅弘<sup>†</sup> 稲田 佳彦<sup>†</sup>  
飯塚 誠也<sup>†</sup> 阿保 達彦<sup>†</sup> 上田 均<sup>†</sup>

<sup>†</sup> 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中 3-1-1

E-mail: †pm9n6cei@s.okayama-u.ac.jp, ††koichi@cl.cs.okayama-u.ac.jp, †††pw2z9792@s.okayama-u.ac.jp

あらまし 本研究では、小論文採点システムにおいて必要となる小論文に関連した文書を取得する方法を開発した。本研究プロジェクトの自動採点の評価軸の1つに「妥当性」がある。妥当性の評価手法として、小論文の内容がWikipediaの文書の内容と、どの程度一致しているかを基準に妥当性スコアを算出する方法を考えている。しかし、Wikipediaの文書は多様であり、小論文で取り上げていない議題に関する文書も多く存在する。そこで本論文では小論文ごとに適切な文書を取得する方法を提案する。いくつかの手法を試した結果、単語ベクトルを使用した方法が、関連した文書を獲得することができたことを報告する。

キーワード 小論文の自動採点, 単語ベクトル, Skip-gram, Wikipedia

## A Study of Extracting Related Documents for Essay Evaluation Modules

Kota MOTOJIN<sup>†</sup>, Koichi TAKEUCHI<sup>†</sup>, Masayuki OHNO<sup>†</sup>, Masahiro TAGUCHI<sup>†</sup>, Yoshihiko  
INADA<sup>†</sup>, Masaya IIZUKA<sup>†</sup>, Tatsuhiko ABO<sup>†</sup>, and Hitoshi UEDA<sup>†</sup>

<sup>†</sup> Graduate School of Natural Science and Technology, Okayama University Tsushima-naka 3-1-1, Kita-ku,  
Okayama-shi, Okayama, 700-8530 Japan

E-mail: †pm9n6cei@s.okayama-u.ac.jp, ††koichi@cl.cs.okayama-u.ac.jp, †††pw2z9792@s.okayama-u.ac.jp

**Abstract** We are developing an automatic Japanese essay-scoring system that is composed of 4 evaluation criteria, comprehensiveness, logical consistency, validity, spelling and grammar. In this paper, we discuss the most powerful approach to extract documents of Wikipedia that relates to the reference texts of the target essay theme for validity evaluation. The reason for using Wikipedia documents for evaluating validity of students' essays is that we assume that validity can be evaluated by the expanded discussions in Wikipedia documents that relates to the essay theme. Experimental results show that the skip-gram based word vector is the best approach to extract relating documents to reference texts among several keyword-based evaluation approaches.

**Key words** Automatic scoring of answers of essay-writing tests, Word vector, Skip-gram, Wikipedia

### 1. はじめに

現在小論文の多くは人手によって採点されている。しかし、小論文の量が多ければ採点者には大きな負荷がかかる上、人手だと一貫した評価も難しい。そこで、採点の基準となる要素を分解し、システムとして構築することで、システムによる一貫した評価が示され採点作業を支援することが可能となる。本研究のプロジェクトではその小論文の自動採点システムの開発を目指している。

小論文のような記述式の問題は2つに大きく分けることが出来る [1]。1つ目は正解の文章が存在するものでこれを短答式

タイプと呼ぶ。2つ目は正解の文章が存在しないものでこれをエッセイタイプと呼ぶ。

海外では小論文の採点システムとして E-rater<sup>(注1)</sup> を採点エンジンに用いた Criterion [2], 記述式問題の採点のために開発された IntelliMetric [3] などがある。前者は特徴量のスコアに基づいた重回帰モデル, 後者は決定木分析に基づくルール発見により小論文を評価している。日本語に関しては、短答式タイプのものについて機械学習を利用した手法 [1], [4] がある。エッセイタイプには Jess [5] がある。Jess は毎日新聞の社説やコラ

(注1) : <https://www.ets.org/erater/about>

ムから学習を行い、統計的手法によって小論文を評価している。また、近年では小論文の要素（議論点）について評価を絞り込み最新手法を構築する研究が行われている [6]。

本研究プロジェクト [7] では、小論文を評価するために 4 つの評価モジュールを構築している。そのモジュールはそれぞれ (1) 理解力、(2) 論理性、(3) 妥当性、(4) 文章力である。その中でも本論文では妥当性に注目する。妥当性の評価手法として、小論文の内容が Wikipedia の文書の内容と、どの程度一致しているかを基準に妥当性スコアを算出する方法を考えている。しかし、Wikipedia には非常に多くの文書があり全ての文書と比較するのは、非常に時間がかかる。そもそも小論文の議論に関係のない文書の場合、一致度を測るのは無意味である。そこで本論文では小論文ごとに適切な文書を取得する方法をいくつか提案する。そして、その方法でどのような文書が取得できたか、また取得した文書を基に実際的小論文の妥当性を評価し、人手で採点した点数とどの程度相関が見られたかについて述べる。

## 2. 自動採点システムの全体像と妥当性モジュール

### 2.1 自動採点システムの全体像

自動採点システムの全体像について説明する。図 1 は自動採点システムの構成について示したものである。

小論文を採点するにあたって、さまざまな評価軸が考えられる。石川 [8] や Steendam [9] の著書には、一般的なよい文書の基準は設定できないと主張している。つまり、すべてのテストに一貫して評価できる評価軸は存在せず状況に応じて各々の評価軸を決める必要があるということである。そこで、本研究プロジェクトでは受講生の記述能力を測る軸として、理解力（小論文が課題に即しているかどうか）、論理性（小論文が論理的に書かれているかどうか）、妥当性（小論文の内容は合っているかどうか）、文章力（誤字脱字などの文字誤りや文法が正しく使えているかどうか）の 4 つの評価軸を設定した。図 1 に記されている 4 つのモジュールが設定した評価軸である。これらの評価軸でそれぞれ小論文を採点し、採点した結果とそれらの合計点を出す。

自動採点システムが採点をする時に使用するデータとして、

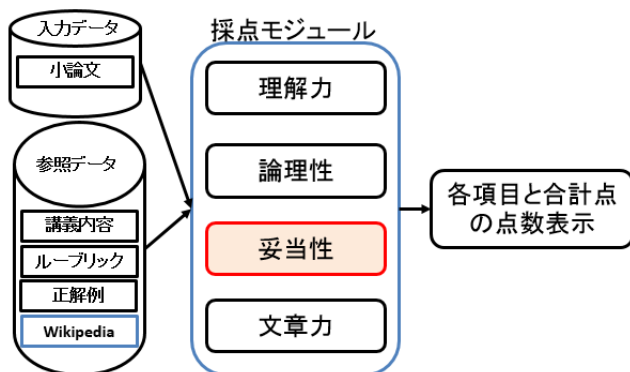


図 1 自動採点システムの構成

講義内容テキスト、ループリック、正解例、Wikipedia がある。これには自動採点システムの開発にともない構築した模擬試験データを使用する。この模擬試験は受講生に講義を受講してもらい、その後に講義に関する各課題に対して論述してもらおうという内容になっている。講義や課題の内容については 4 章で説明する。この時に開かれた講義内容を書きおこしたものを (2500 字程度) を講義内容テキスト、各課題の出題意図と人手で採点する時に使用する評価のポイントが書かれたものをループリック、受講生の回答を入力データの小論文として使用する。正解例に関しては、仮定できる課題については作成し、仮定できない場合は人手で採点した結果一番高い点数を出した回答を正解例とする。

### 2.2 妥当性モジュール

参照データの 1 つである Wikipedia は妥当性評価モジュールで使用する。妥当性評価モジュールでは、小論文の内容は合っているかどうか、すなわち小論文に書かれている内容が世の中でも言われているかどうかを評価する。これは小論文内で記述されている説明の根拠、事例が正しく述べられているかを確認するためである。世の中でも言われているかどうかの参照データとして Wikipedia を使用することにした。Wikipedia の文書を利用する理由として、無料で大量のテキストデータが入手可能であること、間違いはあるかもしれないが多くの人が編集しているため比較的妥当な文章が書かれていると仮定したためである。妥当性評価モジュールでは、小論文と Wikipedia の文書を比較して同じような内容が書かれていればその文章は妥当だと評価するようにした。しかし、Wikipedia の文書は多様な文書がある反面、課題で取り上げるべき議題とはまったく関係ない文書も多く存在する。そのような文書と小論文を比較しても内容が一致しないのは明確である上、時間の無駄である。それを避けるために小論文の課題に関係した文書のみを取得する必要がある。例えば、「多国籍企業とグローバリゼーション」に関する課題があった場合、これらに関連した文書を獲得することが望ましい。

そこで本研究では、課題に即した関連文書の取得方法をいくつか開発した。次章では、開発した関連文書の取得方法について記述する。

## 3. 関連文書取得法

本研究で開発した関連文書の取得法は大きく分けて 2 種類に分類される。それは Wikipedia の文書の本文を基に関連文書を取得するか、本文以外の情報を基に取得するかに分けられる。3.1 節と 3.2 節では Wikipedia の本文以外の情報を基に関連文書を取得する方法を、3.3 節と 3.4 節では本文を基に取得する方法を記述する。

### 3.1 Wikipedia のカテゴリーリンクを使用した方法

本節では Wikipedia の本文以外の情報を基に関連文書を取得する方法の 1 つ目を記述する。カテゴリーリンクとはほぼ全ての Wikipedia の文書をなんらかのカテゴリーに分類したデータである。これを使用することで単なる単語検索では見つからない課題に関連した特定のつながりを持った文書を取得できるだ

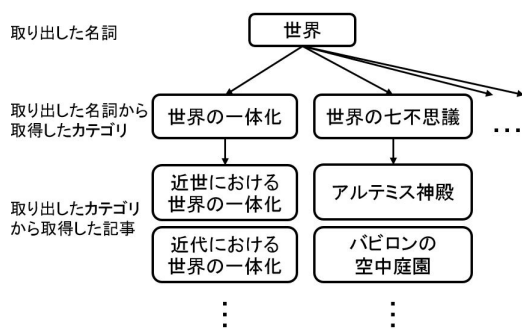


図2 カテゴリーリンクを使用した Wikipedia の文書取得までの流れの例

ろうと考えた。例えば、「マクドナルド」と「トヨタ」は多国籍企業という点で同じカテゴリーに属するが、単なる単語検索でこの2文書を取得するのは難しい。文書を取得するまでの具体的な処理の流れを以下に示す。

まず、小論文の正解例の形態素解析を行う。本研究では、形態素解析に MeCab<sup>(注2)</sup>を用いた。その後、解析したものの中からキーワードとなり得ない数字や「こと」や「もの」のようにそれだけで意味をなさない非自立語を除く名詞のみを取り出す。取り出した名詞を基にカテゴリーリンクの部分一致検索でヒットしたカテゴリーを取得する。その後、ヒットしたカテゴリーに属する全ての Wikipedia の文書を取得する。名詞を取り出しから文書を取得するまでの例を図2に示す。

ここで「国」などの名詞によっては数十万以上もの文書を取得することがあり、キーワードで取得する文書を絞ることができない場合がある。これを防ぐために1つの名詞で1500を超える文書を取得するものに対して、そのキーワードで取得した全ての文書を排除するようにした。

### 3.2 講義内容などでよく使われる名詞を考慮した Wikipedia のタイトル部分一致検索法

本節では Wikipedia の本文以外の情報を基に関連文書を取得する方法の2つ目を記述する。検索の仕方に工夫はあるものの単純なタイトル検索でどの程度の精度で文書を取得できるか確認するためにこの方法を提案する。検索の仕方だが、これは講義内容などからよく使われる名詞でタイトル検索を行うというものである。講義内容などで多く使われる名詞は、その議題で重要な単語でありその単語を含むタイトルの文書もまたその議題に関連した文書であろうと考えたためである。文書を取得するまでの具体的な処理の流れを以下に示す。

まず、講義内容、ループリック、質問文を形態素解析し名詞を取り出し頻度を数え上げる。この時、汎用語が頻度をあげないようにキーワードフィルタリングを行う。フィルタリング規則は2つある。1つ目の規則は文字数であり、1文字の名詞は内容に関わらず削除した。1文字の名詞でタイトルの部分一致検索を行うと大量の文書を取得してしまう恐れがあるためである。2つ目の規則は品詞分類であり、名詞の中でも数、非自立、接尾、代名詞に分類される名詞は削除した。これらの名詞はど

の文章でも使われる名詞であるためである。取り出した名詞集合を1つずつ使って Wikipedia のタイトルの部分一致検索を行う。

次に検索によって取得した各文書に重みをつける。タイトルの部分一致検索を行っていく中で何度も検索される Wikipedia の文書がある。その文書を取り出した名詞の頻度を考慮して重みとして計算する。

$$w = \sum_{i=1}^N frequency_i \quad (1)$$

$w$  は重み、 $N$  はタイトルと部分一致した名詞の総数、 $frequency$  はその名詞が講義内容などで使われた回数である。例えば、 $frequency$  が3の「世界」と  $frequency$  が1の「企業」で「世界の企業」の文書を検索した時の「世界の企業」の重みは  $3 + 1 = 4$  となる。

文字列の長いタイトルは多くの名詞とマッチする可能性があり重みが高くなる。そこでページの重みをタイトルの文字列の形態素の数で割ったものを重要度とする。それによって各文書の重要度を計算した後、重要度の高い上位1000件の文書を本節の方法で取得した文書とする。

### 3.3 講義内容などでよく使われる名詞から頻度ベクトルを作成し、同様の方法で Wikipedia の文書をベクトル化しそれとコサイン類似度を求める方法

本節では Wikipedia の本文の情報を基に関連文書を取得する方法の1つ目を記述する。前節でも述べたように講義内容などで多く使われる名詞は、その議題を説明する重要な要素になりうる。つまり講義内容などと同じようにそれらの単語を多く使っている Wikipedia の文書は講義内容などと同じ議題について論じてあるだろうと考えた。そのために講義内容などから取り出した名詞を基に基本となる頻度ベクトルを作成し、Wikipedia の文書も同様の名詞でベクトル化しコサイン類似度を測って、類似した文書を取り出すという方法を提案する。文書を取得するまでの具体的な処理の流れを以下に示す。

まず講義内容、ループリック、質問文から前節と同様の方法でフィルタリングを行い名詞を取り出す。その中から最も出現頻度の高い上位20個の名詞を取り出す。各名詞の出現頻度を各ベクトルの要素とし、20次元の頻度ベクトルを作成する。次に Wikipedia にも同様の処理を行う。Wikipedia の1文書の本文から講義内容などから取り出した最も出現頻度の高い上位20個の名詞のみを取り出す。各名詞の Wikipedia 本文で出現した回数を各ベクトルの要素とし、20次元のベクトルを作成する。この作業をすべての Wikipedia の文書に対して行う。その後、講義内容などから作成したベクトルと Wikipedia の各文書のベクトルとのコサイン類似度を測る。そして、類似度の高い上位1000件の文書を本節の方法で取得した文書とする。

### 3.4 講義内容などと Wikipedia 本文との単語ベクトル和のコサイン類似度を求める方法

本節では Wikipedia の本文の情報を基に関連文書を取得する方法の2つ目を記述する。たとえ、出現回数が少なくともその議題にのみ現れるような独特な単語は、関連文書の取得に大き

(注2) : <http://taku910.github.io/mecab/>

な役割を果たすと考えられる。また、今まで部分一致の検索の都合上名詞のみに絞って見てきたが、同じ単語にも様々な意味がありどの意味に属するかを判断するために、名詞以外も見る必要がある。本手法では名詞と共起関係になりやすい形容詞と動詞を見ることにした。これらをまとめて、その単語が一般的に使用される頻度を考慮した講義内容などの本文と Wikipedia 本文との単語ベクトルのコサイン類似度を測って、類似した文書を取り出すという方法を提案する。文書を取得するまでの具体的な処理の流れを以下に示す。

まず、ある単語がどのくらい珍しい単語であるかを調べるために Wikipedia から IDF 辞書を作成する。Wikipedia の各文書を形態素解析し、数字を除く名詞、形容詞、動詞に分解する。その後、各単語の IDF 値を計算する。

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

$N$  は Wikipedia の全文書数を示す。 $df(t)$  はある単語  $t$  が出現する文書の数を示す。

次に講義内容と質問文を IDF 辞書を作成した時と同様の規則で単語に分解し、単語集合を作成した後、その単語集合の単語ベクトル和を求める。この時、単語ベクトルは  $nwjc2vec$ <sup>(注3)</sup> を使用する。 $nwjc2vec$  とは、国語研が 1 億程度のコーパスから  $word2vec$  で学習して作成した 300 次元の単語ベクトルのデータであり、160 万程度の単語が格納されている。単語ベクトル和は式 3 によって求める。

$$WordVectorSum = \sum_{i=1}^W IDF(t_i) \times V(t_i) \quad (3)$$

$W$  は取り出した単語集合の総数である。 $V(t)$  はある単語  $t$  の単語ベクトルである。なお、IDF 辞書または  $nwjc2vec$  に存在しない単語の単語ベクトルは 0 とする。

Wikipedia の各文書も同様に単語ベクトル和を求める。その後、講義内容と質問文の単語ベクトル和と Wikipedia 各文書の単語ベクトル和のコサイン類似度を求める。そして、類似度の高い上位 1000 件の文書を本節の方法で取得した文書とする。

#### 4. 評価実験

前章で説明した文書の取得法に関して、実際に取得した文書がどれくらい課題にそった精度の良いものなのかを実験を通して見ていく。まず、4.1 節で本研究の実験で使用したデータについて記述する。その後、4.2 節で前章の各方法でどのような文書を取得したかタイトルを見ながら評価する。最後に 4.3 節で取得してきた文書を基に小論文の採点を行う。

##### 4.1 本研究で使用するデータ

本研究にあたって模擬試験を実施した。模擬試験は受講生に講義を受講してもらい、その後に講義に関する各課題に対して論述してもらおうという内容になっている。この模擬試験を 2016 年 8 月に受講生を岡山大学の学生に限定し、岡山大学内で「グローバリゼーションの光と影」という講義で実施した。出題さ

れた課題を次に示す。

**課題 1:** グローバリゼーションは、世界、または各国の所得格差をどのように変化させましたか。また、なぜ所得格差拡大、または縮小の現象が現れたと考えますか。300 字以内で答えなさい。

**課題 2:** 多国籍企業は、グローバリゼーションの進展の中でどのような役割を果たしましたか。多国籍企業の実例をあげて、250 字以内で答えなさい。

**課題 3:** 文化のグローバリゼーションは、私たちの生活にどのような影響を与えましたか。また、あなたはそれをどのように評価しますか。具体例をあげて、300 字以内で答えなさい。

全てエッセイタイプの課題ではあるが、課題 1 は解答がある程度絞られている問題で、課題 2 と 3 は自分の考えを中心に記述する問題である。本研究では課題提案者による採点が済んだ受講生の小論文を使用した。データの数は課題 1 が 158 個、課題 2 は 154 個、課題 3 は 155 個となっている。

##### 4.2 取得した文書のタイトル一覧

本節では、前章の各関連文書取得法で取得した文書の評価する。評価方法は、取得した文書の上位のタイトルを見て人手で判断する。

課題 2 について各関連文書取得法で取得した文書のタイトルの上位 10 件を表 1 に示す。なお、カテゴリーリンクを使用した方法は文書のランク付けを行っていないため、ランダムに選んだ 10 件の文書を表示する。取得してきた文書の良し悪しに関してだが、講義はグローバリゼーションが世界に及ぼした影響について、金、人、文化の側面而言及しておりそれによって経済格差が広まったという内容である。つまり、グローバリゼーションや経済格差について記述している文書を取得することが望ましい。それを踏まえて表 1 を見てみると、単語ベクトル和で取得した文書が最もよかった。この方法で取得した文書はほとんどどれもグローバルや経済格差のことを追究した文書であり議題に即した文書が取れていたと言える。これに比べて、他

表 1 課題 2 に対して各関連文書取得法で取得した文書の上位 10 件のタイトル (カテゴリーリンクはランダムに選んだ 10 件の文書)

	カテゴリーリンク	タイトルの部分一致検索	名詞の頻度ベクトル	単語ベクトル和
タイトル	アステラス製薬	世界文化社	名古屋国際中学校・高等学校	開発経済学
	網走新聞	世界銀行	大韓民国	経済的不平等
	アメリカンパレル	世界	サイパン島	グローバル資本主義
	ユニット・トラスト	銀世界	ニンジャスレイヤー	自由貿易
	下里公設市場	月世界本舗	最近の出来事 / 2015 年 1 月	空洞化
	谷口恵美	世界一	インド	貧困
	ヘラクレスで上場廃止となった企業一覧	月世界	稲盛和夫	自由貿易協定
	おおすみ半島コミュニティ放送ネットワーク	世界心道教	長坂寿久	進歩的活用理論
	アメリカ領ヴァージン諸島	世界銀行	フランスのファッション	東アジア共同体
	スプレッド取引	世界文化	エリート	マスツーリズム

(注3) : [http://pj.ninjal.ac.jp/corpus\\_center/nwjc/subscription.html](http://pj.ninjal.ac.jp/corpus_center/nwjc/subscription.html)

の3つの方法は議題とは関係ない文書が多く、関連する文書を獲得することに失敗しているといえる。そこで、他の方法との違いから単語ベクトル和で良質な文書が取れた原因を探る。

まず、カテゴリーリンクとタイトルの部分一致検索とは取り出した単語の扱い方が違う。前2つの方法は取り出した単語単体を使って処理していたが、単語ベクトル和の方法では単語同士のベクトルを足しあわせるという単語間の共起関係を見ていた。その結果、前2つでは同じ単語でも違う意味で使われている文書を取得してくることがあったが、単語ベクトル和の方ではそのようなことは少なかった。単語間の共起関係を見るのが有効なのは、タイトルの部分一致検索においても見られる。10位の「世界文化」は講義内容などでよく使われる名詞に複数マッチしており、この文書は議題に即している。10位以下の文書でも複数マッチしたものは議題に即した文書であることが多かった。

頻度ベクトルを使用する方法とはよく使われる単語の扱い方が違う。頻度ベクトルの方では講義内容などでよく使われる名詞を重要視してベクトルを作ったが、他の議題でもよく使われそうな名詞もいくつか取ってきてしまった。対して、単語ベクトル和の方では IDF を用いることで世間一般では使われないある議題においてのみ使われるような単語を重要視したため、汎用語を可能な限り削除できたのであろう。

これらのことから、単語間の共起関係と特定の議題にのみ出現する単語に注目すれば議題に沿った記事が取得できる可能性が高いということが分かる。カテゴリーリンク以外の3つの方法にいえることだが、講義内容に対して質問文の文字数が少ないせいか課題1, 2, 3と取得してくる文書にあまり違いは見られなかった。もし、違いを出すのであれば課題ごとに重要キーワードを設定しておき、その単語の重みを上げることが考えられる。

#### 4.3 単語マッチによる小論文の採点

本節では、前章の各方法で取得した文書で小論文を採点し評価する。採点方法は現在開発中であり、まだ確立していないため今回は簡単な単語マッチによる採点を行う。取得した文書は  $M$  文、小論文は  $N$  文から構成されているので、総当たりで「名詞」、「形容詞」、「動詞」の単語マッチを行いマッチした単語の数を数えたものを式4で定義するスコア  $S$  として出力する。

$$S = \sum_{m \in M} \sum_{n \in N} \text{WordMatch}(\text{Wikipedia}_m, \text{Essay}_n) \quad (4)$$

求めたスコア  $S$  と人手で採点したスコアとの相関を取ることで、取得した文書を評価する。その結果を表2に示す。

結果として、取得した文書は大きく違うものの単なる単語マッチによる採点では各方法ごとにあまり違いは見られなかつ

表2 各方法で取得した文書を用いて採点したスコアと人手のスコアとの相関係数

	カテゴリーリンク	タイトルの部分一致検索	名詞の頻度ベクトル	単語ベクトル和
課題1	0.0815	<b>0.0873</b>	0.0855	0.0631
課題2	0.306	0.314	0.300	<b>0.327</b>
課題3	0.194	<b>0.234</b>	0.211	0.226

た。おそらく簡単な単語マッチでは汎用語のマッチが多くおきているためである。前節の評価した中で最もよかった単語ベクトル和を使用した方法による相関係数は4つの中でも上位に位置するものの、相関として最も良かったのはタイトルの部分一致検索を行う方法であった。タイトルの部分一致検索では、講義内容などで多く使われる名詞を多く使った文書を取り出しているため、その名詞と多くマッチし汎用語の影響を抑えられたのだと考えられる。これらのことから、単語マッチで評価を行っていくのであれば、各課題ごとに重要視する単語に重みをつけて汎用語の影響を抑える工夫をする必要がある。

各課題の相関を見てみると、課題1は課題2, 3よりも低かった。課題1は答える文字数こそ多いものの短答式よりのタイプの問題であるのに対して、課題2, 3はエッセイタイプの問題である。つまり、本手法はエッセイタイプの問題の方が有効であるといえる。

## 5. おわりに

本論文では、採点支援システムの評価軸の1つである「妥当性」を測るための前準備として、質問文の議題に沿った文書を取得する方法をいくつか提案した。取得した文書のタイトルを見てみると、講義内容などと Wikipedia 本文との単語ベクトル和のコサイン類似度を求める方法が最も議題に沿った文書を取得できたことが実験で分かった。しかし、小論文の採点方法を単なる単語マッチにしたせいか、各取得方法による違いがあまり見られなかった。今後は妥当性を評価するための方法について開発を進めていくことにする。

## 文 献

- [1] 石岡恒憲, 亀田雅之, 劉東岳. 人工知能を利用した短答式記述採点支援システムの開発. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, pp. 87–92, 2016.
- [2] 石井雄隆, 近藤悠介. 英語学習者を対象とした自動採点システム—課題と展望—. 外国語メディア学会, pp. 1–11, 2013.
- [3] 石岡恒憲. 小論文およびエッセイの自動評価採点における研究動向. 人工知能学会誌, Vol. 23, pp. 17–24, 2008.
- [4] 寺田凛太郎, 久保頭大, 柴田知秀, 黒橋禎夫, 大久保智哉. ニューラルネットワークを用いた記述式問題の自動採点. 第22回言語処理学会年次大会発表論文集, pp. 370–373, 2016.
- [5] 石岡恒憲, 亀田雅之. コンピュータによる小論文の自動採点システム Jess の試作. 計算機統計学, Vol. 16, No. 1, pp. 3–19, 2003.
- [6] Isaac Persing and Vincent Ng. Modeling Argument Strength in Student Essays. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 543–552, 2015.
- [7] 竹内孔一, 大野雅幸, 泉仁宏太, 田口雅弘, 稲田佳彦, 飯塚誠也, 阿保達彦, 上田均. 小論文の自動採点に向けたオープンな基本データの構築および現段階での自動採点手法の評価. 言語処理学会第23回年次大会発表論文集, pp. 839–842, 2017.
- [8] 石川巧. 「いい文章」ってなんだ?—入試作文・小論文の歴史. 筑摩書房, 2010.
- [9] Elke van Steendam, Marion Tillema, Gert Rijlaarsdam, and Huub van den Bergh. *Measuring Writing: Recent Insights into Theory, Methodology and Practices*. Brill Academic Pub, 2012.