

問われている知識に着目した項目間類似度算出手法の 日本史分野への適用と評価

高木 輝彦^{1,a)} 伊藤 徹郎¹ 高木 正則²

概要: 我々はこれまで、e テスティングにおける類似項目の自動検索を目的として、項目間類似度の算出手法を提案し、情報分野の項目を対象として評価を行ってきた。類似度データや類似項目を用いることで、(1) 類似項目の自動検索、(2) 自動的なアイテム・バンクの構築、(3) 項目間構造の可視化、(4) 同一知識を問う項目の反復学習、(5) 項目作成時の選択肢候補の提示、などの支援が期待できる。これまでの提案手法は、ベクトル空間モデルを軸に、項目内で出現する「単語」をベクトルの要素とする手法と、それらの単語の共起性から推定される「トピック」をベクトルの要素とする手法の2つに分類される。これらの提案手法では、項目内で問われている知識に着目し、それらの知識の出現箇所(問題文、正答、誤答)を自動決定する処理を取り入れている。また、その処理の有効性が示されている。本稿では、先行研究で提案した類似度算出手法の日本史の項目への適用可能性を評価するために、学校教育のICT活用を支援するサービス「Classi」で出題されている日本史の項目を対象とし、類似項目の検索実験を行った。その結果、知識の出現箇所を自動決定する処理がない場合よりもある場合の方が再現率、適合率、F 尺度が高かった。さらに、先行研究で対象とした情報分野の項目よりも日本史の項目の方が再現率、適合率、F 尺度が高く、提案手法が日本史項目にも適用可能であることが示唆された。

キーワード: e テスティング, アイテム・バンク, 類似項目, 項目間類似度

Application and Evaluation of a Test Item Similarity Calculation Method Based on Questioned Knowledge in Japanese History

Abstract: We have proposed methods for retrieving similar items automatically in e-testing by calculating their similarity and conducted experimental evaluations in the information field. Similarity data or similar items can be applied to (1) automatically retrieving similar items, (2) automatically constructing item banks, (3) visualization of structures between items, (4) adaptive testing, and (5) supporting item creation. Our methods were based on a vector space model and were classified into two approaches. In one, each test item was represented by a vector using extracted terms from these test items; in the other, each test item was represented by a vector using topics probabilistically estimated by the co-occurrence relation between terms. We focused on the knowledge questioned in test items, and identified the part where the knowledge occurred (e.g., question, correct choice, wrong choice). Moreover, the method's effectiveness was demonstrated. Here, to evaluate the feasibility of applying our proposed methods to Japanese history, we targeted items about Japanese history tested by "Classi" and conducted experiments on similar item retrieval. The results showed accuracy improvements in recall, precision, and f-measure due to identifying the part. The experimental results also showed greater accuracy improvements in Japanese history than in the information field.

Keywords: e-testing, item bank, similar test item, item similarity

¹ Classi 株式会社
Classi Corp., 163-0415 Nishishinjuku, Shinjuku, Tokyo 2-1-1, Japan
² 岩手県立大学
Iwate Prefectural University, 152-52 Sugo, Takizawa-mura, Iwate 020-0193, Japan
^{a)} teruhiko.takagi@classi.jp

1. 研究の背景と目的

我々はこれまで、e テスティング [1] において類似項目を自動検索することを目的とし、項目間類似度の算出手法を提案してきた [2], [3]. 類似項目とは、「項目内で問われて

いる知識や解決の中心となる知識が一致する項目」と定義する。この知識とは、分野特有の専門用語であり、大きく単名詞と複名詞に分類されることが分かっている [2] (以下、対象知識)。表 1 に対象知識が「弥生土器」の類似項目の例を示す。なお、本研究では多肢選択式の項目を対象とする。

本研究は e テスティングにおける項目の管理の利便性や再利用性を向上させるための 1 つのアプローチとして位置付けられる。例えば、類似度データや類似項目を用いることで、以下の支援が可能となる。

- (1) アイテム・バンク内、または、アイテム・バンク間における類似項目の自動検索 [4].
- (2) クラスタリング手法 [5] などを適用した、自動的なアイテム・バンクの構築.
- (3) 多次元尺度法 (MDS: Multi Dimensional Scaling) [6] などによる項目間構造の可視化.
- (4) 同一知識を問う項目の反復学習における適応型テスト [7], [8].
- (5) 類似項目の選択肢を提示することによる項目の作成 [2].

先行研究では、提案手法の類似度算出の精度を評価するために、コンピュータネットワークに関する項目や初級システムアドミニストレータ試験 [9] (以下、初級シスアド) で出題された項目を対象とした類似項目の検索実験を行った。そして、既存手法よりもその精度の向上が示された。しかしながら、これらの提案手法が情報分野の項目以外で適用可能かどうか明らかになっていない。そこで、本稿では、他分野への適用可能性を評価することを目的とし、学校教育の ICT 活用を支援するサービス「Classi」[10] で、専門用語を問う項目が最も出題されている日本史の項目を対象とした類似項目の検索実験を行う。

表 1 「弥生土器」についての類似項目の例 *1

Table 1 Example of Similar Test Item Questioned about “Yayoi pottery”.

項目 1
薄手で堅く、赤褐色をしていて、東京の向ヶ丘貝塚にて発見されたものとして正しいものはどれか。
1. 縄文土器 2. 青銅器 3. 金属器 4. 弥生土器 (正答)
項目 2
弥生土器の用途について述べた文として正しいものはどれか。
1. 使うものに応じていくつかの基本形があったが、煮炊き用のものを壺という。(正答) 2. 貯蔵用のものを甕という。 3. 盛り付け用のものを高坏という。(正答) 4. 食物を蒸すためのものを石皿という。

*1 出典『Classi Web テスト問題』Classi 株式会社

2. 項目間類似度算出手法

2.1 項目間類似算出手法の概要

先行研究では、ベクトル空間モデル [11] を軸に 2 つの手法を提案している。1 つは、項目に出現する単語の重要度をベクトルの要素とする手法 (以下、単語ベースの手法) [2] である。もう 1 つは、それらの単語の共起性から推定されるトピックの重要度をベクトルの要素とする手法 (以下、トピックベースの手法) [3] である。トピックの推定では LDA (Latent Dirichlet Allocation) [12] を利用している。この 2 つの提案手法において、対象知識の出現箇所 (問題文、正答、誤答) を自動決定する処理を取り入れており、その処理の有効性が示されている。本章では、先行研究で我々が提案した対象知識出現箇所の自動決定手順と 2 つの提案手法について説明する。なお、これらの手順の考案に至った分析や妥当性の検証等については、文献 [2], [3] を参照して頂きたい。

2.2 対象知識出現箇所の自動決定

図 1 に対象知識出現箇所の自動決定手順を示す。図中の i は項目の番号で、項目数 N まで処理が繰り返される。また、図中の名詞とは、単名詞と複名詞に分類される。単名詞はそれ以上分割できない名詞 (例えば、「法隆寺」) で、複名詞は複数の名詞を連結して作られた名詞 (例えば、「ヤマト政権」) である。なお、形態素解析には MeCab[13] を使用し、複名詞への連結処理は termextract[14] を使用している。

まず、正答を形態素解析し、名詞を抽出する。抽出された名詞の数が 1 かつそれ以外の品詞の語が存在しない場合 (図 1: 1)、問題文に否定的な表現が含まれるかを判定する (図 1: 2)。否定的な表現とは「誤っている～」や「間違っただ～」など、問題文の内容と異なる概念や専門用語などを選択させるときに使用される表現である。一致する場合は問題文、一致しない場合は問題文と正答を対象知識の出現箇所とする。一方、正答に複数の名詞や名詞以外の語が存在

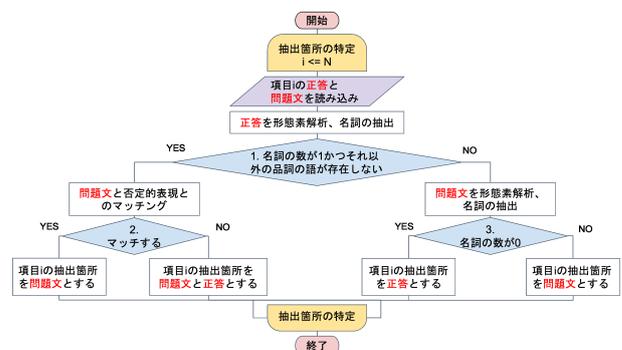


図 1 対象知識出現箇所の自動決定手順 [2]

Fig. 1 Procedure of automatically identifying the part of targeted knowledge occurrence.

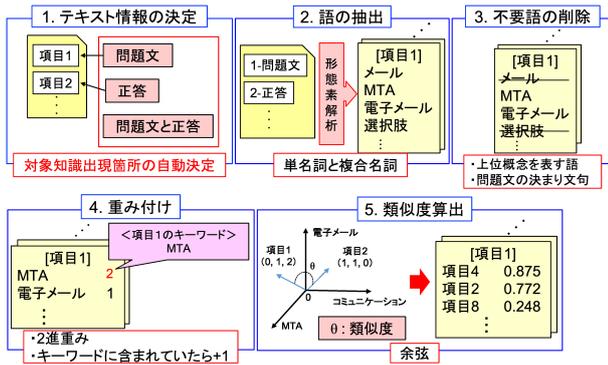


図 2 単語ベースの類似度算出手順 [2]

Fig. 2 Procedure of calculating test item similarity based on term.

在する場合 (図 1: 1), 問題文を形態素解析し, 名詞を抽出する. 抽出された語数が 0 の場合は正答, 1 以上の場合は問題文を対象知識の出現箇所とする (図 1: 3).

2.3 単語ベースの類似度算出手順

図 2 に単語ベースの類似度算出手順を示す. まず, 図 1 の手順に従い, 各項目の対象知識の出現箇所を決定する (図 2: 1). 次に, 項目ごとに決定された出現箇所を形態素解析し, 単名詞と複合名詞を抽出する (図 2: 2). 続いて, 対象知識と関係のない語を不要語として削除する (図 2: 3). 不要語の対象としては, 上位概念を表す語と問題文の決まり文句に含まれる語としている. 項目内では, 対象知識に対して言い換え表現による上位概念を表す語 [15] が多い (例えば, 「電子メール」に対して「メール」). これらの語は広域な意味を持ち, 解決の中心となる知識となることが少ないと考えられるため削除する. また, 項目の決まり文句に含まれる語 (「次」, 「以下」, 「選択肢」など) は, 項目内で問われる知識とは関係がないため削除する.

さらに, 抽出された語に対して重みを付与する (図 2: 4). しかし, 対象知識とその出現頻度や出現分布の関係性が明らかになっていないので, 現状は抽出されたすべての語に対して 1 を付与する (2 進重み [16]). また, 作成者により項目のキーワードのようなメタデータが付与されている場合, それらに含まれる語に対してさらに重み 1 を加える. 最後に, 語の重みによりベクトルで表された項目間の類似度を余弦 [11] により算出する (図 2: 5). x_i, y_i をそれぞれ項目 d_x, d_y から抽出された語 i に対する重み, T を抽出された語の総数とすると, 余弦による類似度 (d_x, d_y) は次のような式で表わされる.

$$\sigma(d_x, d_y) = \frac{\sum_{i=1}^T x_i \cdot y_i}{\sqrt{\sum_{i=1}^T x_i^2 \times \sum_{i=1}^T y_i^2}} \quad (1)$$

このように, 単語ベースの手法は, 予め対象知識が出現する箇所やそれらを表す語を絞り込むという点に特徴がある. 先行研究では, コンピュータネットワークや初級シス

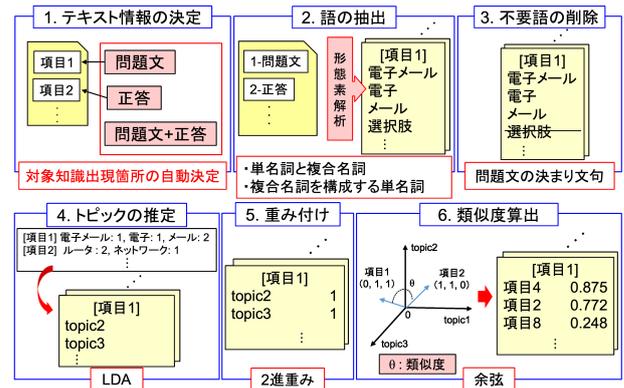


図 3 トピックベースの類似度算出手順 [3]

Fig. 3 Procedure of calculating test item similarity based on topic.

アドに関する項目を対象とした類似項目の検索実験を行った. 実験では, すべての項目情報 (問題文, 正答, 誤答) から語を抽出し, TFIDF [11] や名詞の接続頻度 [17] で語に重みを付与し, 余弦により類似度を算出する既存手法と比較した. 実験結果から, 既存手法よりも類似度の算出精度の向上や対象知識の出現箇所を自動決定することの有効性が示された.

2.4 トピックベースの類似度算出手順

前節で述べた単語ベースの手法では, 対象知識の表記ゆれによる誤検索が課題となり, 項目内で出現する語の内容理解に踏み込んだ処理の必要性が示唆された. これらの課題を解決するために, 項目集合に出現する語の共起性に基づき, 各項目のトピックを推定することが可能な LDA を用いた手法を提案した. これにより, 表記ゆれの語が出現したとしても, その他の語との共起性が類似していれば, 同じトピックの語として認識される.

図 3 にトピックベースの類似度算出手順を示す. まず, 図 1 の手順に従い, 各項目の対象知識の出現箇所を決定する (図 3: 1). 次に, 決定された出現箇所を形態素解析し, 単名詞と複合名詞, また, 複合名詞を構成する単名詞を抽出する (図 3: 2). 例えば, 項目内で「磁気ディスク装置」と「コンピュータ」が出現した場合, 抽出する語は「磁気ディスク装置, 磁気, ディスク, 装置, コンピュータ」とする. この処理を行うことで, 類似項目間で対象知識の表記ゆれが生じていたとしても (例えば, 一方の項目では「磁気ディスク装置」, もう一方の項目では「磁気ディスク」), これらの項目間で同一のトピックが推定されることが期待される. そして, 単語ベースの手法と同様に項目の決り文句に含まれる語を削除する (図 3: 3).

ここで, 抽出された語とその出現頻度を基に, 各項目を共起行列で表し, LDA によりトピックを推定する (図 3: 4). LDA では, 対象とする項目のトピック数は既知であり, 固定されるものと仮定される. 現段階では, 著者が予

め手動で項目から対象知識を抽出し、抽出された対象知識のユニーク数をトピック数として設定している。さらに、推定されたトピックに対して2進重みにより重みを付与する(図3:5)。最後に、トピックの重みを要素とするベクトルで表された項目間の類似度を余弦(式(1))により算出する(図3:6)。

このように、トピックベースの手法は、予め対象知識が出現する箇所を絞り込んだうえで、対象知識やそれらに関連する語の共起性を高める点に特徴がある。先行研究では、初級シスアドに関する項目を対象とした類似項目の検索実験を行い、前節で述べた既存手法や単語ベースの手法よりも、誤検索が大幅に解消され類似度の算出精度が向上し、対象知識の出現箇所を自動決定することの有効性やLDAを用いることの有効性が示された。

3. 研究目標

本研究では、項目間類似度の算出精度を再現率と適合率のマイクロ平均[11]により評価する。再現率は「類似項目が漏れなく検索されているか」、適合率は「類似項目だけが検索されているか」を評価する指標となっており、一般的に、再現率と適合率はトレードオフの関係にある。そこで、再現率と適合率をひとつのスカラ値に変換し、両方を考慮に入れる指標としてF尺度[11]を用いる。ゆえに、F尺度により総合的な項目間類似度の算出精度を評価できる。

ここで、再現率と適合率のマイクロ平均とF尺度の式を示す。項目*i*に対して、類似項目の数を A_i 、検索された項目の数を B_i 、検索された項目のうち類似項目と同じ項目の数を C_i とすると、再現率、適合率のマイクロ平均 \bar{R} 、 \bar{P} の式は以下のように定義される。

$$\bar{R} = \frac{\sum_{i=1}^n C_i}{\sum_{i=1}^n A_i} \quad (2)$$

$$\bar{P} = \frac{\sum_{i=1}^n C_i}{\sum_{i=1}^n B_i} \quad (3)$$

F尺度は二つの値の調和平均であり、再現率、適合率のマイクロ平均に対するF尺度 \bar{F} の式は以下のように定義される。

$$\bar{F} = \frac{1}{\frac{1}{2\bar{P}} + \frac{1}{2\bar{R}}} \quad (4)$$

さらに、「類似項目を上位に検索しているか」を評価するために、検索された項目に閾値を設け、閾値以上の項目のみを抽出し、それらを検索結果として式(2)、(3)、(4)に適用させる。類似項目の検索実験を行い、既存手法や先行研究における提案手法と比較し、これらの評価指標の向上を目指す。

4. 実験・評価

4.1 実験概要

本実験では、先行研究の提案手法において、以下の検証

を目的とする。

(1) 先行研究における提案手法の日本史分野への適用可能性。

(2) 対象知識出現箇所の自動決定を適用することの有効性(図1)。

(1)を評価するために、単語ベースの手法(図2)とトピックベースの手法(図3)により類似項目の検索実験を行った。先行研究では、初級シスアドの項目を対象とした同様の実験を行い、既存手法に比べ提案手法のF尺度が最も高くなった。このことから、先行研究と同程度のF尺度であれば適用可能性と判断する。

さらに、(2)を評価するために、これらの手法において、語の抽出箇所(図2:1, 図3:1)を「問題文, 正答」、また、「問題文, 正答, 誤答」に固定させた手法により、類似項目の検索実験を行った。対象知識出現箇所の自動決定を適用した手法が、出現箇所を固定させた手法に比べ、F尺度が向上していれば有効性が示されたと判断する。なお、LDAによるトピックの推定には、pythonのライブラリであるgensim[18]を用いた。

実験で対象とした項目は、Classiにおいて中学と高校の日本史(古代~平安時代)で出題されている273項目とした。Classiで出題されている項目のうち、専門用語を問う多肢選択式の項目が最も多く存在していたため、日本史の項目を対象とした。上述した手法を用いて、項目ごとにこれらの項目を対象として類似項目を検索し、類似度の高い上位N項目を検索結果とし、再現率と適合率のマイクロ平均(式(2)、(3))、F尺度(式(4))を算出した。今回対象とした項目は1つの項目に対して、類似項目の最大数は3であったためN=3とした。

4.2 類似項目検索実験

実験手順を以下に示す。

(1) 項目の類似項目群への分類

筆者により、各項目の対象知識を判定し、同じ対象知識の項目をまとめて類似項目群(項目数が2以上)を作成する。今回の実験では、273項目のうち150項目で類似項目が存在し、57の類似項目群に分類された。表2に、類似項目群ごとの項目数と対象知識を示す。また、273項目から抽出された対象知識のユニーク数は180であった。後述する手順(4)では、類似項目が存在する150項目を対象とし、これらの分類結果を再現率と適合率のマイクロ平均を算出する際の正解データとする。

(2) 項目間類似度の算出

単語ベースの手法(図2)とトピックベースの手法(図3)、また、それらの手法において語の抽出箇所(図2:1, 図3:1)を「問題文, 正答」、また、「問題文, 正答, 誤答」に固定させた手法により、全項目間の類似

表 2 日本史の項目の類似項目群への分類結果

Table 2 Result of Classifying Test Items into Similar Item Groups.

項目数	類似項目群ごとの対象知識 (57 個)
4	国風文化, 聖徳太子, 撰関政治, 天平文化, 渡来人, 鑑真, 墾田永年私財法, 大宝律令
3	甲骨文字, 大化の改新, 天智天皇, メソポタミア文明, 弥生時代, 厩戸王, 高床倉庫, 初期荘園, 寝殿造, 秦, 聖武天皇, 租, 打製石器, 田堵, 土偶, 班田収授法, 平城京, 平等院鳳凰堂, 弥生土器, 渤海
2	「魏志」倭人伝, 貝塚, 空海, シャカ (釈迦), 平清盛, ナウマンゾウ, 武士団, 法隆寺, ヤマト政権, 開発領主, 漢, 吉野ヶ里遺跡, 遣唐使, 後三条天皇, 公領, 三内丸山遺跡, 持統天皇, 浄土信仰 (阿弥陀信仰), 新羅, 摂政, 中大兄皇子, 天武天皇, 卑弥呼, 仏教, 平安京, 防人, 磨製石器, 万葉集, 庸

表 3 再現率, 適合率のマイクロ平均と F 尺度

Table 3 Micro Average of Recall, Precision, and F-measure.

分野	類似度算出手法	抽出箇所	再現率	適合率	F 尺度
日本史	単語ベースの手法	対象知識出現箇所の自動決定	0.83	0.51	0.64
		問題文, 正答	0.82	0.42	0.56
		問題文, 正答, 誤答	0.56	0.31	0.40
	トピックベースの手法	対象知識出現箇所の自動決定	0.53	0.44	0.48
		問題文, 正答	0.48	0.38	0.43
		問題文, 正答, 誤答	0.35	0.33	0.34
初級シスアド	単語ベースの手法	対象知識出現箇所の自動決定	0.65	0.33	0.44
	トピックベースの手法		0.68	0.37	0.48

度を算出する。類似度を算出する対象は、実験で対象とした 273 項目とした。トピックベースの手法における LDA のトピック数は、2.4 節で述べたように、対象知識のユニーク数である 180 とした。そして、各項目に対して、類似度の高い項目順に整列する。

(3) 閾値による項目の抽出

手順 (1) の結果 (表 2) から、ある項目に対する類似項目の数は最大で 3 項目であった。そのため閾値を上位 3 項目と定め、手順 (2) の結果を基に、各項目に対して上位 3 項目を抽出した。なお、上位から 3 番目と 4 番目の項目が同じ類似度である場合、4 番目の項目も含めて抽出した。

(4) 再現率, 適合率, F 尺度の算出

手順 (1), (3) の結果から類似項目が存在する 150 項目を対象とし、各手法の再現率と適合率のマイクロ平均、また、F 尺度を式 (2), (3), (4) によりそれぞれ算出する。式 (2), (3) において、類似項目とは手順 (1) によって同じ類似項目群に分類された項目とし、検索された項目とは、手順 (3) によって抽出された項目とする。

表 3 に、上述した実験手順により得られた再現率と適合率のマイクロ平均、F 尺度、また、先行研究で初級シスアドを対象に行なった同様の実験から得られたこれらの指標の結果 [3] を示す。なお、表中の数値は小数点第三位で四捨五入した値である。

4.3 実験結果の考察

前節の実験結果から、4.1 節で述べた、2 つの検証項目に

ついて考察する。

4.3.1 先行研究における提案手法の日本史分野への適用可能性

表 3 の日本史と初級シスアドにおいて、抽出箇所が対象知識出現箇所の自動決定による手法 (先行研究における提案手法) に着目すると、日本史の単語ベースの手法の再現率、適合率、F 尺度が最も高い。また、日本史のトピックベースの手法の F 尺度は、初級シスアドの単語ベースの手法より高く、初級シスアドのトピックベースの手法と同程度となっている。これらの結果から、日本史分野の項目において先行研究における提案手法が適用可能であることが示唆された。

今回の実験で、トピックベースの手法に比べ単語ベースの手法の方が類似度算出の精度が向上したのは、日本史と初級シスアドの対象知識の特徴の違いに起因していると考えられる。表 4 に先行研究で対象とした初級シスアドの類似項目群の項目数と対象知識を示す。表 2 の日本史の対象知識と比較すると、日本史の対象知識は歴史上の人物や建物、制度など、表記ゆれが起こる可能性が低い語が多く、それらは 1 つの単名詞として認識されることが多い。2.3 節で述べたように、単語ベースの手法では、抽出された語の重みをベクトルの要素とするので、表記ゆれのない対象知識が多いほど、類似項目間で対象知識がマッチする頻度が高まると考えられる。これらの理由から、単語ベースの手法では類似項目が検索されやすくなり、類似度算出の精度が向上したと考えられる。

4.3.2 対象知識出現箇所の自動決定の有効性

表 3 の日本史の単語ベースの手法とトピックベースの手

表 4 先行研究で対象とした初級シスアドの項目の類似項目群への分類結果 [3]
 Table 4 Result of Classifying Test Items into Similar Item Groups In Previous Study.

項目数	類似項目群ごとの対象知識
8	ウイルス対策, 磁気ディスク装置, ISO9000 シリーズ
7	公開鍵暗号方式
6	EUC, JPEG, システムの信頼性, 退行テスト, プロトタイピング
5	EDI, クライアントサーバシステム, グループウェア, システムアドミニストレータ, システムテスト, デジタル署名, プログラムの著作権, 労働者派遣
4	CORBA, MRP, PDF, QR コード, TCO, インタプリタ方式, スループット, 著作権の帰属, データウェアハウス, パレート図, ヒストグラム, 不正競争防止法, ブラックボックステスト, 連関図
3	CRM, CTI, DNS, ERP パッケージ, E-R 図, GUI 画面, HTML, ISMS プロセスの PDCA モデル, JAN コード, S/MIME, USB, アプリケーション, アプレット, ガントチャート, 管理図, キャッシュメモリ, グラフの使い方, システムの保守, スワッピング, セキュリティ対策, 著作権の保護, データマイニング, ハッシュ, ファイルシステム, フラッシュメモリ, プロポーションアルフォント, ベンチマークテスト, ホットサイト リスク管理体制, ルータ, 個人情報保護法

法のそれぞれに着目する。両手法とも、抽出箇所を問題文や正答、誤答に固定した場合よりも、対象知識出現箇所の自動決定の場合の方が、再現率、適合率、F 尺度が高くなっている。この結果は、先行研究における初級シスアドを対象とした実験と同様の結果となっており、日本史分野の項目においても、対象知識出現箇所の自動決定を適用することの有効性が示唆された。

単語ベースの手法では、対象知識が出現する箇所を絞り込むことで、対象知識とは関係のない語が排除され、それらによる誤検索が少なくなったと考えられる。トピックベースの手法も同様に、対象知識とは関係のない語が排除され、対象知識とそれに関連する語の共起性が高まり、抽出箇所を固定した場合よりも、項目ごとに関連性の高いトピックが推定されたと考えられる。

5. まとめと今後の課題

我々はこれまで、e テスティングにおける類似項目の自動検索を目的として、項目間類似度の算出手法を提案し、情報分野の項目を対象として評価を行ってきた。先行研究における提案手法は、ベクトル空間モデルを軸に2つの手法に分類される。1つは、項目に出現する単語の重要度をベクトルの要素とする手法（単語ベースの手法（図2））である。もう1つは、それらの単語の共起性から推定されるトピックの重要度をベクトルの要素とする手法（トピックベースの手法（図3））である。この2つの提案手法において、対象知識の出現箇所（問題文、正答、誤答）を自動決定する処理（1）を取り入れている。

本稿では、(1) 先行研究における提案手法の日本史分野への適用可能性と、(2) 対象知識出現箇所の自動決定を適用することの有効性を検証するために、Classi で出題されている日本史の項目を対象とし、類似項目の検索実験を行った。実験では、単語ベースの手法とトピックベースの手法、また、そらの手法において、語の抽出箇所（図2:

1, 図3: 1)を「問題文、正答」、また、「問題文、正答、誤答」に固定させた手法を用いた。検索結果を基に、再現率と適合率のマイクロ平均、また、F 尺度を算出し、手法ごとの結果や先行研究の初級シスアドの項目を対象とした同様の実験結果と比較した。

実験結果から、単語ベースの手法の日本史における再現率、適合率、F 尺度が最も高かった。また、日本史のトピックベースの手法の F 尺度は、初級シスアドの単語ベースの手法のより高く、初級シスアドのトピックベースの手法と同程度となった。これらの結果から、日本史分野の項目において先行研究における提案手法が適用可能であることが示唆された。さらに、単語ベースの手法とトピックベースの手法ともに、抽出箇所を問題文や正答、誤答に固定した場合よりも、対象知識出現箇所の自動決定の場合の方が、再現率、適合率、F 尺度が高くなっていた。この結果は、先行研究の初級シスアドの項目を対象とした実験と同様の結果となっており、日本史分野の項目においても、対象知識出現箇所の自動決定を適用することの有効性が示唆された。

今後は、実験結果の原因をさらに考察し、提案手法の改善を行うとともに、情報や日本史以外の分野でも適用可能性を明らかにする。また、今回の実験では、対象知識や類似項目の判断は筆者によるものであったが、対象分野の専門家や Classi を利用している先生や生徒などによるユーザー評価を行い、ユーザー視点による提案手法の課題を明確にする。さらに、1章で述べたように、類似項目や類似度データを用いることで、項目の管理、作成、出題の支援を可能とする方法を検討する。そして、項目の作成、管理、出題のサイクルを半自動的に行うことが可能な新たな e テスティング・システムを実現し、教育現場への導入を目指す。

参考文献

- [1] 植野真臣, 永岡慶三: e テスティング, 培風館 (2009).
- [2] 高木輝彦, 高木正則, 勅使河原可海: 学生が作成した問題の類似度算出手法の提案と評価, 情報処理学会論文誌, Vol. 50, No. 10, pp. 2426-2439 (2009).
- [3] 高木輝彦, 高木正則, 勅使河原可海, 田中健次: e テスティングにおける LDA を用いた項目間類似度の算出, 情報処理学会論文誌, Vol. 55, No. 1, pp. 91-104 (2014).
- [4] 高木輝彦, 高木正則, 勅使河原可海: 知識出現箇所の自動決定アルゴリズムに基づく類似問題検索機能の開発, 人工知能学会先進的学習科学と工学研究会資料, Vol. 58, pp. 33-38 (2010).
- [5] 宮本定明: クラスタ分析入門: ファジィクラスタリングの理論と応用, 森北出版 (1999).
- [6] Borg, I. and Groenen, P.: *Modern Multidimensional Scaling: Theory And Applications*, Springer (2005).
- [7] 池田信一, 高木輝彦, 高木正則, 勅使河原可海: 多肢選択式項目の出題パターンと選択肢の類似性に着目した難易度推定方法の提案と評価, 情報処理学会論文誌, Vol. 54, No. 1, pp. 33-44 (2013).
- [8] 池田信一, 高木輝彦, 高木正則, 勅使河原可海: 類似問題群からの反復学習が可能な適応型テスト出題方式の提案, マルチメディア, 分散, 協調とモバイルシンポジウム論文集, DICOMO2012, pp. 1402-1409 (2012).
- [9] 初級システムアドミニストレータ試験 (AD) : https://www.jitec.ipa.go.jp/1_11seido/h13/ad.html (2022/02/16 閲覧).
- [10] Classi: <https://classi.jp/> (2022/02/16 閲覧).
- [11] 徳永健伸: 情報検索と言語処理, 東京大学出版会 (1999).
- [12] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent dirichlet allocation, *J. Mach. Learn. Res.*, Vol. 3, pp. 993-1022 (2003).
- [13] MeCab: <https://taku910.github.io/mecab/> (2022/02/16 閲覧).
- [14] 専門用語 (キーワード) 自動抽出 Python モジュール termextract: <http://gensen.dl.itc.u-tokyo.ac.jp/pytermextract/> (2022/02/16 閲覧).
- [15] 芳鐘冬樹, 井田正明, 野澤孝之, 宮崎和光, 喜多 一: キーワードの関連用語を考慮したシラバス検索システムの構築, 知能と情報: 日本知能情報ファジィ学会誌: Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 18, No. 2, pp. 299-309 (2006).
- [16] 北 研二, 津田和彦, 獅々堀正幹: 情報検索アルゴリズム, 共立出版 (2002).
- [17] 中川裕志, 湯本紘彰, 森 辰則: 出現頻度と接続頻度に基づく専門用語抽出, 自然言語処理, Vol. 10, No. 1, pp. 27-45 (2003).
- [18] gensim: <https://radimrehurek.com/gensim/models/ldamodel.html> (2022/02/16 閲覧).