

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[†] 伊藤 徹郎[†] 高木 正則[‡]

Classi 株式会社[†] 岩手県立大学[‡]

1. 研究の背景と目的

本研究ではこれまで、e テスティング[1]における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、大きく単名詞と複合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。本研究では多肢選択式の項目を対象とする。このような類似項目をアイテム・バンク内、または、アイテム・バンク間で自動検索可能となれば、同一知識を問う項目の適応的な出題や、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。

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アド試験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」[4]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

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。
a. 縄文土器 b. 青銅器 c. 金属器 d. 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。
a. 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。
b. 貯蔵用のものを甕という。
c. 盛り付け用のものを高坏という。 (正答)
d. 食物を蒸すためのものを石皿という。

出典『Classi Web テスト問題』Classi 株式会社

Application of Test Item Similarity Calculation Method to the Field of Japanese History and Its Evaluation

Teruhiko Takagi[†], Tetsuro Ito[†], Masanori Takagi[‡]

[†]Classi Corp.

[‡]Iwate Prefecture University

2. 項目間類似度算出手法

2.1. 項目間類似度算出手法の概要

先行研究では、ベクトル空間モデルを軸に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[5]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。

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)、問題文と否定的な表現 (「誤っている～」や「間違っていた～」など) のマッチングを判定する (図 1: 2)。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以上の場合には問題文を出現箇所とする (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、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

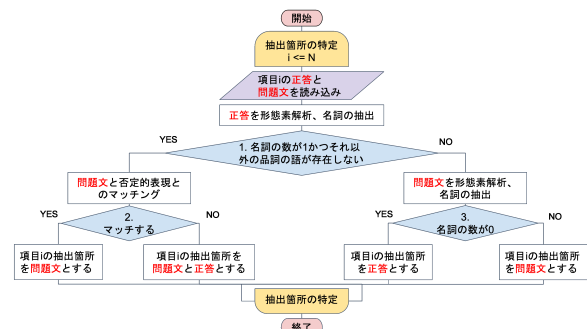


図 1 対象知識出現箇所の自動決定手順[2]

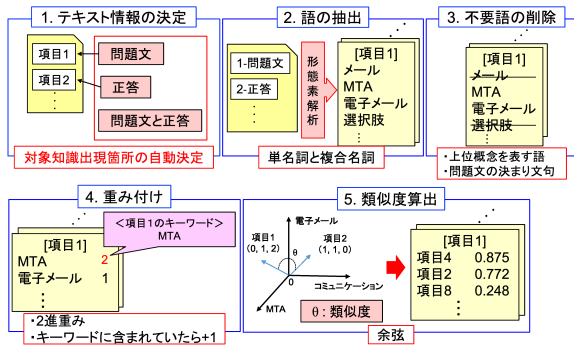


図2 単語ベースの類似度算出手順[2]

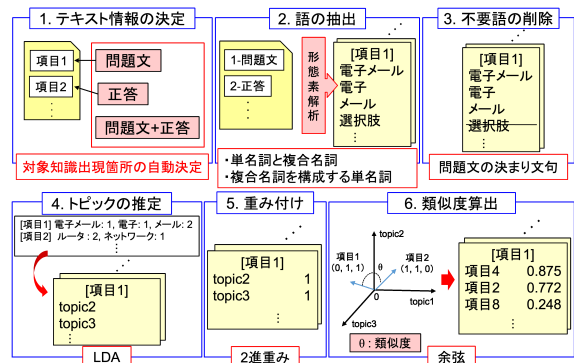


図3 トピックベースの類似度算出手順[3]

出する (図 2: 2) . 続いて, 上位概念を表す語 (「電子メール」に対して「メール」) や問題文の決り文句に含まれる語を削除する (図 2: 3) . さらに, 抽出された語に対して重みを付与し (図 2: 4) , 余弦により項目間の類似度を算出する (図 2: 5) .

2.4. トピックベースの類似度算出手順

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) . そして, 問題文の決り文句に含まれる語を削除する (図 3: 3) . ここで, 語と語の出現頻度を基に LDA でトピックを推定する (図 3: 4) . さらに, トピックに対して重みを付与し (図 3: 5) , 余弦により項目間の類似度を算出する (図 3: 6) .

3. 実験・評価

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を評価するために, Classi の日本史 (古代~平安時代) で出題されている 273 項目を対象とし, 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

表 2 再現率, 適合率の平均と F 尺度

類似度算出手法	科目	再現率	適合率	F 尺度
単語ベース	日本史	0.83	0.51	0.64
単語ベース	初級シスアド	0.65	0.33	0.44
トピックベース	日本史	0.53	0.44	0.48
トピックベース	初級シスアド	0.68	0.37	0.48

度が最も高くなったことから, 先行研究と同程度の F 尺度であれば適用可能と判断する. 実験では, 2 章で述べた提案手法を基に項目間類似度を算出し, 項目ごとに類似度の高い上位 N 項目を検索結果とし, 再現率, 適合率, F 尺度を計算した. 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. 表 2 に実験結果を示す.

実験の結果, 単語ベースの手法の日本史における再現率, 適合率, F 尺度の値が最も高かった. また, トピックベースの手法の日本史における F 尺度は, 単語ベースの手法の初級シスアドより高く, トピックベースの手法の初級シスアドと同程度になった. 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.

4. まとめと今後の課題

本稿では, 先行研究で提案した項目間類似度算出手法の適用可能性を評価するために, 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. 実験結果から, 単語ベースの手法は, 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, 提案手法の日本史の項目への適用可能性が示唆された. 今後は, 実験結果の原因を考察し, 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.

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 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 91-104 (2014).
- [4] Classi: <https://classi.jp/> (2022/01/07 閲覧)
- [5] Blei, D.M, NG, A.Y. and Jordan, M.I.: Latent Dirichlet Allocation, J. Mach.Learn. Res., Vol.3, pp.993-1022(2003).