

令和 5 年 5 月 29 日現在

機関番号：13901

研究種目：挑戦的研究（萌芽）

研究期間：2020～2022

課題番号：20K20821

研究課題名（和文）語学テストにおける自然言語処理手法を活用した敵対項目検出手法の開発と評価

研究課題名（英文）Detecting an enemy items utilizing an NLP approach: development and evaluation for language testing

研究代表者

光永 悠彦（Mitsunaga, Haruhiko）

名古屋大学・教育発達科学研究科・准教授

研究者番号：70742295

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：コンピュータを用いた大規模テストの実施にあたり、出題項目が互いにヒントとなることを避けるために、テストで出題される項目リストに出現する単語語彙の類似度をもとに「敵対項目」のペアを検出する手法を開発し、実際のテストで活用する方法について検討した。言語テスト（英語・日本語）の場合においてはその有効性が示唆されたものの、言語テストではないテストにおいては文脈情報の付加や類似度の集計方法などをさらに検討する必要があることがわかった。

研究成果の学術的意義や社会的意義

昨今、大規模コーパスに基づいた「大規模言語モデル（large language model, LLM）」の検討が進められ、その活用が期待されているが、そのような汎用大規模言語モデルを構築するためには巨大なコーパスと膨大な計算時間が必要であり、本研究課題で検討されるような特定の用途に限定された比較的狭い範囲における応用の事例が少なかった。本研究はLLMを用いずとも、テスト事業者レベルにおいて独自のモデルを用いた敵対項目検出AIエンジンを開発することの可能性を示すものであり、コンピュータを用いたテストの自動化に寄与することが期待される。

研究成果の概要（英文）：When a large-scale testing program which measures examinees' language ability is held based on an item bank, which contains various items of vocabulary, it is mandatory that the 'enemy items', which may be a clue to answer correctly if these items were presented at the same test administration, should not be shown on the same test forms. In this research, the method of preventing from presenting enemy items is proposed and evaluated by empirical test data. An NLP (natural language process) approach and machine learning system were adopted to estimate word similarity in each item of the item bank. Results suggested that it returns proper solutions when the item bank contains the items which measure examinees' vocabulary, however, when items measure non-verbal ability, several external variables should be used in order to get more accurate solutions.

研究分野：テスト理論

キーワード：機械学習 語彙テスト 言語テスト 項目バンク Automated test assembly

1. 研究開始当初の背景

(1) 大規模テストにおける項目バンクの活用と課題

多くの受検者に対して一斉形式で実施する「大規模テスト」について、あらかじめ出題される問題(以下「項目」と表記する)に困難度等の「項目パラメタ」を付与しておき、データベースに登録しておく手法が取り入れられるようになってきている。出題候補となる項目をあらかじめ登録しておく手法をとることにより、出題される項目の「質」を保証することができ、より質の高いテストによる能力の測定が期待できる。文部科学省(2023)の「MEXCBT」に代表されるように、複数のテスト(学力調査やクラスルームテスト等)をまたいで同じ項目バンクを活用する事例も、増えてきている。

テストを実施する側からすれば、ある受検者に出題するテスト版(解答者に出題されるテスト項目を一つにまとめたもの)に、同じようなトピックについて問う複数の項目が含まれていることは、避けなければならない。特に、ある項目の出題文に別の項目の解答へのヒントとなる表現が含まれていることは、テストの公平性に疑義が生じることとなり、あってはならない。しかし、昨今では項目バンクから受検者の能力に応じた項目を自動的に選抜して出題する仕組み(Automated Test Assembly)が応用されつつあり、このような仕組みを実装するうえでは、事前に「互いにヒントとなる項目のペア」を「敵対項目」(enemy items)として特定することが必要とされてきた。

また、テストを実施する側からの問題として、項目バンクの規模が大きくなるにつれ、新規に作成する項目について、項目バンク中に含まれる項目と互いにヒントとならないかをチェックする作業が難しくなるという課題があった。この点は、出題される項目で問われる知識等の能力が、本来測ろうとするトピックをまんべんなくカバーするように作題を進めるという点で重要であり、

(2) 語彙の類似度の検証方法

項目バンクに含まれる問題文をテキストデータと考えるなら、語彙の類似度に着目して敵対項目を検出する方法は以下のように定式化される。

1. 項目バンクに含まれる語彙をスキャンし、それらと類似度の高い単語のリストを作成する。
2. 類似度の高い単語リストが、項目バンク中の他の項目に出現しているかを調べ、出現していれば、類似度で重みづけした「類似度得点」を項目ごとに加算する。
3. すべての項目について1.と2.を繰り返す。

この方法を用いるうえでは、(a)1.で行われる「類似度」を算出するための手法及び(b)2.で行われる「類似度得点」を算出する手法のそれぞれについて検証する必要がある。(a)については、研究開始時点において Word2Vec や BERT といった自然言語処理を用いた語彙の類似度を算出するアルゴリズムが実用となりつつあり(言い換えれば、研究室で設置できる PC 環境で類似度推定が可能となりつつあり) これらの手法を用いることで解決できる一方、(b)については類似度得点の計算方法を数種類試行する必要が生じることがわかった。

2. 研究の目的

以上の背景をふまえ、本研究課題では下記の点を明らかにすることを目的とする。

(1) 大規模テストの項目バンクに適用可能な敵対項目アルゴリズムの開発

想定される主なテストは、言語テストにおける語彙項目を中心とし、必要に応じて言語テストの他分野の項目(文法など)に拡張する可能性を検討する。また、新作項目が既存の項目バンク中にある項目と敵対しているかを検出する目的においては、単語の類似性によらない、非言語的な素養を測るテストへの応用が期待されていることから、言語テストにこだわらずどのような適用が可能かについて検討する。

(2) 敵対項目アルゴリズムの評価

本研究で開発された敵対項目検出アルゴリズムを実際の大規模項目バンクに適用し、「敵対性の高い項目ペア」を含むテスト版を作成した場合、実際の受検者の解答行動にどのような影響が生じるかについて、実データを用いた検証を行う。あわせて、敵対項目の検出に要した時間や計算リソースの大きさから、提案手法が現実のテスト場面にどのように活用できるかについて検討する。

3. 研究の方法

(1) 敵対項目検出アルゴリズムの開発方法

語彙類似度得点を算出する方法としては、研究開始当初において「Word2Vec」が最もよく用いられていた。同方法は「語彙を表現するためのベクトル空間の次元数」や「学習対象となる語彙の近接度(ウインドウサイズ)」といったパラメタを操作することで、互いに似通っているが厳密には異なる結果を得ることとなる。またこれらのパラメタだけではなく、語彙の類似性を学習するための教師信号にあたる「大規模コーパス」についても分析者が与えなければならない。そのため、本研究では計算時間の現実性や分析結果の類似性を考慮し、2種類の次元数(200、

300次元) × 学習モデル2種類 (Skip-gram、CBOW) の4通りの分析を、インターネット上から無償でダウンロード可能な日本語コーパス (Wikipedia 日本語版 + 青空文庫) 及び英語コーパス (Wikipedia 英語版) を用いて類似度算出のための機械学習エンジンを構成した。そして、4通りの結果について、(a) いずれか1つであっても「類似している」と判定された語彙をリスト化し、語彙類似度得点を合算して求める手法と、(b) すべての条件で「類似している」と判定された語彙について語彙類似度得点を合算する方法、の2通りについて類似度を求めた。

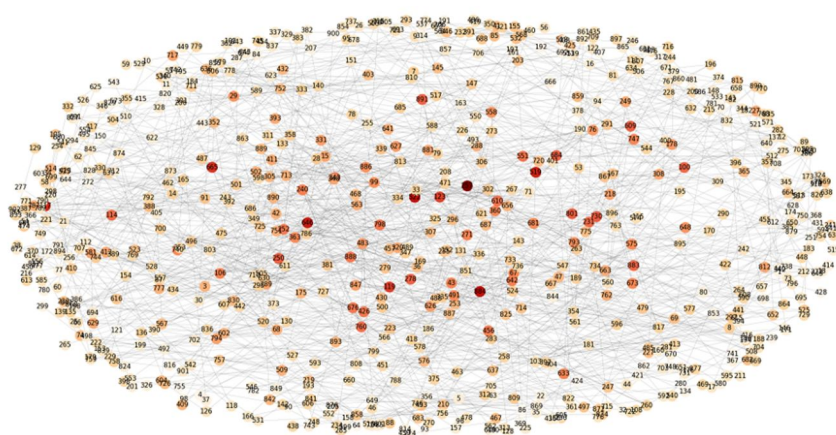
(2) 検出された敵対項目の質的・量的指標による検討

開発された手法により、既存の項目バンク (日本語テスト 1074 項目・英語テスト 1274 項目) に存在するすべての項目を使用して、敵対項目の検出を試みた。項目バンクは「IRT 診断テスト 日本語・英語」を利用し、日本語テストでは高校生までに習得しているであろう日本語語彙を問う項目、英語テストでは高校までの学習範囲の単語に加え、一部、文法を問う項目を対象とした。(a)(b) 双方の手法で構成された敵対項目リストをもとに、項目内容の質的検討として、項目の類似性をエキスパート (作問担当者) に評価させた。またこれらの項目をモニター受検者に解答させ、局所依存傾向 (項目への正解・不正解が、受検者の日本語 / 英語能力ではない要因で「とまって変わる」傾向) が検出されるか検討した (フィールドテスト)。

4. 研究成果

(1) 項目類似度評価の手法について

下図に、類似度評価結果の例を記した (日本語、(b)の手法)。各要素が1項目を表し、色の濃い要素は他の項目と類似度が高い項目ペアを構成していることを示す。また多くの類似ペアが



含まれる項目が図の中央なるように布置されている。

2種類の類似度評価の方法を比較すると、(a)の手法においては類似していると判断される語彙が多くなることから、類似性が少しでもみられる項目において敵対項目と判断されることとなった。一方で、(b)の手法にお

いては、語彙の類似性が高くなければ敵対項目とされるペアとならない傾向がみられた。そのため、実用上では(a)の手法をベースに類似性を検討することが必要とされた。

類似度の評価においては、日本語の場合、エキスパートによる評価で「類似している」とされる項目が概ね検出できている傾向が確認された。しかし、英語の「文法」においては、「類似している」とされない項目が敵対項目として検出される傾向が指摘された。また「語彙」についても、学習に用いられた英語版 Wikipedia の語彙レベルと、英語学習者が重点的に学習する単語の語彙レベルが乖離しているために、十分な類似度判定ができない傾向がみられた。その一方で、敵対項目ペアを含む30項目を用いたフィールドテストでは、敵対項目とされた項目ペアに局所依存傾向が部分的に見られたため、実際にこの仕組みが有効である可能性が示唆された。

本研究課題で提案された手法を用いることで、CAT (コンピュータ適応型テスト) のように、受検者ごとに別々の項目が出題されるテストにおいて、同時出題禁止の項目ペアをあらかじめ機械的に特定することができるようになった。あわせて、類題を指摘する作題支援システムの構築にも有効であることが示唆される。これらの仕組みは「IRT 診断テスト」の実施にも活かされており、今後他のテストにも拡張していくことが期待される。

しかしながら、本手法は万能というわけではなく、いくつかの課題が残される結果となった。一つは手法の汎用性の問題で、単語の語彙の類似性を問うテストに代表される、語彙の類似度が直接解答行動に影響する内容のテストにおいては、一定の性能を発揮することが期待されるが、語彙の類似度以外の要素で正誤が決まるテストにおいては、必ずしもうまく機能するとは限らない傾向がみられた。このことは他のテストの項目バンクにおいても実際に見られた傾向であり、機械的に分類する手法に比べて、項目の内容を深く理解した作問担当者の方がよりの確に敵対性を指摘することができたという意味で、機械的手法の限界が明らかとなったといえる。

もう一つは英語テストにおける学習コーパスに含まれる単語語彙レベルが乖離している問題で、この問題を解決するために、中学校までの学習教材のテキストデータを用いて機械学習エンジンを構成し、類似度を検討したが、用いた学習教材のバリエーションが少なく、また登場する単語の種類も少ないものであったため、十分な成果を得ることができなかった。ある程度の規模の学習教材を用いてより効果的な機械学習エンジンの構成を目指すことは、本研究手法を活用するうえで重要であることが示唆される。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 MITSUNAGA HARUHIKO	4. 巻 59
2. 論文標題 What Kind of Research Is Needed to Make a Large-scale Test Useful for Society? A Review of Recent Studies	5. 発行年 2020年
3. 雑誌名 The Annual Report of Educational Psychology in Japan	6. 最初と最後の頁 116 ~ 127
掲載論文のDOI（デジタルオブジェクト識別子） 10.5926/arepj.59.116	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 光永悠彦
2. 発表標題 自然言語処理を用いた敵対項目検出アルゴリズムの提案とテストへの応用
3. 学会等名 日本テスト学会第19回大会
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 光永 悠彦、西田 亜希子	4. 発行年 2022年
2. 出版社 ナカニシヤ出版	5. 総ページ数 252
3. 書名 テストは何のためにあるのか	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------