

そろばんにおける計算アルゴリズムの解析

工藤杏菜 (社会情報学科、2021446)

論文概要 そろばんは、十七世紀前半から日本で使用されている計算器具である。しかし、現在用いられる計算アルゴリズムは何故使われているのか。これを明らかにするために、過去盛んに用いられ、現在使わなくなった除算アルゴリズムや、Karatsuba 法という情報科学において乗算を速くしたアルゴリズムと比較する。

キーワード そろばん、計算アルゴリズム

1. 研究の目的

そろばんは十七世紀前半に和算書でそろばんが登場して以来、日本で広く使われ続けている計算器具である[1]。そして、そろばんを用いた計算のことを珠算と言う。

しかしこの歴史の長さに反して、現在珠算で普及している計算アルゴリズムは少ないと感じる。例えば、全国珠算教育団体連合会が作成したテキスト[2]は、加算と減算の指導を中心に扱っている。そしてそろばんを用いない暗算で行う乗算として、両落とし法(大きな桁から計算する手法)を簡単に紹介している。両落とし法を珠算に適用することもある。石戸珠算学園が監修したテキスト[3]でも、同様の加算と減算の指導を行っている。さらに、そろばんを使った乗算と除算の解説も行われている。乗算は乗数を盤面に置かない方落とし法(小さい桁から計算する手法)を解説している。そして除算は、商除法と呼ぶかけ算九九を用いた割り算を解説している。このように、乗算のみ二種類扱い、他の演算においては一種類の手法が指導される。

手法の数について確認すると、少ないと感じる理由は現代珠算の立ち位置にある。珠算は、検定、競技どちらも限られた時間内に正答した数で合否や順位が決定する。つまり、より速く正確に計算をするために、新たな手法が提案されても良い環境なのである。そこで、過去に用いられていた手法に関する関連研究を紹介する。

まず山崎ら[4]によって、これまで提案された手法が多く存在していたことがわかる。ここで注目したい点が三つある。一つ目は帰除法と呼ばれる、割算九九を用いた除法が普及していたことがあったこと。二つ目は、乗法において除法のような大きな転換はなかったこと。三つ目は、乗法除法どちらも、特定の条件下でのみ適用する手法が多くあったこと。

そして曾我氏[5]は、帰除法の詳細な解説と帰除法から商除法への転換について自身の意見を述べている。

関連研究からこれまで様々な手法が提案され、用いられていたことがわかる。しかし、帰除法から商除法への転換については両者とも商除法が優れているとするのみで、計算アルゴリズムの観点からは明らかになっていないと考える。また、乗法では計算する順番や、単位の変換など頻繁に用いられる値の処理に対応した手法が多く、除法のような変換が起こらなかったことに疑問を感じる。

以上のことから本研究では、現在採用されている乗法と除法の計算アルゴリズムを再検討することを目的とする。

乗法では、情報科学において提案され実際に採用されている Karatsuba 法を解析し、珠算に採用できないかについて検討する。Karatsuba 法は、加算の回数を増やし、乗算の回数を減らすことで計算の高速化を行っている。現在のそろばんで行う乗法は大きい値の計算であるため、有効だと考える。

除法では、帰除法の解析を通して商除法へ変換した理由を情報科学の観点から明らかにする。これによって近年の珠算で求められる除算アルゴリズムが明らかになり、手法の改良への足掛かりになると考える。

2. 研究の方法

本研究では、そろばんの各手法を情報科学の観点から解析を

行う。そろばんの指導や研究で手法の説明を行う際、問題の具体例と珠の動かし方を解説することが一般的である。そのため、この試みはそろばんの研究においてこれまでにない挑戦となる。具体的には、まず始めに各手法の定式化として疑似コードを作成する。その後、疑似コード内の加算と減算回数の合計を計算の速さ、行数を手法の複雑さの指標とする。そして、作成した疑似コードと設定した指標を基に各手法の解析と比較を行い、本研究の目的に沿って評価する。また、必要に応じて Karatsuba 法をそろばんの特徴に合わせた改善を行う。

3. 研究の結果

始めに、珠算における各手法の疑似コードの作成に成功した。それに伴い手法の解析や比較、評価を行う指標を統一した。

また、現時点において除法の疑似コード作成と解析を終えていない。そのため疑似コードを基にした手法の解析結果は乗法のみ述べる。乗法では、両落とし法の加算と減算の合計回数が方落とし法と Karatsuba 法よりも少なくなった。しかし疑似コードの行数は、珠算に合わせて改良を行った Karatsuba 法が一番少ないという結果を得られた。

除法は、商除法の疑似コード作成を終えて引き続き帰除法の作成に取り掛かっている。疑似コードを全て完成させたのち、解析と比較を行う見通しである。

4. まとめ

本研究は、新たな珠算の計算アルゴリズムの模索と手法転換の理由を明らかにすることを目的として、疑似コードの作成と解析を行った。これによって得られた成果は大きく二つある。

一つは珠算の手法を疑似コードによる表現が可能になったこと。これによって、手法の全体像が把握しやすくなり、明確な基準から評価をすることができる。他に、作成した疑似コードはコンピュータへの実装の助けも期待できる。これらの成果は人間が手法の理解を深めることにも役立つと考えられるため、珠算教育の分野での活躍が見込める。

もう一つは、現在普及している手法が現代の検定や競技には適していたということ。特定の数字や桁を用いた計算が少ない現代の珠算では、どのような入力でも一定の速さが保証される手法が求められる。そして Karatsuba 法は従来の乗算よりも計算を速くはしなかった。その上、値の保存が苦手というそろばんの特性もあることから両落とし法に代わることは難しいと判断する。

珠算の教育と研究、情報科学の発展を期待して除法の解析を引き続き取り組む。

5. 参考文献

[1]吉田光由『塵劫記』、1627年

[2]全国珠算教育団体連盟『たのしいそろばん』、2024年11月閲覧

[3]石戸珠算学園『いしど式で簡単 大人のそろばんドリル 1日10分で計算力・集中力を活性化(コツがわかる本)』、メイツ出版、2019年

[4]山崎与右衛門、戸谷清一、鈴木久男『珠算算法の歴史』、森北出版、1958年

[5]曾我和三郎「帰除法による計算の全容」『帰除法から商除法への転換期とその背景』、全国珠算教育連盟編『珠算春秋』、第6巻 104号、2020年、pp. 50-70.

衆議院議員選挙における投票率と生活満足度の関係

稗田末夢（経済学科 2021315）

論文概要 本研究では、投票に行くことによって投票者自身にメリットが発生するのかを調査した。具体的には投票に行くことで、投票者の生活満足度が高まるかどうかを重回帰分析で明らかにする。逆方向の因果関係に対処するため、操作変数法を用いる。

キーワード 生活満足度、投票率低下、操作変数法

1. 研究の目的

近年問題になっている選挙の低投票率について、その決定要因は様々だと今までの先行研究で明らかにされてきた。[1](桑名(2020))では個人間の所得水準の差は投票率に正の影響があり、長労働時間は負の影響があると明らかにされた。また、[2](稲葉・戸川(2020))でも経済的格差が投票率に負の影響を与えており、身近な人々との付き合いが正の影響を与えると示された。

その他にも環境要因として投票所の数や投票啓発活動、新党の参入などがある。有権者はこれらを考慮して自身の「ベネフィット」が「コスト」を上回ったときに投票に行くようになる。このように、投票率の決定要因については多くの研究蓄積がなされている([3]松林(2023))。

しかし先行研究を調査する中で、その多くが投票に行った後の影響に関しては言及していないことに気が付いた。そこで、本研究では有権者の立場で選挙に行くメリットを明らかにすることを旨とする。

2. 研究の方法

〈データ〉

『第48回衆議院議員総選挙全国意識調査』(2017)

寄託者 明るい選挙推進委員会

本来のサンプルサイズ 2151(アンケート回収率 68.3%)

ここから「わからない」「無回答」は全て除外して1693のデータで分析を行った。

〈分析方法〉

投票に行くことで生活満足度が高まるかどうかを調べるため、Rで重回帰分析を行った。しかし逆方向からの因果関係として、生活満足度が高い人ほど選挙に行く余裕がある、あるいは生活満足度が低い人ほど現状に不満を感じていて選挙に行く、といったことも考えられる。そのような逆方向の因果関係の可能性に対処するため、操作変数法を用いた。

〈被説明変数〉

現在の生活への満足度(1-4の4段階 4が最も満足)

〈注目する説明変数〉

投票参加ダミー

2017年10月の衆院選の投票に行ったかどうかをダミー変数に変換

〈その他の説明変数〉

今後の生活の見通し、現在の政治に対する満足度、男性ダミー、年齢ダミー、大卒ダミー、労働者ダミー、学生ダミー、主婦ダミー、無職ダミー、都市規模ダミー、地域別ダミー

〈操作変数(生活満足度に直接影響を及ぼさず、投票ダミーだけに影響する変数)〉

投票所までの所要時間(居住地から投票所までかかった時間を選択。それぞれの選択肢をダミー変数に変換。1-4の4段階 4が最も長い)

この変数を選んだ理由：投票所までの所要時間は生活満足度

を高める要因にはならず、投票に行く可能性に影響を与えるため、操作変数として適切であることから。

3. 研究の結果

操作変数法(第一段階)

操作変数として用いたtime(投票所までの所要時間)が係数-0.0916かつ0.1%水準で有意だった。これは投票所までの所要時間が増えると、投票に行く確率が統計的に有意に減るという解釈ができる。※他の説明変数の説明はスペースの都合で省略している。第二段階、OLS分析も同様。

Intercept	Estimate	Std.error	T value	Pr(< t)	
time	-0.091639	0.012536	-7.310	4.13e-13	***

(***は0.1%水準で有意であることを示す)

操作変数法(第二段階)

dummy(選挙参加ダミー)の係数は0.5716かつ1%水準で有意だった。これは投票に行く確率が増えると、生活満足度が統計的に有意に増えるという解釈ができる。また、生活満足度の平均は2.479、標準偏差は0.735であるため、その増分0.571は無視できない大きさだと言える。

Intercept	Estimate	Std.error	T value	Pr(< t)	
dummy	0.571661	0.220483	2.593	0.000960	**

(**は1%水準で有意であることを示す)

OLS分析

操作変数を用いずに重回帰分析を行ったところ、係数が大幅に変化し、有意水準も前者より下回る結果となった。ここから、操作変数法を用いて因果関係を統制することが有効的であったと分かる。

Intercept	Estimate	Std.error	T value	Pr(< t)	
dummy	0.075393	0.036928	2.042	0.04135	*

(*は5%水準で有意であることを示す)

4. まとめ

本研究では、選挙に参加することで人々にどのような影響があるのかを調査した。その結果、選挙に参加することで生活満足度が向上するということが明らかになった。以上から電子投票をはじめ、手軽に選挙に参加する政策を推進することは社会と個人両方にメリットがあるため、今後も取り組みを続ける必要があると言えるだろう。

5. 参考文献

[1]桑名(2020)『政治的資源と投票参加意向の関係 ―所得と労働時間に着目した個人間・個人内効果の分析―』:年報社会学論集 33号(2020) 121-132頁

[2]稲葉、戸川(2020)『社会的資本、経済格差、投票率との関係 都道府県データによる考察』:法学紀要(2019) 269頁-285頁

[3]松林(2023)『何が投票率を高めるのか』:株式会社有斐閣

議会議録に対する表構造理解に向けたRAG及びマルチモーダルシステムの構築について

前多陸玖 (社会情報学科 2021344)

論文概要 本論文では議会議録に含まれる表を対象に、マルチモーダルLLMを用いて表を自然言語化することで、従来手法と比較してRAGの精度が向上することを検証する。

キーワード RAG 表構造理解 自然言語処理

1. 研究の目的

近年、大規模言語モデルにおける追加学習なしに外部知識を扱う手法として、RAG(Retrieval-Augmented Generation)[1]が用いられている。RAGでは外部文書をベクトルに変換し、データベースに保存することで、検索を可能にしている。これによりハルシネーションのリスクを軽減させ、信頼性の高い出力を期待できる。一方で、RAGで使用する文書にHTMLや表・画像等が含まれている場合、出力の精度が低下する可能性があり、適切な前処理が必要である。またベクトルインデックス作成の際に使用するEmbeddingモデルやChunkSize、OverlapSizeが出力に影響を与える[2]。

RAGの活用が見込まれる議会議録や有価証券報告書、学術論文には、その文書内に多くの表が含まれており、それらに対する適切な前処理が求められる。本研究では、表の含まれる文書の前処理において、マルチモーダルLLM(M-LLM)を使用することによる、検索における精度への影響を検証する。また、Embeddingモデル間の性能差やChunkSize、OverlapSizeの値の影響を検証することを目的とする。

2. 研究の方法

本研究ではPolinfo4-Minutes-to-Budget-Linking(MBLink)¹のデータセットを用いる。MBLinkは予算に関する市長の発言文と予算表が与えられたときに、発言文に関連する予算表を紐づけることとしたタスクである。今回はそこに含まれる小樽市の令和四年度の予算表46個とそれに対応する発言文46個を対象とする。本研究で使用するRAGの構成は図1に示す通りである。まずは対象となる予算表のHTMLからSeleniumを用いて、表のキャプチャ画像を撮影する。各表の画像をM-LLMであるgpt-4o-2024-05-13に渡し、マークダウン形式に変換する。本研究のRAGはLangChainを用いて実装した。自然言語化された表と、前処理を行わなかった表のHTMLをEmbeddingモデルであるtext-embedding-ada-002を用いてベクトルに変換しChromaDBに格納する。この際ChunkSizeは200から1000まで、200刻みで計測する。

発言文をRAGの入力した際の、文書検索時に提示される参照した文書の上位三件を取得する。その三件内にあらかじめ発言文に対する正解とされる表が含まれていた場合には正解とする。この処理を46個の発言文全てに行い、それぞれのChunk Sizeごとに正解率を求め評価する。

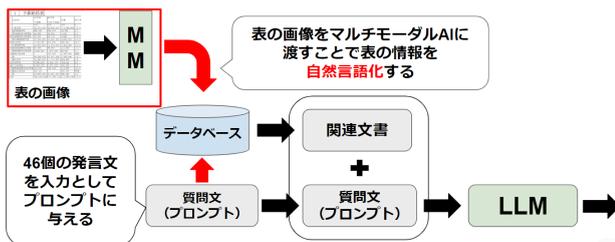


図1:本研究で使用するRAGの構成

3. 研究の結果

各ChunkSizeにおける今回の提案手法における結果と前処理を行わなかった手法における比較結果を表1に示す。

表1:提案手法と従来手法の比較結果

chunk_size	提案手法	通常のRAG
200	54.35%	41.30%
400	52.17%	41.30%
600	56.52%	41.30%
800	56.52%	41.30%
1000	54.35%	41.30%

従来手法の正解率はどのChunkSizeにおいても41.30%であった。一方で提案手法はどのChunkSizeにおいても50%以上の正解率を取り、ChunkSizeが600、800の場合にそれぞれ最大で56.52%を取ることが確認できた。

各embeddingModelにおける、overlapSizeごとの結果を表2に示す。

表2:各embeddingModelにおけるoverlapSizeごとの結果

model	text-embedding-ada-002			text-embedding-3-large			text-embedding-3-small		
	overlapSize 0%	overlapSize 25%	overlapSize 50%	overlapSize 0%	overlapSize 25%	overlapSize 50%	overlapSize 0%	overlapSize 25%	overlapSize 50%
200	54.35%	56.52%	47.83%	45.63%	45.65%	41.30%	58.70%	63.04%	56.52%
400	52.17%	50.00%	50.00%	47.83%	43.48%	52.17%	52.17%	56.52%	58.70%
600	56.52%	43.38%	47.83%	41.30%	36.29%	32.61%	56.52%	50.00%	50.00%
800	56.52%	50.00%	47.83%	45.65%	43.48%	34.78%	54.35%	50.00%	50.00%
1000	54.35%	50.00%	50.00%	45.65%	47.83%	39.13%	52.17%	54.35%	56.52%

性能が高かったのはtext-embedding-3-smallでchunkSizeが200、overlapSizeが25%の場合で、成功率は63.04%であった。

4. まとめ

本研究では、表を含んだ文書においてM-LLMを用いて前処理を行うことで、従来手法との精度比較を目的として実験を行った。その結果、正解率の比較において最大で15%以上の精度改善が見られた。これにより本研究における提案手法の優位性が確認された。

今後の課題として、さらなる精度向上に努める必要がある。具体的にはM-LLMを通じて自然言語化した文書をさらに要約することで、検索の精度をあげる手法を検討している。

5. 参考文献

- [1]Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, Mike Lewis, Wen tau Yih, Tim Rocktaschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021.
- [2]阿部晃弥, 新納浩幸, RAGにおける小説データベースのChunkSizeとOverlapSizeとEmbeddingモデルの効果, 言語処理学会第30回年次大会 発表論文集, 3215, 2024年3月

¹<https://github.com/poliinfo4/Polinfo4-FormalRun-Minutes-to-Budget-Linking>

(様式)

『スポーツ観戦時の CCI による顧客満足度への影響』

小野寺 悠月 (商学科、2021096)

論文概要 本研究は、スポーツ観戦における顧客間インタラクション (CCI) が観戦者の満足度やロコミに与える効果を検証し、観戦スタイルや観客層の違いが CCI の影響にどのように作用するかを明らかにする。また、サービス提供者が CCI を意図的にコントロールすることが顧客体験の向上に寄与する可能性についても検討する。

キーワード 顧客間インタラクション (CCI)、スポーツ観戦体験、価値共創

1. 研究の目的

スポーツ観戦における顧客間インタラクション (Customer-to-Customer Interaction; CCI) が、観戦者の満足度やブランドロイヤルティにどのような影響を及ぼすかを解明することが、本研究の主たる目的である。

近年、顧客間の相互作用 (CCI) が企業の満足度やロコミ、ロイヤルティに与える影響に関する研究がマーケティング分野で注目されている。特に CCI は顧客同士の価値共創 (Value Co-Creation) や価値破壊 (Value Co-Destruction) に関連して、サービス体験の質に大きな影響を与えることが示されている。しかし、日本国内のスポーツ観戦における CCI 研究は限られており、特に観客の特性や観戦スタイルの違いが CCI の効果にどのような変化をもたらすかについては十分に検討されていない。

既存研究では、特定のスポーツ (ゴルフ) において、他の観客の熱意ある応援が観戦者にポジティブな感情的影響を与えることが明らかになっている。(Kyungyeol (Anthony) Kim, Kevin K. Byon, Wooyeul Baek 2019) とところが、ゴルフ以外のスポーツ競技や、観戦理由や観客層が異なる場合の CCI の影響は検証されておらず、その一般化可能性は極めて限定的である。また、顧客セグメントごとで、観戦者の CCI の影響がどのように異なるか、サービス提供者が CCI をコントロールした際の影響の違いについては研究されていない。(Nicholls 2010; 宮井・西尾 2018)

本研究は、異なるスポーツ観戦体験や観客層における CCI の影響を検証し、CCI が観客にとってどのような価値を生むのかを実証的に分析することで、スポーツ・マーケティング論における理論的精緻化およびスポーツ業界における満足度向上のための実務的貢献を目的とする。また、観戦体験に関与するサービス提供者が CCI をコントロールした際の影響についても検討し、スポーツ観戦における新しい顧客体験の可能性を探る。したがって、本研究のリサーチ・クエスチョンは下記のように設定する。

R.Q.1 スポーツ観戦時の周囲の観客との CCI と顧客満足度の関係は競技によってどのような異同があるのか？

R.Q.2 スポーツ観戦時の CCI が顧客満足度と与える影響は、観客セグメント (熱狂的なファン層、雰囲気を楽しみたいファン層など) によって異なるのか？

R.Q.3 スポーツ観戦時の CCI が顧客満足に与える影響は、自然発生的なものと、サービス提供側がコントロールした場合で異なるのか？

2. 研究の方法

本研究では、上記のリサーチ・クエスチョンについて、アンケート調査に基づく定量的分析を行い、CCI が観戦体験に与える影響を明らかにする。アンケート項目は、既存研究に基づいた尺度を用いて設計されており、以下のような項目が含まれる。

1. 他の観客の応援や熱量
2. 他の観客の不適切な行動
3. 観戦体験に伴う知覚価値
(経済的価値、情緒的価値、認知的価値、社会的価値)
4. 顧客満足度

これらの尺度を用いることで、CCI が観戦者の満足度と与える影響のほか、観戦体験がもたらす価値 (知覚価値) を包括的に捉

えることが可能である。データ分析には主に回帰分析と分散分析を用い、CCI の影響の有意性と、観客層や観戦理由による違いを統計的に検証する。これにより、CCI が観客者の体験価値や満足度に対して与えるポジティブ/ネガティブな影響を明確にし、さまざまなスポーツ観戦における一般化可能性を探ることができる。

また、本研究は、サービス提供者が顧客体験をコントロールする際の基礎データとして活用できるよう、CCI の価値共創 (ポジティブな CCI) と価値破壊 (ネガティブな CCI) の二面性に注目し、これらが顧客満足度に及ぼす影響の違いを分析する。

3. 研究の結果

現在、アンケート・データ収集が進行中であり、分析結果はまだ得られていないが、仮説的に下記のような結果を見込んでいる。

1. CCI と顧客満足度の関係は競技によって異なり、雰囲気を楽しみたいファンが多い競技ほどその影響を強く受ける
2. 競技や観戦者の熱量によって CCI の効果が異なることが示され、熱狂的なファンほど CCI にポジティブな影響を受けやすく、雰囲気を楽しむ観客は CCI によって満足度が低下する。
3. サービス提供者が CCI をコントロールすることで、顧客満足度が高まるケースがあることが明らかになる

また、上記のような発見事実があれば、以下の理論的および実務的貢献を主張可能だと考える。まず、理論的貢献として、本研究ではスポーツ観戦における CCI と顧客満足度、価値創造や価値破壊の関係について、競技や顧客セグメントの違いに基づきながら、より詳細に明らかにすることで、スポーツ・マーケティング論における理論的精緻化が可能となる。また、実務的貢献として、スポーツ・チームやイベント運営者に対して、CCI を適切にコントロールすることで、顧客体験を最適化し、顧客満足度を向上させるための具体的な方途の提示が可能となる。

4. まとめ

本研究は、スポーツ観戦における CCI の効果を詳細に分析し、顧客満足度や観戦体験の質の向上に資する新たな知見を提供することで、スポーツ・マーケティング論の領域における理論的貢献を目指している。同時に、観客同士の交流がもたらす情緒的価値や社会的価値、さらには観戦者のロイヤルティにどのような効果があるかを解明することで、スポーツ・イベントの運営や観戦体験向上に役立つ実践的な指針を示したい。さらに、本研究の成果は、スポーツ観戦以外の分野 (例: エンターテインメント、観光業) においても CCI の活用を検討するための基礎的な知見として活用される可能性があると考えている。

5. 参考文献

- Kyungyeol (Anthony) Kim, Kevin K. Byon, Wooyeul Bae (2019), "Customer-to-customer value co-creation and co-destruction in sporting events"
- Nicholls, R. (2010), "New directions for customer-to-customer interaction research"
- 宮井弘之, 西尾チヅル (2018), 『おもてなし消費におけるゲスト側消費者の満足構造の分析』

researchmap 情報を対象とした論文タイトルの類似度を用いた研究者マッチングの研究

椎木太一（商学部社会情報学科 2021190）

論文概要 本稿では、researchmap から論文タイトルを取得し、それを OpenAI Embedding による手法で、文章ベクトルを生成して、ある論文の著者・研究者と、その論文と類似した論文の著者・研究者を提示する手法を述べる。また、OpenAI Embedding と Sentence-BERT との違いは何かを分析・考察する。

キーワード OpenAI Embedding、コサイン類似度、論文タイトル、Sentence-BERT

1. 研究の目的

研究者同士の協力やネットワーク形成は、学問の進展において不可欠である。従来、研究者のマッチングは、学会への参加や共通の知人を介した紹介など、オフラインの場で行われてきた。現在、オンラインによる研究者情報を提供するサイトが増えているが、研究者同士の交流の機会は限られており、新たな交流や共同研究を増やすことが期待されている。

従来の研究者マッチングに関する研究では、論文や研究テーマの類似性を基にした手法が提案されている。例えば、自然言語処理技術を用いて研究者のプロフィール情報から類似度計算を行い、異なる研究者間での協力関係構築の可能性が示されている[1]。このような取り組みは、従来のオフラインでのマッチングに比べ、より広範な研究者間の連携を実現するものとして期待されている。しかしながら、従来研究では、英語タイトルと日本語タイトルを比較するために翻訳することが必要であった。また、現在の大規模言語モデル (LLM) の学習量からすると小さな言語モデルの利用に留まっていた。

そこで、本研究では、大規模言語モデルを活用することで、論文タイトルの類似度と関連付けることで、研究者マッチングを行うことを目的とする。近年 OpenAI の OpenAI Embedding¹ が利用可能となっている。Embedding は、文章や単語をベクトル形式で数値化する自然言語処理の技術である。この Embedding を用いることで researchmap の情報を対象とした研究者マッチングを行う。対象データは、三大学連携を結んでいる小樽商科大学・帯広畜産大学・北見工業大学の研究者の論文タイトルとする。

2. 研究の方法

研究者のデータは論文の一覧が記載されたページの JSON ファイルに格納されている。三大学に所属する研究者のうち、論文タイトルが researchmap 上に掲載されてある研究者 263 人分 (小樽 83 人、帯広 99 人、北見 81 人) の情報をスクレイピングした。対象となる論文は 11, 495 本である。

ここから論文 1 つにつき OpenAI Embedding では 1, 536 次元、SBERT では 384 次元のベクトルを研究者の JSON に記載した。プログラムでの処理は、分析対象となる論文を JSON から抽出し、抽出元を除いた全ての研究者の JSON からコサイン類似度を計算する。そして類似度が近い順に 10 本の論文を Web サイト上に表示した。

OpenAI Embedding は text-embedding-ada-002、SBERT は paraphrase-multilingual-MiniLM-L12-v2 のモデルを用いる。

評価基準は人の目で見たとときの、「タイトル内にあるキーワードと思われる文言の辞書的な意味の近さ (以下①)」「何をメイン・対象とした研究か、テーマ、タイトル全体の類似性 (以下②)」「所属学部・学科・部門名のコサイン類似度 (以下③)」の三項目で評価を行う。論文の自身・本文は考慮しない。①、②ともに 1~5 点の得点をつける。この 3 項目を基に点数をつけ、考察する。計算式は以下の独自の式を用いる。

OpenAI Embedding
 $1.0 \times \text{コサイン類似度} - 0.5 \times \text{③} + (\text{①の点数} + \text{②の点数}) \times 0.1$

Sentence-BERT

$1.2 \times \text{コサイン類似度} - 0.6 \times \text{③} + (\text{①の点数} + \text{②の点数}) \times 0.1$

3. 研究の結果

	氏名	タイトル	類似	①	②	③	点数
対象	佐山 公一	コロナ禍におかれた台湾人観光客の訪日動機に SNS が与える影響 : 北海道、小樽を事例とした一考察	-	-	-	-	-
	小樽						
1	本間 圭一 北見	【調査報告】最北の国立大学における留学生へのコロナ禍の影響	0.882 16	3	3	0.825 67	1.069 325
2	小野 塚 祐紀 小樽	コロナショックへの企業の対応と政策支援措置—サーベイ調査に基づく分析	0.879 68	2	3	0.914 80	0.922 23

図 1 : OpenAI Embedding を用いた類似度分析

	氏名	タイトル	類似	①	②	③	点数
1	王 力勇 小樽	Social interaction in co creating the tourist experience An exploratory study of Chinese visitors to Japan	0.752 82	4	4	0.772 25	1.317 259
2	岩本 博幸 帯広	都市型観光振興方策に関する一考察—滋賀県長浜市文化的施設のコンジョイント分析から	0.728 51	3	2	0.349 34	1.199 542

図 2 : SBERT を用いた類似度分析

図 1 および図 2 に「訪日動機」に関する類似度分析の結果を示す。OpenAI ではコロナに関する論文が多くを占め、SBERT では観光・訪日に関する論文が多かった。そのため①の配点は、SBERT の方を高く付けた。

タイトル全体で見ただけの場合、OpenAI の方では「コロナが影響を与える」という部分に焦点を当てていたが、訪日動機と観光という点にコロナが影響を与えると考慮することは出来なかった。一方 SBERT では、コロナについては考慮されていなかったが、訪日動機と観光についての論文が類似度が高く出力された。したがって、配点に差は出なかった。

研究者の所属学科は、どちらの手法でも異なる大学の研究者を出力していた。そのため異なる大学や学科でありながら、類似する研究をしている研究者を発見するという意味で、有効であると考えられる。

4. まとめ

本稿では OpenAI Embedding を用いた類似度計算を SBERT と比較し、どちらが適切なマッチングが行えるかを明らかにするための比較実験を行った。その結果、OpenAI Embedding と SBERT のどちらが研究者マッチングに適切かは、どの単語に注目するかによって評価が左右されることを確認した。

5. 参考文献

[1] 廣吉和貴

researchmap の情報を用いた研究者マッチング手法の研究

¹ Introducing text and code embeddings | January 25, 2022
<https://openai.com/index/introducing-text-and-code-embeddings/>
2024/9/27 閲覧

(様式)

子育て支援情報サイトを活用した自治体の子育ての議論内容の評価に向けて

黒沢匠 (社会情報学科 2021136)

論文概要 本研究の目的は、妊娠・出産・育児に関する情報サイトを用いて、自治体による子育て支援施策の効果を調査することである。具体的には、自治体による子育て支援施策名、および、議会における子育ての議論に関する発言と、妊娠・出産・育児に関する情報サイトの投稿を紐づけることにより、子育て支援施策の調査を行う。

キーワード 子育て支援施策, 議会会議録, 子育て支援情報サイト

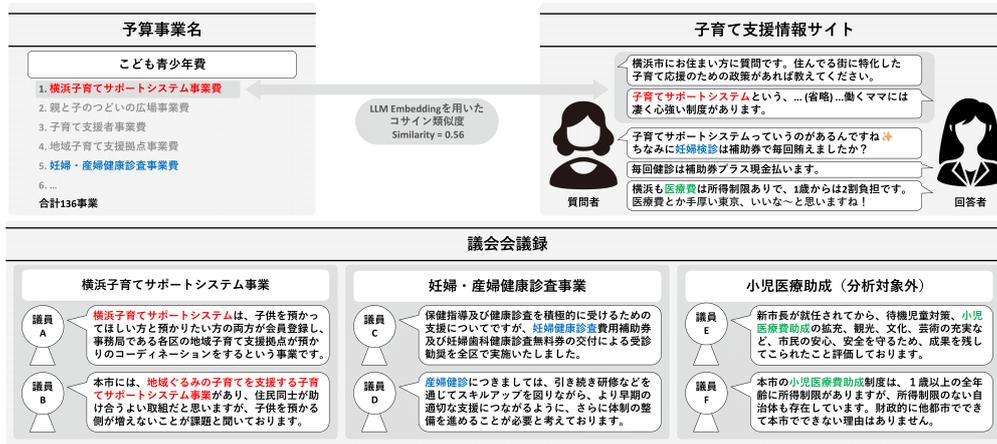


図 1 予算事業名, 子育て支援情報サイト, 議会会議録の紐づけの例

1. 研究の目的

子育て支援情報サイトは、保護者に向けて育児に関する情報を提供するための重要なツールである。しかし、これらの情報は自治体の子育て支援にほとんど活用されていない。また、自治体における子育てに関する議論は、地域のニーズを反映した政策実施を目指しているが、実際にそのニーズを捉えた政策の議論が行われているのか把握しづらい状況である。この課題に対して、子育て支援情報サイトを活用することで、議会の議論が子育ての現場と乖離していないかを検証することが期待される。

最近では、民間が運営する子育て支援サイト活用した研究が進められている。例えば、内田らの研究では、子育て支援情報サイトの検索キーワードを用いることで、子供の成長に合わせて母親の疑問が変化することを示した[1]。また、新井田らは、子育て支援情報サイトの投稿分析し、地方自治体の子育て支援の効果測定を試みている[2]。しかし、地方自治体の事業名や議会における議論と紐づけて分析するまでには至っていない。

そこで本研究では、子育て支援情報サイトの投稿をもとに、自治体における子育て支援の実態を調査するとともに、議会においてどのように子育て支援が議論されているのかを明らかにする。

2. 研究の方法

本研究では、コネヒト株式会社が運営する子育て支援情報サイト「ママリ」のデータを用いて研究を行う。子育て支援情報サイトの投稿、自治体の子育て支援事業名、支援事業の議論内容を紐づけることで、地方自治体の子育て支援についての自治体における議論とその地域における効果と課題を明らかにする。

最初に、調査自治体に対応する子育て支援情報サイトの投稿と、予算事業名を OpenAI API の Embedding を用いてベクトル変換を行い、コサイン類似度を求めることで、子育て支援情報サイトにおける投稿と子育て支援事業名を紐づける。次に、予算事業名をキーワードの表層的な一致による検索を用いることで、議会会議録から事業に関する発言の抽出を行うことで、予算事業名と議会会議録の紐づける。最後に、子育て支援情報サイトの投稿、自治体の子育て支援事業名、支援事業の議論内容を紐づけることで、自治体における子育て支援の実態を調査するとともに、議会

表 1 子育て支援情報サイトのデータ概要

項目名	数値
期間	2019/4/1~2021/3/31
質問数	3,795件
回答数	15,579件
ユーザ数	9,496人

においてどのように子育て支援が議論されているのかを明らかにする。

3. 研究の結果

図 1 を例として、横浜市の分析結果について述べる。例えば「横浜子育てサポートシステム事業」について、母親たちは心強いサービスであると認識し、議会でもさらに良い事業にするための議論が行われていることを確認した。また「小児医療助成」について、母親たちは所得制限があることに課題を感じており、議会でも所得制限の撤廃を求めていることを確認した。このことから、子育て支援事業に関する現状の認識は、現場の母親たちと議会とで大きな乖離がないことが明らかとなった。しかし小児医療助成費の所得制限撤廃は 2021 年 3 月までに撤廃が行われていないことから、事業の課題改善の速度には課題があることが示唆された。

4. まとめ

本稿では、子育て支援情報サイトの投稿をもとに、自治体における子育て支援の実態を調査し、さらに議会において子育て支援がどのように議論されているかを明らかにすることを目的として分析を行った。結果、横浜市の議会では、子育て支援事業に関して母親たちと同様の認識を持っていることが確認された。一方で、議員が課題として認識しているものの、課題改善へ向けた対応に時間がかかるという問題点も明らかになった。

5. 参考文献

- [1] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知. 妊娠・出産・育児に関する情報サイトにおける母親の検索行動の予備的分析. 言語処理学会第 29 回年次大会, 2023.
- [2] 新井田瑠璃, 佐藤栄作, 木村泰知, 内田ゆず. 妊娠・出産・育児に関する情報サイトにおける自治体による子育て支援効果の調査. 言語処理学会第 30 回年次大会, 2024.