

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

新井田瑠璃¹ 佐藤栄作¹ 木村泰知¹ 内田ゆず²¹ 小樽商科大学 ² 北海学園大学

g2020259@edu.otaru-uc.ac.jp kimura@res.otaru-uc.ac.jp

概要

本研究では、妊娠・出産・育児に関する情報サイトのデータを分析し、地方自治体の子育て支援の効果を調査することを目的としている。本研究は、東京都のデータを用いたアノテーションと、東京都以外の46道府県のデータに対するBERTによる自動推定を通じて、機械学習を用いた子育て支援効果の推定の可能性を示した。

1 はじめに

妊娠、出産、育児は、母親にとって大きな生活の変化を伴うため、これらの段階におけるサポートは非常に重要である。地方自治体は多様な子育て支援策を実施しているが、これらの施策が実際に母親たちにとってどのように役立っているかは明らかになっておらず、公開されているデータも自治体ごとに異なり、日本国内の各自治体の効果を比較することが困難である。

従来の研究では、主にアンケート調査やインタビューによって、母親たちのニーズや子育て支援策に対する満足度の調査が行われてきた。例えば、子育て支援政策についての研究としては、ファミリーサポートセンターを中心とした子育て支援の現状や課題について分析を行った研究 [1] や、母子健康包括支援センターの在り方についての研究 [2] がみられる。また、子育て支援政策の財源に関する研究については、松井・今井らによる児童手当制度について分析し、財源について提案する研究 [3] や、山崎による子ども・子育て支援政策全体の財源拠出について提案する主張 [4] もある。

これらの研究では、地方自治体による子育て支援策の効果の定量的な評価も行われているが、ほとんどは対象範囲が限られており、小規模なサンプルに基づいている。対象が限られたアンケートやインタ



図1 妊娠・出産・育児に関する情報サイトを対象とした地方自治体の子育て支援の効果測定の流れ

ビューに基づく研究は、参加者の選択バイアスや回答の偏りが生じる可能性がある。このような評価は、実際の母親たちの生の声や実体験を反映しきれていない可能性があり、子育て支援策の実効性の評価には限界がある。

最近では、オンラインサービスの発達により、母親たちが情報サイトにおいて情報収集をする機会が増えている。情報サイト内の検索データをもとに感情分析を行った内田らの研究 [5] は、母親たちに適切な医療情報を提供する際に役立つ可能性を示すなど、情報サイトのデータを用いた研究が有用であることを示した。

そこで、本研究では、妊娠・出産・育児に関する情報サイトに投稿されているデータを分析し、子育て中の母親たちの声から、地方自治体の子育て支援の効果測定を行う。

図1に、本研究における妊娠・出産・育児に関する情報サイトを対象とした地方自治体の子育て支援

の効果測定の流れを示す。分析では、情報サイトに子育て支援に関する記述が含まれていることを確認した後で、東京都を対象として分析を行うとともに、機械学習のためにアノテーションを行う。東京都のユーザによるデータに付与したアノテーション結果を学習データとして用いて、東京都以外の46道府県のデータをテストデータとして、子育て支援に関する効果の記載がある投稿を自動で推定する。

本研究の貢献は、下記の3点である。

1. 妊娠・出産・育児に関する情報サイトに子育て支援の効果の記述が含まれていることを確認した(3章)
2. 東京都のデータの子育て支援の効果が記述されている投稿に対してアノテーションした(3章)
3. 東京都以外の46道府県のデータに対してBERTによる自動分類を行い、機械学習による分類の可能性を示した(4章)

2 対象データ

本研究では「ママの一步を支える女性向けQ&Aアプリ」としてコネヒト株式会社¹⁾が運営するサービス「ママリ」のデータを用いて研究を行う。このSNSアプリでは、母親たちが妊娠・出産・育児など子育てにまつわる悩みを質問として投稿し、他のユーザから回答を募集し解決を図るシステムとなっている。本研究ではユーザ情報、回答データを用いて、回答者の住む地域と発言内容に焦点を当てて研究を行う。データの概要を表1に示す。

表1 ママリのデータ概要

項目名	数値
期間	2019/4/1~2021/3/31
質問数	4,048,140件
回答数	24,091,131件
ユーザ数	2,818,100人

3 東京都データの分析

3.1 目的

本節では、妊娠・出産・育児に関する情報サイトのなかで、東京都のデータを対象として、子育て支援の効果の記述が含まれているのかを確認をする。具体的には、東京都在住の母親からの質問に対して、子育て支援の効果がある(あるいは、効果がない)こ

1) <https://connehito.com/>

とが記述されている回答とそれ以外の回答に分類するためのアノテーションを行う。

3.2 分析方法

東京都の在住者による回答データ数は2,138,085件である。人手で確認するには件数が多いことから、子育て支援に関連する可能性がある投稿をキーワードにより絞り込み、最終的に人手でアノテーションすることにした。予算に関する単語(キーワード)が含まれる投稿に関連する可能性がある回答の候補とする。キーワードの選択方法は、47都道府県の子育て支援に関する予算項目名、全1,151項目に対してMeCabを用いて形態素解析を行い、40回以上出現した単語とした。子育て支援に関する予算項目名には「特定不妊治療費助成事業」「すこやか子育て支援事業」「家庭的保育事業」などがある。また、40回以上出現した形態素に絞った結果、次の21種類の形態素が選出された²⁾。

21種類の形態素

事業(677回)、支援(356回)、保育(270回)、子育て(159回)、推進(143回)、対策(130回)、児童(128回)、子ども(117回)、補助(92回)、医療(91回)、整備(87回)、地域(76回)、家庭(72回)、助成(71回)、施設(70回)、応援(59回)、給付(51回)、促進(45回)、強化(43回)、こども(43回)、不妊(42回)

アノテーションの対象データは、東京都在住の利用者の回答データ2,138,085件から、下記の条件を両方満たす回答とすることとした。

1. 21種類の形態素のうち一つ以上を含む回答
2. 「子育て」「子ども」「こども」の3つのキーワードのうち、一つ以上含む回答

その結果、対象データは、9,830件となった。

この対象データに対して、表2の基準により「効果あり」「効果なし」「効果あり・効果なしの両方を含む」「その他(関連なし)」に分類する。

3.3 分析結果

3.2の対象データ9,830件の回答データに対して、人手による分類を行った。表3に人手により分類した結果を示す。子育て支援政策に関わる会話を抽出し、「効果あり」「効果なし」「効果あり・効果なし」

2) 出現頻度の高い形態素には「・」などの記号、あるいは、「の」「費」「等」などの一文字の形態素が含まれていたため、が2文字以上かつ記号を含まない形態素に絞ることとした。

表2 回答データの分類基準

効果あり	政策について具体的に称賛する内容、 利用を推奨する内容など 例：「〇〇区のこういった制度がおすすです。」
効果なし	否定的な意見・感想、改善を求める内容(抽象的でも可) ※抽象的でも有用である可能性があることを考慮 例：「〇〇区は待機児童が多いので困っています。」
効果あり・効果なし	効果ありと効果なしのどちらも含む内容 例：「〇〇区は育児応援チケットが2万円分もらえます。 でも待機児童が多く、保育園に入るのが難しいです。」
その他(関連なし)	政策には関係ない内容、関連がみられたが 具体性がなく効果ありに分類されなかった内容 例：「私の地域はごみの分別が細かいです。」

表3 回答データを分類した結果の内訳

関連あり	効果あり	275
	効果なし	68
	効果あり・効果なし	41
関連なし	その他	9,446
合計		9,830

「その他」の4種類に分類した³⁾。

「効果あり」の回答は275件みられ、支援センターや児童館、自治体による子どもの医療費補助についての会話が多くみられた。また、具体的な支援政策名を挙げている会話もいくつか見られ、例として江戸川区の「よちよち応援隊(現：えどがわパパママ応援隊)」の制度についての回答データが数件観察されている。この制度は2019年4月に始まったものであることと、ママリ回答データ自体も2019年4月以降が対象であることから、リアルタイムに子育て支援政策の反応が得られたことになる。「効果なし」の回答は68件であり、特に待機児童の多さや地域の保育園の数についての会話が多くみられた。また、「出産一時金の増額をするよりも、児童手当の所得制限をなくしてほしい」のような具体的な意見も確認できた。このように妊娠・出産・育児に関する情報サイトにおいて、母親たちの子育て支援政策に対する意見や反応が投稿されていることを確認した。

4 46道府県を対象とした自動分類

4.1 目的

本実験では、3章で扱った東京の回答データを学習データとして、BERTを用いて東京都以外の46道府県の回答データを未知データとして分類する。これにより、機械学習を用いた分類の可能性について検

3) 「効果あり・効果なし」については、「効果あり」と「効果なし」のどちらも含むものを集約しているが、一つの回答内に「効果あり」「効果なし」に該当する文章がそれぞれ複数含まれていても「効果あり・効果なし」に該当する1件のデータとして扱うものとする。

討する。

4.2 方法

BERTを用いて、3章で扱った東京の回答データとその分類結果を学習データとして、46道府県の回答データの内容が「子育て支援政策の効果」について記述されているのか、あるいは、記述されていないかを自動分類する。対象の自治体は東京を除いた46道府県であり、データ数の合計は表4に示す。対象データは3.3節と東京都と同じ条件であり、95,638件の回答データである。BERTによる自動分類の結果には「probability」という「子育て支援政策の効果の有無が分かる(分からない)内容である」点における確信度が付与されている。本実験では、各自治体における「子育て支援政策の効果の有無が分かる」確信度が高い上位10件(46道府県の合計は460件)に対して、人手による評価を行う。

表4 46道府県の回答データ数

道府県ID	道府県名	件数
1	北海道	3,771
	・	
	・	
	・	
47	沖縄	1,246
46道府県の合計		95,638

4.3 結果

表5に、BERTによる自動分類した確信度の上位10件を人手で評価した結果として、正解数、不正解数、並びに、46道府県全体の正解率を示す⁴⁾。

表5 正解数、不正解数並びに正解率

正解数	不正解数	正解率
307	153	66.74%

表5から分かる通り、全460件中正解307件に対し不正解153件が確認され、正解率は66.74%であった。正解率は7割近くになっており、高い結果といえる。7割近い正解率であれば、機械学習を用いることで、子育て支援政策の効果を測定できる回答データを見つける負担が軽減できる可能性がある。

4) なお、正解率は少数第3位を四捨五入している。

4.4 考察

4.4.1 正解と不正解のパターン

正解のパターンでは、事例を紹介することが多い。例えば、下記のような例である。

正解例

「●●^aに住んでいます！

… (省略)…

子育て支援は、児童館が徒歩圏内にいくつかあり連れて行きやすいです。また市からの子育て関連のイベントも割とあり充実しています。医療費などは、子供は月1医療機関に付き、200円の負担で受診できます。」

^a 地名を伏せ字にしている。

正解には、上記の例のように子育て支援政策の良い点を紹介している回答データが多数確認された。

次に、不正解となった回答データのうち、特徴的な2つのパターンを示す。

1つ目は、対象地域が異なり不正解となっているパターンである。以下に例を示す。

不正解例 (1)

「最寄り●●駅^aです。子育て世代が多く、支援センターも充実していて私は大好きです。

… (省略)…

保育園もたくさん新設されたおかげで待機児童もなく、幼稚園は定員割れするほど余裕があるので助かります。」

^a 駅名を伏せ字にしている。

内容としては、子育て支援政策の効果を述べる内容である。しかしながら、該当データは、回答データであるが対象地域に●●駅はなく、他県に存在する「●●駅」に住んでいると考えられる。このように、回答データの内容自体は問題がないがユーザが登録している地域と住んでいる地域が異なるため、違う地域について説明している回答データが多くみられた。

2つ目は、政策の効果に直接関係ないため、不正解となっているパターンである。以下に例を示す。

不正解例 (2)

「子育てサロンや支援センターにいつからデビューするか悩んでいます。」

この例においては、「子育てサロン」や「支援センター」など、子育て支援政策に関係する単語は拾えているものの、政策の効果に直接関係しない内容であることが分かる。この結果から、文脈を読み取るという点においては課題があることがわかる。

4.4.2 確信度

ここでは、確信度と正解率について述べる。確信度の詳細については表6に示す。正解、不正解ともに平均値と最高値に大きな差は見られなかったが、最低値においては約2.5ポイントの差がみられた。また、機械学習による分類の結果「子育て支援政策の効果の有無が分かる」と分類されたデータのうち最も高い確信度が「98.98%」であったが、当該データは人力による分類では不正解とされた。最も確信度が高いデータは前述した「回答データの内容自体は問題がないがユーザが登録している地域と住んでいる地域が異なるため、違う地域についての回答データ」であることが判明した。そのため本実験では、機械学習による確信度の閾値を明らかにすることはできなかった。

表6 確信度

	確信度	説明
正解平均値	95.81%	機械学習と人力による分類の結果がともに「子育て支援政策の効果の有無が分かる」と判断されたデータの確信度
正解最高値	98.79%	
正解最低値	85.32%	
不正解平均値	95.17%	機械学習による分類では「子育て支援政策の効果の有無が分かる」と判断されたが人力による分類では異なったデータの確信度
不正解最高値	98.98%	
不正解最低値	82.89%	
全体平均値	95.60%	全体(正解、不正解どちらも含む)の確信度の平均値

5 おわりに

本稿では、妊娠・出産・育児に関する情報サイトを対象として、子育て支援の効果について調査を行った。その結果、妊娠・出産・育児に関する情報サイトに子育て支援の効果の記述が含まれていることを確認した(3章)。また、東京都のデータにおいて子育て支援の効果が記述されている投稿に対してアノテーションした(3章)。さらに、東京都以外の46道府県のデータに対してBERTによる自動分類を行い、機械学習による分類の可能性を示した(4章)。

謝辞

本研究は JSPS 科研費 21H03769, および, 電気通信普及財団の助成を受けたものである。本研究の遂行にあたり, 貴重なデータを提供していただいたコネヒト株式会社様に感謝の意を表します。

参考文献

- [1] 山路憲夫. ファミリーサポートセンターを中心とした子育て支援の現状と課題: 東京都内の主要都市の事例から. 研究年報, Vol. 8, pp. 16–25, 11 2003.
- [2] 日本への「ネウボラ」導入過程と「母子健康包括支援センター」の設置—「切れ目ない支援」政策とは—. 同志社女子大学学術研究年報, 第 71 巻, pp. 63–82, 12 2020.
- [3] 松井圭三, 今井慶宗. 子育て支援政策の考察: 公明党と政府審議会の政策の対比. 中国学園紀要, No. 9, pp. 139–146, 2010.
- [4] 山崎泰彦. 子ども・子育て支援と財源政策の論点. DIO : data information, opinions : 連合総研レポート: 資料・情報・意見, Vol. 34, No. 12, pp. 20–25, 12 2021.
- [5] 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 (NLP2023), pp. Q7–10, 3 2023.