

保健行政のデジタル化における課題抽出のための情報分析基盤の構築

研究分担者 大島 裕明 兵庫県立大学大学院情報科学研究科 准教授

研究要旨

本研究では、新聞記事アーカイブを対象とした情報分析基盤を構築し、保健行政におけるデジタル化の課題抽出を試みた。2020～2023年の毎日新聞記事を分析対象とし、汎用言語モデルBERTや大規模言語モデル（LLM）を活用して、記事の分類・要約・可視化を実施した。分析の結果、業務プロセス、人材・文化、技術、制度、住民対応の5領域における課題が抽出された。新聞記事分析と自然言語処理技術の併用により、網羅的かつ実証的な課題抽出が可能であることが示唆された。

A. 研究目的

保健行政におけるデジタル化推進においては、様々な課題が存在している。これまでのインタビュー調査などによって、様々な課題の存在が明らかにされる一方で、インタビュー調査では明らかとならないような課題についても、より網羅的に分析することも必要になると考えられる。本研究では、そのような課題抽出のための情報分析基盤を構築するために、数年間の新聞記事のアーカイブを分析することができる基盤の構築を行った。

本研究では、特に、COVID-19への対応における保健所に関係する課題抽出を行うことを目的とした。まず、分析対象を、2020年から2023年の毎日新聞の記事とした。テキストマッチによる新聞記事事件数の取得や、ワードクラウドの作成などは容易に行うことが可能である一方で、課題抽出のためには、より高度な自然言語処理技術を活用した分析が必要となる。今回の基盤構築では、汎用言語モデルの一つであるBERTを用いた記事分類と、大規模言語モデルを用いた記事要約ならびに記事分析を行うことができたようにした。さらに、ChatGPTを用いた、現在の大規模言語モデルの持つ知識をベースとした課題抽出についても行った。

B. 研究方法

分析サーバを用意し、そこで、毎日新聞の新聞記事を読み込み、分析するための様々な機能を実装した。具体的には、以下の機能の実装を行った。

- ・ キーワードによる記事検索
- ・ 記事集合のグラフによる可視化
- ・ 記事集合のワードクラウドによる可視化

- ・ 記事や記事集合における重要なキーワードの抽出
- ・ 汎用言語モデルを用いた記事の分類
- ・ 大規模言語モデルを用いた記事や記事集合の要約

実装されたこれらの機能を用いた分析の一例として、「コロナ」という語を含む記事集合の分析と、保健所のデジタル化についての課題についての分析を行った。

（倫理面への配慮）

人を対象とした研究ではないため、倫理面への配慮は特段必要としない。なお、分析データについては、インターネットからアクセスできるようにはなっておらず、また、ChatGPTなどの外部のAPIサービスを利用することによる分析データの流出がないように配慮した。

C. 研究結果

2019年までには、「コロナ」という語を含む新聞記事はほとんど存在していなかった。それに対して、2020年には16,074件、2021年には13,744件、2022年には7,484件、2023年には3,547件の記事が現れており、あらゆる分野のニュース記事において「コロナ」という語が現れていたことが明らかとなった。そのうち、「保健所」という語を含むニュース記事は、2020年には359件、2021年には274件、2022年には140件、2023年には37件存在しており、「コロナ」という語と「保健所」という語の共起性の高いことが示唆された。

大規模言語モデルを用いた保健所のデジタル化に関して、大きく以下の5つの分類についての課題が得られた。

- **業務プロセス・運用面の課題**
アナログ業務の多さ、迅速なデータ分析が困難であること、デジタル業務が属人化することなどを含む。
- **人材・組織・文化の課題**
ITリテラシーのばらつき、現場の抵抗感、データ品質や正確性のばらつき、デジタル人材の不足、教育機会の不足などを含む。
- **技術・システム面の課題**
システムの統一性の欠如、チャットボットや自動応答システムが未活用であること、電子申請や届出が未整備な部分があることなどを含む。
- **制度・ガバナンス・運用設計の課題**
情報共有の非効率であること、DX推進の戦略やビジョンがそもそも不在であることなどを含む。
- **住民・利用者対応に関する課題**
セキュリティ対策が不十分であること、現場の業務とシステム設計に乖離があること、住民参加型DXになっていないことを含む。

D. 考察

本研究における分析基盤の構築から、新聞記事の分析によって、COVID-19における保健所の対応などについて分析が可能となることが示唆された。特に、「コロナ」という語を含む新聞記事が3万件以上存在しており、分析に十分な量のデータがあると言える。一方で、データ量が大量であることから、統計的なデータの概観よりもさらに深い、意味的な分析を行うためには、自然言語処理技術の活用が必要となることも示唆された。

大規模言語モデルを用いた課題の分析については、ある程度網羅的な課題について得られた。その一方で、課題を裏付けるデータとの紐付けなど、さらに深掘りした分析が必要であると考えられる。

E. 結論

新聞記事のアーカイブの分析や大規模言語モデルの利用によって、保健行政におけるデジタル化推進のための課題抽出が行えることが示唆された。今後、構築された分析基盤を用いて、さらに深い分析を継続する予定である。

F. 健康基本情報

(総括研究報告書にまとめて記載)

G. 研究発表

1. 論文発表
特になし。
2. 学会発表
特になし。

H. 知的財産権の出願・登録状況

1. 特許取得
特になし。
2. 実用新案登録
特になし。
3. その他
特になし。