

◆研究背景

社会科学は、自然科学と比べて理論体系や研究アプローチが多様化する傾向があり、文献レビューにおいても総合する難しさが指摘されてきました。その代表例として、公会計 (Public Sector Accounting: PSAR) の領域は、財務会計・管理会計・行政管理・経済学・政治学というような複数の観点が交錯する学際的領域であり、従来のレビュー手法では著者が論文を読み、トピックを著者の主観に基づいて分類するため、著者の持つバイアスを小さくしながら全体像や主要なトピックを把握することが困難でした。特に定量的・定性的などの多様な方法論が混在した研究領域である場合、メタアナリシスなどの統合作業も難しい状況です。さらに、旧来の LDA などのトピックモデリングを用いた研究では、単語間の関係性のみに着目しており、文脈の理解に乏しい状況でした。

一方、自然言語処理の分野では、BERT (Bidirectional Encoder Representations from Transformers) に代表される大規模言語モデル (LLM) が登場し、文脈理解に基づいた高精度な文献分類やトピック抽出が可能になっています。

本研究はこのような背景のもと、BERT ベースのトピックモデリング (BERTopic) を活用して、公会計研究の 40 年間 (1980 ~ 2019 年) の動向と構造を明らかにすることを目的としています。

◆研究内容

本研究では、以下の 2 つのリサーチクエスチョン (RQ) を設定し、大規模言語モデル BERT を用いたトピック分類を実施しました。

RQ1: BERT ベースのトピックモデリングにより、公会計研究にはどのような主要トピックを顕在化できるのか?

RQ2: それらのトピックを踏まえ、今後の研究課題はどのように整理されるべきか?

具体的には、ABDC ランク*3 上位の会計系ジャーナル 11 誌から「public sector」、「government」等のキーワードで抽出した公会計関連論文 306 本 (1980~2019 年) を収集し、全文を対象に形態素解析や前処理を実施した後、BERTopic を用いて文献のクラスタリングと主題分類を実施しました。分析の結果、7 つのトピックが識別され、特に以下の 6 つが主要テーマとして抽出されました (①監査と債券市場、②医療費管理、③業績と管理会計、④公共汚職、⑤公共サービス、⑥公共・環境アカウンタビリティ、⑦その他)。これらを時系列でトレンドを確認したところ、1980 年代から 2010 年代にかけて「監査と債券市場」、「医療費管理」は減少傾向にありますが、「業績と管理会計」、「公共・環境アカウンタビリティ」は 2000 年代までは増加傾向でその後は同論文数程度を維持、「公共サービス」、「公共汚職」は増加傾向であることを確認しました。

◆今後の展開

本研究の今後の展開としては以下の 3 点が期待されます。

1. Emerging Technology 分野との接続

公会計研究の領域では、AI・XBRL*4・ブロックチェーンといった「新興技術」分野に関するトピックが十分に抽出されませんでした。これは、公会計分野における技術研究の不足を示唆しており、既存トピックと技術応用の交差領域が重要な研究テーマとなる可能性があります。

2. 他分野・他領域への応用展開

BERTopic による文献レビューは、公会計に限らず、ESG 投資*5、サステナビリティ報告、監査研究など他の会計領域にも応用可能です。特に、多様な方法論が採用されている分野や、比較的新しい文献数の少ない分野でも、全文分析によって高精度な分類が期待されます。

3. 政策・教育・実務への知識転移

本手法は、研究者だけでなく、政策立案者・実務家・教育者に対しても、研究トレンドやギャップの可視化を通じた実務知識の活用支援として貢献できる可能性があります。

◆研究費

本研究は、JSPS 科研費 JP21H00762、JP22K01811、JP25K05431、JST さきがけ JPMJPR2365 の支援を受けて実施されました。

◆論文情報

タイトル: Topic classification of public sector accounting research: A BERT-based approach

著者: Kuroki, Makoto and Hirose, Yoshitaka

掲載雑誌: Journal of Emerging Technologies in Accounting

DOI: <https://doi.org/10.2308/JETA-2024-019>

◆用語説明

*1 大規模言語モデルBERT: 2018年にGoogleにより論文で発表されたNLP(自然言語処理)モデルで、Transformer双方向エンコーダ表現(Bidirectional Encoder Representations from Transformers)の略。深層学習と呼ばれる学習方法のモデルの一種で、過去のNLP(自然言語処理)モデルと異なり、文章を文頭と末尾の双方向から事前学習するように設計されている。また、学習に使用することができるデータが大量に存在し、様々なタスクに対して柔軟な対応が可能という特徴がある。

*2 LDA: Latent Dirichlet Allocation (潜在ディリクレ配分法)とは、膨大なテキストデータから「隠れたテーマ(トピック)」を自動的に発見するための統計的手法で、データがどのトピックをどの程度含んでいるかを確率的に推定する。

3 ABDC ランク: Australian Business Dean Council (オーストラリアビジネス学部長評議会)は、ビジネス研究と学術における高い水準を提供するために設立され、ABDC ジャーナルリストは、A、A、B、C の4段階に分類されている。研究者が論文掲載に適したジャーナルを特定するのに役立つ。

*4 XBRL: eXtensible Business Reporting Language の略で、XML を財務報告用に特化させた言語。有価証券報告書のうち、XBRL で作成されるドキュメントは財務諸表の貸借対照表や損益計算書、株主資本等変動計算書、キャッシュフロー計算書など。

*5 ESG 投資: Environment (環境)、Social (社会)、Governance (ガバナンス)の頭文字を表しており、環境問題の一種である気候変動など、ESG 要素を考慮する投資手法。

以上