

令和 6 年 6 月 5 日現在

機関番号：11301

研究種目：基盤研究(B)（一般）

研究期間：2020～2023

課題番号：20H01714

研究課題名（和文）機械学習によるテキストデータの分析に基づく新たなFD・IRモデルの開発

研究課題名（英文）Development of a new FD/IR model based on analysis of text data using machine learning

研究代表者

松河 秀哉（Matsukawa, Hideya）

東北大学・高度教養教育・学生支援機構・講師

研究者番号：50379111

交付決定額（研究期間全体）：（直接経費） 13,600,000円

研究成果の概要（和文）：本研究は、複数の大学から収集した大量の授業評価アンケートの自由記述データに基づいて、機械学習により自由記述を自動的に複数のカテゴリに分類することができるシステムを開発した。このシステムを用いて、各大学において、授業評価アンケートの自由記述の分類や、その他の教学データを紐づけた分析を容易に行えるようにすることで、エビデンスに基づいてFD(Faculty Development)やIR(Institutional Research)の実践を推進する仕組みを実現した。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、大学での教育活動に対する学習者の反応として、重要な情報が含まれることが分かっていながら、その量が膨大であるために、ほとんど分析・活用されてこなかった授業評価アンケートの自由記述を、機械学習を活用することで自動的に分類できるようにしたことである。これにより、各記述にどのようなことが書いてあるのか、ある観点を含む自由記述は全体の中にどの程度含まれるのかといった分析を、誰でも簡単に素早く行えるようになった。こうした分析に基づく教育活動の現状把握や教育改善を、大学全体として行うことも容易となるため、今後はそうした活動をとおした教育の質の向上が期待される。

研究成果の概要（英文）：In this study, we developed a system that employs machine learning to automatically categorize free descriptions of course evaluation questionnaires from a substantial dataset collected from multiple universities. This system enables universities to efficiently classify free descriptions of course evaluation questionnaires and integrate them with other educational data for in-depth analysis. By leveraging this system, we established a framework that supports evidence-based practices in Faculty Development (FD) and Institutional Research (IR), thereby advancing the implementation of these initiatives.

研究分野：教育工学

キーワード：授業評価アンケート 自由記述 機械学習 FD IR トピックモデル

1. 研究開始当初の背景

大学における授業評価アンケートの実施率は、1991年の大綱化以降増加の一途をたどり、2016年度の実施率は99.2%と、ほとんどの大学で実施されるようになった(文科省2016)。しかしアンケートの結果の活用については、2013年度の段階で、集計・分析結果を報告書等で学内外に公表している大学が18.2%と低調な状況が続いていた(文科省2013)。報告書の内容についても、授業や科目群ごとに、アンケートの各項目の頻度や平均値をまとめたものが主流であり(東北大学2018など)、項目間の関係など、より詳細な分析はほとんど扱われていなかった。

授業評価アンケートは授業担当教員にとって授業のフィードバックとしてたしかに重要であるが、それ以上に、個々の授業の壁を越えて、全体をとりまとめて分析することで、科目群レベルや、カリキュラムレベル、さらには大学全体の傾向など、非常に多くの情報を引き出せる可能性を秘めている。そうした分析は、研究レベルではいくつか見られるものの(星野ほか2006など)、実践レベルでは活用されているとは言えない状況であった。特に授業評価アンケートの自由記述に関しては、あらかじめ設定された項目に比べて、より多様で利用価値が高い情報を含んでいると思われるにもかかわらず、主に自由記述の分量に起因する、手作業による分析の困難さから、1つのクラスのレベルを超えて、どのような記述がどの程度見られるのかといった分析が行われることはなかった。

そうした状況において、代表者らはこれまでに、機械学習の一手法である、トピックモデルと呼ばれる技術を用いて、授業評価アンケートの自由記述の自動分類に取り組んできた。トピックモデルは、文書に含まれる話題抽出を行う教師無し学習の一種であり、文章中の各文がどのような話題に該当するのかを自動的に分析することができる。代表者らはこの手法を用いて、自由記述の中にどのような話題が、どの程度含まれるかに関して、十分な精度で分析することに成功し、さらに各自由記述とその自由記述が書き込まれた科目の情報を紐付けることにより、科目群の中でどのような話題が多いのか、科目群間で話題に差はあるのかといった、FDやIRに繋がる様々な分析が可能になることを示すことができた(松河ほか2018a)。この成果を受け、代表者らは、挑戦的研究(萌芽)(平成29~31年度)「自由記述の自動分類に基づいた授業評価の分析と大学における教育改善への包括的活用」において、トピックモデルを用いて2つの大学の授業評価アンケートの自由記述を分析し、科目群ごとの自由記述の特徴を抽出して科目群別のFD活動に対して情報提供を行ったり、大学間の自由記述の特徴を比較することでIRの活動を推進したりすることが実現しつつあった。

これらの研究を進める中で、トピックモデルを用いた授業評価アンケートの分析に関して得られたのは、以下の2点の知見であった。

1. トピックモデルを用いて授業評価アンケートの自由記述を分析する場合、大学ごとに分析するよりも、複数大学のデータをまとめて分析した方が、データ量が増えるため、分類の精度があがる。
2. 複数大学のデータをまとめて分析すると、各大学の自由記述が同じ基準に従って分類されることになるため、各大学内での授業や科目群などの比較に加え、大学間の話題の差についても量的な比較が可能となる。

これらの知見から導かれるのは、より広範囲な大学から、より大量の授業評価アンケートの自由記述を収集し、大量のデータに対応するために分析とフィードバックの自動化をより一層進めることで、高精度な分類結果へのアクセシビリティが高まり、データに基づいたFDやIRの実践を一層推進させることができるのではないかという問いであった。本研究はこうした背景のもと、複数の大学から大量の授業評価アンケートデータを収集し、分析することで、この問い実証的に応えようとするものであった。

2. 研究の目的

本研究では、前述の問いに対応するため、2つの目的を設定した。第一の目的は、複数の大学から授業評価アンケートの自由記述データを大規模に収集したうえで、トピックモデルを中心とした機械学習の技術を用いて、授業評価アンケートの自動分類モデルや、分類結果のフィードバックシステムを開発することであった。第二の目的は、開発したモデルやシステムを用いた各大学による自由記述の分析を核として、各大学が固有なデータを紐付けた、より深い分析を独自に行い、そのエビデンスに基づいてFDやIRの活動を推進する、新たなFD・IRモデルを開発することであった。

3. 研究の方法

本研究では、研究目的に対応して、以下の2つの課題を設定し、研究を遂行した。

課題1：機械学習を活用した自由記述の自動分類・フィードバックシステム開発

この課題では、機械学習を活用して授業評価アンケートの自由記述を自動的に分類し、その結果についても自動的にフィードバックするシステムの開発を目指した。代表者らはこれまでに、トピックモデルによる分析の大部分を自動化するソフトウェアを開発していたが(松河ほか



図3 webシステムの概念図

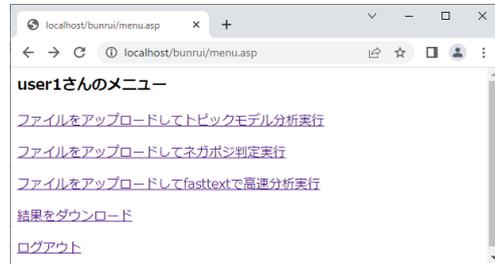


図4 最終的なwebシステムのメニュー画面

先に示したシステムの入出力をweb化し、分析がサーバ上で実行されるようにすることで、分析対象となる自由記述が格納されたエクセルファイルをwebにアップロードするだけで、各自由記述に該当するトピックの番号が付加されたエクセルファイルをダウンロードできるようにするものである。システムのweb化によって、自身のPCに特別なソフトウェアをインストールする必要がなくなり、より容易に自由記述の自動分類を行うことが可能となった。

以上の開発により、トピックモデルによる専門的な知識を持たなくとも、自由記述の自動分類を実行できる環境が整ったわけであるが、このシステムを用いて分析を進めていく上で、一つの問題が生じた。それは、コロナ禍による授業のオンライン化にともない、多くの教員の授業スタイルが変化し、結果として授業評価アンケートの自由記述にも変化が見られたことである。すでに開発していたモデルは、コロナ禍以前に収集した授業評価アンケートの自由記述に基づいて開発されていた。しかし、そこには動画の質や、オンデマンド/リアルタイムといった授業形態など、オンライン授業に関係する観点は全くといっていいほど含まれていなかったため、従来のモデルを用いて、コロナ禍以降の授業の分析を行っても、授業の実態に即した結果が得られないことが予想された。

そこで、課題1では、各大学からコロナ禍になって以降の2020年度の自由記述も含めて、約38万件の自由記述を収集し、トピックモデルによって学習を行って、自由記述をオンライン授業に対応する観点を含んだ160種類のトピックに分類することが可能なモデルを新たに作成し、webシステムに組み込んだ(松河ほか 2021b)。

webシステムにはこのほかに、fastTextを用いて、分析対象となる自由記述がポジティブな記述なのか、ネガティブな記述なのかを判断する機能も実装した(松河ほか 2022, 2023)。さらに、トピックモデルによる160種類の分類結果のうち似通ったものを67種類に統合した上で、それを正解データとしてfastTextに学習させることで、より高速で解釈が容易な分類モデルを作成し、webシステムに組み込んだ(松河ほか 2023)。webシステムは最終的には図4に示したような分析機能を実装するに至った。

課題2については、研究期間がちょうどコロナ禍と重なったため、特に前半は、研究会等をおした、分析の事例や研究の知見の共有を行うことは困難であった。しかし例えば2022年に東北大学教育評価分析センターが発行したCIR Insights(図5参照)では、本科研の取り組みとして作成した、コロナ禍以降の授業評価アンケートの分類モデルを活用した分析結果が活用された。そのほか、授業評価アンケート以外の教育調査の自由記述の分析に、本研究で開発した、各記述がネガティブなものかポジティブなものかを判断するモデルが適用できるかどうかを検証したところ、授業評価アンケート以外の教育調査の自由記述であっても、ある程度正確にその記述がネガティブなものかポジティブなものかを判断することが可能であることが明らかとなり、本研究で開発したシステムの適用可能性を拡張することができた。

コロナ禍の影響がほぼ落ち着いた、科研の最終年度となる2023年度においては、いくつかの対面でのイベントを通して、本科研の取り組みを大会的に伝えることができた。まず、名古屋大学で開催された「名古屋大学教育発達科学研究科MDS教育推進室第1回講演会」において、「教育分野におけるテキストデータの分析 - トピックモデルの活用を中心に - 」と題した講演を行い、教育データの分析にトピックモデルを用いることの利点と限界について論じる機会を得た。

日本教育工学会2024年春季全国大会においては、「機械学習を用いた授業評価アンケートの自由記述の自動分類システムの開発とその活用」と題した自主企画セッションを開催した(松河ほか 2024)。このセッションでは、本研究で開発したシステムの概要が紹介されると共に、京都外国語大学や大阪大学においてシステムを用いて行われた分析の事例も紹介され、システムを用いたFDやIRのイメージをセッションの参加者に伝えることができた。

また、大学コンソーシアム京都が主催する「第29回FDフォーラム」においては、「授業評価アンケートの自由記述の自動分類とその応用」と題したワークショップ形式の実践を行う機会が得られた。この実践では、30名弱の参加者に、自身のPCと分析対象となる自由記述データを持った状態で来場してもらい、全員にIDとパスワードを配布することで、開発したwebシステムにアクセスを許可して、自由に分析を試してもらった。分類結果とその他の観点を組み合わせたクロス表の作成方法などのチュートリアルも提供した結果、参加者からは、webシステムをおした分析の容易さや、分析の精度、分類結果が付与された自由記述の活用方法に関する見通しなど様々な観点からの感想を得ることができた。今後も同様の活動をとおして、本システムによ



図5 CIR Insights での分析結果の活用事例
<http://www.cir.ihe.tohoku.ac.jp/info/cirinsights.html> より引用

る分析結果を活用したエビデンスに基づくFDやIRの取り組みの普及に努めていきたい。

参考文献

星野敦子, 牟田博光 (2006) 大学の授業における諸要因の相互作用と授業満足度の因果関係. 日本教育工学会論文誌, 29(4):463-473

松河 秀哉, 大山 牧子, 根岸 千悠, 新居 佳子, 岩崎 千晶, 堀田 博史 (2018a) トピックモデルを用いた授業評価アンケートの自由記述の分析. 日本教育工学会論文誌, 41(3):233-244

松河秀哉, 大山牧子, 根岸千悠, 新居佳子, 岩崎千晶, 堀田博史, 串本剛, 川面きよ, 杉本和弘 (2018b) トピックモデルによるテキスト分析を支援するソフトウェアの開発. 日本教育工学会論文誌, 42(Suppl):37-40

松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛, 杉本 和弘 (2020) 学習済モデルを活用したトピックモデルによるテキスト分析支援ソフトウェアの開発. 日本教育工学会 2020 年秋季全国大会講演論文集, 51-52

松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛, 杉本 和弘 (2021a) トピックモデルによるテキスト分析支援 web システムの開発. 日本教育工学会 2021 年春季全国大会講演論文集, 15-16

松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛 (2021b) 大規模データに基づいた授業評価の自由記述分類モデルの開発. 日本教育工学会 2021 年秋季全国大会講演論文集, 15-16

松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛 (2022) 機械学習による授業評価の自由記述のネガポジ判定. 日本教育工学会 2022 年春季全国大会講演論文集, 1-2

松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛 (2023) 授業評価アンケートの自由記述を自動分類するための web システムの開発. 日本教育工学会 2023 年秋季全国大会講演論文集, 123-124

松河 秀哉, 根岸 千悠, 村上 正行 (2024) 機械学習を用いた授業評価アンケートの自由記述の自動分類システムの開発とその活用. 日本教育工学会 2024 年春季全国大会講演論文集, 7-8

文科省 (2013) 大学における教育内容等の改革状況について (平成 25 年度).

文科省 (2016) 大学における教育内容等の改革状況について (平成 28 年度).

東北大学 学務審議会 評価・改善委員会 (2018) 平成 30 年度 第 2 学期 学生による授業評価アンケート実施結果報告書.

東北大学高度教養教育・学生支援機構 教育評価分析センター (2022) CIR Insights (Vol.14) 2021 年冬号 <http://www.cir.ihe.tohoku.ac.jp/sys/wp-content/uploads/2022/03/120c7739c863333e69485fb327c85a3.pdf> (参照日 2024.05.09)

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 松河 秀哉 串本 剛 杉本 和弘	4. 巻 8
2. 論文標題 東北大学の全学教育における授業形態の影響 - 授業評価アンケートの結果から -	5. 発行年 2022年
3. 雑誌名 東北大学高度教養教育・学生支援機構紀要	6. 最初と最後の頁 3-12
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 -

1. 著者名 松河秀哉, 山内保典, 佐藤智子, 中川学, 縣拓充, 中村教博, 串本剛, 杉本和弘, 渡邊文枝	4. 巻 第7号
2. 論文標題 オンライン授業の現状と学生の評価 基礎ゼミ受講者へのアンケート結果を中心に	5. 発行年 2021年
3. 雑誌名 東北大学 高度教養教育・学生支援機構 紀要	6. 最初と最後の頁 3-21
掲載論文のDOI（デジタルオブジェクト識別子） なし	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14件（うち招待講演 3件/うち国際学会 1件）

1. 発表者名 松河秀哉
2. 発表標題 授業評価アンケートの自由記述の自動分類とその応用
3. 学会等名 大学コンソーシアム京都第29回FDフォーラム（招待講演）
4. 発表年 2024年

1. 発表者名 松河秀哉
2. 発表標題 教育分野における テキストデータの分析 - トピックモデルの活用を中心に -
3. 学会等名 名古屋大学教育発達科学研究科MDS教育推進室 第1回講演会（招待講演）
4. 発表年 2024年

1. 発表者名 松河秀哉
2. 発表標題 ChatGPTの大学教育への影響と活用
3. 学会等名 福島大学 2023 年度 全学FD 研究会（招待講演）
4. 発表年 2023年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛
2. 発表標題 授業評価アンケートの自由記述を自動分類するためのwebシステムの開発
3. 学会等名 日本教育工学会 2023年秋季全国大会
4. 発表年 2023年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛
2. 発表標題 大規模言語モデルを活用した自由記述の分類手法の提案
3. 学会等名 日本教育工学会 2024年春季全国大会
4. 発表年 2024年

1. 発表者名 松河 秀哉, 根岸 千悠, 村上 正行
2. 発表標題 機械学習を用いた授業評価アンケートの自由記述の自動分類システムの開発とその活用
3. 学会等名 日本教育工学会 2024年春季全国大会
4. 発表年 2024年

1. 発表者名 村上 正行, 岡田 玖美子, 松河 秀哉
2. 発表標題 初年次教育における授業アンケートの自由記述の分析
3. 学会等名 日本教育工学会研究会(2024年5月)
4. 発表年 2024年

1. 発表者名 Hideya MATSUKAWA, Chiaki IWASAKI
2. 発表標題 Development of a Formative Assessment System for Reports Using Large Language Models and Rubrics
3. 学会等名 ISET2024 (国際学会)
4. 発表年 2024年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛
2. 発表標題 授業評価に対するネガポジ判定モデルの他データへの援用可能性の検討
3. 学会等名 日本教育工学会 2022年秋季全国大会
4. 発表年 2022年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛
2. 発表標題 教師なし学習による分類結果を教師データとした授業評価アンケートの自由記述の自動分類
3. 学会等名 日本教育工学会 2023年春季全国大会
4. 発表年 2023年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛
2. 発表標題 大規模データに基づいた授業評価の自由記述分類モデルの開発
3. 学会等名 日本教育工学会 2021年秋季全国大会
4. 発表年 2021年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛
2. 発表標題 機械学習による授業評価の自由記述のネガボシ判定
3. 学会等名 日本教育工学会 2022年春季全国大会
4. 発表年 2022年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛, 杉本 和弘
2. 発表標題 学習済モデルを活用したトピックモデルによるテキスト分析支援ソフトウェアの開発
3. 学会等名 日本教育工学会 2020年秋季全国大会
4. 発表年 2020年

1. 発表者名 松河 秀哉, 大山 牧子, 根岸 千悠, 村上 正行, 川面 きよ, 渡辺 雄貴, 江本 理恵, 富永 陽子, 串本 剛, 杉本 和弘
2. 発表標題 トピックモデルによるテキスト分析支援web システムの開発
3. 学会等名 日本教育工学会 2021年春季全国大会
4. 発表年 2021年

〔図書〕 計0件

〔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
研究分担者	川面 きよ (Kawazura Kiyo) (20782064)	成城大学・グローバル研究センター・研究員 (32630)	
研究分担者	村上 正行 (Murakami Masayuki) (30351258)	大阪大学・全学教育推進機構・教授 (14401)	
研究分担者	渡辺 雄貴 (Watanabe Yuki) (50570090)	東京理科大学・教育支援機構・教授 (32660)	
研究分担者	江本 理恵 (Emoto Rie) (60400181)	北海道大学・高等教育推進機構・教授 (10101)	
研究分担者	串本 剛 (Kushimoto Takeshi) (60457835)	東北大学・高度教養教育・学生支援機構・准教授 (11301)	
研究分担者	根岸 千悠 (Negishi Chiharu) (60726610)	京都外国語大学・共通教育機構・講師 (34302)	
研究分担者	大山 牧子 (Oyama Makiko) (70748730)	神戸大学・大学教育研究センター・准教授 (14501)	

6. 研究組織（つづき）

	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
研究 分 担 者	冨永 陽子 (Tominaga Yoko) (70775361)	岩手大学・研究支援・産学連携センター・客員准教授 (11201)	

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