

地域鉄道に関するInstagramの投稿を対象とした LLMによる利用実態の把握

Determining the purpose of local railways usage in Instagram posts using large language models

藤本雄紀[†] 櫻井淳[‡] 安彦智史^{††} 池辺正典[‡]
Yuki Fujimoto[†] Jun Sakurai[‡] Satoshi Abiko^{††} Masanori Ikebe[‡]

[†] 大阪産業大学 デザイン工学部

[‡] 文教大学 情報学部

^{††} 仁愛大学 人間学部

[†] Faculty of Design Technology, Osaka Sangyo University.

[‡] Faculty of Information and Communications, Bunkyo University.

^{††} Faculty of Human Studies, Jin-ai University.

要旨

地域間移動における二次交通手段として期待される地域鉄道だが、利用者数は減少傾向にある。地域鉄道を有効活用した公共交通施策の策定が求められているが、その検討の根拠となる利用実態やイベント等施策の実施効果の把握が困難である。本研究では、地域鉄道に関するInstagramの投稿に着目し、LLM（大規模言語モデル）を用いた投稿内容の分類をおこない、地域鉄道の利用実態を明らかにする。

1. はじめに

地方の観光地において、観光地間を繋ぐ二次交通の不便さや周遊観光におけるアクセス手段不足といったモビリティ（移動利便性）の課題がある。こうした中、大量輸送が可能かつ、定時性、速達性に優れることから、地域鉄道の活用が注目されている。日本版持続可能な観光ガイドライン[1]においても、環境負荷の小さい交通機関の利用促進プログラムがあることが推奨されており、持続可能な観光の観点からも、地域鉄道の活用が求められている。しかしながら、国土交通省の調査によると、地域鉄道の利用者は急速に減少しており、ピーク時（平成3年度）から令和元年度にかけて22%もの減少が確認されている[2]。そのような状況下において、地方自治体は地域鉄道を有効活用した公共交通施策の策定が求められている。その一案として、地域の課題やニーズに即したサービスの提供などが考えられるが、その検討の根拠となる地域鉄道を利用した移動の実態（利用実態）、観光客の移動の実態（潜在需要）、イベント等施策の実施効果の把握が困難であるといった課題が存在している。主に経済・経営・商学分野で地域鉄道の存続または廃線による経済的效果を分析する研究や[3]、全国の地域鉄道95社に対し、「需要希薄型」、「需要潜在型」、「需要堅調型」に分類し、それぞれに応じた活性化方策を提案している研究[4]があるが、いずれの分析も鉄道の利用率や運賃、総延長といった諸元データに基づくものであり、利用者の意見を対象に大規模かつ継続的に収集・分析した研究は見当たらない。

実態を把握するための方策として、「地域モビリティ確保の知恵袋2022」では、地域モビリティ（移動利便性）確保のために、デジタル技術を駆使した具体的な計画・マネジメントに関する取り組みのポイントが紹介されている[5]。目的別に携帯電話位置情報やICカード記録などのデジタルデータの活用方法などが例示されているが、ここにSNSからの利用者の生の声を加えることができると、より実態に即した分析が可能となると考えられる。

SNSから抽出できるデータの一例として、地域鉄道に対する要望や利用目的、イベント等施策の感想などが挙げられるが、雑多な内容が含まれることから、収集したデータから適切に分類し、第三者にわかりやすい形で可視化する必要がある。既存研究にはソーシャルセンサとしてSNSから防災情報を抽出する研究[6]などがあるが、地域鉄道を対象とした場合に、適切に潜在需要や利用特性を抽出する手法は明らかとなっていない。以上を踏まえ、本研究では、SNS上の地域鉄道に関する投稿に対し、LLMを用いた利用実態の把握を試みる。

2. 利用実態の抽出実験

本実験では、各地方鉄道の利用実態を把握するため、Instagram の投稿データを対象として分析する。地域鉄道の利用に関する投稿に対し、生活交通としての日常的な利用に関する投稿であるか、観光目的での利用に関する投稿であるかを判別し、各路線の利用傾向を把握できるかを検証する。

2.1. 対象路線

分析対象とする地方鉄道として、路線特性と地域性の異なる2箇所を選定した。まず、観光目的での利用が主である石川県能登地方の路線であるのと鉄道は、車両本体の内外装を装飾した観光列車が走行していることや、アニメ作品の舞台となった駅舎などがあり、観光目的での利用が多い鉄道である。もう一方は、兵庫県神戸市兵庫区から三田市・小野市方面へ延びる路線である神戸電鉄を選定した。生活交通としての利用が主であるが、有馬温泉などの観光地の最寄り駅も有しており、日常・観光両方の利用が含まれる鉄道である。

2.2. データの収集

各路線の利用実態を反映するデータとして、Instagram API を用いて各路線名のハッシュタグを含む投稿を収集した。収集データは、投稿テキスト、投稿日時、いいね数、投稿に添付された動画像ファイルである。のと鉄道は2024年1月1日に能登半島地震にて被災し、それ以降災害に関する投稿が増加したため、収集期間は2023年8月20日から2023年12月31日までに投稿された699件を対象とした。神戸電鉄は2023年08月21日から2023年09月28日までに投稿された699件を対象とした。これに対し、人目によって日常利用・観光利用・その他の投稿に分類した正解データを作成した。正解データの内訳は、のと鉄道は日常利用が31件、観光利用が451件、その他が217件、神戸電鉄は日常利用が353件、観光利用が96件、その他が250件であった。

2.3. LLM を用いた投稿内容の分類

LLM を用いて Instagram の投稿データを分類する際には、1つの投稿に対して1つのプロンプトを作成することとした。膨大な数のプロンプトを実行することとなるため計算コストを低減させることや、パラメータ調整やファインチューニングなどの今後のカスタマイズ性を考慮し、ローカル LLM による分析とした。ローカル LLM には各社から様々なモデルが公開されているが、パラメータ数が異なる日本語に対応している LLM モデルを選定し、その性能を比較する。プロンプトには、投稿テキスト、投稿ページ URL を指定し、日常利用・観光利用・その他のの中から一意に利用目的を判定することと、そう判断した理由を出力するよう記述した。そして、人目によって判定した正解データと、LLM が output した結果を比較し、F 値によって性能を評価する。

3. 結果と考察

LLM で分類した結果を表1に示す。のと鉄道においては、LLM のパラメータ数が増えるにつれ、F 値が高い傾向があることが確認できる。一方、神戸電鉄においては、Qwen2.5 のモデルが最も高い結果が得られた。また、のと鉄道と神戸電鉄を比較すると、のと鉄道に関する投稿の方が正確に分類できており、神戸電鉄は全体的にスコアが低い結果となった。

のと鉄道の結果に着目すると、Gemma2 では7割程度の精度で日常・観光・その他の投稿に分類できており、LLM による分析は、地域鉄道の利用傾向の把握の一助となることが示された。投稿内容を確認すると、「【2023夏】のと旅その2】君ソムラッピング車と稻穂この旅の目的の一つが、君ソムラッピング車と稻穂を撮ること。」や、「愛称の「能登さくら駅」というネーミングそのままに、春はホーム全体が桜に覆われるのと鉄道・能登鹿島駅。ずっと気になっていた桜の駅に、今春ついに訪問できました。」といったような、のと鉄道のラッピング車両や駅舎など、のと鉄道を利用すること自体を観光目的とする投稿が多くみられ、投稿文中にも旅や訪問といった観光に関する単語も頻出していることから、精度よく分類できたと考えられる。

表1 各地域鉄道に関する投稿の分類性能の結果

対象鉄道	投稿数	LLM モデル	パラメータ数	正解率	適合率	再現率	F 値
のと鉄道	699 件	Llama3 ELYZA-JP	80 億	0.468	0.743	0.468	0.421
		Qwen2.5	140 億	0.670	0.682	0.670	0.673
		Gemma2	270 億	0.784	0.769	0.784	0.772
神戸電鉄	699 件	Llama3 ELYZA-JP	80 億	0.403	0.759	0.403	0.296
		Qwen2.5	140 億	0.583	0.662	0.583	0.569
		Gemma2	270 億	0.459	0.665	0.459	0.376

一方、神戸電鉄では、LLM のパラメータ数を高めたとしても比較的分類精度が低く、最大でも 6 割弱程度の結果となった。投稿を確認すると、日常利用に関する投稿の例では、「神鉄粟生線は鈴蘭台から鈴蘭台西口間において線路内土砂流入のため運転見合わせ※大雨で法面が崩壊している模様です。」といったような鉄道の運行状況に関する投稿が多い傾向があった。運行情報や遅延情報など、必ずしも利用者の行動を直接的に反映しない投稿が多く含まれていたことが、分類精度が低くなつた一因と考えられる。また、のと鉄道に比べて日常・観光利用共に、神戸電鉄そのものに愛着を持つのではなく、神戸電鉄を利用した先にある周辺の観光地や店舗、風景などに興味関心を持つ投稿が多く、そういった投稿がその他の分類として誤判定されている事例も確認できた。さらに、鉄道車両に着目した投稿、所謂「撮り鉄」の投稿が多く散見され、ノイズとなつていることがわかつた。

このように、特徴的な路線や取り組みを実施しており、地方鉄道そのものに魅力を感じられる鉄道に対しては適切な利用傾向の把握ができるが、ノイズとなる投稿が多く発生しうる路線に対してはテキストベースでの分析では限界があることがわかつた。そのため、動画像の内容を解釈することができるマルチモーダル LLM の適用や、LLM のファインチューニング、その他の SNS での適用も検討する必要がある。また、ユーザの投稿頻度や位置情報、投稿内容を分析してユーザ属性を把握し、ノイズとなるユーザを除去した上で分析することなど、分析手法の工夫も必要であることがわかつた。

4. おわりに

本研究では、地域鉄道の利用実態の把握のために、Instagram の投稿に対して LLM を用いた利用目的の分類を行つた。取り組みが特徴的な鉄道であれば、日常・観光利用の分類は可能であったが、多様な利用目的が含まれる鉄道に対しては、単に LLM を適用するだけでは難しいことがわかつた。今後は、分類性能の向上のためのファインチューニングや、ユーザ属性に着目した分析を加えるなどしたうえで、対象を全国の地域鉄道に拡張し、全国の地域鉄道の特徴を明らかにすることを目指す。

謝辞

本研究は JSPS 科研費 JP24K21021, JP24K05887 の助成を受けたものである。

参考文献

- [1] 観光庁、日本版持続可能な観光ガイドライン、<<https://www.mlit.go.jp/kankochō/content/810000951.pdf>>、2024.1.31 参照。
- [2] 国土交通省、地域鉄道の現状、<<https://www.mlit.go.jp/tetudo/content/001573729.pdf>>、2024.10.31 参照。
- [3] 藤田知也：“地方圏における鉄道需要に関する一考察—パネルデータによる実証分析—”，交通学研究、日本交通学会、Vol.62, pp.141-148, 2019.
- [4] 那須野育大：“地域鉄道活性化方策の提案—事業者の収益改善に資する要因分析—”，戦略経営ジャーナル、国際戦略経営研究学会、Vol.9, No.1, pp.39-58. 2021.
- [5] 国土交通省、地域モビリティ確保の知恵袋 2022、<<https://www.mlit.go.jp/sogoseisaku/soukou/content/chiebukuro2022.pdf>>、2024.10.31 参照。
- [6] 鳥海不二夫：“ソーシャルメディアにおける災害情報”，災害情報、日本災害情報学会、Vol.16, No.2, pp.139-142, 2018.