一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
IN2024-53 (2024-10)

旅行活動ツイートのインスタンス的可視化のための VLM による Image-to-Text を用いた機械分類の精度比較

藤大　友都† 　服部　　峻† 　砂山　　渡†

† 滋賀県立大学　〒522–8533 滋賀県彦根市八坂町 2500
E-mail: †on23yfujidai@ec.usp.ac.jp, ††{hattori.s,sunayama.w}@e.usp.ac.jp

あらまし　旅行活動データ利活用の一つとして可視化が考えられ，新たな旅行先の決定や旅行需要の発見に寄与すると考えるが，SNS のデータは膨大で，旅行以外にも多種多様な情報が含まれる．これまでの我々の研究では，SNS から旅行活動データを網羅的に収集するため，投稿データの本文テキストのみを用いて，機械学習で旅行活動か否かを判定する手法の精度比較を行って来た．一方で，旅行活動データにおいては，本文テキストだけでなく，その旅行先での写真も添付されている場合もあり，重要な構成要素である．そこで本稿では，本文テキストに加え，Vision-Language モデルにより投稿に紐づく画像をテキストデータに変換したものも活用して，機械学習で旅行活動か否かを分類する．
キーワード　分類, SNS 分析, Image-to-Text, 観光, 可視化

Comparison of Effects of VLM-based Image-to-Text on Machine Classifications for Visualizing Tourism Tweets as Instances

Yuto FUJIDAI†, Shun HATTORI†, and Wataru SUNAYAMA†

† The University of Shiga Prefecture　2500 Hassaka-cho, Hikone, Shiga 522–8533 Japan
E-mail: †on23yfujidai@ec.usp.ac.jp, ††{hattori.s,sunayama.w}@e.usp.ac.jp

Abstract　One potential application of travel activity data is visualization, which can aid in the decision-making process for new travel destinations and the identification of travel demand. However, social media data is vast and contains a wide variety of information, not solely related to travel activity. In our previous research, we focused on comprehensively collecting travel activity data from social media and compared the accuracy of machine learning methods for determining whether a post pertains to travel activity or not, using only the post's text. However, in the context of travel activity data, posts often include not only text but also photos at the travel destinations, which are important components. In this paper, we classify whether a post is related to travel activity by utilizing both the post's text and text data converted from the post's images based on a Vision-Language Model (VLM).
Key words　Classification, SNS Analysis, Image-to-Text, Tourism, Visualization

1. ま え が き

近年，旅行活動を SNS に投稿するユーザが増えている．JTB の調査 [1] によると，実施した国内旅行の計画にあたり情報収集を行った媒体として，「SNS やブログ，動画サイト」を選択した割合が，若い世代になるにつれて高く，Z 世代では 3.5 から 5 割を占めている．また，旅行先を選択する際，SNS の情報を重視すると回答した層は全体の 2 割，自分自身の SNS に投稿することを意識して旅行先を選択した層が全体の 1 割を占めており，旅行活動において SNS は必要不可欠な存在となっている．じゃらんリサーチセンター [2] による日本国内の観光ニーズに関する調査からは，より新しく，未知の観光スポットや地域を旅行先として好むユーザが少なからず存在する．

このような旅行活動データの利活用の一つとして，地図へのマッピングなど，可視化が考えられる [3]〜[9]．可視化を行うことで，ある観光エリアにおけるスポットごとの旅行活動の特性が見えるようになり，観光エリアにおいて体験できる活動が明確になる．可視化を通して，新たな旅行先の決定や旅行需要の発見に繋がり，旅行活動が活発になるのではないかと考えた．

旅行活動データの可視化には大別して，具体例一つ一つを全て可視化するインスタンスごとの可視化である「インスタンス的可視化 [3]〜[6]」と，インスタンス群を何らかの形で集約した集合知的な可視化 [7]〜[9] とがある．SNS にしかない旅行活動はマイナーな投稿が少数しかないため，本稿では，それらも

図 1　旅行活動を正しく認識できない場合
Fig. 1　In case of not recognizing travel activity properly.

（集約によって）省略されることなく可視化されるインスタンス的可視化を目指す．

しかし，SNS に蓄積されるデータは日々膨大になり続けており，旅行以外にも多種多様な情報が含まれている．旅行活動データを収集する際，収集の仕方を工夫しなければ，旅行と関係のない情報が混ざり，可視化の妨げになることが考えられる．例えば，彦根の旅行活動を可視化する場合，彦根城や四番町スクエアの観光が想定される．しかし，図 1 のように彦根駅前での街頭演説や，住民の日常生活の投稿など，旅行活動ではないものまで可視化されると，旅行活動が正しく認識できなくなる．

我々はこれまでに，Twitter のテキストデータのみに着目し，旅行活動ツイートの分類を行ってきた [3]〜[5]．しかし，その分類精度には限界があった．一方で，旅行活動データにおいては，本文テキストだけでなく，その旅行先での写真も添付されている場合もあり，重要な構成要素である．そこで，本稿では JSVLM（Vision-Language Model）[10] による Image-to-Text を使用し，従来のテキストデータに加えて投稿に紐づく画像データを合わせて利用することで，機械学習による旅行活動ツイート判定精度の比較を行う．最終的な目標を，「旅行活動ツイートのインスタンス的可視化」と定め，本稿ではその前段階に当たる「旅行活動ツイートの分類精度向上」について比較検討を行う．

本稿の以降の構成は以下の通りである．まず 2 章で関連研究を紹介する．次に，3 章で Image-to-Text を用いた旅行活動ツイート判定精度の比較実験の方法について詳述し，4 章で実験結果をまとめる．最後に，5 章で今後の研究課題について述べる．

2. 関 連 研 究

本章では，関連研究について紹介する．

2.1 Twitter に投稿された画像の分類に基づくツイート文の傾向分析

森野ら [11] は，Twitter の投稿から救助要請の情報を抽出するため，ツイートテキストと画像との関連性を見出すことを目指した．具体的には，令和 2 年 7 月豪雨で「救助/救援」を含むツイートに紐づく画像を分類し，それを基にテキストの特徴を調査した．この研究においては画像を人手で分類しており，本稿とは Twitter の画像特徴をテキスト化して分類するという点で違いがあると考えている．

2.2 深層学習によるインスタグラム画像からの流行抽出

西田ら [12] は，SNS の一種であるインスタグラムの画像を大量に収集し，深層学習を用いた対象物の詳細分析により流行を判定する手法を検討した．対象物の切り出し（物体検出）には YOLO [13]（主に YOLOv3 と YOLOv1）を使用しているが，最新の YOLO の事前学習済みモデル [14] でも切り出し不可能な物体（のクラス）が存在し，流行の網羅的な抽出は不可能である．追加的に fine-tuning を行うなど，出来る限りインスタンス的な物体検出に近付け，流行を漏れなく捉えるという研究課題が残っていると考えられる．また，本稿で重視する，旅行活動における動作や行動は YOLO で取得することができない．

以上の関連研究を踏まえ，本稿では既存の様々な機械学習を用いて，収集したツイートを「彦根の旅行活動」と「彦根の非旅行活動（日常生活や，広告や個人の願望など活動ではないもの）」に分類する実験を行った．具体的なツイートは以下である．

- 彦根の旅行活動：「親が遊びに来たので、彦根城下で昼食。 #ほっこりや #比内地鶏 #彦根 #親子丼」
- 彦根の非旅行活動：「らー麺潮騒です！ 今日から朝ラーメンスタートです 是非お越しください！ #滋賀ラーメン #彦根ラーメン #彦根 #朝ラー #Twitter #siga_biwako」

3. 実 験 方 法

我々はこれまでに，観光地における旅行活動ツイートの収集手法として，ツイートテキストを使用し，「地名」，「位置情報」，「特徴語」の 3 手法を組み合わせたルールベースによるフィルタリング手法により，旅行活動ツイートの網羅的な収集精度の実験 [3] や，SVM や BERT など，従前の機械学習を用いた旅行活動ツイート判定精度の比較実験 [4] を行った．その結果，SVM や BERT は適合率が高く再現率が低いため，取りこぼす旅行活動ツイートが多くなっていることが分かった．また，ナイーブベイズは再現率が高いが，適合率が低いため，旅行活動ツイートを正しく認識する精度が低いという結果となった．

最終的な目標である「旅行活動ツイートのインスタンス的可視化」のために必要なツイートは「観光スポットを示す場所で，旅行活動を示す行動」である．このため，まずは実際に収集される投稿から「旅行活動と関係のない投稿」や，「題材とする観光地での投稿ではあるが旅行活動ではない一般的な活動」をできるだけ取り除きつつも，旅行活動ツイートを精度よく網羅的に収集する手法を検討する．

旅行活動がより活発となる行楽シーズンを対象に，2022 年 10 月のツイートを収集した．その結果，「彦根」を含むツイートを 15944 件収集した．そのうち，画像を含むツイートを 200 サンプリング件抽出し，機械分類の精度比較に用いる．はじめに，200 件のツイートを人の手で「彦根の旅行活動ツイート (a)」，「彦根の（旅行ではない）一般的な活動ツイート (b)」，「彦根を含むだけの（活動とは関係のない）ツイート (c)」の 3 つに分類した．このうち，可視化したいのは「彦根の旅行活動ツイート (a)」であるため，機械学習では「彦根の旅行活動ツイート (a)」か「彦根の旅行活動ツイートではない ($b+c$)」の分類を行う．一方，実際に可視化されるのは何らかの活動を行っている

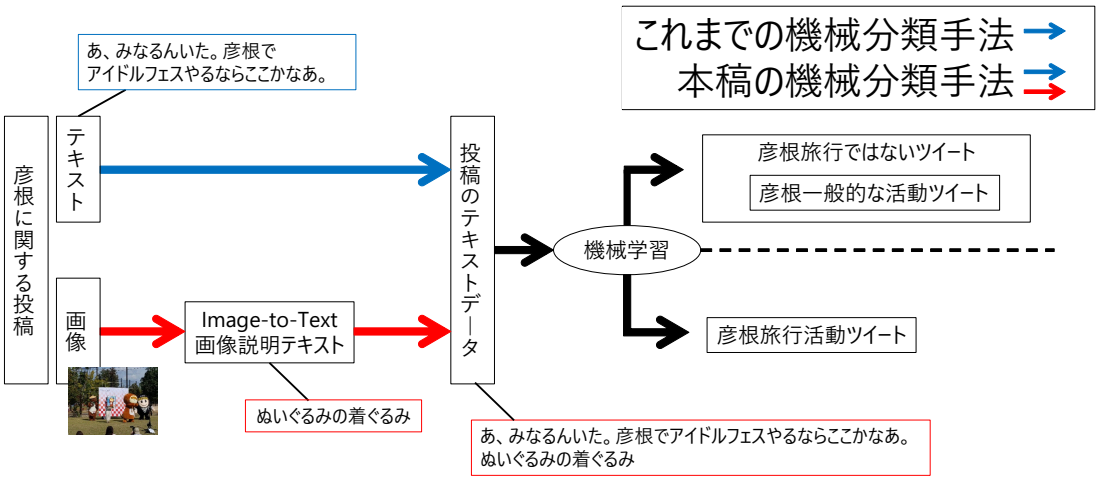

図 2　旅行活動ツイートのインスタンス的可視化のための機械分類手法の流れの比較

Fig. 2　Flow comparison between the previous methods [3]〜[5] and the proposed method of machine classification for visualizing Tourism Tweets as Instances.

表 1　図 3 を JSVLM の入力画像とした時のデータ型毎の出力結果

Table 1　Textual output per data-type by inputting Fig. 3 to JSVLM.

ツイートテキスト（JSVLM には入力していない）	今日の夕食は彦根駅前の麺・食処「八千代」さんにて。 せっかく滋賀県に来たので、近江牛のすき焼き鍋御膳をいただきました。 親子丼などは食べたことがありますが、すき焼きは初めてで満足です。 ごちそうさまでした
int8	うどん定食
fp16	ご飯、味噌汁、その他の食べ物が入った木製のテーブル
fp32	ご飯、味噌汁、その他の食べ物が入った木製のテーブル

「彦根の旅行活動ツイート (a)」と「彦根の（旅行ではない）一般的な活動ツイート (b)」となるが，「彦根の（旅行ではない）一般的な活動ツイート (b)」は適切な可視化の妨げとなるノイズとなるため，「彦根の（旅行ではない）一般的な活動ツイート (b)」が「彦根の旅行活動ツイート (a)」に誤って分類されないようにしたい．そこで，機械学習が「彦根の旅行活動ツイート (a)」と分類したツイートのうち，実際には「彦根の（旅行ではない）一般的な活動ツイート (b)」であったツイートの割合をノイズ率とし，計算式を以下に示す．

$$\text{ノイズ率} = \frac{(x)}{(x)+(y)} \tag{1}$$

- x：彦根旅行活動と予測 かつ 実際のラベルは一般的活動 (b) であったツイート数
- y：彦根旅行活動と予測 かつ 実際のラベルは彦根旅行活動 (a) であったツイート数

ツイートに紐づく画像の内容の説明をテキスト化（Image-to-Text）するため，JSVLM（Japanese Stable VLM）[10] を用いた．JSVLM は 1 つの画像から適切に表現する説明が欲しい時に扱うモデルで，日本語に特化しているという特徴がある．JSVLM にはデータ型が 3 種類あり int8，fp16，fp32 の順にデータ量が多くなり，精度が向上する．本稿では 3 種類のデータ型を全て実行し，それらの精度についても比較した．また，JSVLM による Image-to-Text の出力サンプルとして，入力画像を図 3 に，データ型毎の結果を表 1 に示す．

図 3　Image-to-Text である JSVLM への入力サンプル

Fig. 3　An example of input images to a Image-to-Text, e.g., JSVLM.

機械学習には，これまでの研究で適合率が比較的高い SVM，BERT と，再現率が比較的高いナイーブベイズを使用した．それぞれ交差検定を行い，正解率，再現率，適合率，F1 値，ノイズ率を算出した．以前の研究ではツイートテキストのみを入力していたが，本稿では先述の JSVLM による Image-to-Text を活用する．我々の以前の研究との比較を図 2 に示す．

4. 実験結果

SVM による旅行活動ツイート分類結果を表 2 に，ナイーブベイズの実験結果を表 3 に，BERT の実験結果を表 4 に示す．

表 2　旅行活動ツイート分類に関する SVM の実験結果
Table 2　Experimental results by SVM for tourism tweet classification.

		彦根の旅行活動ツイート			彦根の非旅行活動ツイート			
	正解率	適合率	再現率	F1 値	適合率	再現率	F1 値	ノイズ率
画像なし	0.620	0.224	0.655	0.333	0.913	0.614	0.734	0.050
int8	0.680	0.388	0.733	0.508	0.896	0.665	0.763	0.029
fp16	0.680	0.365	0.756	0.492	0.913	0.660	0.766	0.000
fp32	0.655	0.329	0.700	0.448	0.896	0.644	0.749	0.034

表 3　旅行活動ツイート分類に関するナイーブベイズの実験結果
Table 3　Experimental results by Naive Bayes for tourism tweet classification.

		彦根の旅行活動ツイート			彦根の非旅行活動ツイート			
	正解率	適合率	再現率	F1 値	適合率	再現率	F1 値	ノイズ率
画像なし	0.545	0.447	0.463	0.455	0.617	0.602	0.609	0.073
int8	0.600	0.541	0.529	0.535	0.643	0.655	0.649	0.042
fp16	0.615	0.506	0.551	0.528	0.696	0.656	0.675	0.023
fp32	0.600	0.506	0.531	0.518	0.670	0.647	0.658	0.023

表 4　旅行活動ツイート分類に関する BERT の実験結果
Table 4　Experimental results by BERT for tourism tweet classification.

		彦根の旅行活動ツイート			彦根の非旅行活動ツイート			
	正解率	適合率	再現率	F1 値	適合率	再現率	F1 値	ノイズ率
画像なし	0.655	0.553	0.603	0.577	0.730	0.689	0.709	0.021
int8	0.680	0.588	0.633	0.610	0.748	0.711	0.729	0.000
fp16	0.670	0.565	0.623	0.593	0.748	0.699	0.723	0.020
fp32	0.705	0.647	0.655	0.651	0.748	0.741	0.745	0.018

実験の結果，使用した全ての機械学習において，Image-to-Text を追加することにより精度が 5% 程度向上したことを確認できた．判断材料となるテキストが増えたことで，旅行活動がより明確になったものと考えられる．また，データ型が増えるに伴い，適切に状況を捉えられるようになったが，機械学習による判定精度が変わるほどではなかった．

彦根の旅行活動ツイートを基準とすると，SVM と BERT は適合率よりも再現率が高く，ナイーブベイズは適合率と再現率は同程度であった．これは我々が以前に行ったテキストのみでの分類 [4] と同じ傾向であった．

ノイズ率についてはどの機械学習，データ型であっても低い値を示した．我々は，以前の研究でノイズ率の目標値を 0.1 から 0.2 以下と定めており [4]，本稿の実験結果は目標値を下回っている．これは，ノイズとなる「彦根の（旅行ではない）一般的な活動ツイート (*b*)」が 100 件中 3 件しかなかったことが原因である．今後，「彦根の（旅行ではない）一般的な活動ツイート (*b*)」が増えた場合にノイズ率がどのように変化するかを検証する必要がある．

5.　今後の研究計画

本稿では，SNS から旅行活動データをノイズなく網羅的に収集するため，既存の様々な機械学習を用いて，JSVLM（Vision-Language Model）による Image-to-Text を使用し，テキストデータに加えて投稿に紐づく画像データを合わせて利用することで，機械学習による旅行活動ツイート判定精度の比較を行った．その結果，画像説明テキストを追加することで旅行活動ツイートをより効果的に収集することができることが分かった．

今後の研究計画として「Image-to-Text による観光地特定精度の向上」，「写真からの行動推定」の 2 点を計画している．

5.1　Image-to-Text による観光地特定精度の向上

JSVLM による Image-to-Text は，図 3 や表 1 の入出力結果より，全体的な状況説明が必要な時に有用であることが分かった．しかし，図 3 には存在しない料理が表 1 で出力された．また，図 4 の出力結果（表 5）は自動車の車種名を出力したが，実際の車種名は図 4 にある通り「ダイハツ タント ファンクロス」である [15]．以上の結果より，JSVLM は固有名詞等の詳細に弱いと考えられる．そこで，観光地の写真を JSVLM に事前学習させた上で，Image-to-Text を行うことで，観光地を特定できるのではないかと考えている．

5.2　写真からの行動推定

本稿では，Image-to-Text は画像を説明させるのみに使用した．3 章で述べた通り，旅行活動ツイートの可視化（マッピング）には，場所と行動の情報が必要である．しかし，テキストデータには直接的に明示されていない場合，投稿者が何を行ったかを推測する必要があったため，旅行活動ツイートとしての抽出が困難であった．一方，写真からはテキストだけでは伝えられない情報が多く含まれており，テキストデータと組み合わせることで行動推定の精度が向上できると考えている．図 5 のように，「彦根城で写真を撮った」という行動の情報など，写真から得られる情報も活用することで，より効果的な可視化を目指す．

表 5　図 4 を入力画像とした時のデータ型毎の出力結果

Table 5　Textual output per data-type by inputting Fig. 4 to JSVLM.

ツイートテキスト （JSVLM には入力していない）	こんにちは 滋賀ダイハツハッピー彦根店です！ ハッピー彦根店にファンクロスの試乗車が届きました グレード→ファンクロス カラー→サンドベージュメタリック 室内も広々空間多機能な車内…(続きはブログで) #滋賀ダイハツ #滋賀 #ダイハツ #彦根 https://bit.ly/3Dh23cW
int8	ダイハツ キャスト アクティバ
fp16	駐車場に駐車したダイハツムーヴキャンバスカスタム
fp32	駐車場に駐車したダイハツムーヴキャンバスカスタム

図 4　自動車に関する投稿に対する画像

Fig. 4　An example of images in tweets on automobiles.

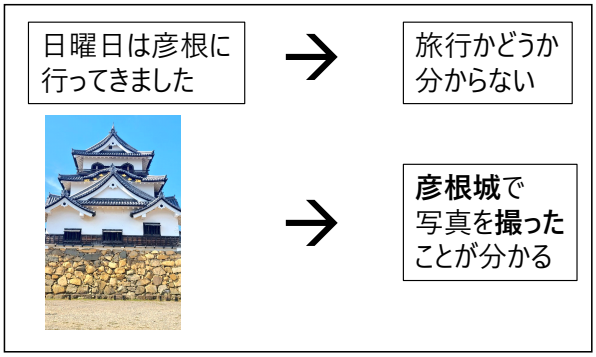

図 5　ある投稿の本文テキストに基づく行動推定とある投稿に添付された画像（写真）に基づく行動推定との比較

Fig. 5　Comparison between behavior inference based on a post's text and behavior inference based on a post's image.

文　　献

[1] 公益財団法人日本交通公社, “国内旅行における SNS・写真に対する意識／実態 ～JTBF 旅行実態調査トピックス～,” https://www.jtb.or.jp/research/statistics-tourist-sns-pictures2022/ (参照 2024-09-03).

[2] 国内宿泊旅行ニーズ調査, リクルート じゃらんリサーチセンター, https://jrc.jalan.net/surveys/corona_investigation/ (参照 2024-09-03)

[3] 藤大 友都, 服部 峻, 砂山 渡, “旅行活動ツイート可視化のためのフィルタリングと Artisoc エージェント化,” 第 15 回データ工学と情報マネジメントに関するフォーラム, 5b-3-2 (2023).

[4] 藤大 友都, 服部 峻, 砂山 渡, “機械学習を用いた旅行活動ツイート判定精度の可視化視認性への影響,” 第 16 回データ工学と情報マネジメントに関するフォーラム, T5-B-1-04 (2024).

[5] Shun Hattori, Yuto Fujidai, Wataru Sunayama, Madoka Takahara, “Effects of Machine Learning and Multi-Agent Simulation on Mining and Visualizing Tourism Tweets as Not Summarized but Instantiated Knowledge,” MDPI Electronics, Vol.13, No.16, 3276 (2024).

[6] 永澤 勇樹, 吉田 京平, 服部 峻, “モバイル端末における旅行記の理解支援のための行程抽出と地図化,” 電子情報通信学会 モバイルネットワークとアプリケーション研究会 (SIG-MoNA), 信学技報, Vol.114, No.31, MoNA2014-4, pp.19–24 (2014).

[7] Taro Tezuka, Takeshi Kurashima, and Katsumi Tanaka, “Toward tighter integration of web search with a geographic information system,” Proceedings of the 15th International Conference on World Wide Web (WWW'06), pp.277–286 (2006).

[8] Hiroshi Kori, Shun Hattori, Taro Tezuka, and Katsumi Tanaka, “Automatic Generation of Multimedia Tour Guide from Local Blogs,” Proceedings of the 13th International MultiMedia Modeling Conference (MMM'07), LNCS Vol.4351, Part I, pp.690–699 (2007).

[9] 渡邉 小百合, 吉野 孝,“観光地間の類似性を基にした向上点発見のための観光情報可視化システム,” マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, 2016, pp.1357–1362 (2016).

[10] Stability AI, “Japanese Stable VLM - stability.ai,” https://ja.stability.ai/blog/japanese-stable-vlm (参照 2024-09-03).

[11] 森野 穣, 安尾 萌, 松下 光範, 藤代 裕之, “Twitter に投稿された画像の分類に基づくツイート文の傾向分析 ―令和 2 年 7 月豪雨のツイートデータを対象に― ,” 第 13 回データ工学と情報マネジメントに関するフォーラム, I25-1 (2021).

[12] 西田 奈生, 金本 玲花, 松本 尚, “深層学習によるインスタグラム画像からの流行抽出,” 情報処理学会研究報告 数理モデル化と問題解決（MPS）, Vol.2020-MPS-127, No.15, pp.1–6 (2020).

[13] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16), pp. 779–788 (2016).

[14] Ultralytics, “Detect - Ultralytics YOLO Docs,” https://docs.ultralytics.com/tasks/detect/#models (参照 2024-09-07).

[15] ダイハツ,【公式】タント ファンクロス トップページ, https://www.daihatsu.co.jp/lineup/tanto_funcross/ (参照 2024-09-06).