

# Twitter を利用した旅行者の状況推定と観光ルート推薦

竹下 知宏<sup>†</sup> 上野 史<sup>††</sup> 太田 学<sup>††</sup>

<sup>†, ††</sup> 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中 3-1-1

E-mail: <sup>†</sup>pkx63txu@s.okayama-u.ac.jp, <sup>††</sup>{uwano, ohta}@okayama-u.ac.jp

あらまし 本稿では、Twitter における観光体験のツイートに着目し、ツイートから観光時の移動手段と観光目的を旅行者の状況として推定する手法を提案する。また、その推定結果を用いて、ユーザの移動手段と観光目的に合った観光ルートを推薦する手法を提案する。実験では、分類器を用いてツイートを移動ツイート、非移動ツイートに分類し、移動ツイートをさらに移動手段に分類した。また、観光スポット名を含むツイートを観光目的に分類した。その結果、移動手段の推定で BERT は他の分類器と比較して高い精度となり、どの移動手段においても正しく分類されたツイート数が誤って分類されたツイート数よりも多かった。観光目的の推定でも BERT の分類精度が最も高かったが、学習データ内で最も多い観光目的である「その他」へ誤って分類されるツイートが多かった。一方、Twitter ユーザのツイートの分類結果を利用して観光ルートを生成した結果、ユーザが過去の観光体験で優先していた観光目的に近い観光スポットを含む観光ルートが生成された。また、移動手段の優先順位に応じて「車」もしくは「鉄道」での移動を想定した観光ルートが生成された。

キーワード Twitter, 観光体験, 状況推定, 観光ルート推薦

方法について説明する。そして、4 節で観光特性の決定方法と推薦する観光ルートの生成手法について述べ、5 節でまとめる。

## 1 はじめに

近年、モバイル通信機器の普及とソーシャルネットワークサービスサービスの登場に伴い、インターネット上での知識や経験、価値観の共有が活発となっている。中でも Twitter は、ツイートと呼ばれる 140 文字以内の文章や画像、動画を不特定多数の人物が気軽に投稿できるサービスであり、「いつ」、「どこで」、「誰が」、「何を」しているというような個人の体験をリアルタイムに発信することができる特徴をもつ。そのため、旅行者のツイートからは、観光の際の移動手段や観光目的を知ることが可能である。これらをツイートから自動で推定することができれば、観光時によく使う移動手段や重視する観光目的といったユーザの特性を把握し、それに合った観光ルートを推薦できる。

本研究では、その人が「何を」しているかといった様子を状況と定義し、旅行者の状況として、観光時の移動手段と観光目的を推定する。具体的には、ユーザの観光体験を示すツイートを収集し、それらに人手で状況を示すラベルを付与して分類器を作成した後、その分類器で未知のツイートの状況を分類する。

本研究ではまた、分類結果を用いてユーザに合った観光ルートを推薦する。そのため、ツイートの分類によってユーザの旅行時の状況を推定した結果から、そのユーザにどのような観光目的、移動手段が多いかといった観光特性を決定する。次に、決定した観光特性に応じた観光スポットの評価値を利用して、観光ルートを生成する。実験では、岡山県内の観光スポットを含む観光ルートを生成し、被推薦者の観光特性やルート内の観光スポットの特徴に基づいて、そのルートの有用性について考察する。

本稿では、まず 2 節で関連研究について述べる。次に、3 節で旅行者のツイートを利用して移動手段や観光目的を推定する

## 2 関連研究

### 2.1 状況推定

NEC [1] は、人が持ち歩くスマートフォンの加速度センサのデータを分析することで、「静止中」、「会議中」、「歩いている」、「車で移動中」、「電車で移動中」の 5 つの現在状況を推定するソリューションを開発した。また、笠原ら [2] はバスや列車が減速するバス停や交差点、駅を環境制約として利用し、GPS の移動軌跡から徒歩や列車、路線バスといった移動モードを推定する手法を提案した。旅行者の GPS 移動軌跡を対象に移動モードを推定した結果、「タクシー」の再現率が 17.4%、「路線バス」の再現率が 68.0%、「列車」の再現率が 43.1%、「徒歩」の再現率が 93.6% となった。これらの研究では、加速度センサや GPS といったユーザの協力が必要な情報から、主に移動手段を含む移動中の状況を推定している。一方本研究では、一般に公開されているツイート文を収集し、ツイートの内容から移動手段を推定すると同時に、観光体験における食事や土産といった 5 種類の観光目的を推定する。

石野ら [3] は震災情報に関するツイートを利用して、Twitter ユーザの震災時の行動経路を抽出する手法を提案した。この研究では、機械学習によって東日本大震災のツイート文に「移動元」、「移動先」、「移動手段」のタグを自動で付与することで経路を抽出した。実験では、「移動」、「避難」など行動経路を表す際に使われやすい単語が含まれるツイートを対象とし、Conditional Random Field (CRF) によってタグを付与した。その結果、「移動元」の推定精度は 38.2%、「移動先」の推定精度は 48.1%、「移動手段」の推定精度は 81.8% となった。一方本研究

では、ツイートから旅行者の観光時における移動手段と観光目的を推定する。

## 2.2 観光体験の推薦

開地ら [4] は Web 上の観光情報をテキストデータとして取得し、単語の分散表現を用いてユーザの潜在的興味を発見して、それに基づいて観光地を推薦するシステムを試作した。彼らは Wikipedia の見出し語と、Yahoo!知恵袋の旅行に関するカテゴリで投稿された質問文を利用して観光情報を収集し、観光地データベースを作成した。単語の分散表現は、観光地データベースから作成したコーパスを Word2Vec で学習することで獲得した。推薦システムのインタフェースにユーザの好きな観光地と嫌いな観光地をそれぞれ 3 件まで入力させ、それらの単語ベクトルを足し引きしたものを潜在的興味とした。そして、それとデータベースにある他の観光地の分散表現との  $\cos$  類似度に基づいて観光地を推薦した。一方本研究では、観光地ではなく、被推薦者の過去のツイートの状況推定の結果から旅行者の興味を判定し、観光スポットを推薦する。

新井ら [5] は観光体験に関するツイートを収集し、それらのツイートから各観光スポットについて、時間帯スコア・カテゴリスコア・共起スコアの 3 種類のスコアを算出した。時間帯スコアはその観光スポットがどの時間帯によく訪れられているかを表すスコアである。カテゴリスコアは、観光スポットを含むツイートを手がかり語によって「食事」、「景観」、「行動」、「土産」の 4 種類のカテゴリに分類し、旅行者がその観光スポットを訪れる目的を割合で表したスコアである。共起スコアは他の観光スポットと共に訪れられやすい観光スポットであるかどうかを示すスコアである。新井らはこれらのスコアの和が最大となる観光ルートを推薦する手法を提案した。本研究は、中野ら [14] が新井らの提案を一部改変して実装した観光ルート推薦システムを利用する。

## 3 旅行者の状況推定

### 3.1 状況推定の概要

本研究では旅行者の移動手段に着目し、移動中の状況として「鉄道」、「車」、「バス」、「歩き」、「その他の移動手段」の 5 種類の移動手段を定義する。また、旅行者は観光スポットで、ご当地グルメを食べたり、有名な景色や建造物を眺めたりする。そこで、新井ら [5] が観光スポットの観光目的を表すカテゴリスコアに利用していた「食事」、「景観」、「行動」、「土産」の 4 種類と「その他」を合わせた計 5 種類を、ユーザの観光目的として定義し、それを旅行者の観光スポットでの状況とみなす。本研究では機械学習を利用して、ユーザの移動を表すツイートを個々の移動手段に分類し、観光スポット名が含まれるツイートを観光目的に分類する。

### 3.2 状況推定の対象ツイート

本稿では、観光スポット名を含むツイートを「観光ツイート」と呼ぶ。状況推定の対象とするツイートは、観光ツイートと、観光ツイートを投稿したユーザの同日の他のツイートとする。

本研究ではこれらを「観光タイムライン」と呼ぶ。なお、特定の時間に自動的にツイートを投稿する Bot アカウントや速報等を伝えるニュースアカウントのツイートを除くため、本研究では「Twitter for iPhone」等のスマートフォンやタブレット端末の Twitter 公式アプリから投稿されたツイートのみを分類の対象とする。Bot アカウントやニュースアカウントのツイートは、「Twittbot」や「Twitter Web App」等のサードパーティ製のアプリケーションや PC で主に投稿されているため、これらを除くことができる。さらにリプライやリツイート、外部サイトへの URL が付与されたツイートも、ユーザの観光体験とは無関係であることが多いため使用しない。

### 3.3 ツイートの分類

ツイートの状況を分類するための手法と分類器、前処理について説明する。

#### 3.3.1 分類の概要

まず収集した観光タイムラインの全ツイートを「移動ツイート」と「非移動ツイート」の 2 種類に分類する。「移動ツイート」に分類されたツイートは、さらに「鉄道」、「車」、「バス」、「歩き」、「その他の移動手段」の 5 種類に分類する。一方観光スポット名が含まれる観光ツイートに関しては「食事」、「景観」、「行動」、「土産」、「その他」の 5 種類に分類する。

#### 3.3.2 使用する分類器

ツイートの分類には Support Vector Machine (SVM) [6] と、Long Short-Term Memory (LSTM) [7], Bidirectional Encoder Representations from Transformers (BERT) [8] の 3 種類の分類器を用いる。

SVM は、マージン最大化の考えに基づき、汎用性の高い識別を行うことができる。また、SVM は、線形分離が不可能でない場合にも利用することができる。

LSTM は、再帰型ニューラルネットワークの一種であり、短期的な依存関係だけでなく、長期的な依存関係を学習することができる。

BERT は、機械学習のアーキテクチャの一種である Transformer が組み込まれた自然言語処理モデルである。Transformer が文章中の離れた単語間の関係を学習することで、優れた文脈の理解を実現している。

#### 3.3.3 ツイートの前処理

SVM, LSTM の分類器を利用する際、まず、ツイート内の画像の URL や、括弧等の記号といった語をストップワードとして除く。次に形態素解析エンジンである JUMAN++ [9] によってツイート文を意味を持つ最小の言語単位に分解する。その後、形態素解析したツイート文を Paragraph Vector の実装の 1 つである Doc2Vec [10] を用いて数値ベクトルに変換する。Doc2Vec は似た文章からは似たベクトルを生成する。

また、BERT の分類器を利用する際は、SentencePiece [11] と日本語の Wikipedia で事前学習したモデル [12] を用いて、あらかじめストップワードを除いたツイート文を単語に分割する。その後、同じモデルを用いて単語分割したツイート文を数値ベクトルに変換する。

表1 移動ツイートと非移動ツイート

	移動ツイート	非移動ツイート	計
学習データ	2,023	11,438	13,461
テストデータ	205	1,403	1,608
計	2,228	12,841	15,069

表2 移動ツイートの移動手段の内訳

	鉄道	車	バス	歩き	その他の移動手段	計
学習データ	1,039	189	196	330	269	2,023
テストデータ	61	55	22	38	29	205
計	1,100	244	218	368	298	2,228

表3 観光ツイートの観光目的の内訳

	食事	景観	行動	土産	その他	計
学習データ	353	866	706	135	1,764	3,824
テストデータ	71	125	73	16	211	496
計	424	991	779	151	1,975	4,320

### 3.4 ツイートの分類実験

3.3節で説明した分類方法でツイートを分類し、その精度を評価する。具体的には、実験データを学習データとテストデータに分割し、学習データで SVM と LSTM、BERT を学習し、テストデータの分類精度を求める。

#### 3.4.1 実験データ

実験ではまず、国内の観光スポットやイベント情報をまとめたサイトであるじゃらん net[13] から、日本全国の観光スポット名 6,226 件を取得した。そしてこの中の中国地方の 473 件のスポット名を Twitter API で検索し、得られた 2020 年 10 月 26 日～2020 年 11 月 4 日に投稿されたツイートのユーザ ID を 17,099 件取得した。この中の 282 件のユーザ ID と先述した 6,226 件の観光スポット名を利用して観光タイムラインを収集した。その結果得られた 282 ユーザの 3,032 件の観光タイムラインに含まれるツイート 15,069 件を実験データとし、これらを学習データとテストデータに分割した。学習データは 249 ユーザの 2,672 件の観光タイムラインにある 13,461 件のツイート、テストデータは 33 ユーザの 360 件の観光タイムラインにある 1,608 件のツイートとなった。これらを著者が手作業で移動ツイート、非移動ツイートに分類した内訳を表 1 に示す。また、表 1 の学習データの移動ツイート 2,023 件と、テストデータの移動ツイート 205 件を著者が手作業で移動手段に分類した結果は表 2 のようになった。一方、観光ツイートは表 1 の学習データに 3,824 件、テストデータに 496 件の計 4,320 件含まれており、これらを著者が手作業で観光目的に分類した結果、表 3 のようになった。

#### 3.4.2 移動ツイートと非移動ツイートの分類結果

表 1 のテストデータ 1,608 件を分類器を用いて、移動ツイートと非移動ツイートに 2 値分類したところ、表 4～表 7 のような結果となった。表 4 は各分類器による移動ツイートと非移動ツイートの 2 値分類の分類精度を示す。表 5、表 6、表 7 はそれぞれ SVM、LSTM、BERT で移動ツイートと非移動ツイートの分類結果を示す。表 4 から BERT の分類精度が最も

表4 移動ツイートと非移動ツイートの 2 値分類の分類精度

	SVM	LSTM	BERT
精度	0.883	0.863	0.938

表5 SVM による移動ツイートと非移動ツイートの分類結果

		SVM		計
		移動ツイート	非移動ツイート	
手作業	移動ツイート	83	122	205
	非移動ツイート	66	1,337	1,403
計		149	1,459	1,608

表6 LSTM による移動ツイートと非移動ツイートの分類結果

		LSTM		計
		移動ツイート	非移動ツイート	
手作業	移動ツイート	92	113	205
	非移動ツイート	108	1,295	1,403
計		200	1,408	1,608

表7 BERT による移動ツイートと非移動ツイートの分類結果

		BERT		計
		移動ツイート	非移動ツイート	
手作業	移動ツイート	161	44	205
	非移動ツイート	56	1,347	1,403
計		217	1,391	1,608

高かった。また、表 5～表 7 から、各分類器の移動ツイートの再現率、適合率に関しても、BERT がそれぞれ  $161/205=0.785$ 、 $161/217=0.742$  となり最も高くなった。しかし、「駅のホームで立ち食い蕎麦を食べてみた」という内容の非移動ツイートは、どの分類器でも移動ツイートに分類された。このツイートは「駅」という単語を含んでいることなどから、移動ツイートに分類されたと考えられる。

#### 3.4.3 移動手段の分類結果

表 2 の学習データ 2,023 件を用いて学習した分類器で、移動ツイートに分類されたツイートを、移動手段に分類した結果、表 8～表 11 のようになった。表 8 は各分類器による移動手段の分類精度を示し、表 9、表 10、表 11 はそれぞれ SVM、LSTM、BERT で移動手段の分類結果を示す。表 9 は表 5 の SVM で移動ツイートに分類された 149 件のツイート、表 10 は表 6 の LSTM で移動ツイートに分類された 200 件のツイート、表 11 は表 7 の BERT で移動ツイートに分類された 217 件のツイートを分類した結果である。表 9～表 11 から、「鉄道」のツイートはどの分類器でも多くが正しく分類された。また「車」のツイートは、SVM、LSTM では他の移動手段に分類されることが多かったのに対し、BERT は 1 件を除き正しく分類した。これは BERT が他の分類器とは異なり、文中に「高速降りた」等の「車」での移動を示す表現をもつツイートを正しく「車」と分類できたためである。また、表 8 で最も精度の高い BERT では、表 11 から、「鉄道」の再現率は  $41/47=87.2\%$ 、「車」の再現率は  $40/41=97.6\%$ 、「バス」の再現率は  $19/22=86.4\%$ 、「歩き」再現率は  $32/34=94.1\%$  となった。なお、笠原らの GPS の移動軌跡を用いた移動手段推定[2]では、「列車」の再現率は 43.1%、「タ

表 8 移動手段の分類精度

	SVM	LSTM	BERT
精度	0.389	0.270	0.668

表 9 SVM による移動手段の分類結果

		SVM					計
		鉄道	車	バス	歩き	その他	
手作業	鉄道	38	0	0	0	2	40
	車	10	5	0	1	0	16
	バス	3	2	8	1	0	14
	歩き	1	1	2	2	0	6
	その他	2	0	0	0	5	7
	非移動ツイート	40	9	6	4	7	66
計		94	17	16	8	14	149

表 10 LSTM による移動手段の分類結果

		LSTM					計
		鉄道	車	バス	歩き	その他	
手作業	鉄道	31	1	5	0	2	39
	車	2	6	1	6	6	21
	バス	1	1	6	1	1	10
	歩き	1	1	1	8	1	12
	その他	6	1	0	0	3	10
	非移動ツイート	43	14	9	24	18	108
計		84	24	22	39	31	200

表 11 BERT による移動手段の分類結果

		BERT					計
		鉄道	車	バス	歩き	その他	
手作業	鉄道	41	0	0	4	2	47
	車	0	40	0	0	1	41
	バス	0	2	19	1	0	22
	歩き	0	2	0	32	0	34
	その他	1	2	0	1	13	17
	非移動ツイート	21	10	6	7	12	56
計		63	56	25	45	28	217

クシー」の再現率は 17.4%,「路線バス」の再現率は 68.0%,「徒歩」の再現率は 93.6%であった。

#### 3.4.4 観光目的の分類結果

表 3 に示した観光ツイートのテストデータ 496 件を観光目的に分類した結果は、表 12～表 15 のようになった。表 12 は各分類器による観光目的の分類精度を示し、表 13、表 14、表 15 はそれぞれ SVM、LSTM、BERT での観光目的の分類結果を示す。表 13～表 15 から、SVM や LSTM ではほとんどが正しく分類できなかった「行動」や「土産」のツイートが、BERT では比較的正しく分類されていることがわかる。しかし、どの分類器においても、誤って「その他」に分類されるツイートが多くみられた。この要因としては表 3 の学習データに「その他」のツイートが多いことが挙げられる。これは、観光ツイートの中には、観光スポット名は含むが、内容がその観光スポットを訪れる予定や、訪れたいという願望のみで状況としての観光目的を含まないツイートが多かったためである。本実験では、単純に観光スポット名を含むツイートを観光ツイートとして集め

表 12 観光目的の分類精度

	SVM	LSTM	BERT
精度	0.534	0.490	0.681

表 13 SVM による観光目的の分類結果

		SVM					計
		食事	景観	行動	土産	その他	
手作業	食事	28	5	3	3	32	71
	景観	4	57	2	0	62	125
	行動	4	9	6	0	54	73
	土産	3	1	0	0	12	16
	その他	6	23	6	2	174	211
計		45	95	17	5	334	496

表 14 LSTM による観光目的の分類結果

		LSTM					計
		食事	景観	行動	土産	その他	
手作業	食事	28	8	15	3	17	71
	景観	8	66	10	2	39	125
	行動	8	14	12	1	38	73
	土産	5	3	3	1	4	16
	その他	11	35	26	3	136	211
計		60	126	66	10	234	496

表 15 BERT による観光目的の分類結果

		BERT					計
		食事	景観	行動	土産	その他	
手作業	食事	49	5	3	2	12	71
	景観	2	88	9	3	23	125
	行動	1	7	35	0	30	73
	土産	2	2	1	7	4	16
	その他	2	15	27	8	159	211
計		56	117	75	20	228	496

分類したが、観光タイムラインにある予定や願望を表す表現を含むツイートを除けば、「その他」のツイートの多くを減らすことは可能である。

## 4 観光ルート推薦

本研究では、3 節で説明した旅行者の観光タイムラインのツイートを分類した結果を利用して、移動手段や観光目的の優先順位といったユーザの観光特性を決定し、その特性を利用して観光ルートを推薦する。なお提案する観光ルート生成手法は、[14] で中野らが使用した、観光ルートを擬似全探索する手法を改変したものである。また本研究では、[5] で新井らが使用した観光スポットの評価値を改変したものを提案し、観光ルート生成に利用する。

### 4.1 中野らの擬似全探索による観光ルートの生成

中野らは、観光ツイートから観光スポットの評価値を算出した結果を利用し、観光ルート推薦システムによって観光ルートを生成した [14]。このシステムでは、被推薦者はまず出発地点、到着地点、出発時刻、到着時刻を入力する。入力後、シス

テムのインタフェース上に表示された観光スポットから訪見したい観光スポットを1つ選択する。システムは、ここで選択された観光スポットを含んだ観光ルートを作成する。中野らは観光ルート生成手法の1つとして、決められた時間内に探索を打ち切る擬似全探索を利用した。これは、全探索では観光ルートの生成にかかる時間が膨大になるためである。この観光ルート生成手法では、まず「出発地点→選択スポット→到着地点」の初期ルートを作成し、そのルートに、入力された時間内に訪問可能な観光スポットを追加していく。この際、移動時間と観光スポットでの滞在時間の合計が、出発から到着までにかかる時間を超えないようにし、指定した到着時刻までに到着地点に着く観光ルートを作成する。なお、移動手段は「車」のみであり、移動時間は、スポット間の直線距離1kmを2分40秒で移動するとみなして算出する。この擬似全探索を10秒で打ち切り、生成された観光ルートの中で、各観光スポットの評価値の和が最も大きい観光ルートを推薦する。

## 4.2 新井らの観光スポットの評価値

新井らは、Twitter API によって観光スポットの観光ツイートを集集し、それらを用いて各観光スポットの時間帯スコア、カテゴリスコア、共起スコアの3つのスコアを算出した[5]。

時間帯スコアは1日を8等分に区切り、ある観光スポットの全観光ツイート数に対する時間帯毎のツイート数の割合を求める。例えば、ある観光スポットの観光ツイート50件のうち、12時から15時までの間に投稿されたツイートが10件のとき、その観光スポットの12時から15時の時間帯スコアは、 $10/50 = 0.2$ となる。

カテゴリスコアは、ある観光スポットの観光ツイートを、手がかり語を用いて「食事」、「景観」、「行動」、「土産」の4カテゴリに分類し、各カテゴリに分類された観光ツイートの総数に対する割合を求める。例えば、ある観光スポットの4カテゴリのいずれかに分類された観光ツイート50件の中に、「食事」カテゴリに分類されたツイートが25件あったとき、その観光スポットの「食事」のカテゴリスコアは、 $25/50 = 0.5$ となる。また、手がかり語には Wikipedia と Weblio 類語辞典の単語や表現を用いた。なお、1つのツイートが複数のカテゴリに分類されることもある。

共起スコアはある観光スポットの観光ツイートを投稿したユーザが、同じ日に別の観光スポットの観光ツイートを投稿する確率から算出する。例えば、観光スポットAを訪れたユーザ50人のうち5人が同じ観光タイムラインで観光スポットBも訪れた場合、観光スポットAに対する観光スポットBの共起スコアは $5/50 = 0.1$ となる。

新井らは、式(1)によって観光スポットの評価値 $v_1$ を算出した。

$$v_1 = s^T + s^C + s^{CO} \quad (1)$$

ここで、 $s^T$ 、 $s^C$ 、 $s^{CO}$ はそれぞれ観光スポットの時間帯スコア、カテゴリスコア、共起スコアを表す。またカテゴリスコア $s^C$ は、被推薦者の選択スポットの4種類のカテゴリスコアのう

ち、スコアが最大のカテゴリスコアであり、共起スコア $s^{CO}$ は選択スポットとの共起確率である。

中野ら[14]は新井らが定めた観光スポットの評価値を使用した。カテゴリスコアは使用しなかった。しかし、本稿で提案する観光ルート推薦では、被推薦者の観光目的を考慮するのにカテゴリスコアが必要となるため、新井らの定義した式(1)の評価値を改変して観光ルートの生成に利用する。

## 4.3 ユーザの観光特性を利用した観光ルート推薦

3節で述べたユーザの状況推定結果を利用して、ユーザの移動手段と観光目的の優先順位を考慮した観光ルートの推薦手法を提案する。具体的にはまず、ユーザが観光目的として何を重視しているか、よく使う移動手段は何かというユーザの観光特性を、3節で述べたツイートの分類結果に基づいて決定する。そして、その観光特性を用いて推薦する観光ルートの移動手段や、観光スポットの評価値を変えて推薦ルートを生成する。

### 4.3.1 ユーザの観光特性

ユーザの観光タイムライン内のツイートの分類結果から、ユーザの観光目的と移動手段の優先順位を表すユーザの観光特性を決定する。観光特性は、「特性：観光目的」と「特性：移動手段」の2種類とする。

「特性：観光目的」は、ユーザの観光ツイート内で「その他」以外の各観光目的に分類されたツイートの数に基づいて決める。例えば「景観」に分類されたツイートが最も多かった場合、「景観」を「特性：観光目的」の第1観光目的とし、「食事」に分類されたツイートが2番目に多かった場合、「食事」を第2観光目的とする。第3観光目的、第4観光目的も同様に決める。

「特性：移動手段」は、ユーザのツイート内で、「その他の移動手段」以外の各移動手段に分類されたツイート数の多寡によって「特性：観光目的」と同様に決める。

### 4.3.2 移動手段の優先順位を用いた観光ルートの生成

本稿では、「車」での移動を想定した観光ルートの他に、「鉄道」での移動を想定した観光ルートを推薦する。具体的には、ユーザの「特性：移動手段」において、「車」の移動手段の優先順位が「鉄道」よりも高い場合には、「車」での移動を想定した観光ルートを推薦する。逆に「鉄道」の優先順位のほうが高い場合は、「鉄道」での移動を想定した観光ルートを推薦する。また、観光ルート推薦システムのインタフェースに「移動手段を変更」ボタンを加え、移動手段の異なる観光ルートも生成できるようにする。

本研究の「鉄道」での移動を想定した観光ルートでは、観光スポットの最寄り駅を利用して次の観光スポットへ移動する。また、観光スポットと最寄り駅間を歩いた際の移動時間と、駅での待ち時間、「鉄道」による駅間の移動時間の総和を計算し、指定した到着時刻までにルート内のすべての観光スポットを訪れることができる観光ルートを生成する。駅の位置情報は駅データ.jp[15]に記載されている岡山県内の162の駅のそれを使用し、駅の電車等の発車時刻や停車時刻はYahoo!路線情報[16]に記載されている時刻表から取得する。なお、スポット間で最寄り駅が同じ場合、そのスポット間の移動は「歩き」が指定さ

表 16 被推薦者の観光目的, 移動手段の優先順位

	観光目的	移動手段
ユーザ 1	行動 > 食事 = 景観 > 土産	車 > 鉄道
ユーザ 2	食事 > 行動 > 景観 > 土産	鉄道 > 車

表 17 観光ルートの推薦条件

出発地点	出発時刻	到着地点	到着時刻	選択スポット
岡山駅	11:00	岡山駅	20:00	美観地区 (倉敷)

れ, その移動時間は, スポット間の直線距離を 1km あたり 15 分で移動することを想定し算出する. また, スポットと最寄り駅間の「歩き」による移動時間も同様にして算出する.

#### 4.3.3 観光目的の優先順位を用いた観光スポットの評価値

式 (1) の  $s^C$  は, 被推薦者の選択スポットでスコアが最大のカテゴリのカテゴリスコアであるが, 本研究では全カテゴリのカテゴリスコアをユーザの観光目的の優先順位に応じて利用することを提案する. 具体的には, ユーザの「その他」を除いた観光目的の種類と, 新井らの観光スポットのカテゴリスコアのカテゴリが同一であることを利用し, 式 (2) で「特性: 観光目的」を用いて観光スポットのカテゴリスコア  $s^{C2}$  を算出する. これによって, 観光ルート中の観光スポットのカテゴリが偏らないようにしたうえで, ユーザの観光目的に沿った観光ルートを推薦する.

$$s^{C2} = \sum_{j=1}^4 (4 - j + 1) \cdot p_j \quad (2)$$

式 (2) で“4”はカテゴリの種類数,  $j$  は第 1 から第 4 までのユーザの観光目的の優先順位,  $p_j$  はユーザの第  $j$  観光目的と同名のカテゴリのカテゴリスコアを表す. 本研究では式 (1) の  $s^C$  を  $s^{C2}$  に変更したものを, 観光スポットの評価値とする.

### 4.4 観光ルートの生成実験

#### 4.4.1 推薦条件

ユーザの観光特性を利用して岡山県内の観光ルートを推薦する. 本実験では, 中野ら [14] と同様に, 2019 年 11 月 11 日～2019 年 11 月 15 日に収集した岡山県にある 51 の観光スポットの観光ツイート 12,172 件から, 式 (1) の観光スポットの評価値  $v_1$  および式 (2) のカテゴリスコア  $s^{C2}$  を算出した. 実験では, 表 11, 表 15 の BERT によるツイートの分類結果の中のツイートを投稿した 2 ユーザを被推薦者とする. この 2 ユーザのツイートの分類結果から, それぞれの「特性: 観光目的」と「特性: 移動手段」の優先順位は表 16 のようになった. また, 観光ルート推薦の条件である出発地点, 到着地点, 出発時刻, 到着時刻, 選択スポットは表 17 の通りである.

#### 4.4.2 推薦ルートの生成結果

ユーザ 1, ユーザ 2 に推薦する観光ルートの生成結果は, それぞれ図 1, 図 2 のようになった. ユーザ 1 には「車」での移動を想定した観光ルートが推薦され, ユーザ 2 には「鉄道」での移動を想定した観光ルートが推薦される. また, 推薦システムのインタフェース上の「移動手段を変更」ボタンでユーザ 2 の「車」での移動を想定した観光ルートを生成した結果, 図 3

表 18 観光スポットのカテゴリスコア

	食事	景観	行動	土産
東山公園	0.014	0.938	0.944	0.000
石山公園	0.796	0.370	0.241	0.093
岡山城	0.301	0.206	0.714	0.081
倉敷珈琲館	0.000	0.000	0.000	0.000
アリオ倉敷	0.781	0.040	0.233	0.017
美観地区 (倉敷)	0.604	0.361	0.286	0.064
文の助茶屋	1.000	0.000	0.000	0.000

のようになった.

図 1 と図 3 からユーザ 1 とユーザ 2 の「車」での観光ルートでは, 図 1 のユーザ 1 の観光ルートには「東山公園」が, 図 3 のユーザ 2 の観光ルートには「石山公園」と「岡山城」が含まれており, これらが両者で異なる観光スポットである. 観光スポットのカテゴリスコアをまとめた表 18 をみると, 「東山公園」はユーザ 1 の第 1 観光目的である「行動」と第 2 観光目的である「景観」のカテゴリスコアが高い. 一方, 「石山公園」はユーザ 2 の第 1 観光目的である「食事」のカテゴリスコアが高く, 「岡山城」はユーザ 2 の第 2 観光目的である「行動」のカテゴリスコアが高い. このことから生成された観光ルートは, ユーザの優先する観光目的に合った観光スポットを含んでいることがわかる.

また, 図 2 のユーザ 2 の「鉄道」での観光ルートには「倉敷珈琲館」と「アリオ倉敷」が含まれている. 表 18 より, 「アリオ倉敷」はユーザ 2 の観光目的の優先順位に沿ったカテゴリスコアである一方, 「倉敷珈琲館」はどのカテゴリスコアも 0 であり, 観光スポットの評価値  $v_1$  も低いが, ユーザ 2 の観光ルートに入った. また「池田動物園」のように付近に他の観光スポットがなく, その結果移動距離が長くなる観光スポットは, 「車」での観光ルートでは図 1 と図 3 のいずれにも含まれていたが, 図 2 の「鉄道」での観光ルートには含まれなかった.

観光スポットの評価値  $v_1$  の算出に利用するカテゴリスコアを, 式 (1) の  $s^C$  とし, 表 17 の条件によって生成した「車」での観光ルートは図 4 のようになった. このとき観光スポットのカテゴリスコア  $s^C$  は, 選択スポットである「美観地区 (倉敷)」のカテゴリスコアの中で最も高い「食事」のカテゴリスコアである. この観光ルートを図 3 の第 1 観光目的が「食事」であるユーザ 2 の「車」での観光ルートと比較すると, 図 4 では「文の助茶屋」が含まれているが, 図 3 では「石山公園」が含まれており, 訪問順を除けばこれらのスポットのみが異なる. 表 18 から, 「文の助茶屋」は「食事」のカテゴリスコアのみ 1 で, 他のカテゴリスコアが 0 である. 一方, 「石山公園」は「食事」のカテゴリスコアが 0.796 で一番大きい, 「景観」, 「行動」のカテゴリスコアもそれぞれ 0.370, 0.241 と 0 ではない. このことから, 式 (2) の  $s^{C2}$  によって, 図 3 では 1 カテゴリのみのカテゴリスコアが高い観光スポット (「文の助茶屋」) ではなく, 被推薦者の観光目的の優先順位に近いカテゴリスコアをもつ観光スポット (「石山公園」) が含まれるようになったといえる.



		到着時刻	出発時刻
0	岡山駅		11:00
1	池田動物園	11:04	11:54
2	東山公園	12:05	13:05
3	美観地区（倉敷）	13:52	14:52
4	岡山県立博物館	15:33	16:11
5	岡山県立美術館	16:12	17:12
6	讃岐の男うどん	17:13	18:13
7	お城茶屋	18:16	19:16
8	岡山駅	19:20	

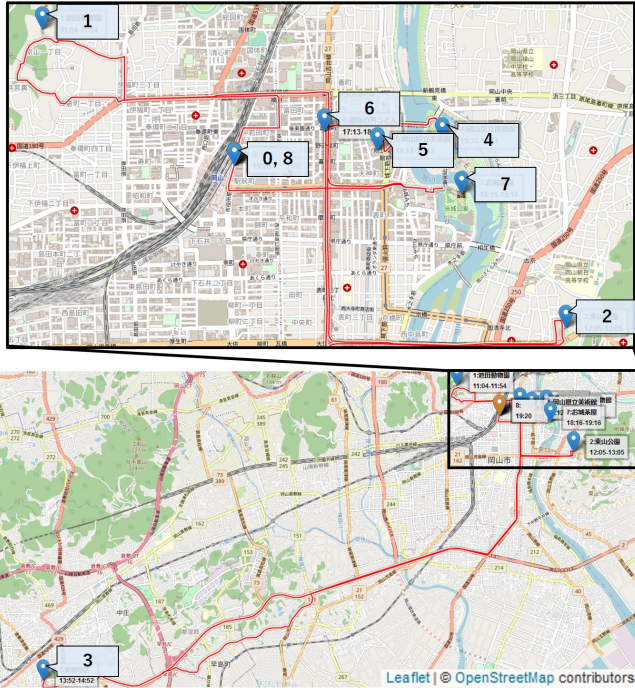


図1 ユーザ1に推薦された「車」での移動を想定した観光ルート

## 5 まとめ

本稿ではツイートの文章を分析することでユーザーの旅行時の状況を推定する手法を提案した。提案手法は、まずツイート文を移動ツイートと非移動ツイートに分類し、移動ツイートはさらに「鉄道」、「車」、「バス」、「歩き」、「その他の移動手段」の移動手段に分類する。また、観光スポット名を含むツイートを「食事」、「景観」、「行動」、「土産」、「その他」の観光目的に分類する。分類の精度は、著者が手作業で分類したテストデータを用いて評価した。また、分類器はSVMとLSTM、BERTの3つを比較した。その結果、移動ツイートと非移動ツイートの分類ではBERTの精度が0.938となり最も高かった。移動ツイートに分類された非移動ツイートが、全体の1/4以上を占めるテストデータの移動手段の分類では、精度が最も高いBERTで0.668となった。観光目的の分類でもBERTの精度が最も高く0.681であった。また学習データ内に「その他」の観光目的のツイートが多いために、どの分類器でも誤って「その他」に分類されるツイートが多かった。

本稿ではまた、Twitterユーザーの状況推定の結果を用いて、ユーザーの観光時の移動手段と観光目的の優先順位を決め、それ

		発着時刻	乗車路線
0	岡山駅	11:00 発	
移動	岡山-倉敷	11:05 発 11:16 着	伯備線
1	倉敷珈琲館	11:29 着 12:29 発	
移動	徒歩		
2	美観地区（倉敷）	12:31 着 13:31 発	
移動	徒歩		
3	アリオ倉敷	13:46 着 14:50 発	
移動	倉敷-岡山	15:02 発 15:19 着	赤穂線
移動	岡山-城下	15:29 発 15:33 着	東山線
4	石山公園	15:36 着 16:36 発	
移動	徒歩		
5	お城茶屋	16:41 着 17:41 発	
移動	徒歩		
6	岡山県立博物館	17:47 着 18:25 発	
移動	徒歩		
7	岡山県立美術館	18:31 着 19:31 発	
移動	城下-岡山	19:37 発 19:47 着	東山線
8	岡山駅	19:48 着	

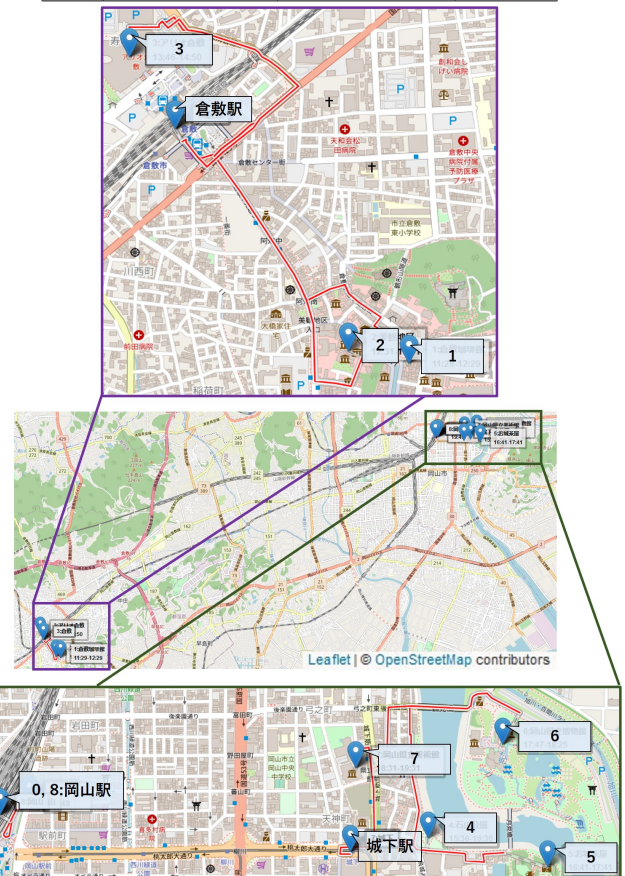


図2 ユーザ2に推薦された「鉄道」での移動を想定した観光ルート

らを利用して観光ルートを推薦する手法を提案した。実験の結果、ユーザーが優先する観光目的に近い観光スポットが含まれる観光ルートが生成されることを確認した。また、移動手段の優先順位に応じて「車」もしくは「鉄道」での移動を想定した観光ルートが推薦された。

今後の課題としては、ユーザーの移動手段や観光目的の推定精度の向上とともに、「バス」等他の移動手段を想定した観光ル



		到着時刻	出発時刻
0	岡山駅		11:00
1	池田動物園	11:04	11:54
2	美観地区（倉敷）	12:33	13:33
3	岡山県立博物館	14:14	14:52
4	岡山県立美術館	14:53	15:53
5	石山公園	15:54	16:54
6	お城茶屋	16:55	17:55
7	岡山城	17:55	18:55
8	讃岐の男うどん	18:57	19:57
9	岡山駅	19:59	

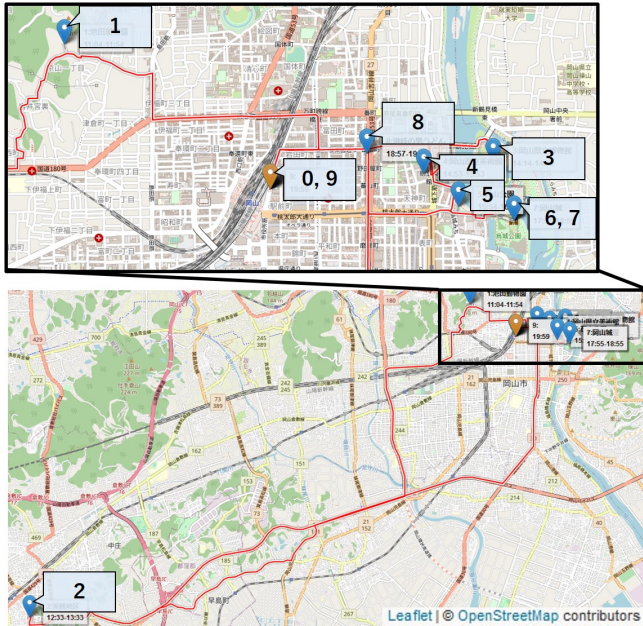


図3 ユーザ2の「車」での移動を想定した観光ルート

		到着時刻	出発時刻
0	岡山駅		11:00
1	池田動物園	11:04	11:54
2	美観地区（倉敷）	12:33	13:33
3	岡山城	14:12	15:12
4	お城茶屋	15:12	16:12
5	岡山県立博物館	16:13	16:51
6	文の助茶屋	16:53	17:53
7	岡山県立美術館	17:55	18:55
8	讃岐の男うどん	18:56	19:56
9	岡山駅	19:58	

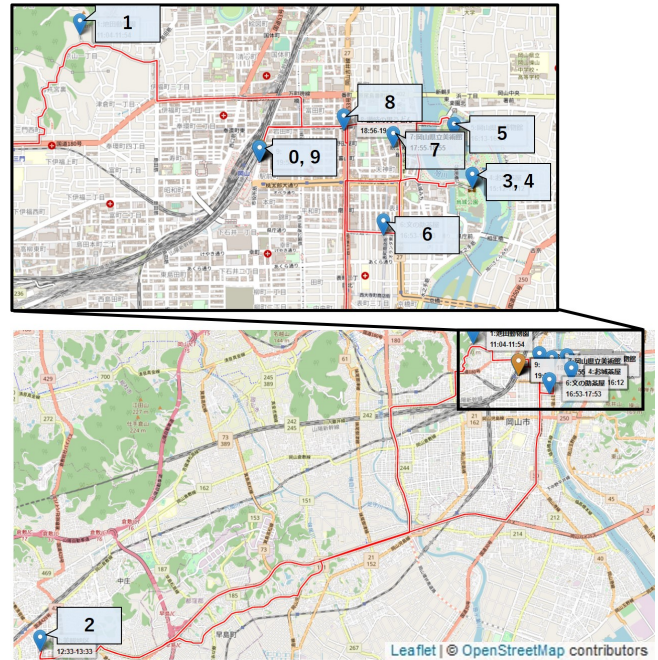


図4 1種類のカテゴリスコアから生成した「車」での移動を想定した観光ルート

トの推薦などが挙げられる。

## 文 献

- [1] “NEC プレスリリース”, <http://www.nec.co.jp/press/ja/1202/2002.html>
- [2] 笠原秀一, 飯山将晃, 美濃導彦, “環境制約を用いた旅行者 GPS 軌跡からのオンライン移動モード推定手法”, 電子情報通信学会技術研究報告, vol. 115, no. 245, pp. 43-48, 2015.
- [3] 石野亜耶, 小田原周平, 難波英嗣, 竹澤寿幸, “Twitter からの被災時の行動経路の自動抽出および可視化”, 言語処理学会 第 18 回年次大会, pp. 907-910, 2012.
- [4] 開地亮太, 檜垣泰彦, “潜在的興味に基づく観光地推薦システムの試作”, 電子情報通信学会技術研究報告, vol. 115, no. 138, pp. 29-34, 2015.
- [5] 新井晃平, 新妻弘崇, 太田学, “Twitter を利用した観光ルート推薦の一手法”, 第 7 回データ工学と情報マネジメントに関するフォーラム, G7-6, 2015.
- [6] Corinna Cortes and Vladimir Vapnik, “Support-Vector Networks”, Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.
- [7] Felix A. Gers, Jurgen Schmidhuber, and Fred Cummins, “Learning to forget: Continual prediction with lstm”, Neural Computation, vol. 12, Issue 10, pp. 2451-2471, 2000.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint arXiv:1810.04805, 2018.
- [9] “日本語形態素解析システム JUMAN++”, <http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN%2B%2B>
- [10] Quoc Le and Tomas Mikolov, “Distributed Representations of Sentences and Documents”, Proceedings of the 31st International Conference on Machine Learning, pp. 1188-1196, 2014.
- [11] Taku Kudo and John Richardson, “SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing”, arXiv preprint arXiv:1808.06226, 2018.
- [12] “BERT with SentencePiece for Japanese text”, <https://github.com/yoheikikuta/bert-japanese>
- [13] “じゃらん net 観光ガイド”, <https://www.jalan.net/kankou>
- [14] 中野翔子, 新妻弘崇, 太田学, “ユーザが入力した地点の近傍を優先する観光ルート推薦の一手法”, 第 12 回データ工学と情報マネジメントに関するフォーラム, J3-4, 2020.
- [15] “駅データ.jp”, <http://www.ekidata.jp>
- [16] “Yahoo!路線情報”, <https://transit.yahoo.co.jp>