

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

藤大 友都[†] 服部 峻^{††} 砂山 渡^{††}

[†] 滋賀県立大学 工学部 電子システム工学科 〒 552-8533 滋賀県彦根市八坂町 2500

^{††} 滋賀県立大学 先端工学研究院 〒 552-8533 滋賀県彦根市八坂町 2500

E-mail: [†]ton23yfujidai@ec.usp.ac.jp, ^{††}{hattori.s,sunayama.w}@e.usp.ac.jp

あらまし 近年, SNS の利用者増加に伴い, SNS に蓄積されるデータは多種多様, 日々膨大になり続けているため, 旅行の感想データが埋没してしまい, SNS は旅行先を決定するきっかけとして有効活用できていない. そこで, 本稿では新たな旅行需要の発見と旅行先の決定を支援するため, 旅行活動データを有効活用し, 旅行活動へと促進することを目的とする. まず, 膨大なデータの中から, ある場所の旅行活動データだけを精度良く, かつできる限り網羅的に収集するフィルタリング手法について検討する. 次に, 収集された個人個人の旅行活動データを用いて, 旅行活動の特徴を理解できるよう, マルチエージェントシミュレーションである Artisoc 上にエージェント化する.

キーワード ソーシャルメディア, SNS, 可視化, 情報抽出, 観光

1 はじめに

旅行したことを SNS に投稿するユーザは多い. SNS には JTB [1] など旅行サイト上では書かれなようなネガティブな感想も自由に, そして評価を正直に投稿できるという点, またガイドブックでは可能としていた体験が, 今は体験できないといった情報のリアルタイム性といった利点が存在する. しかし, SNS の情報を利用して旅行先を決定するユーザは少ない. JTB グループ [2] によると, 旅行先を決定するための情報源として, 「友人からのおすすめ・体験談」が 18.4% を占める一方, 「個人の SNS・ブログ」とするユーザは 1.6% と少ないという調査結果がある. 確かに, SNS には旅行活動のデータだけでなく, 日常生活や広告なども含まれており, 検索を工夫しなければ旅行活動データだけを取り出すことは難しい. また, SNS 検索が旅行先の決定に特化されていない. このままでは SNS 投稿を活用するニーズが近年高まっているにもかかわらず, SNS にしかない旅行の情報を取りこぼしてしまう. 例えば, 観光地の地元住民しか行かないような穴場スポットは, 旅行サイトやガイドブックに掲載されない. これらの情報を取りこぼさないように拾い集め, なおかつ有効活用するために可視化することにより, 旅行活動がもっと活発になり景気が上向くのではないかと考えた.

そこで, 本稿では SNS の 1 種である Twitter に着目し, 膨大なツイートの中からある場所 (例えば「彦根」) を中心とした旅行における活動データだけを精度良く, かつできる限り網羅的に収集するためのフィルタリング手法について検討する. また, 収集された旅行活動データを用いて, Twitter ユーザが「ある場所」を訪れたとき, どのスポットによく行っているのか, どのような活動を行っているかを可視化し, 「ある場所」における旅行活動の特徴を捉えられるようにする. ツイートの可視化と

して, 人工社会シミュレーションである Artisoc [3] にエージェント化する. Artisoc は, シミュレーション上に人工社会を構築し, 個々のエージェントの相互作用による社会現象を再現することで, ある原因に対する影響を評価するソフトウェアである. 最終的には, ある地域における各スポットで振る舞うユーザの活動から, 旅行の目的地決定を支援できるシステムを構築することを目的とする.

本稿の以降の構成は次の通りである. まず, 2 章で関連研究について, 3 章で提案システムについて詳述する. 次に, 4 章で評価実験を行い, 提案システムの有効性を検証する. 最後に, 5 章で本稿をまとめ, 今後の研究課題についても述べる.

2 関連研究

本章では, 関連研究について述べる.

2.1 観光情報ツイートの抽出に関する研究

観光情報ツイートの抽出手法としては, キーワード検索を行うことが一般的である. 渡邊ら [4] は観光地の感想を抽出する方法として, 観光地名の特徴語を含むツイート, 観光地名入りツイートの前後に呟かれたツイート, 観光地名入りツイートに対するリプライ, 観光地名を含まない画像付きツイートの 4 つに着目し, 観光地名なしツイートからの観光地に関する感想の抽出を試みた. しかし, 感想には行動が含まれておらず, 具体的にどのような行動を行ったかは明らかになっていない.

そこで本稿では, 最終的な目標を「旅行活動データの可視化」とする. 具体的には, 「彦根城は興味深かった」というような感想ツイートではなく, エージェント化するために必要な「カフェでコーヒーを飲んだ」というような行動を含むツイートを収集する. このような旅行活動ツイートを精度良く, 網羅的に収集するためのフィルタリング手法を検討する.

また、松本ら [5] は Web 上の膨大なデータから、観光推進や地域振興などで活用するために地域情報を収集し、有益な知見・知識を抽出することを目的に、下関地域で投稿された観光に関するツイートデータの抽出を試みた。抽出には、語と語のつながりを直観的に理解することのできる共起ネットワークを用いており、抽出した結果から、新たな観光資源やニーズがどの程度読み取れるか検討した。その結果、観光客が増加する可能性のある観光資源をツイートデータから発見することができた。しかし、大型連休やスポーツイベントが行われている時期以外は観光に関するツイートが少ないという問題があった。そこで、本稿では、汎用的な TFIDF を用いて、観光地（彦根など）の特徴語抽出手法を提案し、一般的に旅行が活発でない時期を対象に適用して、その精度を検討する。また、抽出結果を魅力的に可視化することで旅行先の決定を支援する点でも先行研究と異なる。

2.2 Twitter の可視化に関する研究

本庄ら [6] は、1 ユーザに着目し、ユーザのツイートに含まれるキーワードの特徴語を用いて類似性の高いユーザの検出を試みた。また、可視化の方法として、「対象ユーザを中央に配置し、囲むようにフォロワーを配置するモデル」、「つぶやきを時計の針の動きに合わせて表示し、時間的な関係性を可視化するモデル」、「対象ユーザのつぶやきに対して他のユーザがどのような反応をしたかを可視化するモデル」を提案し、その効果を検証した。可視化には、Artisoc を用いており、着目したユーザをエージェントとすることで、他のユーザエージェントとの関係性を見ることが可能である。しかし、本研究ではユーザ同士の関係ではなく、ユーザ 1 人 1 人の行動を可視化している点で異なる。

3 提案手法

提案システムの全体的なシステム構成図を図 1 に示す。本稿で提案するシステムは、「旅行活動ツイートのフィルタリング」と「Artisoc への可視化」の 2 つに大別される。

3.1 旅行活動ツイートの取得

旅行活動ツイートを収集する方法として、従来研究 [4] も参考に、「観光地の地名と旅行活動を含むツイート」、「観光地付近を位置情報に持つツイート」、「観光地の特徴語を含むツイート」で網羅的に旅行活動ツイートを収集した上で、「旅行活動単語」でフィルタリングする手法を提案する。本稿における旅行とは「観光地に移動し、観光地内の各スポットで何らかの活動を行うこと」を指す。また、「観光地内の各スポットにおいて何らかの活動を行うための単語」を旅行活動単語と定義する。旅行活動単語は、ツイートから旅行活動であると絞り込むことができる単語として、独自に選定したものである。旅行活動単語を表 1 に示す。

3.1.1 「観光地名」「位置情報」「特徴語群」による網羅的なツイート収集

一般的に観光地の投稿を検索する場合、まずは目的とする観

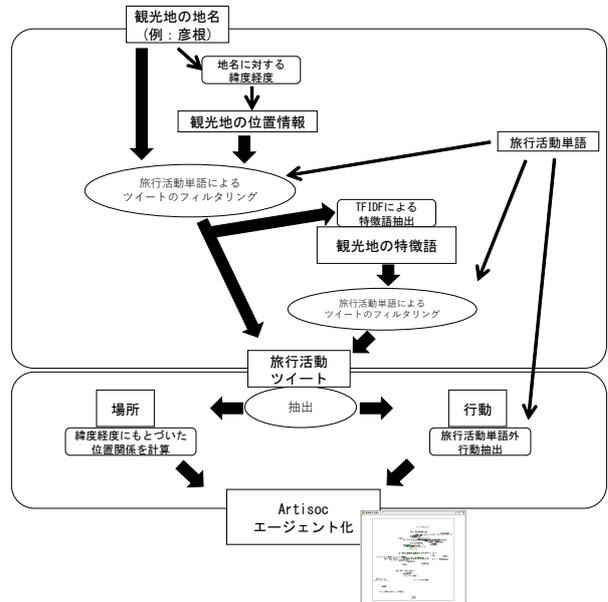


図 1 提案システムの流れ図

表 1 旅行活動単語

食べる	飲む	買う
移動する	乗る	寝る
撮る	帰る	着く
見る	観光する	楽しむ
登る	遊ぶ	会う

光地の地名で検索することが想定される。また、旅行活動を行ったユーザも、投稿するツイートには観光地の地名を含める可能性が高いと考えられる。

次に、観光地名による検索だけでは、観光地の地名を含まないが、観光地における旅行活動を行ったツイートを収集することができない。そこで、従来研究でも示されている通り、目的とする観光地付近の位置情報を含むツイートには、旅行活動が含まれている可能性が高いと考えられる [4]。

最後に、位置情報による検索は、その場所を訪れたユーザを抽出するのに有効であるが、近年はセキュリティ対策の観点から位置情報設定を切断するユーザが存在する。この場合、位置情報による検索ができない。そこで、観光地の地名や位置情報の代わりに、その観光地における特徴語を検索条件として収集することも考えられる。例えば観光地として「彦根」を検索する場合、特徴語は「彦根城」や「ひこにゃん」が考えられる。文中での単語の抽出のために、形態素解析エンジンである MeCab を活用する。また、MeCab のシステム辞書として、mecab-ipadic-NEologd [7] を用いる。これは、新語・固有表現に強く、語彙数も多いため、SNS のような新語が出現しやすい文章からの単語抽出に適している。特徴語の計算方法として、文書の特徴語を抽出できる TFIDF 法を用いる。観光地名、位置情報によって収集されたツイート群と、日本語で書かれた一般的なツイート群とを比較し、観光地の旅行活動ツイートを特徴づける単語を抽出するため、ある観光地の地名 p に対する単語 w の TFIDF 値を以下の式で計算する。

$$TFIDF_P(w) = \frac{TF_P(w)}{TF_P} \cdot \frac{TF_N}{TF_N(w)+1} \quad (1)$$

ただし、 TF_P をある観光地の地名 p に関するツイートの数、 $TF_P(w)$ をある観光地の地名 p に関しておりかつ単語 w を含むツイートの数、 TF_N を日本語で書かれたツイートの数、 $TF_N(w)$ を日本語で書かれたツイートのうち、単語 w を含むツイートの数とする。

3.1.2 旅行活動単語によるツイートのフィルタリング

3.1.1 項で収集されたツイートから、旅行活動ツイートだけを精度良く収集できるようにするため、フィルタリングを行う。本文中に表 1 に示す旅行活動単語が含まれているツイートを抽出する。次に旅行活動を投稿したユーザは、該当のツイートの前後にも何らかの旅行活動ツイートを投稿していることが考えられる。このため、抽出されたツイートを投稿したユーザの、前後 3 日間に投稿されたツイートを再度収集し、旅行活動ツイートを抽出する。従来研究 [4] では前後 3 時間としていたが、本稿ではより網羅的に収集するため、前後 3 日間とする。

3.2 Artisoc への可視化

本提案システムでは、前節で収集された個人個人の多種多様な旅行活動ツイートをを用いて、観光地の旅行活動の特徴が理解できるような可視化の手法について提案する。

Artisoc は、複数のエージェントを用いた仮想実験を行うことができるソフトウェアである。ここでいうエージェントとは、自分と自分の周囲の状況を把握し、一定のルールで自律的に行動する主体のことで、現実世界では人間や生物などに当てはめることができる。このソフトウェアを用いることで、「鳥の群れ」や「高速道路の自然渋滞」といった現象を可視化できる [3]。

収集したツイートを Artisoc に可視化するためには、ツイートデータを Artisoc が持つデータ型に整形する必要がある。このため、ツイートから場所と行動に関する情報を抽出する。場所については、予め題材とする場所（彦根市内など）の観光スポットをリスト化しておき、ツイートを観光スポット名が含まれていれば抽出する。場所が抽出できなかった場合、可視化は行われぬ。行動については、表 1 に示す旅行活動単語が含まれていれば抽出する。ただし、「行かない」「飲まず」といった単語の後ろに否定を表す助動詞がある場合など、抽出した単語が否定表現であれば抽出しない。また、旅行活動単語でない単語の場合でも、観光スポットに対する行動を表す単語である可能性があるため、ツイート本文に行動を示す単語（動詞）があれば追加で抽出する、行動拡張抽出を行う。この場合も否定表現は抽出しない。行動拡張抽出を行った結果、行動を抽出できなかった場合は、可視化時に行動は表示されず、観光地内のスポットに移動する部分のみ可視化される。

本稿では、エージェントとして Spot エージェントと User エージェントを作成する。図 2 に可視化のイメージ図を示す。まず、それぞれのエージェントが持つ情報、働きを示す。その上で、シミュレーション全体の流れについても述べる。

3.2.1 Spot エージェント

Spot エージェントは、観光地名を彦根とすると、彦根城や四

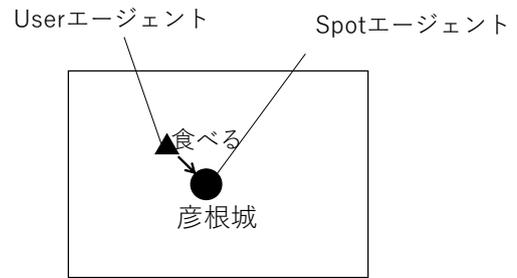


図 2 Artisoc 可視化のイメージ図

番町スクエアといった各スポットの名称と、対応する座標情報を持っている。ツイート本文に観光地内のスポット名が含まれている場合、User エージェントに座標情報を提供する。また、Spot エージェント自身もシミュレーション開始時に座標情報をもとに表示する。

3.2.2 User エージェント

User エージェントは、旅行活動を行ったユーザの数だけ存在し、ツイート本文、ツイート本文に含まれる観光地内スポット、ツイート本文に含まれる行動を示す単語、ツイートを投稿した日時の情報を持っている。あらかじめツイートから場所（観光地内のスポット）と、行動（観光地内スポットでの行動）を抽出する。シミュレーション上で時間が進み、ツイート投稿日になったら行動を表示し、Spot エージェントが持つ場所の座標に向かって移動する。行動が抽出されていない場合、行動は空白である。移動完了後、一定時間が経つと User エージェントは非表示となる。

3.2.3 可視化の流れ

シミュレーション全体の流れは以下の通りである。

- Step 1. シミュレーション開始時に、Spot エージェントは各スポットの名称と対応する座標情報、User エージェントはツイート本文とツイート投稿日時を読み込む。その後、Spot エージェントは各スポットの座標に名称と共に出現し、User エージェントは自身の持つ観光地内スポット名を使って Spot エージェントから座標を受け取る。
- Step 2. 時間を進め、User エージェントが持つ日時と一致すると、User エージェントは観光地内スポットの近くに行動を表示しながら出現し、観光地内スポットに向かって少しずつ移動する。
- Step 3. User エージェントが観光地内スポットに到達し一定時間が立つと User エージェントは消滅する。
- Step 4. すべてのツイートの可視化が完了すると、シミュレーションを終了する。

4 評価実験

本章では、前章で示した提案システムの有効性を検証するため、「彦根」という観光地の旅行活動ツイートを収集し、収集された旅行活動ツイートをを用いて Artisoc にエージェント化した。

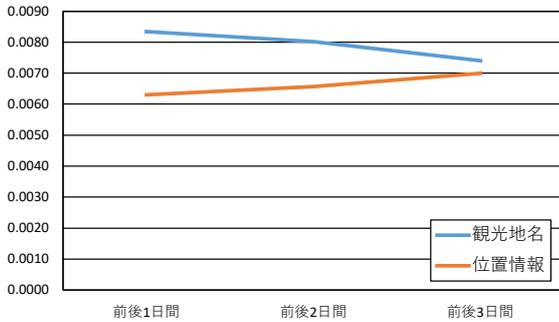


図3 前後ツイート収集の範囲と適合率の関係

表2 観光地名による旅行活動ツイート収集結果

	検索数	正解数	適合率
観光地名	1701	272	0.1600
観光地名+旅行活動単語	47	14	0.2979
観光地名+前後3日間	5708	42	0.0074

抽出する期間は新型コロナウイルスによる旅行控えの影響を考慮し、2019年2月1日から2019年2月10日とした。一般に2月ほどの観光地も観光客が少ない閑散期であるが、閑散期だからこそ観光客に知られていない観光スポットがあるのではないかと考えたためである。さらに、得られたシミュレーション結果が有効的、魅力的に可視化できているかを確かめるため、アンケート調査を行った。

4.1 旅行活動ツイートの取得

表2に観光地名として「彦根」を検索した時のツイート数と実際に著者の目で判断した旅行活動ツイートの正解数、適合率を示す。ただし、観光地名のみで収集された1701ツイートのうち、100ツイートをサンプリングして正解数、適合率を計算している。

次に、表3に位置情報として、彦根市の範囲である北緯35.253802°、東経136.251852°から半径5kmで囲われたツイートを収集した時のツイート数と実際に第1著者の目で判断した旅行活動ツイートの正解数、適合率を示す。ただし、位置情報のみで収集された1341ツイートのうち、100ツイートをサンプリングして正解数、適合率を計算している。

観光地名、位置情報と前後1日間から3日間に範囲を変化させた時の適合率の関係を図3に示す。観光地名検索では範囲を広げるにつれて適合率が減少したものの、位置情報検索では適合率が増加していることが確認できた。最後に、TFIDF法を用いた特徴語の抽出について述べる。観光地名検索によって得られた58ツイート(異なり数)、位置情報検索によって得られた12ツイート(異なり数)の合計70ツイート(延べ数)と、2019年2月1日11時59分00秒からの1分間において日本語で書かれた一般的なツイート(12880ツイート)とを比較し、彦根を旅行活動するツイートを特徴づける単語を抽出した。彦根を特徴づける単語の上位10単語と、TFIDF値を表4に示す。

表4より、特徴語として「ベンチ」「滋賀県」「Vidal」「IC」を特徴語とする。ただし、「彦根」、「彦根市」、「彦根駅」に関し

表3 位置情報による旅行活動ツイート収集結果

	検索数	正解数	適合率
位置情報	1341	0	0.0000
位置情報+旅行活動単語	27	1	0.0370
位置情報+前後3日間	1705	12	0.0070

表4 観光地「彦根」の特徴語の抽出結果

特徴語	TFIDF 値
彦根	7912.00
彦根市	3312.00
ベンチ	2024.00
彦根駅	1840.99
滋賀県	1748.00
Vidal	1104.00
ヒコアマゾン	735.99
ゴーストロンガー	735.99
滋賀	735.99
IC	735.99

表5 観光地「彦根」の特徴語によるツイート収集結果

	TFIDF 値と順位	検索数	正解数	適合率
ベンチ	2024.00	21002	0	0.000
ベンチ+旅行活動単語	3位	680	0	0.000
滋賀県	1748.00	330	7	0.021
滋賀県+旅行活動単語	5位	4	1	0.250
Vidal	1104.00	140	0	0.000
Vidal+旅行活動単語	6位	2	0	0.000
IC	735.99	22712	0	0.000
IC+旅行活動単語	10位	558	1	0.018
彦根城		416	112	0.269
彦根城+旅行活動単語	圏外	11	4	0.364

ては観光地名「彦根」としてすでに検索しているため、「ヒコアマゾン」、「ゴーストロンガー」についてはユーザ名のため、「滋賀」は「滋賀県」ですでに検索しているため、特徴語としない。また、比較として彦根から連想される「彦根城」についても検索を行う。表5に「ベンチ」「滋賀県」「Vidal」「IC」「彦根城」をそれぞれ検索した場合の検索数、正解数、適合率について示す。ただし、それぞれの単語の、旅行活動単語で絞り込む前の正解数、適合率については、検索数から100件サンプリングしてから得られた結果である。また、「IC」「Vidal」については日本語のツイートのみを収集している。フィルタリングの結果、どの結果も適合率が低かった。これは、旅行活動単語の選定が十分ではなく、旅行活動ではないツイートを多く収集してしまったことが原因と考えられる。実際、いくつかの旅行活動単語は、旅行活動でないツイートのみを収集していた。また、「登る」は一般的に山に対する単語であるため、平地の観光地には出てくるのが少ないといった地域性も考慮する必要があると考えられる。今後の改良により、観光地に合わせた旅行活動単語の選定が必要であると考えている。

また、位置情報による旅行活動ツイート収集については、旅行活動単語で絞り込む前の時点で、サンプリングによる正解ツイートが収集されなかった。このため、位置情報による収集は

表 6 旅行活動ツイート群から観光スポットと行動を抽出した結果

	ツイート数
旅行活動ツイート	94
観光スポット抽出ツイート	94
旅行活動単語による行動抽出ツイート	29
行動拡張抽出による行動抽出ツイート	64

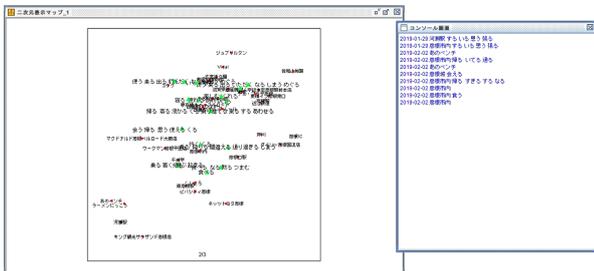


図 4 2019年2月3日の Artisoc 画面

効果的ではない可能性が考えられる。今後、位置情報に変わる新たな収集手法や、検索する位置情報の調整が必要であると考えている。

一方で、特徴語抽出については一般的な単語が上位に含まれたことで、適合率が低くなってしまった。考えられる原因としては、旅行活動ツイートが少なく、特定の単語が多ければ反応しやすい状態であること、一般的な名詞が特徴語になっている点が挙げられる。特に一般的な名詞について、例えば「ベンチ」は TFIDF 上位である。これは、彦根市の観光スポットである「あのベンチ」に関連する単語であると考えられる。しかし、ベンチが一般的な単語であるため、旅行活動と関係のないツイートを多く収集してしまった。また、「ベンチ」を含むツイートが旅行活動であっても、彦根以外の地域のベンチとも共起してしまっただけで、彦根以外での旅行活動ツイートも検索された。今後は、「あのベンチ」など1つの複合名詞として特徴語抽出ができるようにすること、また「特定地域に依存する旅行活動ツイートに多く出現し、地域に依存しない一般的なツイートにはあまり出現しない」かつ「特定地域以外の任意の地域に依存するツイートにも出現しない」特徴語を求める必要がある。

4.2 Artisoc への可視化

4.1 節で旅行活動ツイートとして得られた合計 94 ツイート（異なり数）の正解データを可視化する。表 6 に観光スポットの抽出ができたツイート数、旅行活動単語による行動抽出ができたツイート数、旅行活動単語以外に追加で行動抽出ができたツイート数を示す。

以上の結果を用いて、Artisoc にエージェント化した。図 4 では、シミュレーションが 2 月 3 日になった直後の Artisoc の画面を示している。ツイート投稿日が 2 月 3 日のツイートがエージェントとして発生し、各エージェントが持つ観光スポットの座標に向かって進んでいる。図 5 では、図 4 から少し経過した Artisoc の画面を示している。いくつかのエージェントが観光スポットに到達し、コンソール画面に日付、観光スポット、行動が表示されている。

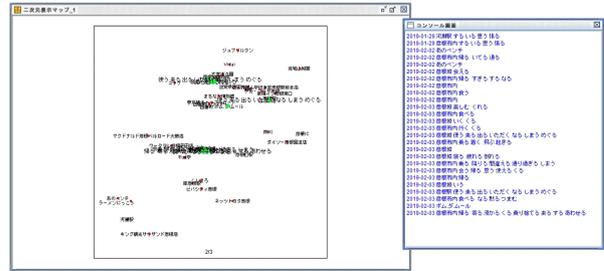


図 5 2019年2月3日から少し経過した Artisoc 画面

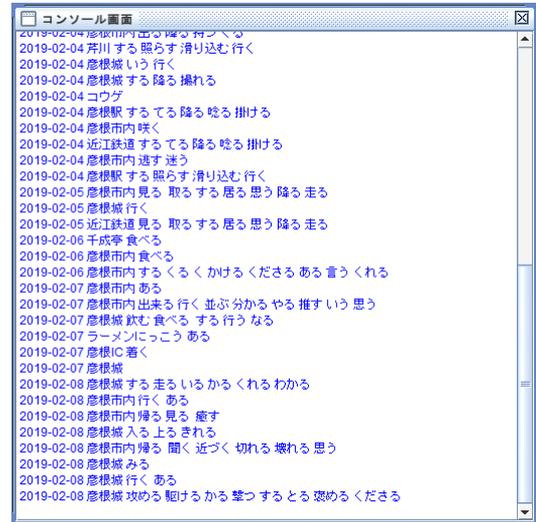


図 6 2019年2月4日から2月8日の Artisoc コンソール画面

可視化を行った結果、日付ごとのユーザーの特性が確認できるようになったことで、旅行先を決定し、旅行活動の促進に繋がる可能性が高いと考えられる。しかし、観光地内スポットが増えることで領域内に大量に Spot エージェントが出現し、視認性が下がる。Spot エージェントは現実の緯度経度に合わせて配置をしているため、彦根城と彦根城博物館など、現実の位置関係に近いエージェントはシミュレーション上でも近くに表示されるという問題がある。今後は Spot エージェントの数に応じたエージェント同士の座標を調整する機能や、各 Spot エージェントの目的地に向かう User エージェントが発生する間のみ Spot エージェントを表示するといった改良が必要である。

次に、図 6 に示すコンソール画面から旅行行動拡張の考察を行う。現在はツイート文に動詞が存在すればすべて抽出している。このため、抽出した動詞が観光地内スポットを指していない可能性が考えられる。また、「する」「く」「ある」「くださる」といった単語も動詞であるが、これらの単語は単体では何を行ったのか読み取ることができないため、旅行行動としては不適切である。今後の課題として、観光地内スポットを主語とする動詞のみを抽出することや、意味を読み取ることができない単語は抽出しないように改善する必要がある。

4.3 システム評価アンケート調査の実施

滋賀県立大学の学生 81 名に対して、2023 年 1 月 19 日から 1 月 26 日までの期間でシステムの評価実験としてアンケート

表 7 旅行活動ツイートテキストデータの各評価項目の内訳 (N = 81)

1: 全くそう思わない, 2: あまりそう思わない,
3: どちらでもない, 4: 少しそう思う, 5: とてもそう思う

質問事項	1	2	3	4	5	平均値
彦根を魅力的に感じた	3	25	21	31	1	3.02
彦根を旅行したくなった	5	31	26	17	2	2.75
彦根地域で行きたい場所を見つけた	7	21	20	27	6	3.05
彦根に関して新たな発見があった	8	19	19	27	8	3.10

調査を実施した。4.1 節で収集された旅行活動ツイートのテキストデータと、4.2 節で可視化した Artisoc のシミュレーション画面を比較し、どの程度新たな旅行需要の発見や旅行先の決定支援、有効的、魅力的な可視化に繋がったかどうかの検証を行った。調査の方法は、テキストデータと、Artisoc のシミュレーション動画を見てもらい、それぞれのデータに対して質問に 5 段階で評価してもらう形式で行った。テキストデータは 4.1 節で収集された 94 ツイートを、日付とツイートのみを記載したデータである。テキストデータの一例を以下に示す。

● 2019-01-29, 今日河瀬駅近くの彦根工業高校へ出張。河瀬駅ではトイレの様式化の工事をされていた。滋賀県内の JR は、数年前にすべての駅にエスカレーターが設置されたが今回はすべての駅に様式トイレの設置がされるみたいだ。これは良いことだと思う。新快速設置の要望の端が張られていた。難しんじゃないかな? <https://t.co/nYUt3wVW5m>

● 2019-01-31, "I'm at ちゃんぽん亭総本店 彦根駅前本店 in 彦根市, 滋賀県 <https://t.co/xaJPAgWESN>"

● 2019-02-02, みんな今どこ? 俺、彦根 (笑)

● 2019-02-02, #滋賀ライダー あのベンチ

<https://t.co/gwLORWLxbS>

Artisoc シミュレーション動画は図 4 に示す Artisoc のシミュレーション動画データを見てもらった。質問は「彦根を魅力的に感じた。」、「彦根を旅行したくなった。」、「彦根地域で行きたい場所を見つけた。」、「彦根に関して新たな発見があった。」の 4 つで、それぞれ 5 段階評価である。次に、行きたい場所があった場合は、その場所についても回答してもらった。最後に、テキストデータ、動画データで気になる点、改善点を自由回答で記入してもらった。表 7 にテキストデータの、表 8 に Artisoc 動画データの質問と回答した人数を示す。平均値は、評価を 1 (全くそう思わない) から 5 (とてもそう思う) に点数化し、平均値を算出した。また、テキストデータにおける主な行きたい場所は、「彦根城」、「ラーメンにっこう」、「千成亭」、Artisoc シミュレーション動画における主な行きたい場所は、「彦根城」、「ラーメンにっこう」、「彦根城博物館」であった。

アンケート調査の結果、現在の提案システムでは、目的とする旅行需要の発見、旅行先決定支援、魅力的な可視化のすべての観点の評価がテキストデータの評価を下回った。

ここからは、テキストデータ、Artisoc シミュレーションの気

表 8 Artisoc シミュレーション動画の各評価項目の内訳 (N = 81)

1: 全くそう思わない, 2: あまりそう思わない,
3: どちらでもない, 4: 少しそう思う, 5: とてもそう思う

質問事項	1	2	3	4	5	平均値
彦根を魅力的に感じた	10	27	30	14	0	2.59
彦根を旅行したくなった	13	25	27	16	0	2.57
彦根地域で行きたい場所を見つけた	12	28	19	18	4	2.68
彦根に関して新たな発見があった	7	24	24	22	4	2.90

になる点、改善点の収集結果を示す。テキストデータの気になる点、改善点を示す。

- 文章だけだと微妙、画像がほしい。
- 旅行の楽しい要素が伝わりにくい。
- 内容の無い投稿が多い、要らない情報が多い (特に I'm at など)

これらは、フィルタリング手法の改善点とする。特に、内容の無い投稿については、エージェント化した際に観光地内スポットの抽出はできるため可視化されるが、行動が含まれていないためその場所で行っているのかわからない。しかし、可視化できているため不必要なデータではない。今後は場所、行動ともに内容の充実したツイートをできるだけ多く抽出できるよう、改善を行う。

次に、Artisoc シミュレーションの気になる点や、改善点を自由回答により収集した結果を示す。

- 文字同士の重なりを解消する。
- 各 Spot を訪れた人数を表示する。
- 実際の地図を重ねて表示する。
- 任意の Spot を選択すると周囲の有名地点を表示する (User が集まるエリアに存在する Spot を可視化)。
- 移動した形跡が多いルートを推薦ルートとして示す。

この結果を受けて、より魅力的な可視化ができるよう改善を行う。特に、任意の Spot から周囲の Spot を表示するという改善案は、元々訪問予定のスポットから足を伸ばしてもう一箇所訪問するという新たな旅行需要の発見に繋がるため、テキストデータの評価を上回ることができると考える。

5 まとめと今後の課題

本稿では、旅行活動データだけを精度良く、かつできる限り網羅的に収集するフィルタリング手法と、収集された旅行活動データから旅行の目的地決定を支援できるようなエージェント化の手法について検討した。その上で、エージェント化したシミュレーション画面を評価するアンケート調査を実施して、エージェント化による旅行需要の発見と旅行先決定支援の効果について検証した。

その結果、観光地「彦根」に対し、観光地名検索から 58 ツイート (異なり数)、位置情報検索から 12 ツイート (異なり

数)を得た。また、TFIDF法により彦根を特徴づける単語を抽出した結果、「ベンチ」「滋賀県」「Vidal」「IC」が特徴語として抽出された。さらに、特徴語に「彦根城」を追加して旅行活動ツイートを再収集した結果、35ツイート(異なり数)を更に得ることができた。フィルタリングにより得られたツイートをを用いて、Artisocにエージェント化するためにデータ整形と行動拡張抽出を行った結果、行動情報を持つツイートが31ツイートから75ツイートに増加した。一方で、エージェント化したシミュレーション画面は、旅行需要の発見、旅行先決定支援、魅力的な可視化の観点でテキストデータの評価を下回った。その原因として、文字同士の重なりや、実際の地図を重ねて表示、各Spotを訪れた人数の表示などが挙げられた。今後の研究で改善を図って行く。

最後に、フィルタリングとArtisocに関する今後の展望を述べる。本稿では、膨大なツイートの中からある場所を中心とした旅行活動データを抽出するためのフィルタリング手法について検討したが、現在は旅行活動でないツイートの方が多く収集されており、フィルタリング手法としては実用的ではない。その原因として、旅行活動単語で絞った結果取りこぼしたツイートが存在する点が挙げられる。例えば、「彦根城で何かをする」のようなツイートは、旅行活動単語を含んでいないため、絞り込みでは取りこぼしてしまう。しかしながら、「する」を旅行活動単語に含むと、「する」という単語は汎用性が高いため適合率が下がる。現在の旅行活動単語は単語に重点を置いているため、今後は「修飾語と被修飾語」「主語と述語」のような係り受け関係を考慮するよう改善することが考えられる。また、特徴語抽出についてもすでに述べた通り、特定地域にのみ頻出する単語を求める必要がある。今後は、現状のTFIDF法のIDFの有効性について検討し、さらに明確な区分分けができることが期待される。

また、Artisocはエージェント同士の相互作用による現象を再現するソフトウェアでありながら、現在はまだエージェント同士の相互作用がない状態である。Spotエージェントに関しては、シミュレーション開始時にSpotエージェント同士が接近していれば、距離を取ることで視認性を向上させるといった点が考えられる。Userエージェントに関しては、今後機械学習なども取り入れながら、社会における複数の旅行者の相互作用による旅行活動の「予測」に取り組むたいと考えている。また、これらの改善に加え、複数旅行者がよく通るルートを「推薦」するエージェントを追加することや、任意のSpotエージェントを選択すると周囲のSpotエージェントを表示するといった機能を追加することで、旅行需要発見や旅行先決定により貢献できると考えられる。

文 献

- [1] 株式会社JTB, “JTB公式 | 海外旅行・国内旅行 | ツアー・ホテル・旅館・宿の予約サイト,” <https://www.jtb.co.jp/> (参照 2023-01-01).
- [2] 株式会社JTB, “「コロナ禍の生活におけるインターネットやSNSからの“情報”に対する意識と旅行」に関する意識調査 (2021-04-09),” <https://press.jtbcorp.jp/jp/2021/04/sns.html>

- (参照 2022-12-01).
- [3] artisoc4 - MAS コミュニティ - 構造計画研究所, <https://mas.kke.co.jp/artisoc4/> (参照 2022-12-27).
 - [4] 渡邊 小百合, 吉野 孝, “観光地名なしツイートからの観光地に関する感想の抽出手法,” 情報処理学会論文誌, Vol.59, No.1, pp.43-51 (2018).
 - [5] 松本 義之, 井上 仙子, “下関地域における Twitter を利用した観光情報分析,” バイオメディカル・ファジィ・システム学会誌, Vol.22, No.2, pp.59-66 (2020).
 - [6] 本庄 由佳, 高嶋 恵子, 向 直人, “マルチエージェントを利用した Twitter における人間関係の可視化,” 第 16 回 MAS コンペティション (2016).
 - [7] Toshinori Sato, GitHub - neologd/mecab-ipadic-neologd: Neologism dictionary based on the language resources on the Web for mecab-ipadic, <https://github.com/neologd/mecab-ipadic-neologd> (参照 2023-01-28).