

パーソナライズされた観光地推薦のための、 性格特性を考慮することの効果検証

伊藤 桃[†] 今井 美希[†] 榎 美紀^{††} 小口 正人[†]

[†] お茶の水女子大学 〒112-8610 東京都文京区大塚 2-1-1

^{††} IBM Research - Tokyo 〒103-8510 東京都中央区日本橋箱崎町 19-21

E-mail: [†]{momo-i,miki,oguchi}@ogl.is.ocha.ac.jp, ^{††}enomiki@jp.ibm.com

あらまし 近年、訪日外国人の増加が見受けられる。それに伴い、有名な観光スポットは簡単に様々な Web サイトなどから情報を取得できるようになり、AI を使った観光スポット推薦システムなども増えてきた。主流はユーザの趣味嗜好情報からスポットを推薦するようなシステムである。しかし、そのような既存の推薦システムは、似かよった趣味嗜好のスポットを推薦する傾向があり、ユーザにとって単一的な推薦になってしまう。本研究では、新たなアプローチで観光スポットを推薦したいと考え、後述するように Personality Insights [1](以下 PI) を用いてユーザの性格特性の情報から推薦するようなシステム構築をすることを目指す。本論文では、システムを構築するためにまず前提となる、スポットごとにユーザの性格特性に傾向があることを検証する。

キーワード SNS, PersonalityInsights, Twitter, 観光地推薦, 情報抽出, 性格推定

1 はじめに

近年、訪日外国人の増加が見受けられる。2020 年の東京オリンピック開催も影響し、年々増加傾向にある。[2] それに伴い、有名な観光スポットは簡単に様々な Web サイトなどから情報を取得できるようになり、AI を使った観光スポット推薦システムなども増えてきた。主流はユーザの趣味嗜好情報からスポットを推薦するようなシステムである。しかし、そのような既存の推薦システムは、似かよった趣味嗜好のスポットを推薦する傾向があり、ユーザにとって単一的な推薦になってしまう。例えば、体を動かすことが好きか、食べることは好きか、などの簡単なアンケートをユーザに行わせ、その情報から体を動かすことが好きなユーザにはアクティビティ体験が行えるようなスポット、食べるのが好きなユーザには飲食店などを推薦するようなシステムである。本研究では、新たなアプローチで観光スポットを推薦したいと考え、PI を用いてユーザの性格特性の情報から推薦するようなシステム構築をすることを目指す。性格を考慮することで、単に趣味趣向情報から推薦するようなシステムでは訪れることがなかった場所の発見ができ、より観光地選択肢を増やすことのできるようなシステムを提案する。本論文では、システムを構築するためにまず前提となる、スポットごとにユーザの性格特性に傾向があることを検証する。本論文の構成は以下のとおりである。2 章で関連研究について述べ、3 章では提案システムの概要を紹介し、4 章で本研究で使用する PI について具体的に説明する。5 章で観光スポットや駅の性格特性の傾向の検証、6 章で神社や寺の性格特性の傾向の検証を行い、最後に、7 章で本稿をまとめる。

2 関連研究

本研究では、Twitter にて情報収集を行いそこで得たテキスト情報からユーザの性格特性を判断する。Jalal ら [3] や、Gou ら [4] が Twitter のテキスト情報から性格を判断するその正確性について説いている。Jalal らは、空港の待ち時間などその場にはないと分からないローカルな情報を Twitter のユーザから収集する qCrowd というシステムを構築したく、Twitter の情報からローカル情報を提供をしてくれるユーザか否かを判断するモデルを作成した。結果としては、65%以上の確率で正しくユーザをモデルが判別した。また、Gou らも、同じように Twitter の情報から最も主流な心理学的な性格の指標である Big Five Personality に基づいてユーザの性格特性を判定するモデルを作成しており、正確性を立証している。また本研究では、テキスト情報から性格特性を判断するのに独自のモデルではなく、PI という IBM のサービスを用いる。このサービスは Twitter などのユーザが書いたテキスト情報をインプットとして、性格の特性を、Big five(個性)、Needs(要求)、Value(価値感)の3つの次元に分割して出力する [5]。PI を用いた結果を検証している研究論文として、富永 [6] の論文を挙げさせていただく。この論文では、Twitter のユーザの時間とともに変化する人格を、PI を用いて分析している。結果として、人格特性の変化が感覚的に理解のできる要因によって変化していて、PI の正確性を裏付けるような結果であると解釈している。先述した従来の研究から、Twitter から取得したデータから正確性の比較的高い性格特性の判断ができることが言える。また本研究と従来研究の相違点としては、PI を用いた観光スポットの推薦に焦点を当てているという点である。性格特性に基づいてパーソナライズされた観光スポットを推薦することでユーザの行動

選択肢を増やすことが可能となると言える。

3 提案システム

3.1 提案システムの概要

本研究で提案する概要図を図 1 に示す。

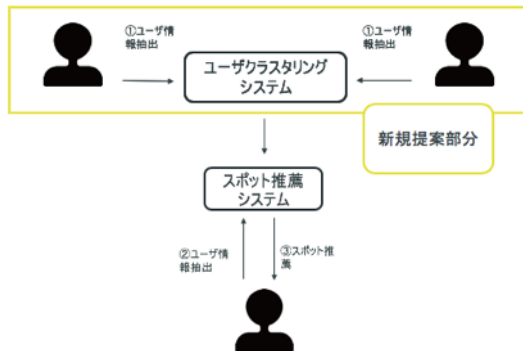


図 1: 提案システムの概要

スポット推薦システム以下については、後述の先行研究を引き継いだ形となっている。まず、大量のユーザ情報をあらかじめ抽出し、ユーザクラスタリングシステムにてユーザを性格特性に基づいてグループに分ける。そのクラスタリングシステムのデータに基づき、新たなユーザは自身のユーザ情報からグループにクラスタリングされ、そのグループが性格的によく訪れるスポットをスポット推薦システムによって推薦される、というものである。

3.2 先行研究

訪日外国人の趣向に合わせたイベント情報提供のためのソーシャルメディア活用の検討 (DEIM 2019) [7] という題で、今井らが発表している。図 2 にシステム概要図と流れを示す。

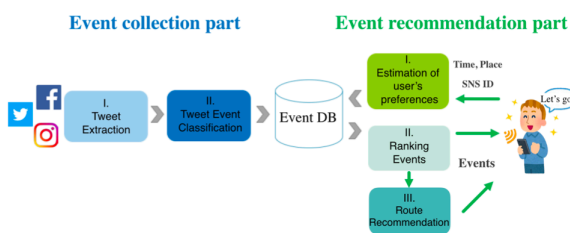


図 2: 先行研究システムの概要

まず、イベントを収集する部分では、観光客に有益なイベント情報をソーシャルメディアから抽出し、保存する。イベントを種類ごとにカテゴライズする。イベントを推薦する部分については、まず、各ユーザの SNS のアカウントを分析することでユーザの趣味趣向を検出する。そして、ユーザにその時、その場のイベント情報を提供するというものである。

4 PI から分かる性格特性

PI が推定する性格特性を表 1 に示す。性格の基本的な次元が 5 つであるという Big Five Model の特性項目と、Kevin Ford の Universal Needs Map に沿った Needs (欲求) 分析、Schwartz の価値概説 (Schwartz Value Survey) に沿った Values (価値観) 分析の特性項目からなる。

表 1: PI が推定する性格特性

Big Five (個性)	facet (小分類)
Agreeableness (協調性) 人当たりの良い・温情のある vs. 冷たい・不親切	altruism, cooperation, modesty, morality, sympathy, trust
Conscientiousness (誠実性) 勤勉・まめな人 vs. 楽天的・不注意	achievement, striving, cautiousness, dutifulness, orderliness, self-discipline, self-efficacy
Extraversion (外向性) 外向的・エネルギーッシュ vs. 孤独を好む・控えめ	activity, level, assertiveness, cheerfulness, excitement, seeking, friendliness, gregariousness
Neuroticism (感情起伏) 繊細・神経質 vs. 情緒安定な・自信家の	anger, anxiety, depression, immoderation, self-consciousness, vulnerability
Openness (知的好奇心) 好奇心が強い・独創的 vs. 着実・質実心強い	adventurousness, artistic interests, emotionality, imagination, intellect, liberalism

Needs (欲求)	Challenge (挑戦), Closeness (親密), Curiosity (好奇心), Excitement (興奮), Harmony (調和), Ideal (理想), Liberty (自由主義), Love (社会性), Practicality (実用主義), Self-expression (自己表現), Stability (安定性), Structure (仕組)
Value (価値観)	conservation (現状維持), hedonism (快楽主義), openness_to_change (変化許容性), self-enhancement (自己増進), self-transcendence (自己超越)

5 スポットごとの性格特性の傾向の検証

提案システムを実現するため、本論文ではまず、スポットごとに性格特性に傾向が見られるか否かを検証する。各スポットを訪れるユーザの性格特性数値からスポットを性格特性の傾向が違おうであろう、2 グループに分けて検定や scikit-learn ライブラリ [8] を用いて、ランダムフォレストモデルを作成し、実際に機械的に性格特性をもとにして訪問スポットを推定できるか検証する。

5.1 検証用データセット概要

まず、Twitter の公式 API [9] を用いて、キーワードに複数の関東圏の観光スポットを設定しデータを収集した。キーワードに設定した各スポットについてツイートしているユーザを、そのスポットに訪れたユーザと仮定し、設定したキーワードで見つけた日本人ユーザの過去の全てのツイートを PI を通して表 1 の各特性の数値を検証に用いる。数値は、具体的には 0 から 1 の連続値である。なお、PI は 100 単語以上からなるテキストからしか性格特性を算出しない制限がある。よって 100 単語以上のツイートをしているユーザに限定している。また、観光スポットは、複数の観光情報サイトから得られたスポットから後述する取得期間内に一定以上のツイート数を得られたスポットを選び、正しく性格特性を判断できないと思われる箇所 (URL 表記など) は適宜ツイートを削除するなどして加工している。データ取得期間は、2019 年 11 月 17 日から同年 12 月 11 日までである。まず、各スポットを表 2 の通りに駅名グループとレジャー施設名グループに分けてこの 2 グループについて検証を行う。この 2 グループに分けた理由としては、駅は目的なく誰でも訪れるが、レジャー施設はユーザは目的を持って訪れるので、違いが出やすいのではないかと考えたからである。

5.2 効果量算出

上述の駅名・レジャー施設の 2 グループ間に性格特性の差異

表 2: データセットスポットグループ表

駅名	レジャー施設
吉祥寺	上野動物園
後楽園	ラフォーレ原宿
代官山	よみうりランド
高円寺	東京ドーム
三鷹	東京スカイツリー
お台場	ディズニーランド
下北沢	赤レンガ倉庫
東京駅	コスモワールド
秋葉原	ガラスの森美術館
自由が丘	ツインリンクもてぎ
浅草	中禅寺湖
	那須どうぶつ王国

があるか検証するため、各性格特性値ごとに Cohen's d の効果量を算出する。この値の絶対値が大きいと、その PI 項目が比較グループにより値が異なると言えるため、各グループの特徴量として用いることができる事を意味する。一般的に効果量の絶対値が 0.2 以上から比較グループに差がある効果が存在する。また、正の項目のものがレジャー施設グループに対して駅名名グループが正に効果があるものである。表 3 に算出結果を示す。

表 3: Cohen's d 効果量絶対値降順結果 (上位 5 抜粋)

PI 項目名	効果量
big5_openness	0.507833
need_love	-0.303105
need_closeness	-0.261188
value_self_transcendence	-0.212750
need_harmony	-0.206479

表の上から PI 項目に関する詳しい説明をする。'big5_openness' は知的好奇心の強さや、新しい物事や考えに対して開放的かどうかを表すもので、'need_love' は社会との接触を求めやすいか否か、'need_closeness' は所属感を求めやすいか否か、'value_self_transcendence' は生活において挑戦をしたりしやすいか否か、'need_harmony' は協調性に関する値となっている。結果から、比較的レジャー施設に訪れるユーザは所属や社会の接触を求めやすく協調性の高いユーザが多い傾向があると言える。

5.3 ランダムフォレスト分類モデル作成

説明変数に PI 性格特性 52 項目、目的変数に駅名グループかレジャー施設名かを設定した。データのボリュームとしては、合計 3,734 名の性格特性データを使用している。また、学習データとテストデータの割合は 3:7 とした。混同行列結果を表 4 の通り示す。

約 61%の精度が得られた。50%以上の精度が得られたことで駅とレジャー施設には何らかの性格特性の違いがあることが窺える。

表 4: ランダムフォレスト混同行列

		予測	
		駅	レジャー
正解	駅	382	177
	レジャー	257	305

5.4 ランダムフォレストモデル特徴量重要度

先述したランダムフォレストモデルの特徴量重要度を表 5 に示す。特徴重要度とは、簡単にそのモデルを作成した際にそれぞれの特徴量がどれほどモデルに寄与したかを数値で表したものである。1 に近いほどモデルによく寄与した特徴量である。

表 5: ランダムフォレストモデル特徴量重要度

PI 項目名	効果量
big5_openness	0.144386
need_closeness	0.065096
need_love	0.060057
big5_neuroticism	0.050645
value_hedonism	0.050387

上位の 'big5_openness', 'need_closeness', 'need_love' は先ほどの効果量の結果と被っており、整合性の取れた結果となった。

5.5 階層クラスタリングモデル作成

次に、駅とレジャースポットの各々に対して、PI 性格特性 52 項目を用いて階層クラスタリングを実施した。図 3 に結果を示す。

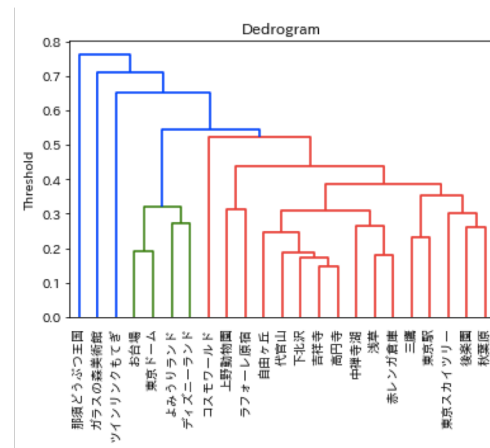


図 3: スポット階層クラスタリング結果

駅のスポットが近く固まっている印象がある。レジャースポットに関しては、那須、箱根、もてぎなど距離的に遠いスポットが性格特性も他と遠い関係にあり、興味深い結果となった。また、駅名グループスポットのみで階層クラスタリングを行った結果を図 4 に、クラスタリング結果を色ごとに地図にプロットしたものを図 5 に示す。緑色のクラスタは、更に深緑色で代官山、下北沢、明るい緑色で吉祥寺、高円寺をプロットした。

同じクラスタは比較的場所が近いことが伺える。

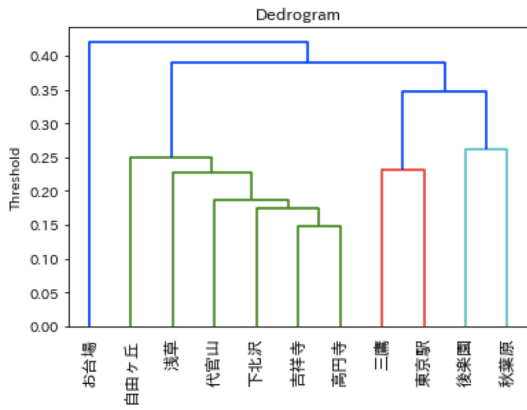


図 4: 駅グループスポット階層クラスタリング結果

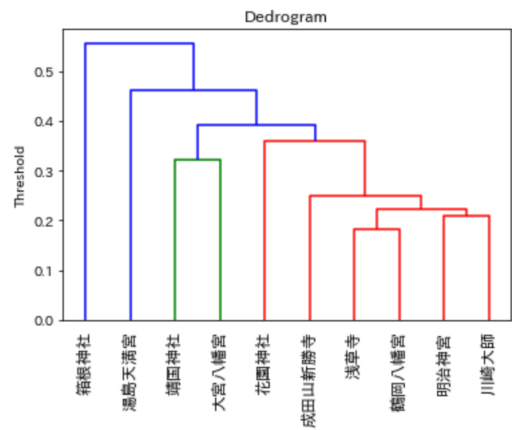


図 6: 初詣スポット階層クラスタリング結果



図 5: 駅グループスポット地図プロット結果

離れてクラスタリングされた印象があったので、箱根にあるスポットと他スポットで比較検証をするべく考えた。

6.3 効果量算出

先述した指標である *cohen's d* という効果量を算出する。また、正の項目のものが箱根神社以外のスポットグループに対して箱根神社のグループが正に効果があるものである。表 6 に算出結果を示す。

表 6: Cohen's d 効果量絶対値降順結果 (上位 5 抜粋)

PI 項目名	効果量
big5_openness	-0.474621
need_closeness	0.406022
big5_agreeableness	0.332959
need_harmony	0.316028
need_love	0.305521

結果から、箱根神社に訪れるユーザは先ほどのレジャーと駅に分けた際のレジャー施設グループのユーザに傾向が似ていると言える。箱根神社に訪れるユーザは、比較的所属や社会の接触を求めやすく協調性の高いユーザが多い傾向があるとみられる。箱根神社は身近に訪れるには遠く、レジャースポットの要素が含まれていると考えられるため、ほかのレジャー施設グループのユーザと傾向が似ていることに違和感はないと思われる。

7 まとめと今後の予定

本研究では、性格特性を考慮した観光地推薦システムを実現すべく、前提となるスポットごとの性格の違いを様々なデータセットと手法で検証を行なった。スポットを駅とレジャー施設とで比較した検証では、レジャー施設に訪れるユーザの方が、所属や社会の接触を求めやすく協調性の高いユーザが多い傾向があると言え、感覚的に腑に落ちる結果となった。また、駅名のスポットごとの階層的クラスタリングを実施した結果を地図にプロットした検証では、スポットの位置によってクラスが固まった印象があり、駅に訪れるユーザの中でも、例えば沿線

6 神社、寺スポットによる性格特性の傾向の検証

先述の検証セクションと同様の要領で、観光スポットとなる都内の神社や寺に訪れるユーザを、データ収集時期に合わせ、初詣に行くユーザを想定し各スポットで違いがあるかどうかを検証した。

6.1 検証用データセット概要

先述の検証セクションと同様の手順でデータセットを作成した。各神社や寺をキーワードに設定し、設定したキーワードを含むツイートを行なったユーザの過去のツイートから PI 項目ごとの数値を算出したものを用いて検証を行う。データ収集期間は、2020 年 1 月 1 日から同年 1 月 6 日であり、取得ユーザは合計で 896 名である。スポットは複数の初詣神社サイトの中からデータを収集し、一定のユーザが得られたスポットをデータとして使用する。

6.2 階層クラスタリングモデル作成

PI 性格特性 52 項目を用いて、スポットを階層クラスタリングした。図 6 に結果を示す。

年齢性別など関係なく、様々なユーザが訪れるであろう明治神宮や川崎大師などと離れた位置に、学問にまつわる神社として知られる湯島天満宮があったり、位置的に他スポットと比較して遠い位置にある箱根神社が離れてクラスタリングされた印象がある。先のセクションと現セクションを通して、箱根にあるスポットが共通して階層的クラスタリングを実施した際に、

ごとの性格特性の傾向がありそうであると言える。総じて、スポットごとに性格特性の違いの傾向は見られたので、今後は具体的に提案したシステムの実現に向けて研究を進めていきたい。それとともに、より様々なバリエーションのスポットデータを収集し、新たなスポットごとの性格傾向を見つけたい。

文 献

- [1] Ibm watson personality insights. <https://cloud.ibm.com/catalog/services/personality-insights>.
- [2] 日本政府観光局. https://www.jnto.go.jp/jpn/statistics/data_info_listing/pdf/200117_monthly.pdf.
- [3] Jalal Mahmud, Michelle X Zhou, Nimrod Megiddo, Jeffrey Nichols, and Clemens Drews. Recommending targeted strangers from whom to solicit information on social media. pp. 37–48, 2013.
- [4] Liang Gou, Michelle X Zhou, and Huahai Yang. Knowme and shareme: understanding automatically discovered personality traits from social media and user sharing preferences. pp. 955–964, 2014.
- [5] 那須川哲哉, 上條浩一, 山本真大, 北村英哉. 日本語における筆者の性格推定のための言語的特徴の調査. 言語処理学会第 22 回年次大会発表論文集, pp. 1181–1184, 2016.
- [6] 富永登夢, 土方嘉徳. Twitter ユーザの受け取るフィードバックと人格特性の変化の関係に関する調査と分析. 知能と情報, Vol. 31, No. 1, pp. 516–525, 2019.
- [7] 今井美希, 工藤瑠璃子, 榎美紀, 小口正人. 訪日外国人の趣向に合わせたイベント情報提供のためのソーシャルメディア活用の検討. *DEIM Forum 2019*.
- [8] scikit-learn. <http://scikit-learn.org/stable/>.
- [9] Twitter search api. <https://dev.twitter.com/rest/public/search>.