性格特性を考慮した 高パーソナライズ性スポット推薦システムの一検討

伊藤 桃[1]　榎 美紀[2]　小口 正人[1]

概要：昨今新型コロナウイルスが，世界中に多大な影響を与え続けている．驚異的な感染力により未だにウイルスの収束も見えず，現在も外出を控えた日常生活を送っている人が多いだろう．しかしながら，収束後はその反動から，以前より観光業界の需要が再び盛り上がりを見せることは容易に想像でき，その際に，多様なスポット推薦システムの需要も共に高まってくると考える．現在既に，様々な観光スポットは簡単に Web 上から情報を取得できるようになり，AI を用い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，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，Personality Insights(以下 PI)[1] を用いユーザの性格特性情報からスポット推薦を行うシステムの構築を目指す．性格を考慮することで，より観光地選択肢を増やすことのできるシステムを提案したい．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した．様々なスポットを性格特性によって全体的にクラスタリングした結果，性格特性からいくつか同系統のスポットがクラスタリングされた．しかしながら，スポットによってはユーザごとの訪問目的の違いから，性格値の標準偏差が大きいものがあり，今後詳しく分析進め，システムの実現を進めていく必要がある．

A study of a highly personalized spot recommendation system that takes personality traits into consideration

MOMO ITO[1]　MIKI ENOKI[2]　MASATO OGUCHI[1]

1. はじめに

昨今新型コロナウイルスが，世界中に多大な影響を与え続けている．驚異的な感染力により未だにウイルスの収束も見えず，現在も外出を控えた日常生活を送っている人が多いだろう．しかしながら，収束後はその反動から，以前より観光業界の需要が再び盛り上がりを見せることは容易に想像でき，その際に，多様なスポット推薦システムの需要も共に高まってくると考える．現在既に，様々な観光スポットは簡単に Web 上から情報を取得できるようになり，AI を用い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，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，PI を用いユーザの性格特性情報からスポット推薦を行うシステムの構築を目指す．性格を考慮することで，より観光地選択肢を増やすことのできるシステムを提案したい．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 章では提案するシステムについての説明をし，4 章では本研究で使用する PI について具体的に説明する．5 章で実験にあたり使用したデータセットについての概要を説明し，6 章ではデータをクラスタリングした結果をまとめる．7 章ではスポット別の性格値の標準偏差の違いについて延べ，8 章で本稿をまとめる．

[1] お茶の水女子大学
Ochanomizu University
[2] IBM Research - Tokyo

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い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 と言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．また，Hrazdil [7] の論文についても触れたいと思う．この論文では，CEO と CFO のビッグファイブの性格特性を測定し，これらの 5 つの特性に基づいてリスク許容度の測定値を計算している．結果として，彼らのリスク許容度特性の推定は，CEO のリスク許容度と監査手数料との関連付けを示した．具体的には，CEO のリスク許容度が高いほど監査手数料が大幅に高くなること，および CEO のパーソナリティが，PI を使用することで報酬ポートフォリオによって引き起こされるリスクをとるインセンティブを超えて，クライアントのリスクの監査人の評価に影響を与えることがわかった．これは，PI の正確性または有用性を証明するものであると言える．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

図 1 に提案システムと流れを示す．

- SNS から事前に大量のユーザテキスト情報を抽出
- PI によりテキストデータをパーソナリティ数値配列に変換
- PI 数値を入力として使用してクラスタリングモデルを作成し，いくつかのクラスタグループに分割
- 問合せユーザのテキストデータを抽出し，そのデータを PI を通じてパーソナリティ数値配列に変換
- 問合せユーザが所属するクラスタリンググループを決定し，クラスタリンググループが頻繁に訪れる場所を推薦

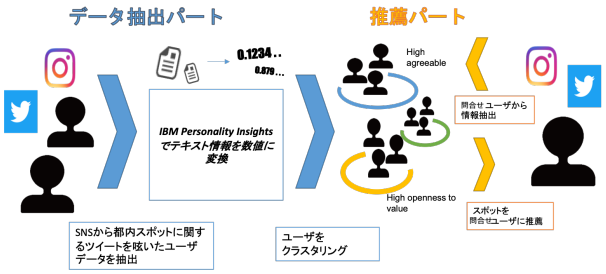

図 1 提案システム

本研究では，推薦パートではなく，データ抽出パートに焦点を当てていく．

4. Personality Insights から分かる性格特性

性格特性を数値化するために，IBM Personality Insights(以下 PI)[5] というツールを用いる．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．PI とソーシャルメディアを用いた結果を検証している研究論文も数多く存在し，ソーシャルメディアから取得したデータから正確性の比較的高い性格値の判断ができることが言える．

表 1 PI によって推定できる性格特性

Big Five(個性)	数値の大小による意味
協調性	人当たりの良い・温情のある vs 冷たい・不親切
誠実性	勤勉・マメな人 vs 楽観的・不注意
外向性	外交的・エネルギッシュ vs 孤独を好む・控えめ
感情起伏	繊細・神経質 vs 情緒安定な・自信家の
知的好奇心	好奇心が強い・独創的 vs 着実・警戒心が強い

5. 検証用データセット概要

まず，Twitter の公式 API[8] を用いて，キーワードに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スポットは，公益財団法人東京観光財団と呼ばれる観光を推進する公式の団体による Web サイト「GO TOKYO」[9] に載っているスポットである．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の中で目視で「行った」と呟いているユーザ，あるいは実際に行っていないと知り得ない情報と共に呟いているユーザに

絞り，その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 の各特性の数値を得，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，スポットは，2020 年 10 月 22 日, 2020 年 12 月 20 日, 2021 年 2 月 22 日に取得し，その間に一定以上のユニークユーザ数を得られたスポットに絞っている，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．

6. クラスタリング

前章の性格特性を全ての期間のデータを用いて，スポットごと訪問ユーザの性格値の中央値を用いて，最初に図 2 のように scipy ライブラリ [12] の dendrogram による階層的クラスタリングを実施した．各スポットごとの重複ユーザデータは除いている．また，結果はスポットをカテゴリごとに色分けをしている．黄色が映画館カテゴリ，黄緑色が神社や寺カテゴリ，水色が専門販売店カテゴリ，茶色が商店街カテゴリ，ライム色がアミューズメントカテゴリ，赤色がデパートカテゴリ，紫色が川や橋カテゴリ，ピンク色が公園カテゴリ，黒色が美術館や博物館カテゴリ，緑色が飲食店カテゴリ，グレー色がホテルカテゴリである．カテゴリが比較的まとまってクラスタリングされている部分があるが，これはそのカテゴリのスポットに訪れるユーザの性格に特徴があると考える．

まず特徴的なのは，黒色の美術館，博物館カテゴリが右側に固まってクラスタリングされている所である．また，ピンク色の公園カテゴリも特徴的に右側に固まっている．このように偏ってクラスタリングされているカテゴリもあれば，大半のスポットカテゴリはバラバラにクラスタリングされている様子もうかがえる．次は，実際にクラスタリンググループにいくつか分け，どの性格値に特徴があるのか詳しく分析する．

グループを分けるにあたり，次に用いたのは，pyclustering ライブラリ [10] の x-means である．クラスタリングのライブラリである scikit-learn[11] の k-means をベースに，最適なクラスター数を自動で決定し，出力するものである．この結果，8 つのグループにクラスタリングされ，それぞれのグループのスポット群は表 2 のようになった．

また，それぞれのクラスタリンググループのセンターとなる性格値を図 3 のように示す．

表 2 クラスタ番号とそのスポット群

クラスター番号	スポット名
0	'神田明神', '大丸東京店', 'グランスタ', '丸の内オアゾ', '築地本願寺', '築地場外市場', '護国寺', 'ピックカメラ池袋', '不忍池', '国立科学博物館', '舎人公園', '光が丘公園', '隅田川', '江戸東京博物館', '荒川河川敷', 'ららぽーと豊洲', 'ゲートブリッジ', '豊洲ぐるり公園', '木場公園', '葛西臨海公園', '渋谷ストリーム', '井の頭恩賜公園', '中野ブロードウェイ', '砧公園', '日本科学未来館', '城南島海浜公園'
1	'乃木神社', 'サンシャイン水族館', 'アニメイト池袋', 'サンシャインシティ', 'スカイツリータウン', '渋谷ロフト', '渋谷センター街', '渋谷マークシティ', 'WOMB', '竹下通り', '表参道ヒルズ', 'アクアパーク品川', '愛宕神社', 'リキッドルーム'
2	'GINZASIX', '東急プラザ', '新丸ビル', '丸ビル', 'COREDO 室町', '日枝神社', 'コンラッド東京', '西武池袋本店', '新宿ゴールデン街', '目黒川', '六本木ヒルズ', '等々力渓谷'
3	'国立映画アーカイブ'
4	'秋葉原電気街', 'アニメイト秋葉原', 'ポケモンセンターメガトウキョー', 'ダイバーシティ東京', 'デックス東京ビーチ', 'アクアシティお台場'
5	'銀座三越', '水天宮', '日本橋三越', '飛鳥山公園', '吾妻橋', '京王百貨店 新宿'
6	'三菱一号館美術館', '日比谷シャンテ', '赤坂 ACT シアター', '東京芸術劇場', '旧岩崎邸庭園', '上野の森美術館', '増上寺'
7	'東京ステーションギャラリー', '根津美術館', '国立新美術館', '東京都庭園美術館'

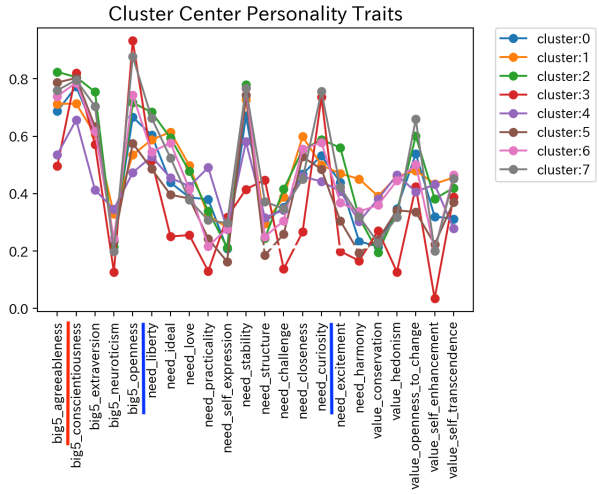

図 3 クラスタグループセンター性格値

まず，図 3 の赤線部分項目である big5_agreeableness を見てみると，クラスタ 3,4 のグループが他グループに比べ低いことがわかる．この性格値は協調性を表し，値が高い人は他人とうまくやって行こうとするタイプとし，低い人は人より自分の興味を優先するタイプを表す．クラスタ 3 や 4 のグループに属しているスポット群を見てみると，映画館であったり，アニメイトやポケモンセンターなど自身の趣味によって訪れるスポットが多い．また，ダイバーシティやアクアシティ，電気街といったスポットも，個人の趣味に関するイベントに参加する目的で訪れたとみれるユーザが多数見受けられ，そのようなユーザの趣味に関するツイートが big5_agreeableness の値の低さにつながっているのではないかと考察する．

また，青線部分項目である need_curiosity や big5_openness と言った性格値がクラスタ 3,7 のグループが比較的高い．この性格はそれぞれ探究心や知的好奇心度を表し，属しているグループのスポット群をみてみると，先ほども挙がった映画館や，クラスタ 7 のスポット群は美術館のスポット群になっており，このようなスポットに訪れるユーザが探究心や知的好奇心度が比較的高いことは感覚的に腑に落ちる．このように PI によって求めた性

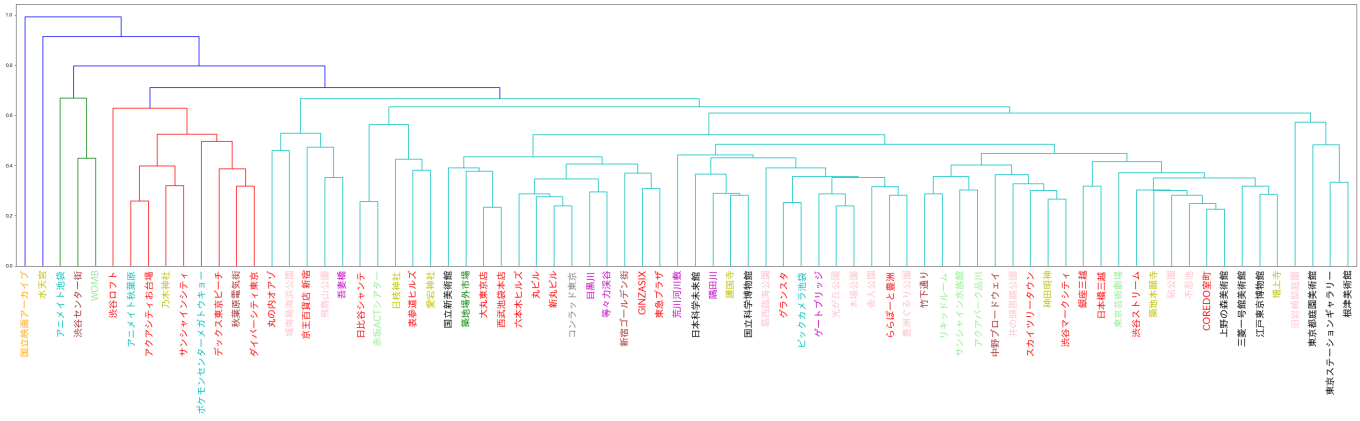

図 2　スポットクラスタリング

格値で，クラスタリングをするとある程度感覚的に腑に落ちる形で結果が得られ，提案システムの実現ができ得ると考える．

7.　スポット別性格値標準偏差

前章で，性格値によるクラスタリングを行ったが，クラスタ 0,1,2 のグループは比較的スポット群の数が多く，性格値に特段特徴も見られなかった．この原因として，クラスタ 0,1,2 のグループは性格のばらつきが大きいスポットがまとまった可能性がある．スポットごとの性格のばらつきを見るために，スポットごとの性格値標準偏差の大小を見る．以下表 3,4 に，スポットごとの性格値の標準偏差平均の上位下位 5 位を載せる．

表 3　スポット別全性格値標準偏差
平均昇順上位 5

スポット名	全性格値標準偏差平均
国立映画アーカイブ	0.238542
東京ステーションギャラリー	0.251342
日比谷シャンテ	0.263196
東京都庭園美術館	0.263236
赤坂 ACT シアター	0.264242

表 4　スポット別全性格値標準偏差
平均昇順下位 5

スポット名	全性格値標準偏差平均
渋谷センター街	0.288051
竹下通り	0.290899
目黒川	0.291056
サンシャインシティ	0.293577
渋谷ロフト	0.295444

この結果をみると，ユーザの性格値に比較的ばらつきがない上位のスポットは映画館や美術館といった訪問目的のはっきりしたスポットが多い．それに対して下位の結果は，商店街や通り，川や複合施設など，そのスポット自体に目的がないスポット，あるいは複合施設など訪問目的がそのユーザによってバラバラであるようなスポットである．そして，この下位 5 つのスポット群は全て表 2 で特徴のあまりみられなかったクラスタ 0,1,2 のいずれかのグループに属しており，反対に上位 5 つのスポット群は全てクラスタ 0,1,2 のグループ以外に属している．このことから，スポットごとに訪問ユーザの性格特性にばらつきの大小があることが分かり，今後詳しく分析を進める必要がある．

8.　まとめと今後の課題

本研究では，性格特性を考慮したスポット推薦システムを実現すべく，PI で求めた性格値をスポットごとにクラスタリングすると，感覚的に腑に落ちるようなクラスタリング結果となった．しかしながら，スポットごとに訪問目的の違いから，ユーザの性格値にばらつきの大小があることも確認した．今後は例えば訪問目的別による性格値の違いの検証をするなど，より詳しいスポット別の分析を進め，それらを加味しながら提案したシステムの実現を進めていきたい．

9.　謝辞

本研究は一部，JST CREST JPMJCR1503 の支援を受けたものである．

参考文献

[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_listin
[3] MAHMUD, Jalal, et al. Recommending targeted strangers from whom to solicit information on social media. In: Proceedings of the 2013 international conference on Intelligent user interfaces. 2013. p. 37-48.
[4] GOU, Liang; ZHOU, Michelle X.; YANG, Huahai. KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014. p.

955-964.
[5] 那須川哲哉, et al. 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, 2016, 1181-1184.
[6] 冨永登夢, 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 31.1 (2019): 516-525.
[7] Hrazdil, Karel, et al. "Measuring executive personality using machine-learning algorithms: A new approach and validity tests." Journal of Business Finance and Accounting Conference Paper. 2019.
[8] Twitter Search API. https://dev.twitter.com/rest/public/search
[9] Information on https://www.gotokyo.org/
[10] pyclustering. https://pypi.org/project/pyclustering/
[11] scikit-learn. http://scikit-learn.org/stable/
[12] scipy. https://www.scipy.org/