

観光地推薦のための性格特性を考慮した

パーソナライズ性向上の一検討

伊藤桃 (指導教員: 小口正人)

1 はじめに

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旨とする。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。

2 提案システム

本研究で提案する概要図を図 1 に示す。

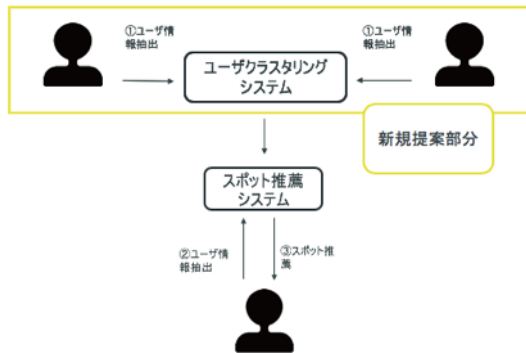


図 1 提案システムの概要

スポット推薦システム以下については、先行研究 [2] 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 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・まじめな vs. 無気力・不注意	achievement, striving, cautiousness, dutifulness, orderliness, self-discipline, self-efficacy
Extraversion (外向性) 外向的・エネルギッシュ vs. 内向的・控えめ	activity, assertiveness, cheerfulness, excitement, seeking, friendliness, gregariousness
Neuroticism (感情不安定) 繊細・神経質 vs. 情緒安定な・自信家の	anger, anxiety, depression, immorality, self-consciousness, vulnerability
Openness (知的好奇心) 好奇心が強い・独創的 vs. 着実・警戒心が強い	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), Ideas (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 (組織)
Values (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)

4 データセット概要

まず、TwitterAPI[3] 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、そのユーザの過去のツイート全てを PI を通して図 1 の各特性の数値を検証に用いる。数値は、0-1 の連続値である。なお、100 単語以上のツイートをしているユーザに限定した。また、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検証を行う。この 2 グループに分けた理由は、駅は目的なく誰でも訪れるが、レジャー施設はユーザが目的を持って訪れるので、比較すると違いが出るのではないかと考えたからである。

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

5 スポットごとの性格特性の傾向の検証

5.1 効果量算出

サンプルサイズに依存することのない指標である cohen's d という効果量を算出する。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

