

季節によって変化する観光名所プロフィールの自動生成法 ～TwitterとWikipediaを情報源として～

房 冠深 亀井 清華 藤田 聡
(広島大学 大学院工学研究科 情報工学専攻)

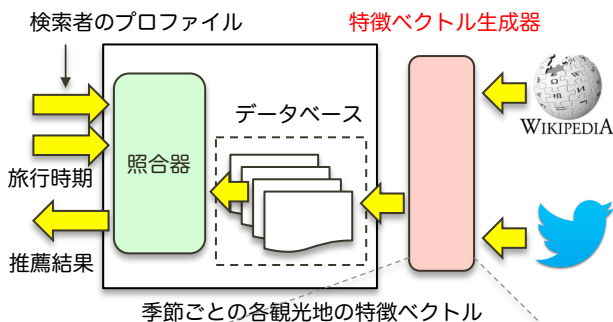
概要 季節によって大きく変化する観光名所のプロフィールをTwitterとWikipediaの記述から自動的に生成する手法を開発した。この手法を用いることで、旅行時期と旅行者の好みにマッチした適切な観光地推薦がおこなえるようになる。

◎ 背景と研究目的

経産省が推進するクールジャパン戦略や急激な円安などを背景として、年間1000万人を超える外国人観光客が日本を訪れるようになってきた。外国人観光客の多くは、一般的なガイドブックには載っていない穴場情報を強く求めているが、そのような場所を旅行者の興味と希望する旅行時期に合わせて適切に提示する観光地推薦システムは実現されていない。

本研究では、そのような推薦機能を実現する上で鍵となる、**季節に応じた観光地プロフィールの自動生成手法**を開発した。提案手法はTwitterとWikipediaのみを情報源としており、Wikipediaに記載されている観光地であればどんなにマイナーな観光地であっても推薦対象とすることができる。

◎ 観光地推薦システムの構成



◎ 提案手法の流れ

ステップ1：Wikipediaの各観光地の説明文からTF-IDF値を使って特徴ベクトルを生成（単語集合は観光地ごとに異なる）

ステップ2：Twitterから観光に関するツイートを収集して季節ごとにまとめ、TF-IDF値により特徴ベクトルを生成

ステップ3：ステップ2で求めたベクトル中の値を用いて、各観光地ベクトルに含まれる単語の重みを選択的に増加させる

◎ 提案手法の特徴

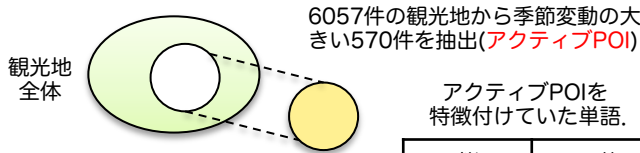
1. 特徴ベクトルが人手を介さず自動的に生成できる
2. それまでツイートされたことのない観光地のベクトルも、他の観光地のトレンドに合わせてハイライトされる
3. 季節ごとのベクトルをあらかじめ取得しておけば、季節に依存させる度合いをユーザの側で制御できる

◎ データセットについて

Wikipediaの「日本の観光地一覧」から6057件分の観光地の説明文をダウンロードし、形態素解析(Mecab)により名詞を抽出。抽出された各名詞にTF-IDFで重み付け（ステップ1）

2013年10月から2014年4月までにつぶやかれた5千万件のツイートをStreaming APIで取得し、観光地名によるフィルタリングで約50万件に絞り込む。得られたツイートを月単位でまとめ、月ごとの特徴ベクトルを作成（ステップ2）

◎ 実験結果



12月→2月	195件
10月→2月	149件
9月→2月	102件
1月→2月	44件
11月→2月	39件
9月→1月	32件

桜	299件
名所	177件
カフェ	118件
イベント	114件

最大の振幅を与える二地点、秋→春で振れる観光地が多いが、冬→春、秋→冬も観測される。

仁和寺(京都)と類似した観光スポット(類似度の高い順)

	11月	2月	Wikipediaのみ
1	下呂温泉合掌村	姫路城	大覚寺
2	大覚寺	八戸公園	曼珠院
3	再度公園	高遠城	観世音寺
4	いやし・もてなし神山街道	津山城	荒川堤
5	梅ヶ瀬溪谷	百十郎桜	妙心寺



秋は紅葉、春は桜で観光地が正しく特徴付けられている

◎ 今後の展開

ユーザの検索履歴などからユーザプロフィールを取得する部分を実装し、観光地推薦システムとして完成させる。別途開発中の経路推薦システムや口コミを考慮したホテル推薦システムなどと組み合わせ、統合旅行支援システムとして公開する。