

FIT のプロフィールに合った旅行先推薦システム

李明昊† 山田敬三† 高木正則† 佐々木 淳†

岩手県立大学ソフトウェア情報学部†

1. はじめに

近年の岩手県の統計[1]によると、岩手県においての外国人旅行者数が急激に増加している。平成 29 年度の総外国人旅行者数は約 26 万人となっており、平成 28 年度と比べ約 10 万人増加している。特にこの中でも、団体客より海外個人旅行 Free Independent Traveler（以降 FIT と称する）が著しく増加している。

また、岩手県の平成 29 年度の旅行者数の内訳として、台湾が約 60%を占める 16 万人、中国が約 11%を占める 3 万人となっている。つまり、中国語を使用する旅行者数が全部の 70%を占めている。

今後このような中国語圏 FIT たちに対し、適切な対策や情報提示が求められる。

現存する観光推薦システムの多くは観光地の訪問人数や評価順などを基準にして推薦を行っている。この方法で推薦される観光地は必ずしも、ユーザにとって適切である保証はなく、ユーザはさらに選別する必要がある。また、同じ観光地がずっと上位に上がるという悪循環が生じている[2]。

本研究では、上記の課題を改善できる、中国語圏の FIT を対象とした、個人のプロフィールに合った岩手県の旅行先推薦するシステムを提案する。

2. システムの提案

2.1. システム概要

図 1 に本研究で提案するシステム利用の流れを示す。本システムでは、ユーザのプロフィールと岩手県を観光した旅行者のプロフィールを比較し、プロフィールが類似していればいるほど、ユーザとの趣味趣向が近いと考えられる。その旅行者が訪問した観光地を推薦することにより、よりユーザが求めるものに近い観光地が推薦される可能性が高まる。システムの流れとして、まずユーザが自分のプロフィールを入力する。そして、ネットより有用情報を自動的に集め、各投稿者のプロフィールと目標となる観光地を抽出する。抽出したプロフィールとユーザのプロフィールを比較し、類似度を算出する。最後にこの類似度を元に投稿

者をランキングし、上位投稿者が訪問した観光地を推薦する。

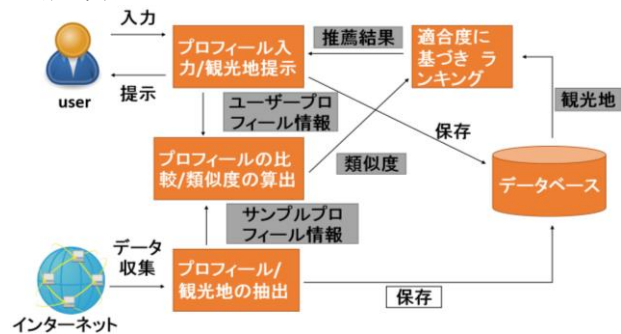


図 1. システムの流れ

2.2. 情報収集

本研究では旅行中・旅行後の情報を必要とするため、旅行中でも随時情報を発信できる SNS と観光地の評価や旅行時の詳細が詳しく書かれている観光情報サイトを情報の収集元とする。

2.3. 有用情報の判定

収集した情報には信憑性や魅力が低いものも含まれている可能性があるため、投稿に対して有用であるかどうかの判別を行う必要がある。判別の方法としては、投稿した日にち、写真付きかどうか、位置情報は有効かどうかなどが上げられる。ここで、SNS から収集した情報で有用な投稿を 1 ヶ月以内の写真付き・位置情報有効の投稿と定義し、観光情報サイトから収集した情報で有用な投稿を評価 3 以上の投稿と定義する。

さらに、本研究では収集した日本語の投稿には MeCab[3]、中国語の投稿には SnowNLP[4]を用いて形態素解析を行い、旅行中・旅行後の判別を行う必要がある。この判別を行うためには、MeCab・SnowNLP が使用する独自の辞書を作り、特定のキーワードを抜き出す必要がある。例として、“行った”という言葉は旅行後、“今来た”という言葉は旅行中、“行きたい”という言葉は旅行前のことをそれぞれ表す。このように事前に言葉の定義をしておくことにより、判別が可能となる。

2.4. プロフィールの抽出手法

ここで 2.3 のプロセスで選別し残った投稿に対して、もう一度 MeCab・SnowNLP を用いて形態素解析を行う。このプロセスでは、投稿者のプロ

Travel destination recommendation system suitable for Foreign Independent Tourist's profile

† Minghao LI † Keizo YAMADA, † Masanori TAKAGI

† Jun SASAKI

† Faculty of Software and Information Science, Iwate Prefectural University

フィールドを抽出する. 判別方法としては, 2. 3 と同じく, MeCab・SnowNLP が使える独自の辞書を用いて, 言葉に対して定義を行う.

投稿本文	観光地名	投稿者情報	年齢	人数	出身
投稿1: 一人で小岩井農場に二十歳旅行してきました♥(≥▽≤)	小岩井農場	女性	20歳	単独	日本
投稿2: 高校最後の夏休みを利用して家族と岩手山に登ってきました!	岩手山	?	17~18歳	グループ	日本
投稿3: 我今天在不来方城跟女朋友求婚了	不来方城	男性	22歳以上	複数	中国

図2プロフィール抽出例

プロフィール抽出例として, 図2 投稿1 に対して旅行先を小岩井農場と抽出し, 投稿者のプロフィールを 20 歳日本人で一人旅行をしていると抽出できる. また, 顔文字を使っていることにより女性と推測する. 同様に投稿2 に対して旅行先を岩手山と抽出し, 投稿者のプロフィールを 17~18 歳の日本人で家族なので複数人での旅行と抽出できる. 投稿3 に対しても同様に SnowNLP を使い観光地とプロフィールを抽出できる. ただし, 今後の課題として, “二十歳の時に来たことがある” などの文章が出現したときに二十歳として判別しないようにする必要がある.

2.5. プロフィール項目の定量化

項目	点数	備考
性別	0/5	一致する場合5点、しない場合0点
年齢	0/2.5/5	±1歳5点、±3歳2.5点、以外0点
出身	0~5	一致する場合5点、しない場合0点
同行者数	0/2.5/5	人数が近ければ点数が高い

図3プロフィールの定量化例

ユーザのプロフィールと投稿者のプロフィールを比較して, 類似度を算出するためには, 比較するプロフィール項目を定量化しなくてはならない.

定量化の方法としては, 図2プロフィールの定量化例で示している通り, 各プロフィール項目には 0~5 までの点数が振り分けされる. ユーザと

投稿者のプロフィールを比較するときに, ユーザと投稿者のプロフィールが一致すればするほど, 点数が高くカウントされる. 具体的に, 性別は選択肢が男性や女性しかないのので, 一致すれば 5 点, 一致しなければ 0 点としてカウントする. 年齢は ±3 歳以内ならば 5 点, ±10 歳以内なら 2.5 点, それ以外は 0 点とカウントする.

3. おわりに

本研究は, 中国語圏 FIT が岩手県への観光時において, 自分のプロフィールにあった適切な観光地推薦が得られることを目的とし, FIT のプロフィールに合った旅行先推薦システムを提案してきた. 本システムはまだ研究途中であるため, 今後は具体的に, ユーザと投稿者の類似度をより正確に測れるように, 使用するプロフィール項目について検討していきたい. また, MeCab・SnowNLP を使用した形態素解析, プロフィール抽出を実現可能にするため, MeCab・SnowNLP が使用する辞書を作成していきたい.

4. 参考文献

[1] 岩手県ホームページ, 平成 29 年版岩手県観光統計概要について:

http://www.maff.go.jp/j/seisan/tyozyu/higai/h_zy_okyo2/h26/

[2] 神宮敏弘: 推薦システムのアルゴリズム:

<http://www.kamishima.net/archive/recsysdoc.pdf>

[3] MeCab:

Yet Another Part-of-Speech and Morphological Analyzer:
<http://taku910.github.io/mecab/>

[4] SnowNLP 中文语法分析:

<http://www.voidcn.com/article/p-sqedkrzz-bch.html>