

# SNS による FIT と GIT の分類手法の提案

喬 冠語<sup>†,a</sup>      廣田 雅春<sup>‡,b</sup>      荒木 徹也<sup>‡,c</sup>      遠藤 雅樹<sup>‡,d</sup>  
石川 博<sup>†,e</sup>

† 首都大学東京大学院システムデザイン研究科

‡ 岡山理科大学総合情報学部

‡ 群馬大学理工学部

‡ 職業能力開発短期大学校基盤ものづくり系

a) [qiao-guanyu@ed.tmu.ac.jp](mailto:qiao-guanyu@ed.tmu.ac.jp) b) [hirota@mis.ous.ac.jp](mailto:hirota@mis.ous.ac.jp)

c) [tetsuya.araki@gunma-u.ac.jp](mailto:tetsuya.araki@gunma-u.ac.jp) d) [endou@uitech.ac.jp](mailto:endou@uitech.ac.jp)

e) [ishikawa-hiroshi@tmu.ac.jp](mailto:ishikawa-hiroshi@tmu.ac.jp)

**概要** 近年、観光庁の調査によると、訪日外国人の旅行者の中でも、パッケージツアーなどによる団体旅行 Group Inclusive Tour (GIT) だけではなく、個人で旅行を行う Foreign Independent Travelers (FIT) での旅行形態が増加している。FIT 旅行者の情報収集は主に SNS やブログが中心である。FIT 旅行者の満足度を向上させるためには、観光に関する適切な情報を推薦することが望ましい。そのためには、旅行者の中から FIT 旅行者と GIT 旅行者を分類する必要がある。そこで、本研究では、Support Vector Machine (SVM) を用いて、Weibo の観光情報が FIT 旅行者による投稿か、GIT 旅行者による投稿かを分類する手法を提案する。

**キーワード** 観光, ユーザー分類, FIT, GIT, Weibo

## 1 はじめに

海外旅行に行く際、パッケージツアーを利用した GIT と呼ばれる旅行形態が一般的であった。しかし、海外旅行に慣れた人の場合や、旅行の時間を十分に確保可能な場合には、パッケージツアーを利用せずに、個人で旅行の手配を行う FIT と呼ばれる旅行形態が増加している。FIT の場合、旅行中のスケジュールを自由に組めるため、グループで旅行する GIT よりも個人の目的や嗜好に合わせた旅行が期待できる。

FIT 旅行者の情報収集は、SNS やブログなどの Web 上の情報が中心である。旅行を計画している人に対して、適切な情報を推薦することは、観光客を増加させるために重要である。しかし、これらの情報源には GIT 向けのものも含まれており、FIT 旅行者にとって必ずしも有益であるとは限らない。そのため、FIT 旅行者が有益な観光情報を取得するためには、観光情報の投稿が FIT 旅行者によるものか、GIT 旅行者によるものかを分類する必要がある。

本研究では、訪日中国人の旅行者に着目する。観光局<sup>1</sup>の発表では、中国からの FIT 旅行者の割合は、2012 年の 28.5% から、2018 年の 64.7% と大きく増加している。そのため、中国は、FIT 旅行者への旅行情報の推薦の需要が高い国のひとつであると考えられる。

本研究では、中国の主要な SNS のひとつである Weibo

の旅行情報が FIT 旅行者による投稿か GIT 旅行者による投稿であるかを分類する手法を提案する。

## 2 関連研究

橋田ら [1] は、投稿されたツイートから観光地に対して意見を含むツイートと含まないツイートの分類を行うために、畳み込みニューラルネットワークを用いたツイートの分類手法を提案している。

佐伯ら [2] は、ツイートの投稿時間や、付与された位置情報に着目することで、外国語を用いるユーザの属性が短期滞在外国人か、長期滞在外国人かを分類し、長期滞在外国人が日本での滞在を開始した日を推定する手法を提案している。

本研究では、研究 [1] のツイート分類手法と研究 [2] の長短期滞在外国人の分類手法を参考に、Weibo から取得した観光情報の分類を行った。

## 3 提案手法

### 3.1 データセット

本研究では、Weibo API<sup>2</sup>を用いて、日本国内で投稿のみ取得した。投稿には、ユーザ ID、テキスト、位置情報、画像、投稿時間などの情報が含まれる。

### 3.2 観光投稿の抽出

SNS に投稿されるデータには、観光とは無関係な投稿も含まれている。そのため、はじめに、観光に関する投稿のみを抽出する必要がある。

Copyright is held by the author(s).

The article has been published without reviewing.

<sup>1</sup>平成 30 年国地域別外国人行動特性調査報告書

<sup>2</sup><https://open.weibo.com/wiki/API>

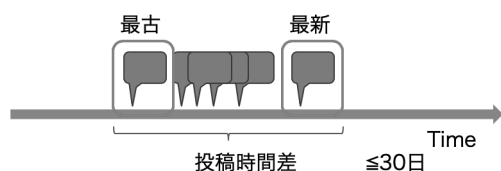


図1 投稿時間差の判定図

団体	ガイド	個人旅行	独り	東京	大阪	京都	奈良	富士山	北海道	他の観光地	投稿件数
----	-----	------	----	----	----	----	----	-----	-----	-------	------

図2 特徴量項目の例

本研究では観光を食事・景観・行動であると仮定して、それぞれのユーザの全投稿に対して、以下の四つの条件を満たした投稿を観光投稿として抽出した。

- ・「美味しい」「寿司」「食べた」などの食事に関する特徴を持つ投稿.
- ・「紅葉」「雨」「絶景」などの景観に関する特徴を持つ投稿.
- ・「登った」「買い物」「拝観できた」などの観光行動に関する特徴を持つ投稿.
- ・「図1」のように、設定した期間内の投稿時間が最も古い投稿と、最も新しい投稿の投稿時間差が30日以内のデータ.

在中国日本国大使館<sup>3</sup>によると、観光ビザの場合に一度の渡航で滞在可能な期間は30日である。そのため、投稿時間差が30日以内の場合は、観光による滞在が主であると考えられる。

### 3.3 分類器生成

本研究では、人手によって特徴量を生成した。生成した特徴量を以下に示す。

- ・FITを示す特徴として「個人旅行、独り」とGITを示す特徴として「団体、ガイド」をそれぞれを示す特徴量として扱った。
- ・中国で有名な旅行サイト mafengwo(注2)の調査を元に、GIT旅行者が頻繁に訪れる観光地を選定してから、GITを示す特徴量として「東京、大阪、京都、奈良、富士山、北海道」、その他の観光地名等は「他の観光地」として、FITを示す特徴量として扱った。
- ・在日期間内の投稿件数が多いユーザーはGITの傾向があることがデータセットを生成段階で判明したため、在日期間内の投稿件数を特徴量として扱った。

<sup>3</sup>[https://www.cn.emb-japan.go.jp/consular\\_j/joho170504-2-j.htm](https://www.cn.emb-japan.go.jp/consular_j/joho170504-2-j.htm)

本研究では、データセットの元に、生成した特徴量を用いてSVMの分類器を生成した。

## 4 実験

本研究では、685人のユーザーの2018年12月から2019年2月の3ヶ月分の観光投稿を使用して、実験を行った。

特徴量の項目例を図2で示す。

表1 実験データ

	GIT	FIT	判別不能	合計
train	39	133	307	479
test	17	70	119	206

表2 実験結果

	Precision
GIT	0.69
FIT	0.89
判別不能	0.73
平均値	0.78

分類を行う際、GITかFITのいずれの投稿であるかを人手で判別できなかった投稿に関しては、「判別不能」とした。SVMの他クラス分類手法 one-versus-rest で3クラス分類を行った。作成したデータを表11に示す。

特徴量を標準化して実験を行った。表22に実験結果を示す。

## 5 おわりに

本研究では、SVMを用いて、Weiboへの投稿がFITによる投稿かGITによる投稿かを分類する手法を提案した。実験結果として、提案手法を用いることでFITとGITへの分類が可能なることを示した。今後の課題として、特徴量の改善による分類性能の向上、別のSNSに対する実験を予定している。

### 参考文献

- [1] 橋田 修一, 田村 慶一, 酒井 達弘: 畳み込みニューラルネットワークを用いた観光ツイートの分類手法, The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018.
- [2] 佐伯 圭介, 遠藤 雅樹, 江原 遥, 廣田 雅春, 横山 昌平, 石川 博: 日本での滞在期間による長期滞在外国人の東京での訪問先の差異の分析, 研究報告データベースシステム (DBS), 2015-DBS-162, 15, 1-8, 2015-11-19.