

外国人観光客の相談相手となりうる
Twitter ユーザの推薦に関する研究

筑波大学
図書館情報メディア研究科
2016年3月
田中 匠

目次

第1章	はじめに	1
1.1	本研究の目的と意義	1
1.2	背景	1
1.2.1	Twitterについて	2
1.2.2	日本の観光の現状	2
1.3	論文の構成	4
第2章	関連研究	5
2.1	ジオタグ付きツイートを用いた研究	5
2.2	観光に関する研究	6
2.2.1	位置情報を用いた観光に関する研究	7
2.2.2	外国人観光客を対象とした研究	9
2.3	ユーザ推薦に関する研究	10
2.4	本研究の位置づけ	11
第3章	日本を訪れた経験のある 英語ユーザの収集	13
3.1	ユーザ推薦の手順	13
3.2	日本でつぶやいたTwitterユーザの収集	13
3.3	英語を用いるユーザの選別	14
3.4	推薦に適さないユーザの除去	15
3.4.1	在日米軍関係者	15
3.4.2	Bot	16
3.5	有用なユーザについての予備調査	16
3.5.1	目的	17
3.5.2	方法	17
3.5.3	判定対象	17

3.5.4	結果	17
3.5.5	考察	18
3.5.6	有用なユーザの基準	18
第4章	ユーザ推薦システムの構築	19
4.1	実装方法	19
4.2	地域の付与	20
4.2.1	地域の分割	21
4.2.2	地域情報の付与	22
4.3	システムにより推薦されたユーザのツイート例	22
4.4	システムの評価	23
4.4.1	目的	24
4.4.2	方法	24
4.4.3	結果	24
4.4.4	考察	24
第5章	実験：観光客の相談相手と なりうるユーザの推薦	27
5.1	提案手法の仕様	27
5.2	実験1：既存サービスに基づくユーザ検索手法との比較	28
5.2.1	目的	28
5.2.2	比較手法1：既存サービス（ツイート検索）に基づくユーザ検索	28
5.2.3	方法	29
5.2.4	結果	30
5.2.5	考察	31
5.3	実験2：従来の文書検索技術を利用したユーザ検索手法との比較	31
5.3.1	目的	31
5.3.2	提案手法の変更点	32
5.3.3	比較手法2：文書検索技術に基づくユーザ検索	33
5.3.4	方法	33
5.3.5	結果	34
5.3.6	考察	34

第6章 おわりに	38
6.1 まとめ	38
6.2 今後の課題	39
謝辞	41
参考文献	42
発表論文	45

表 目 次

2.1	中嶋ら [1] が旅行者のツイートの判定に用いた特徴	7
2.2	新井ら [2] が旅行者のツイートの判定に用いた特徴	8
2.3	佐伯ら [3] が分析した言語一覧	10
3.1	言語ランキング (単位: 万)	14
4.1	写真, チェックイン付きツイートの条件	20
4.2	地域の分布	22
4.3	関東地方の X さんのツイート	23
4.4	近畿地方の Y さんのツイート	23
4.5	推薦するユーザの推薦の精度	25
5.1	実験 1 の結果	30
5.2	実験 2 の結果	35
5.3	発見したユーザの評価の平均	36

目 次

1.1	日本への外国人訪問者数の推移	3
3.1	ユーザ推薦の手順	14
3.2	在日米軍に関わる地域でつぶやかれたツイートの割合とそのユーザ数 . .	16
4.1	日本の一般的な地域区分	21
5.1	提案手法のインタフェース	28
5.2	比較手法1のインタフェース	29
5.3	変更後の提案手法のインタフェース	32
5.4	比較手法2のインタフェース	33

第1章 はじめに

1.1 本研究の目的と意義

本研究では、英語を話す外国人観光客が日本を観光する際の相談相手となりうる Twitter ユーザの推薦を行う。外国人観光客は、推薦された Twitter ユーザの中から相談者を選び、Twitter¹上でコミュニケーションをとることで、情報収集や問題の解決を行う。

現在、観光についての情報収集や相談を行う場としては、TripAdvisor²や Yelp³といったような旅行口コミサイトがある。また、Google ローカルガイド⁴により、Google マップ⁵上で様々なスポットのレビューが投稿されている。これらのサービスでは、特定の観光地や場所についての様々な情報やレビューが投稿され、旅行の計画をたてる際の情報収集にとっても有用だといえるが、これらのサービスで扱っている場所についての情報収集・相談しかできない、リアルタイムな相談が難しいといったような問題がある。現地で気軽に情報収集を行えない外国人観光客にとって、この問題は非常に大きく、トラブルに繋がったり、満足のいく旅行ができない可能性がある。そこで本研究では、リアルタイムに自由な相談が可能な Twitter を相談に用いることを提案する。しかし Twitter のツイート検索機能だけでは、相談者となる Twitter ユーザを探しだすのは困難であるため、相談相手の候補を推薦することが本研究の目的となる。

1.2 背景

本研究の背景として、Twitter と日本の観光の現状について述べる。

¹<https://twitter.com/>

²<http://www.tripadvisor.jp/>

³<https://www.yelp.co.jp/>

⁴<https://www.google.com/intl/ja/local/guides/>

⁵<https://maps.google.co.jp/>

1.2.1 Twitter について

本研究では Twitter を利用する。Twitter とは、ツイートと呼ばれる 140 字の短文を投稿し、共有することで、不特定多数のユーザとコミュニケーションをとることのできるマイクロブログである。Twitter 公式ページで発表している情報⁶によると、2015 年 9 月での Twitter の月間アクティブユーザ数は 3 億 2000 万人と非常に多く、多種多様な趣味や目的をもったユーザがいるため、求められる情報を発信するユーザの推薦を行う研究は多く行われている。

Twitter の利用については、PC やスマートフォンから投稿、閲覧が可能となっており、特にスマートフォンを用いた Twitter の利用は全体の 80 % にのぼり、活発に行われている。また、Twitter ではツイートに緯度経度からなる位置情報を付与することができる。この位置情報が付いたツイートは、ジオタグ付きツイートと呼ばれ、特に外出先や旅行先で、写真を付けて投稿されることが多い。さらに、このジオタグ付きツイートは、Instagram⁷ や Foursquare⁸ などのサービスと連携したものも多くなっており、これらのツイートをを用いて、観光スポットの推薦など、観光に関する研究が活発に行われている。ここで、Instagram とは画像共有を目的とした SNS であり、前述のような写真付きツイートを行う方法のひとつである。また、Foursquare とは位置情報を用いた SNS であり、ベニューと呼ばれる特定の場所を訪れた際に自分が訪れたスポット情報を発信する（チェックインする）ことで、訪れた場所を友人などと共有できるサービスである。

その他にも、Twitter を用いた研究は多岐にわたって行われている [4]。

1.2.2 日本の観光の現状

近年、インバウンドという言葉が話題になっている。インバウンドとは、日本を訪れる外国人観光客を指すが、日本はこのインバウンドが日本人海外旅行者（アウトバウンド）よりも大幅に少ないことから、2003 年にビジット・ジャパン・キャンペーン⁹を打ち出した。その結果、図 1.1 のように、2005 年はリーマン・ショック、2011 年は東日本大震災で落ち込んではいるが、日本を訪れる外国人観光客は増加傾向にある。また、さらに 2020 年に開催予定の東京オリンピックでは、これまでにないほどの規模で外国からの観光客が日本を訪れることが予想されている。

⁶<https://about.twitter.com/ja/company>

⁷<https://instagram.com/>

⁸<https://foursquare.com>

⁹<http://www.jnto.go.jp/jpn/services/coop-vjc.html>

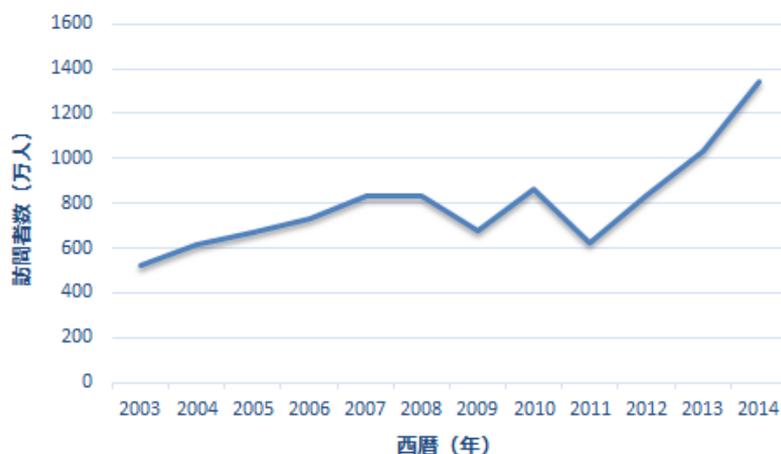


図 1.1: 日本への外国人訪問者数の推移

そのような状況から，日本はさらに外国人観光客を呼びこもうと力を入れているが，2014年の世界各国・地域への外国人訪問者数ランキング¹⁰によると，日本は未だに世界で22位，アジアでは7位となっている．この理由のひとつとして，日本人の英語力が低いことが考えられる．EF 英語能力指数¹¹によると，日本は英語能力ランキングで26位となっており，日本では英語が通じない場面がまだまだ多いことが，外国人が観光を行う際の障壁となっている．実際に，原の研究 [5] では，日本を訪れた外国人観光客に対するアンケート調査や外国人観光客のGPSログの解析を行っており，その中の，旅行情報サイト TripAdvisor の口コミ分析では，外国人観光客の口コミの否定的意見の多くに「(日本の観光を行う際の) 日本語の必要性」があったという結果を得ている．

本研究では，以上のような外国人観光客が日本を観光する際の課題を解決する方法として，英語を使うことのできる相談役を提供することを提案する．その相談相手として Twitter ユーザを推薦し，相談の場としてコミュニケーションが容易な Twitter を利用することで，顔見知りでないユーザにも気軽に相談を持ちかけることが可能だと考えられる．Twitter ユーザの推薦には，ジオタグ付きツイートを用いることとし，各 Twitter ユーザのツイートを行なった場所の分布や写真を添付したツイートの頻度，チェックインを行なったツイートの頻度などを利用する．

¹⁰https://www.jnto.go.jp/jpn/reference/tourism_data/visitor_statistics.html

¹¹<http://www.efjapan.co.jp/epi/>

1.3 論文の構成

本論文の構成を以下に示す．2章では，観光やユーザ推薦に関する研究について述べる．3章では，ユーザ推薦を行うまでの手順，日本を訪れた経験のある英語ユーザの収集，有用なユーザについての予備調査の結果について述べる．4章では，地域別ユーザ推薦システムの実装とその評価結果について述べ，5章では，被験者によるユーザ推薦実験について述べる．最後に，6章でまとめを述べる．

第2章 関連研究

本章では、本研究と関連が強いジオタグ付きツイートを用いた研究や、観光に関する研究、ユーザ推薦に関する関連研究について述べ、本研究がどのような位置づけとなるかを述べる。

2.1 ジオタグ付きツイートを用いた研究

ジオタグ付きツイートは、どこでツイートが行われたかが判定できるため非常に有益な情報といえる。このジオタグ付きツイートを分析することでユーザのアクティビティの推測を行ったり、ツイート文と位置情報を合わせることでツイートした場所についての評判や口コミの収集をすることができる。このようなジオタグ付きツイートを用いて観光に関する研究も盛んに行われているが、これについては次節で述べる。

ソーシャルメディアの分析を行う Semiocast 社の調査 [6] によると、2010年6月22日の24時間における全世界のツイート中のジオタグ付きツイートの割合は0.6%と、非常に低い割合となっている。この結果に対し橋本ら [7] は2種類の方法でツイートの収集を行い、ジオタグ付きツイートについて分析を行っている。

Twitter には、ツイートを収集する方法として大きく分けてクエリの条件に合うツイートを現在から過去に遡って取得を行う REST API¹²と、収集を開始した時点からツイートをリアルタイムに取得を行う Streaming API¹³の2つの方法がある。この Streaming API にはさらに、以下の3つの方法に分けられる。

- Public stream : Twitter の全ツイートが対象
- User streams : 指定した1ユーザのツイートが対象
- Site streams : 指定した複数ユーザのツイートが対象 (利用するには Twitter 社の許可が必要)

¹²<https://dev.twitter.com/rest/public>

¹³<https://dev.twitter.com/streaming/overview>

そしてさらに，Public stream は以下の3つの方法に分けられる．

- filter : 指定したクエリの条件に合うツイートを取得
- sample : Twitter の全ツイートの中からサンプリングした数%のツイートを取得
- firehose : Twitter の全ツイートを取得 (利用するには Twitter 社の許可が必要)

橋本ら [7] はこの filter と sample を使い，2011年7月15日から261日間のツイートをを用いてジオタグ付きツイートの割合を調査した．その結果，sample で取得したツイートのうちの，言語判定が日本語となっているツイートをを用いた結果，ジオタグ付きツイートの割合は0.18%と報告している．さらに，filter を用いて，日本国土全体をカバーする複数のバウンディングボックスを指定して日本で行われたジオタグ付きツイートの収集を行なった結果を，先ほどのジオタグ付きツイートの割合0.0018で割ると，おおむね全世界のツイート数に日本人のユーザ比率を掛けた値と近い値を示したと報告している．

このように，ジオタグ付きツイートの割合は非常に少ないことが分かっているため，ジオタグが付いていないツイートの発信位置を推定する研究も行われている．Chengらの研究 [8] では，大量のジオタグ付きツイートを分析して抽出した地域ごとの特徴語をもとに，ジオタグを付けずにツイートを行うユーザの地域推定を行なっている．さらに，ユーザが絶えず移動することを考慮して，上村らの研究 [9] では，ユーザごとの地域推定ではなくジオタグの付いていないツイートごとに発信地域の推定を行なっている．上村らは，地域ごとに特徴語を出すだけでなく，特徴語の時間変化に着目してリアルタイムに特徴語を更新していく手法を提案した．

本研究では，橋本らと同様に filter を利用し，日本国土全体をカバーするバウンディングボックスを用いて日本で行われたジオタグ付きツイートの収集を行う．また，相談者がどの地域でよくツイートをする，どの地域についての観光を行なった経験があるのかということを得るためにジオタグ付きツイートをを用いているが，相談者として適しているのかということ趣旨とするため，本研究ではジオタグが付いていないツイートの地域推定は行わないこととする．

2.2 観光に関する研究

観光に関する研究で最も盛んなものとしては，観光スポット，観光ルートの推薦がある．そのなかでも，位置情報を用いた研究は非常に多い．また，本研究は外国人観光客

を対象としている．ここでは，位置情報を用いた観光に関する研究と，外国人観光客を対象とした研究を紹介する．

2.2.1 位置情報を用いた観光に関する研究

観光スポットや観光ルートを推薦する研究にも，色々なものがある．中嶋らの研究 [1] では，Twitter の REST API の位置情報検索を用いて任意の名所付近で行われたツイートを収集，ツイートに現れる特徴と Foursquare や Instagram のサービスから旅行者のツイートを判定，それらのツイートを「食事」「景観」「行動」の3つに分類し，その分類をもとに旅行者の好みに合わせた観光ルートの推薦を行っている．中嶋らの研究で行われた旅行者のツイートの判定に用いた特徴は，表 2.1 の通りである．

表 2.1: 中嶋ら [1] が旅行者のツイートの判定に用いた特徴

特徴	ツイート例
スポット + なう	「スカイツリーなう」
スポット + 画像 URL	「清水寺の様子 pic.twitter.com/...」
「I'm at」 + スポット	「I'm at 東京スカイツリー ...」
「@」 + スポット	「... @ 東京スカイツリー」
instagram.com を含む URL	「清水寺に参拝してきました instagram.com/...」

また，新井らの研究 [2] でも同様にツイートを用いた観光ルートの推薦を行っているが，あらかじめ観光スポットを収集しておくことで，ジオタグ付きツイートだけでなくすべてのツイートを用いているという点が異なっている．新井らは観光スポットの収集に，Google Places API¹⁴のプレイス検索を用いて指定した地点周辺の観光スポット候補を集め，Yahoo!知恵袋の質問検索 API¹⁵で観光スポット候補名を含む質問数がどれくらいあるかにより観光スポットとするかどうか決定している．収集した観光スポット名を含むツイートを検索し，ジオタグ付きでないツイートも利用したことで，収集できたツイートは増えていたと報告している．また，収集したツイートの分類は新井らの研究に基づき「食事」「景観」「行動」「土産」の4つに分類を行なっている．新井らの研究で行われた旅行者のツイートの判定に用いた特徴は，表 2.2 の通りである．

本研究で推薦する相談者を決める際の手がかりのひとつとして，旅行を頻繁に行っていることが挙げられる．それに際し，旅行を行なっているツイートの判定は，表 2.1，

¹⁴<https://developers.google.com/places/>

¹⁵<http://developer.yahoo.co.jp/webapi/chiebukuro/chiebukuro/v1/questionsearch.html>

表 2.2: 新井ら [2] が旅行者のツイートの判定に用いた特徴

URL 中に “swarmapp.com” または “4sq.com” を含む
URL 中に “tou.ch” を含む
URL 中に “instagram.com” を含む
URL 中に “.jpg” または “.png” を含む
観光スポットを訪れたとみなせる表現が存在する

表 2.2 を参考に行う．ただし，現在ロケタッチ (tou.ch) はサービスを終了しており¹⁶，Foursquare (4sq.com) のチェックインサービスは完全に Swarm に移行されている．具体的な方法については，4章で述べる．

奥らの研究 [10] では，ツイートがどのスポットについて言及しているのか不明確なことを解決するために，あらかじめジオタグ付きツイートと観光スポットを対応付けしてから，観光スポットの推薦を行っている．

奥らは Foursquare から収集した観光スポットと，写真共有サイト Panoramio の API¹⁷ を利用して取得したスポット名をタイトルに持つ写真を用いて，ツイートと観光スポットを対応付けを行なっている．また，この対応付けには，OC-SVM[11] と呼ばれる領域判別問題の解法の一つを用いている．

観光スポットを推薦する研究すべてが，あらかじめ決められた観光スポットの中から推薦する場所を決めているわけではない．前田らの研究 [12] では，前述のような観光スポットとツイートを対応付けて観光スポットの評価を行う研究と大きく異なり，Twitter ユーザのジオタグ付きツイートの分布から観光地の抽出を行なっている．さらに，各ユーザの移動数や移動距離の計算結果からその観光地の魅力の大きさを，その地点でツイートされたツイート文の TF-IDF の分布からその観光地が他にはない魅力があるかどうかを分析している．前田らは最終的に，話題に固有性があり他にはない大きな魅力を持った観光地や，話題の固有性は低いが魅力はある便利な場所，話題の固有性は高いが魅力は低い経由地となっている場所などの抽出を行なった．

過去の行動履歴から訪れる場所を推薦する研究もある．Kurashima らの研究 [13] では，トピックモデルによるユーザの興味に基づく推薦に，ユーザの行動範囲を考慮する条件を加えた，ジオトピックモデルを提案している．このジオトピックモデルにより，過去の行動履歴にもとづいてユーザが次に訪れる場所を推薦を行なった．同様に，Chen ら

¹⁶<http://blog.tou.ch/>

¹⁷<http://www.panoramio.com/api/data/api.html>

の研究 [14] では、Twitter のような中国のマイクロブログ SINA¹⁸ を用いた場所の推薦を行っている。Chen らは、意味階層カテゴリグラフを用いて、ユーザの興味に基づいた類似ユーザから未訪問の場所を推薦する手法を提案している。

2.2.2 外国人観光客を対象とした研究

1.2.2 節で述べたように、日本の観光において外国人観光客の呼びこみは重要な課題となっている。村上らの研究 [15] では、英語ブログを分析することで外国人から見た日本の旅行のイメージを調べている。村上らが利用したのは、IgoUgo¹⁹、TravelBlog²⁰、TravelPod²¹ の 3 つのサービスで、北海道、東京、石川について、レビューの出現頻度の高い語の分析や手作業の分析を行っている。その結果、3 箇所とも全体的に好印象なレビューが多かったと報告している。しかし、分析したブログの記事件数はすべて合わせても 100 記事程度で、データが足りていないと考えられる。

本研究と同様に、Twitter のジオタグ付きツイートを用いて外国人の行動を分析している研究としては、佐伯らの研究 [3][16][17] がある。佐伯らは、表 2.3 のように、インターネット上で多く利用される 10 言語から日本語を除いた 9 言語を分析対象として、場所ごとの使用言語の違いや訪問順序を調査し、おすすめの観光地や観光ルートを推薦している [3]。

また、ユーザに付与された言語情報とツイートの言語情報から外国人ユーザを抽出し、さらにツイート頻度を利用して、外国人ユーザを在日外国人ユーザと訪日外国人ユーザに分類。分類した日本人ユーザ、在日外国人ユーザ、訪日外国人ユーザについて、特定の観光スポットを訪問する頻度の違いを分析している [16]。その結果、在日外国人ユーザには米軍関係者と思われる特徴が現れたという結果を得ている。ここで、在日外国人ユーザと訪日外国人ユーザの分類には、一定期間ツイートが無い外国人ユーザを訪日外国人ユーザとする手法と、分析する期間を分割し、分割した区間の間にツイートがあった区間がいくつあるかによって訪日外国人ユーザか、在日外国人ユーザかの判断を行う手法の 2 種類を用いており、後者は 1 ユーザあたりの実行時間が短くなるが、前者の方が精度が少し高くなったと報告している。

さらに、在日外国人ユーザの滞在開始日を推定し、滞在期間別に東京の任意のエリアを訪れた割合を調査することで、それぞれの特徴について考察を行っている [17]。その

¹⁸<http://overseas.weibo.com/>

¹⁹既にサービスを終了

²⁰<https://www.travelblog.org/>

²¹<http://www.travelpod.com/>

表 2.3: 佐伯ら [3] が分析した言語一覧

言語名	略称
英語	en
中国語簡体字	zh-cn
中国語繁体字	zh-tw
スペイン語	es
ポルトガル語	pt
ドイツ語	de
アラビア語	ar
フランス語	fr
ロシア語	ru
韓国語	ko

結果，来日してから日が浅い在日外国人ユーザは訪日外国人ユーザと近い傾向にあるが，来日から時間が経過するにつれて訪日外国人ユーザとの差異が生じる傾向があることが確認された。

佐伯らの研究では，訪日外国人ユーザの訪問先と日本人ユーザ，在日外国人ユーザの訪問先に差異があることを示したが，日本人ユーザや在日外国人ユーザが外国人観光客の相談役として適さないわけではないと考えられる．そのため本研究では，日本人ユーザ，在日外国人ユーザ，訪日外国人ユーザの分類は行わない。

2.3 ユーザ推薦に関する研究

ユーザ推薦は，非常に多くのユーザ間でコミュニケーションが行われる Twitter で非常によく行われている研究である。

田島らの研究 [18] では，何らかのイベントに参加した時など，短期間だけ持続的に情報収集をする際に，同じような状況にあるユーザを入力したクエリに関する情報を発信する可能性のある Twitter ユーザ（レポータと呼ぶ）として推薦を行なっている．推薦された Twitter ユーザを短期的にフォローすることによって，同じ状況にあるレポータから情報が得られる．情報量や有益性が保証されていない従来のハッシュタグを用いた情報収集方法と比べ，この研究の手法が優れていることを示した．本研究も，旅行前もしくは旅行中の短期的な情報収集を行うためのユーザ推薦であるが，推薦後にコミュニケーションを取ることが前提であるという点で異なる。

稲垣らの研究 [19] では、入力したクエリを含むブログエントリを投稿したブログを抽出し、それらのブログの特徴を用いてブログのランキングを行うシステムを提案している。この研究は、正確にはユーザ推薦を行う研究ではないが、クエリを含むブログのランキングを行う過程でブログの熟知度のランキングを用いており、本研究における、クエリに関連した日本の観光に対する熟知度の高いユーザの推薦ととても近いねらいとなっている。

稲垣らはあらかじめ以下の様な作業を行っており、これらを用いてブログのランキングシステムを実現した。

1. ブログでよく言及されるトピックを抽出
2. トピックごとに、そのトピックについて熟知しているブログの集合（熟知グループ）を作成
3. 各熟知グループのブログに対して熟知度スコアを算出

稲垣らの研究は、あらかじめブログに対して熟知度スコアを算出している点において、外国人観光客の相談相手として適しているかを測るために、あらかじめ複数の手がかりを用いて Twitter ユーザの評価値の算出を行なっている本研究とアプローチが類似している。しかし、本研究では、相談内容を特定の観光のジャンルやスポットに限定させないために、あらかじめ特定の観光のジャンルやスポットごとにユーザの評価値を算出することは行わない。

2.4 本研究の位置づけ

本研究では、外国人観光客が日本の観光を行う際の相談相手の推薦を行う。

日本人であれば日本語を話すことができるため、基本的にその場、もしくはその周辺でコミュニケーションをとって容易に情報を集めることが可能だが、日本語がわからない外国人にとっては、レビューサイトに乗らないような現地の詳細な情報を集めたり、とっさに旅行中のトラブルに対応することは難しい。

推薦する相談相手として Twitter ユーザを推薦し、相談を Twitter 上で行うことで、旅行口コミサイトやレビューサイトでは難しかった、自由な話題でリアルタイムな相談が可能となる。このように、最終的に外国人観光客は推薦された Twitter ユーザの発信するツイートから情報を得るだけでなく、推薦された Twitter ユーザと Twitter 上でコミュ

ニケーションを行うことで、日本を観光する際の短期的な問題の解決を行うことが本研究のねらいとなる。ユーザ推薦に用いるデータはジオタグ付きツイートとし、旅行者のツイート判定は、中島ら [1] と新井ら [2] が旅行者のツイートの判定に用いた特徴を参考に行う。また、外国人の観光に有用かどうかという観点でユーザを評価するため、日本人ユーザ、在日外国人ユーザ、訪日外国人ユーザの分類は行わないこととする。

第3章 日本を訪れた経験のある 英語ユーザの収集

本章では、ユーザ推薦を行う前段階として行なった、日本でつぶやいた Twitter ユーザの収集、英語を用いるユーザの選別、推薦に適さないユーザの除去について、および、有用なユーザについての予備調査について説明する。

3.1 ユーザ推薦の手順

本研究で、外国人の観光に有用な Twitter ユーザを推薦する手順は、以下の通りである。また、その流れを図 3.1 に示す。

1. Twitter から日本で行われたツイートを収集し、そのデータから日本でつぶやいた Twitter ユーザを収集する。
2. 日本でつぶやいた Twitter ユーザの言語判定を行い、英語を用いるユーザを選別する。
3. 英語を用いるユーザから、推薦に適さないユーザを除去し、残ったユーザについて有用性判定を行って有用なユーザを推薦する。

以下では、それぞれの処理について説明する。ただし、有用なユーザの推薦については 4 章で述べる。

3.2 日本でつぶやいた Twitter ユーザの収集

日本でつぶやいた Twitter ユーザの収集には、日本国内で行われた位置情報付きツイートのみを用いる。収集には、Twitter の非公式ライブラリである Twitter4j²²を用い

²²<http://twitter4j.org/>

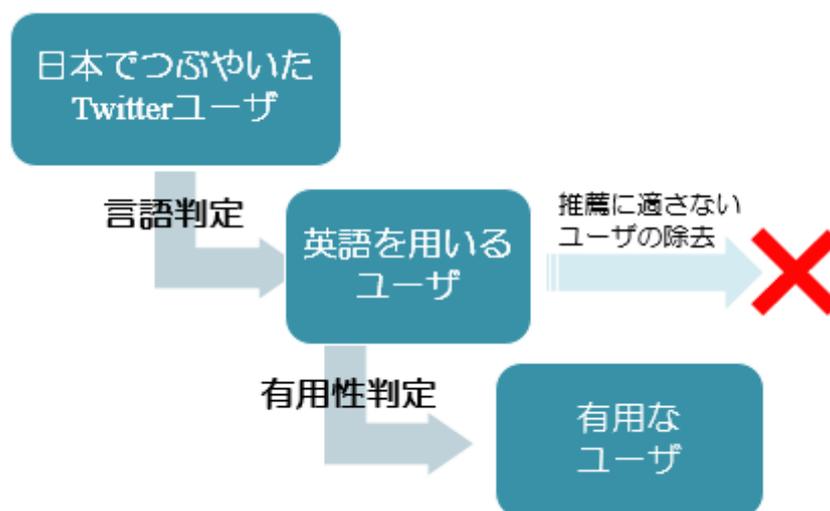


図 3.1: ユーザ推薦の手順

て、Twitter Streaming API により 2014 年 7 月 21 日から 2015 年 7 月 30 日の期間で約 1.1 億ツイート、約 150 万ユーザを取得した。

3.3 英語を用いるユーザの選別

Twitter では、ユーザとツイートに対して自動で言語情報が付与されており、この言語情報を用いて、日本でつぶやいた Twitter ユーザから英語を用いるユーザを選別する。そこでまず、日本でつぶかれたツイートの言語情報において、英語と日本語がどれくらいあるのか調査を行った。収集したツイートとユーザに付与されていた言語のランキングの 3 位までを表 3.1 に示す。ただし、この調査を行った段階では、ツイート数は約 1 億、ユーザ数は約 140 万となっている。

表 3.1: 言語ランキング (単位: 万)

	ツイート	ユーザ
total	10196	140
1	ja(9840)	ja(129)
2	und(126)	en(9)
3	en(115)	-

また、ツイートの付与された言語の und とは言語不明となっているもので、URL や UserName のみのツイート内容のものに付与されていた。やはり、日本語の次に多い言

語は英語となっており、言語が英語だと判定されたユーザの割合は全体の 6.4% を占めている。

そこで、英語を用いるユーザの条件を、Twitter ユーザに対する言語判定が英語かつ、日本における英語でのつぶやきを N ツイート²³以上行なっているユーザとする。

3.4 推薦に適さないユーザの除去

推薦に適さないユーザとして、在日米軍関係者と Bot の除去を行う。Twitter における Bot とは、人が考えてツイートを行うのではなく、設定された規則に従って自動化されたツイートを継続的に行うアカウントのことである。

以下では、それぞれを除去する理由と、その方法について述べる。

3.4.1 在日米軍関係者

在日外国人ユーザには、米軍関係者と思われる特徴が現れたという関連研究の結果から、日本で英語を用いるユーザに在日米軍関係者が非常に多く含まれている可能性が高い。また、在日米軍関係者は通常の外国人観光客とは境遇が大きく異なるため、相談役には適さない可能性がある。そこで、在日米軍関係者と思われるユーザを抽出し、その特徴について以下の手順で調査を行った。結果を図 3.2 に示す。

1. 英語を用いるユーザから、つぶやいた英語ツイート数の上位 1,000 ユーザを抽出。
2. 防衛白書²⁴を参考に、在日米軍に関わる地域を抽出。
3. 抽出した各ユーザについて、ユーザがつぶやいた全ツイートの内、在日米軍に関わる地域でつぶやかれたツイートの割合を調査。

図 3.2 より、在日米軍関係者と思われるユーザとそうでないユーザとが明確に区別でき、在日米軍関係者と思われるユーザは、ツイートの大部分が在日米軍に関わる地域からであることが分かる。また、それらのユーザのツイートを見ると、ほとんどが仕事や日常についてのつぶやきであり、観光に有用なツイートが非常に少ないことが分かる。そこで、在日米軍に関わる地域で行われたツイートの割合が 9 割以上のユーザを、在日米軍関係者とみなし、推薦に適さないユーザとして除去する。

²³調査に基づき $N = 10$ とした。

²⁴<http://www.mod.go.jp/j/publication/wp/>

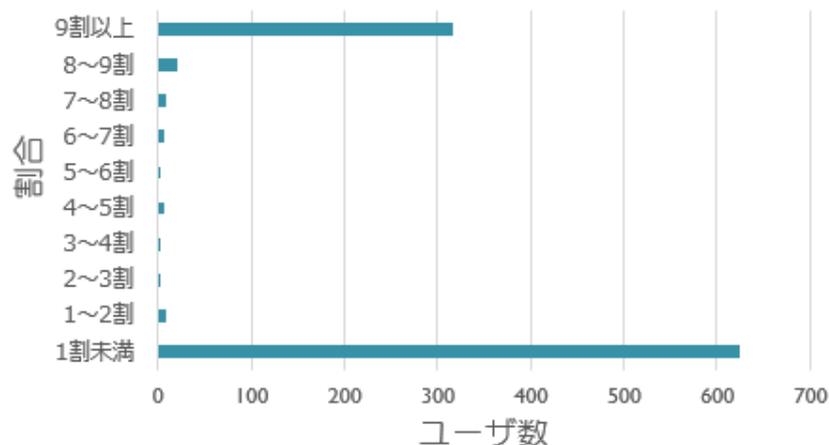


図 3.2: 在日米軍に関わる地域でつぶやかれたツイートの割合とそのユーザ数

3.4.2 Bot

Bot は機械的にツイートを行うだけで、基本的にコミュニケーションをとることはできない。さらに、観光に有用なツイートも行っていないため、推薦に適さないユーザとして除去する。方法として、以下の調査に基づき、ツイート場所が `null` の割合が 4 割以上のユーザを、推薦に適さないユーザとして除去する。

1. 英語を用いるユーザから、つぶやいた英語ツイート数の上位 1,000 ユーザを抽出。
2. 抽出した各ユーザの全ツイートの内、ツイート場所が `null` のツイートの割合を調査。

さらに、ユーザ ID が以下の条件に合うユーザについて、機械的に同じようなツイートを繰り返すユーザであることが分かったため、除去を行う。

- 最初が「@wind_」
- 最後が「_bot」、「_game」
- 「jobs」、「oozora」、「oracle」を含む

3.5 有用なユーザについての予備調査

外国人の観光に有用な Twitter ユーザとはどのようなユーザなのか、まずは人手で判断を行った。

3.5.1 目的

外国人の観光に有用な Twitter ユーザの判定に一貫性があるか検証する。また、日本人と外国人が、判定を行った場合の差について調査を行う。

3.5.2 方法

各 Twitter ユーザのツイートを見て、外国人の観光のための有用性をユーザ単位で評価する。評価基準は、「非常に有用」、「有用」、「有用でない」、「全く有用でない」の4段階である。被験者は日本人の第一著者と外国人の留学生の2名である。

3.5.3 判定対象

判定対象は、日本を訪れたことのある英語ユーザから、推薦に適さないユーザを除いたユーザの内の、英語のツイート数の上位70ユーザとなり、英語もしくは日本語ツイートを1ユーザあたり最大300件提示して判断する。ただし、書き出しが以下のツイートを除く。

- 会話文：「@」
- リツイート：「RT」
- 引用：「”@」(半角)、「”@」(全角)

3.5.4 結果

4段階評価における Cohen の 値 [20] は0.34となり、Landisらの基準 [21] によると軽度の一致 (Fair) となった。また、「有用」と「非常に有用」、「有用でない」と「全く有用でない」をまとめた2段階評価の 値は0.74となり、かなりの一致 (substantial) となった。つまり、4段階の判定には個人差があるが、それぞれをまとめた2段階の判定の相関は高く、判定基準に一貫性があることが確かめられた。

3.5.5 考察

4段階評価に差が出た理由として、チェーン店と個人経営の店の区別がつかないなど、日本人と外国人の留学生との間に日本の地名やスポットの理解度や重要度の違いがあったことが考えられる。

3.5.6 有用なユーザの基準

有用なユーザについての予備調査をもとに作成した、有用なユーザの基準は以下の4つとなる。

- A) 観光についてつぶやく頻度が高い
 - B) 観光についての有益な情報をつぶやく
 - C) 多くの場所に行っている
 - D) 日本の滞在期間が長い
-

第4章 ユーザ推薦システムの構築

本章では、ユーザ推薦システムの実装方法と、ユーザ推薦時に利用するユーザの地域情報の付与方法、システムの評価を行った結果について述べる。

4.1 実装方法

日本を訪れたことのある英語ユーザから推薦に適さないユーザを除いたユーザについて、有用性判定を行い、ユーザを推薦する。有用性判定では、予備調査をもとに作成した判定基準を以下の手がかりを用いて判定する。

- A) 写真，チェックイン付きツイートの頻度
- B) ツイートの平均語数
- C) ツイートを行なった場所の数
- D) 日本の滞在期間

各手がかりを 0,1,2,3 の 4 段階で評価し、その評価値の合計 (0~9) をユーザの評価値とする。そして、評価値のランキングを行い、上位のユーザを推薦する。

A では、ユーザが日本で行った直近 100 件の英語か日本語ツイートについて、写真，チェックイン付きツイートの頻度が多いほど高評価となる。この手がかりでは、写真，チェックイン付きツイートは旅行時のツイートであるという仮定に基づき、ユーザがどれくらい旅行を行なっているかを測る。写真，チェックイン付きツイートの判定には、4.1 の条件を用いる。swarmapp.com が含まれるものは、チェックインを行なっていることを示し、twitter.com/[ユーザ ID] が含まれるものは、そのユーザが twitter に写真や動画をアップロードしたことを示す。また、instagram.com が含まれるものは、instagram に投稿した写真と連携を行なったことを示す。ここで、Twitter では <http://t.co/...> のように、URL が短縮されていることが多いため、元の URL に復元してから判定を行なっている。

表 4.1: 写真, チェックイン付きツイートの条件

URL 中に “swarmapp.com” を含む
URL 中に “twitter.com/ [ユーザー ID]” を含む
URL 中に “instagram.com” を含む

B では, ユーザが日本で行った写真, チェックイン付きの英語ツイート, 最大 100 ツイートについて, 空白文字で分割し, 平均ツイート語数をカウントする. URL など 1 語とカウントし, 平均ツイート語数が多いほど高評価となる. この手がかりでは, 平均ツイート語数が長いほど, ツイートに観光についての有益な情報が含まれている可能性が高いという仮定に基づき, ユーザの旅行時のツイートが有益なツイートかどうかを測る.

C では, ユーザが日本で行なった英語か日本語のツイートの場所数が多いほど高評価となる. ただし, ツイート場所が NULL の地域を除く. また, 場所の粒度は市, 区単位とする. この手がかりでは, ユーザが様々な場所を訪れ, ツイートを行なった場所数が多いほど, 観光に興味があり観光スポットについてよく知っているという仮定に基づき, ユーザの観光への興味やその知識量を測る.

D では, データの収集期間を分割し, ユーザがツイートをを行なった区間が多いほど高評価となる. ここで, 利用するデータの収集期間が約 12ヶ月であるため約 1 月ごとの 12 分割とする. この手がかりでは, 日本を訪れた期間が長ければ長いほど, 日本の様々なことについてよく知っているという仮定に基づき, 日本についての知識量を測る. ここで, たとえば 3 区間ツイートがあるとき, 約 3ヶ月にまたがって日本に居続けた場合と何度も日本を訪れた結果 3 区間となった場合が考えられる. 約 3ヶ月にまたがって日本に居続けた場合のほうが日本にいる期間は長いから, 経験として蓄積される知識はこちらの方が多いが, 何度も日本を訪れた結果 3 区間となった場合は, 約 3ヶ月にまたがって日本に居続けた場合よりも日本に興味があるため何度も訪れているということが考えられ, 観光についての情報を多く持っている可能性がある. そこで, 今回はこれらの判別は行わないこととする.

4.2 地域の付与

本研究では, 外国人の観光に有用なユーザを, 主にツイートしている地域ごとに分けることで, 求める地域に詳しいユーザを推薦する. そのために, ユーザに主にツイート

している地域情報の付与を行う。

4.2.1 地域の分割

付与する地域は、図 4.1 のような日本で一般的な地域区分²⁵を参考に決定した。北海道地方と東北地方をひとつに、中国地方と四国地方をひとつに、九州地方と沖縄地方をひとつにすることで、地域該当なしを含めると、付与する地域情報は以下の7種類となった。

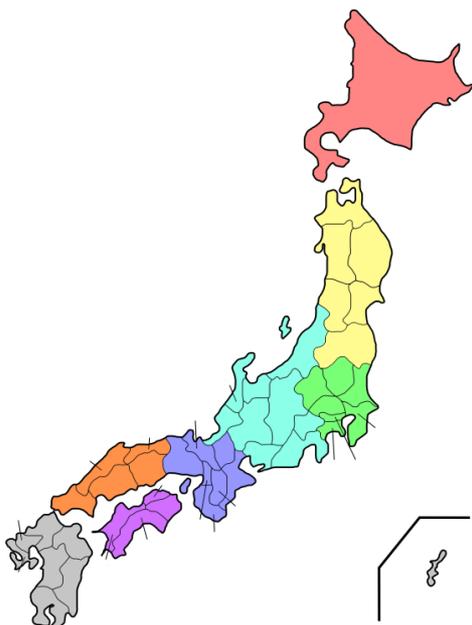


図 4.1: 日本の一般的な地域区分

- 北海道・東北地方：北海道，青森県，岩手県，宮城県，秋田県，山形県，福島県
- 関東地方：茨城県，栃木県，群馬県，埼玉県，千葉県，東京都，神奈川県
- 中部地方：新潟県，富山県，石川県，福井県，山梨県，長野県，岐阜県，静岡県，愛知県
- 近畿地方：三重県，滋賀県，京都府，大阪府，兵庫県，奈良県，和歌山県
- 中国・四国地方：鳥取県，島根県，岡山県，広島県，山口県，徳島県，香川県，愛媛県，高知県

²⁵<https://ja.wikipedia.org/wiki/都道府県>

- 九州・沖縄地方：福岡県，佐賀県，長崎県，熊本県，大分県，宮崎県，鹿児島県，沖縄県
- 該当なし

4.2.2 地域情報の付与

各ユーザについて，ユーザの日本での全つぶやきの半分以上を占めている地域を，地域情報として付与する．ただし，つぶやいた地域が分散して，全つぶやきの半分以上を占めている地域が無いユーザについては「該当なし」を付与する．具体的な方法としては，逆ジオコーディングサービス²⁶を利用して，各ユーザのツイートについているジオタグから都道府県名を特定，地域の特定を行なった．

外国人の観光に有用かどうかのユーザに対する評価値が8以上，つまり4つの手がかりの評価の平均が2以上となった529ユーザに地域情報を付与した結果，表4.2の様な分布となった．

表 4.2: 地域の分布

地方	ユーザ数
北海道・東北地方	14
関東地方	348
中部地方	39
近畿地方	61
中国・四国地方	6
九州・沖縄地方	31
該当なし	30

評価値が0～12となるなか，各地域において，評価値が8以上となった有用と考えられるユーザを抽出することができた．しかし，その6割以上が関東地方のユーザとなり，北海道・東北地方と中国・四国地方のユーザが非常に少ない結果となった．

4.3 システムにより推薦されたユーザのツイート例

構築したユーザ推薦システムにより上位にランキングされたユーザのツイート例を表4.3，表4.4に示す．ただし，ツイート中のユーザ名とURLは@***と[URL]に置き換え

²⁶<http://www.finds.jp/wsdocs/rgeocode/index.html> ja

ている。

表 4.3: 関東地方の X さんのツイート

I'm at 東京ディズニーランド (Tokyo Disneyland) - @*** (浦安市, 千葉県) w/ 3 others [URL] [URL]
神楽坂祭! this was the kid parade... professional parade later this evening (@ 神楽坂 w/ 5 others) [URL] [URL]
found it! the last お神輿 of 深川祭! #shrinehoppers [URL]
#shrinehoppers with @*** (@ 新小岩 香取神社) [URL]
プリクラ with @*** and @***!! (@ 上野アメ横商店街・アメヤ横丁 - @ameyoko w/ 2 others) [URL] [URL]
you can come here to pray for good luck in marriage #shrinehoppers (@ 延命地藏尊) [URL] [URL]

表 4.4: 近畿地方の Y さんのツイート

Iaido demonstration at Burtokuden #Iaido #Kyoto #Japan @ 京都武道館 [URL]
On the way to Kangosan's peak. @ 金剛山山頂 [URL]
The murky depths of the I-IPA. #Beard #craftbeer #Japan @ Craft Beer Base [URL]
Belive it or not, this is also a kind of Sakura! #緑 #桜 #造幣局 #nofilter #Sakura @ 大阪造幣局桜の通り抜け [URL]
At Kyoto today. #nofilter #Kyoto #maiko #Japan @ 祇園 Gion, Kyoto [URL]
Craft Beer Live 2015 @ 湊町リバープレイス [URL]

関東地方の X さんは、神社、お寺、お神輿などに興味があり、#shrinehoppers というハッシュタグをつけて、主に関東地方を巡っている。一方近畿地方の Y さんも、日本の文化に関わる場所を訪れたり、ビールのイベントに何度も参加している。2 人とも、それぞれの地方をメインに様々な観光場所を訪れたりイベントに参加していて、有用なユーザといえる。

4.4 システムの評価

ここでは、システムにより推薦されたユーザの精度について評価を行なった結果について述べる。

4.4.1 目的

A~Dの4つの手がかりを用いて作成したランキングを提案手法として、各手がかり1つを除いた3つの手がかりを用いて作成したランキング4つを比較手法として、合計5つのランキングを比較することにより、システムの評価を行う。これにより、各手がかりの有用性や特徴を調べる。

4.4.2 方法

作成した5つのランキングの上位10, 20ユーザについて、推薦するユーザが外国人の観光に有用なツイートをしているか、第一著者が人手で判定を行い、推薦の精度を測る。ただし、推薦するユーザとして各手がかりの平均評価が2以上、つまり提案手法では評価が8以上、比較手法では評価が6以上のユーザという条件を設ける。

4.4.3 結果

地域別のランキング上位10, 20ユーザ中の有用なユーザの割合をPrecision@10(P@10), Precision@20(P@20)とし、推薦の精度を評価した結果を表4.5に示す。

4.4.4 考察

まず、北海道・東北地方と中国・四国地方では、条件に合うユーザが20ユーザに満たない場合が多くあった。この原因は、根本的にその地域でツイートを行なっているユーザ数が少ないためである。ただ評価値の条件を緩めるだけでは精度を低下させてしまうため、今後これらの地域の推薦ユーザを増やすことも検討したい。

提案手法と比較手法を比較すると、写真、チェックイン付きツイートの頻度を除いた比較手法(A)とツイートを行なった場所の数を除いた比較手法(C)の精度が低くなっている。これは、写真、チェックイン付きツイートの頻度やツイートを行なった場所の数という手がかりが、有用なユーザを判定するのに特に重要であることを示している。日本の滞在期間を除いた比較手法(D)では、P@20の精度は比較手法よりも低いにもかかわらず、P@10の精度は提案手法よりも高くなっている。この原因として、日本の滞在期間が短いため得られたツイート量も少なく、結果が安定しなかったということと、ユーザの日本の観光に対する詳しさまで評価に含められなかったことがある。短

表 4.5: 推薦するユーザの推薦の精度

		提案手法	以下の手がかりを除く（比較手法）			
			A	B	C	D
北海道・東北地方	P@10	0.8	1.0	0.9	0.7	0.9
	P@20	-	-	-	0.75	-
関東地方	P@10	1.0	1.0	0.9	0.9	1.0
	P@20	0.95	0.95	0.95	0.9	0.95
中部地方	P@10	1.0	0.9	1.0	0.8	1.0
	P@20	1.0	0.8	1.0	0.8	0.95
近畿地方	P@10	0.8	0.8	0.9	0.8	0.8
	P@20	0.9	0.85	0.95	0.7	0.85
中国・四国地方	P@10	-	0.8	0.8	0.8	-
	P@20	-	-	-	-	-
九州・沖縄地方	P@10	1.0	0.8	1.0	0.9	1.0
	P@20	1.0	0.85	0.95	0.95	0.9
平均	P@10	0.920	0.883	0.917	0.817	0.940
	P@20	0.963	0.863	0.963	0.821	0.913

期間で様々な観光地を周り、短期間で大量のツイートを行うような、日本の観光に関して詳しくない可能性のあるユーザの評価の判断は難しいため、今後の課題となる。ただし、ツイート量が多い場合は滞在期間という手がかりは有効である。最後に提案手法とツイートの平均語数を除いた比較手法（B）を比較すると、ほとんど差が見られなかった。今後、ツイートの平均語数を除いた比較手法（B）を改善するもしくは本当に必要かどうかの調査を行う必要がある。

また、人手でユーザのツイートを見て有用か有用でないかを判断する際に、色々な観光地や店舗を熟知していて間違いなく役に立つと思われるユーザと、少しは役に立つであろうユーザを同列に扱ってしまっているのは問題だと考えられるため、今後評価を行う際には段階的な評価を行う。

提案手法で推薦されたが、不正解となったユーザの特徴としては、人が手動で投稿している会社や店舗の公式アカウントが多く見受けられた。これらのユーザは、Botではなく人の手によって書かれたツイートを行うため、一般ユーザとの区別が難しい。ただし、今回は公式アカウントは不正解としているが、特定の情報については詳しいことが多く、外国人観光客の疑問に親切に答えてくれる可能性も十分に考えられるため、必ず

しも不要なユーザではないと考えられる。

最後に、各手がかりを除くランキングだけに現れたユーザには、以下の傾向がみられた。

- A：観光についてのツイートが少なく、日常的なツイートが多い。
 - B：チェックイン付きツイートが多い。
 - C：特定の場所でしか活動せず、観光に関心が薄い。
 - D：短期間に同じ場所でツイートしていることが多く、会話やリツイートを除いた日本でのツイートの全体量が少ない。
-

第5章 実験：観光客の相談相手となりうるユーザの推薦

本章では、ユーザ推薦システムの有効性を検証するための実験について説明する。

実験には、提案手法となるユーザ推薦システムの他に、2種類の比較手法を使用する。これらの比較手法には、オープンソースの検索エンジンである Apache Solr²⁷を利用している。

5.1 提案手法の仕様

実験に際して、実験のためのインタフェースを実現し、提案手法自体にもいくつかの変更を加えた。まず、4章のシステムの評価で意味が無いとされたツイートの平均語数 (B) という手がかりを削除し、写真、チェックイン付きツイートの頻度 (A)、ツイートを行なった場所の数 (C)、日本の滞在期間 (D) という手がかりのみを用いることとする。さらに、入力として地域だけでなく、フリーワードの検索クエリを必要とする。

手がかり A, C, D の評価値に、この検索クエリを含むツイート数に応じて各ユーザに評価値を加算する。そして、選択した地域で最終的な評価値が高い順に検索クエリを含むツイートが1ツイート以上ある最大20ユーザのランキングを出力する。ただし、この検索クエリの照合では、大文字小文字の区別は行なわない。

図5.1に、提案手法のインタフェースを示す。

クエリを入力し地域を選択すると、選択した地域の20ユーザが表示される。ユーザIDをクリックすると、Twitterのユーザページにジャンプすることができ、Show tweetsをクリックすると、本研究で収集している各ユーザの日本で行われた直近の最大300ツイートを表示する。

²⁷<http://lucene.apache.org/solr/>



図 5.1: 提案手法のインタフェース

5.2 実験 1：既存サービスに基づくユーザ検索手法との比較

Twitter においてユーザを検索するには、クエリに適合するツイートを検索し、そこからユーザページを確認するという手順を踏む必要がある。この実験では、この既存のツイート検索に基づいたユーザ検索手法と提案手法の比較を行う。

5.2.1 目的

提案手法と既存サービス（ツイート検索）に基づくユーザ検索手法の比較することで、提案手法の検索の有効性や検索のしやすさについて評価を行う。

5.2.2 比較手法 1：既存サービス（ツイート検索）に基づくユーザ検索

比較手法のデータは提案手法と同じものを用いる。つまり、3章で抽出した、日本を訪れた経験のある英語ユーザのツイートを利用する。これらのユーザの各ツイートのツイート文、ツイート URL、ユーザ ID とユーザ名、ユーザ URL、投稿日時、投稿場所を一つの文書として、Apache Solr²⁸に検索対象文書として入力している。つまり、文書数は、利用したツイートの数と同じとなる。

²⁸<http://lucene.apache.org/solr/>

検索インターフェースは Apache Solr 準拠のものを利用する。図 5.2 に、比較手法 1 のインターフェースを示す。

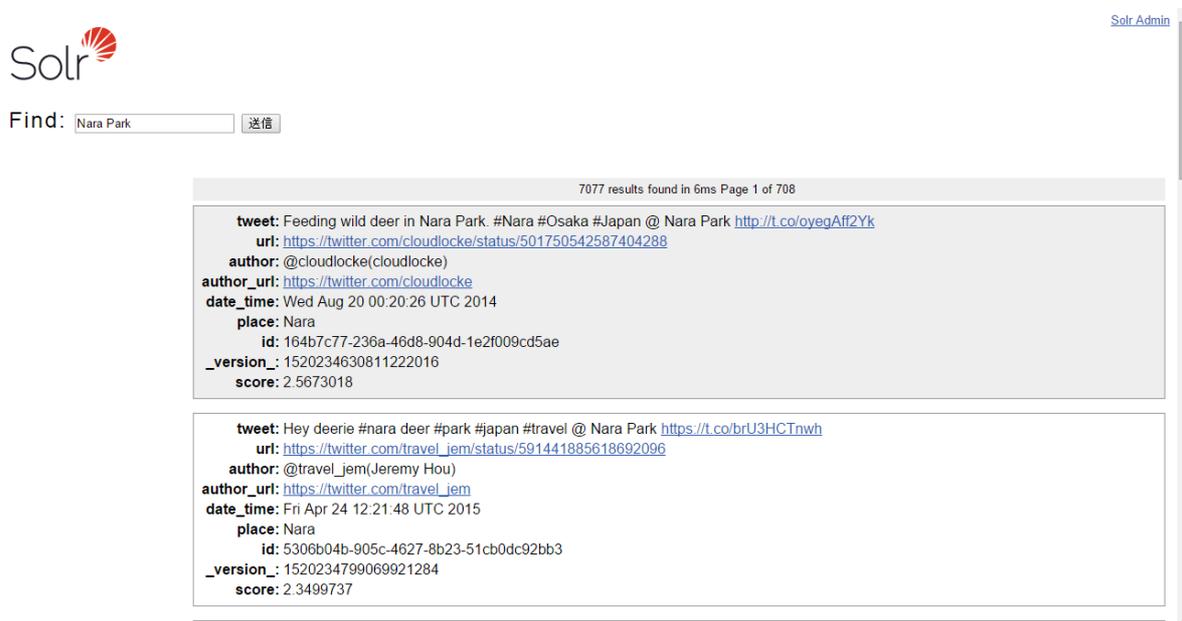


図 5.2: 比較手法 1 のインターフェース

クエリを入力すると、Apache Solr により文書がランキングされ、表示される。

5.2.3 方法

提案手法と比較手法 1 とを利用して、相談者となりうる Twitter ユーザを探す。地域は、4 章のシステムの評価で 20 以上のユーザが抽出されていた関東地方、中部地方、近畿地方、九州・沖縄地方の 4 つのエリアとし、被験者は 1 エリアにつき 3 箇所の観光場所を決めて、1 箇所の観光場所につき最大 3 ユーザを相談者として検索する。最終的に検索する相談者は 4 地域 × 3 箇所 × 3 ユーザで、最大 36 ユーザとなる。また、システムごとに Search Effectiveness (検索の有効性)、System Usability (使いやすさ)、Total System Evaluation (総合評価) の評価を行う。評価基準は、1:poor, 2:fair, 3:average, 4:good, 5:excellent の 5 段階である。被験者は外国人の留学生で、20 代男性と 20 代女性の 2 名である。

被験者実験の手順を以下に示す。ただし、システムを利用した順番による結果の差を無くすために、被験者に提示する手法は、提案手法と比較手法 1 とを使用する順番をそれぞれ入れ替えている。

1. 外国人向けの日本の観光情報サイト japan-guide.com²⁹ を利用し、特定のエリア（関東地方など）から観光場所を3箇所選択する。
2. 提案手法を用いて、選択した特定の観光場所の相談者として各3人のユーザを検索し、選んだ理由と共に結果用紙に記入する。ただし、規定時間内（20分以内）に各観光場所で3人、計9人のユーザが見つからなかった場合は、9人以下でも次の作業に移行する。
3. 比較手法1を用いて、作業2で行なった作業を繰り返す。
4. 4つのエリアについての作業の終了後に、提案手法と比較手法1の評価および良い点と悪い点について、アンケートに記入する。

5.2.4 結果

各被験者の結果を表5.1に示す。

2人の被験者共に、提案手法の Search Effectiveness（検索の有効性）が、比較手法1を上回った。また、被験者2は、System Usability（使いやすさ）と Total System Evaluation（総合評価）が、提案手法と比較手法1で同じとなったが、被験者1は、System Usability（使いやすさ）と Total System Evaluation（総合評価）においても、提案手法が比較手法1を上回っている。しかし、見つけたユーザ数については、被験者1の比較手法1、被験者2の提案手法と比較手法1では36ユーザ全員見つけられているが、被験者1の提案手法では35ユーザと、ある観光場所において2人のユーザしか見つからないという結果となった。

表 5.1: 実験1の結果

	被験者1		被験者2	
	提案手法	比較手法1	提案手法	比較手法1
Search Effectiveness	4	3	5	4
System Usability	4	3	4	4
Total System Evaluation	5	3	4	4
見つけたユーザ数	35	36	36	36

²⁹<http://www.japan-guide.com/>

5.2.5 考察

総合的に見ると，提案手法の評価が比較手法1を上回っていることが確認できた．提案手法について，アンケートで「ユーザの色々なツイートを読むことで，詳細な情報が得られた」といったような意見が得られた．また，比較手法1については，アンケートで「ユーザごとにツイートが統一されておらず不便」といったような意見が得られ，ツイートのランキングから有用なユーザを判断することは難しいことが分かる．これにより，従来の Twitter からの，ツイート検索に基づくユーザ検索よりも，提案手法のユーザ検索が優れているといえる．

また，被験者1の提案手法で，ある観光場所において2人のユーザしか見つからなかったという問題があった．この原因は，提案手法でその観光場所をクエリとした際に，2人のユーザしか推薦されていなかったためである．提案手法では，出力するユーザには「検索クエリに適合するツイートが1件以上ある」という条件を設けていたため，マイナーな観光場所を選択した場合，推薦されるユーザが3人に満たない場合がある．選択した観光場所について各3人のユーザを相談者として探すという被験者実験において，検索システムとして相談者となりうるユーザを3人未満しか出力しないのは問題と考え，次の実験では，推薦するユーザ数が少なくならないよう改善する．

5.3 実験2：従来の文書検索技術を利用したユーザ検索手法との比較

この実験では，各ユーザのツイート内容すべてを結合し，それぞれ1つの文書とした場合に，文書検索技術を用いてユーザを検索する手法と，提案手法との比較を行う．実験1は既存のツイート検索に基づいた検索手法と提案手法の比較であり，検索のしやすさも評価対象となるため，提案手法と比較手法1のインターフェースは異なっていた．本実験は，検索結果のユーザランキングのみについて評価を行うため，提案手法と比較手法2のについては，同一のインターフェースを採用する．

5.3.1 目的

提案手法と，ツイート内容すべてを結合し，それぞれを1つの文書とした場合に，検索エンジンを用いてユーザを検索する手法とを比較することで，従来の文書検索技術に基づくユーザ検索に対する提案手法の検索の有効性について評価を行う．

5.3.2 提案手法の変更点

提案手法では、出力するユーザには「検索クエリを含むツイートが1ツイート以上ある」という条件を設けていたため、実験1において、マイナーな観光場所を選択した場合、推薦されるユーザが3人に満たない場合があった。検索クエリに適合するツイートがあるユーザと無いユーザでは大きな隔たりがあるが、シンプルなクエリの照合により実現しているため、観光場所についてのツイートも書き方によっては適合しない場合や、そもそも観光場所に訪れたことがあってもツイートを行っていない場合もある。そこで、検索クエリに適合するツイートがあるユーザが20人に満たなかった場合、検索クエリに適合するツイートがあるユーザの下位に、その地域のユーザを評価値の高い順に20ユーザになるまで追加を行う。つまり、検索クエリに適合するツイートがないユーザでも、その地域で評価値が高く、その地域に詳しいと考えられるユーザであれば相談者となりうるユーザとして推薦する。

また、インタフェースについても、若干の変更を加えた。実験1で表示する各ユーザのツイートは、日本で行われた直近の最大300ツイートだけだったが、各ユーザの検索クエリに適合するツイートも、別枠で表示を行う。

図5.3に、変更を行った後の提案手法のインタフェースを示す。

Num	TweetText	DateTime	TweetPlace
1	WHAT TIME IS IT? (the other deers were gr8) #narapark #nara #ohayjaytravels #japanfriendshiptrip @... https://t.co/0PvsGbc2Y5	2015-07-14 00:06:32	Osaka-shi Naniwa
2	stupid deer that ate our map... #narapark #nara #ohayjaytravels #japanfriendshiptrip @ 奈良公園 Nara Park https://t.co/oukoflNz5	2015-07-14 00:04:46	Osaka-shi Naniwa

Num	TweetText	DateTime	TweetPlace
1	ugh Kyoto i miss you already ☹️ #kyoto #ohayjaytravels #japanfriendshiptrip @ Gion, Kyoto, Japan https://t.co/sz4KEQXwm7	2015-07-21 23:36:39	Kyoto City Nakagyo Ward
2	last night in Osaka so of course we're reppin' #dotonburi #osaka #ohayjaytravels #japanfriendshiptrip... https://t.co/SQa8VH9Q0	2015-07-16 23:05:42	Osaka City Chuo Ward
3	BUTTER BEER!!! 🍷 #wizardingworld #harrypotter #universalstudiosjapan #USJ #osaka #ohayjaytravels... https://t.co/Lady1TEZp2	2015-07-15 20:13:05	Osaka-shi Konohana
4	Wizarding World of Harry Potter ☺️ only got to go on the Forbidden Journey ride but it was seriously... https://t.co/KmX6AF8ltw	2015-07-15 20:10:38	Osaka-shi Konohana
5	spent the day at USJ but I was too chicken to go on the main/biggest ride they had because it had... https://t.co/LrciTTX2t	2015-07-15 20:05:54	Osaka-shi Konohana
6	YAAASSSSS ☺️ this is not just takoyaki, it has neji and super soft boiled egg 🍣 #dotonburi #osaka... https://t.co/S9XtUob7uJ	2015-07-14 22:35:21	Osaka-shi Yodogawa-ku
7	finally got our butts down to dotonburi to EAT EVERYTHING #dotonburi #osaka #ohayjaytravels... https://t.co/LG952+H.2t	2015-07-14 22:28:24	Osaka-shi Yodogawa-ku
8	more todajji #todajji #nara #ohayjaytravels #japanfriendshiptrip @ Nandai-Mon Gate, Todai-ji, Nara https://t.co/EP1YotAOIV	2015-07-14 00:12:25	Nara
9	got to see the largest Buddha statue in all of Japan, doesn't look it, but it was huuUUUGE #todajji... https://t.co/1TKnKFnYc	2015-07-14 00:09:58	Nara
10	WHAT TIME IS IT? (the other deers were gr8) #narapark #nara #ohayjaytravels #japanfriendshiptrip @... https://t.co/0PvsGbc2Y5	2015-07-14 00:06:32	Osaka-shi Naniwa
11	stupid deer that ate our map... #narapark #nara #ohayjaytravels #japanfriendshiptrip @ 奈良公園 Nara https://t.co/oukoflNz5	2015-07-14 00:04:46	Osaka-shi Naniwa

図 5.3: 変更後の提案手法のインタフェース

5.3.3 比較手法 2：文書検索技術に基づくユーザ検索

実験 1 と同様に，比較手法のデータは提案手法と同じものを用いる．つまり，3 章で抽出した，日本を訪れた経験のある英語ユーザのツイートを利用する．これらの各ユーザのツイートをすべて結合し，それぞれを一つの文書として，Apache Solr³⁰に検索対象文書として入力している．つまり，文書数は，利用したユーザの数と同じとなる．

検索インターフェースは「地方」の選択が無くなっている以外は提案手法と同一である．図 5.4 に，提案手法のインターフェースを示す．

Num	TweetText	DateTime	TweetPlace
1	Green as can be #vscopotph #vscophilippines #vscobeau #vsconian #vscogoodshots @ Nara Park https://t.co/vc5lp6zJFz	2015-05-13 20:22:19	Nara
2	Treats for my dear deer 🌿 @ Nara Park https://t.co/3QJqNSU9xM	2015-04-26 21:38:10	Nara
1	Brick walls are love 🍷 @ Nara Station https://t.co/dz2M2RTX5s	2015-05-24 20:29:00	Nara
2	Brick walls are love 🍷 @ Nara Station https://t.co/WvRLckCKyE	2015-05-13 22:35:50	Nara
3	Brick walls are love 🍷 #vscopotph #vscophilippines #vscobeau #vsconian #vscogoodshots @ Nara Station https://t.co/QQpeNsTshM	2015-05-13 20:24:56	Nara
4	Green as can be #vscopotph #vscophilippines #vscobeau #vsconian #vscogoodshots @ Nara Park https://t.co/vc5lp6zJFz	2015-05-13 20:22:19	Nara
5	Torii gate on our way up to the mountain from Fushimi Inari 🌿 @ Fushimi Inari https://t.co/KASl8T4F5G	2015-05-05 19:55:27	Taito-ku
6	Torii gate on the way up to the mountain from Fushimi Inari 🌿 #Kyoto @Thousand Torii Gate (Fushimi... https://t.co/GWJkP33z	2015-05-02 22:34:56	Kyoto-shi Shimogyo
7	The temple of the crystal skull👁️#IndianaJones #DisneySea @ Tokyo Disney Sea Indiana Jones https://t.co/yTU4xJU0w	2015-05-01 21:37:43	Machida City
8	Hogwarts Castle 🏰 #UniversaStudiosOsaka @ Universal Studios Osaka, Harry Porter Castle https://t.co/hTFC0urpds	2015-04-28 21:17:37	Osaka-shi Konohana
9	Treats for my dear deer! 🌿 @ Nara Deer Park https://t.co/MwKCVL9Qn	2015-04-27 17:37:32	Nara
10	Treats for my dear deer 🌿 @ Nara Park https://t.co/3QJqNSU9xM	2015-04-26 21:38:10	Nara

図 5.4: 比較手法 2 のインターフェース

5.3.4 方法

提案手法と比較手法 2 とを利用して，相談者となりうる Twitter ユーザを検索する．地域は，4 章のシステムの評価で 20 人以上のユーザが抽出されていた関東地方，中部地方，近畿地方，九州・沖縄地方の 4 つのエリアとし，被験者は 1 エリアにつき 3 箇所の観光場所を決めて，1 箇所の観光場所につき最大 3 人のユーザを相談者として検索し，さらに各ユーザについて評価を行う．最終的に検索する相談者は 4 地域 × 3 箇所 × 3 ユーザで，最大 36 ユーザとなる．また，システムごとに Search Effectiveness (検索の有効性)

³⁰<http://lucene.apache.org/solr/>

の評価を行う。ユーザとシステムとの評価基準は、1:poor, 2:fair, 3:average, 4:good, 5:excellent の 5 段階である。被験者は外国人の留学生で、20 代の男女 5 名である。

被験者実験の手順を以下に示す。ただし、システムを利用した順番による結果の差を無くすために、被験者に提示する手法は、提案手法と比較手法 2 とを使用する順番をそれぞれ入れ替えている。

1. 外国人向けの日本の観光情報サイト [japan-guide.com](http://www.japan-guide.com)³¹ を利用し、特定のエリア（関東地方など）から観光場所を 3 箇所選択する。
2. 提案手法を用いて、選択した特定の観光場所の相談者として各 3 人のユーザを検索し、各ユーザの評価と選んだ理由とを結果用紙に記入する。ただし、規定時間内（20 分以内）に各観光場所で 3 人、計 9 人のユーザが見つからなかった場合は、9 人以下でも次の作業に移行する。
3. 比較手法 2 を用いて、作業 2 で行なった作業を繰り返す。
4. 4 つのエリアについての作業の終了後に、提案手法と比較手法 2 の評価および良い点と悪い点について、アンケートに記入する。

5.3.5 結果

各被験者の結果を表 5.2 に示す。Search Effectiveness（検索の有効性）では、提案手法のほうが良いと評価した被験者が 2 人、同じと評価した被験者が 2 人、比較手法 2 のほうが良いと評価した被験者が 1 人となり、提案手法が比較手法 2 よりも若干上回る結果となった。しかし、見つけたユーザ数では、提案手法のほうが多くなった被験者が 2 人、比較手法 2 の方が多くなった被験者が 3 人となり、全員の被験者の見つけたユーザの合計は、提案手法が 126 ユーザ、比較手法 2 が 134 ユーザと、比較手法 2 が提案手法を上回っている。

5.3.6 考察

Search Effectiveness（検索の有効性）では提案手法の方が良い結果だが、見つけたユーザの数は比較手法 2 の方が良い結果になった理由は、今回提案手法に加えた変更が原因と考える。実験 1 では、提案手法では検索クエリに適合するツイートがあるユーザしか

³¹<http://www.japan-guide.com/>

表 5.2: 実験 2 の結果

		Search Effectiveness (検索の有効性)	発見したユーザ数					合計
			評価					
			poor	fair	average	good	excellent	
被験者 1	提案手法	2	10	2	5	6	0	23
	比較手法 2	2	10	7	4	7	0	28
被験者 2	提案手法	4	0	4	13	19	0	36
	比較手法 2	3	1	3	17	11	0	32
被験者 3	提案手法	3	0	1	10	13	1	25
	比較手法 2	2	1	2	6	11	0	20
被験者 4	提案手法	3	0	1	15	10	2	28
	比較手法 2	3	0	0	21	10	0	31
被験者 5	提案手法	2	2	2	8	1	1	14
	比較手法 2	3	3	8	11	1	0	23

推薦していなかったが、推薦するユーザ数が 20 ユーザに満たない場合には、検索クエリに適合するツイートが無いユーザも推薦した。しかし、結局は推薦に追加したユーザは被験者に相談者として選ばれることが少なく、結果として提案手法で見つけたユーザ数は少なくなった。実際に、どのような場合に相談者が見つからなかったか調べてみると、被験者が観光場所として比較的マイナーな場所を選んだ時が多かった。例を挙げると、関東地方の「Oze National Park」や中部地方の「Hida Fork Village」などである。このような場合には、検索クエリに適合するツイートがあるユーザが少なくなるため、相談者が見つからなかったと考えられる。

発見したユーザの質について考察するため、各評価を 1:poor, 2:fair, 3:average, 4:good, 5:excellent のような値として、各被験者が発見したユーザの評価の平均を、表 5.3 に示す。

表 5.3 より、すべての被験者において、提案手法を用いて見つけたユーザの評価値は高くなっていた。また、評価が excellent のユーザが見つかったのは、提案手法だけであった。ここで、評価が excellent のユーザは、特定の観光場所での経験についての詳しいツイートや観光場所付近の有用な写真を何度も行なっていたとコメントされていた。提案手法において追加した、検索クエリに適合するツイートの無いユーザが相談者としてあまり選ばれなかったとすると、提案手法で推薦する、検索クエリに適合するツイートのあるユーザの質はかなり高いことが分かる。また、提案手法において相談者としてのユーザが見つからない場合を減らすには、この検索クエリに適合するツイートのあるユーザ

表 5.3: 発見したユーザの評価の平均

		評価値の平均値
被験者 1	提案手法	2.30
	比較手法 2	2.29
被験者 2	提案手法	3.42
	比較手法 2	3.19
被験者 3	提案手法	3.56
	比較手法 2	3.35
被験者 4	提案手法	3.46
	比較手法 2	3.32
被験者 5	提案手法	2.79
	比較手法 2	2.43
被験者全員 の平均	提案手法	3.11 *
	比較手法 2	2.92

* t 検定で有意差あり (両側検定で有意水準 5%)

を増やせば良い。そのためには、データの収集期間を延ばしてデータ量を増やすという単純な方法や、検索クエリの完全一致だけではなく語形変化にも対応した照合を行うという方法が考えられる。

最後に、被験者実験で得られたシステムの問題点について、被験者たちの各コメントとその考察について以下で述べる。

- 九州・沖縄地方の有田 (Arita) というクエリで探した際に、成田 (Narita) が含まれるユーザばかりが推薦された。

名前が短くあまり有名でない観光スポットでは特にこの問題が発生しやすく、この問題について、場所を表すクエリを用いた検索システムの照合方法について改善を行う必要がある。

- いくつかのツイートが英語ではなく日本語で書かれていたため、読むことができなかった。

推薦者には、普段英語を利用する日本人や、非常に日本に慣れているため日本語を含むツイートを行う外国人が含まれてしまった。日本語がわからない外国人観光客に対しての推薦システムであるため、表示するツイートを英語で書かれたものに限定するといった対策が必要である。

- Twitter のユーザ情報を見ることができないユーザが数人いた。

本研究でツイートを収集し始めた後に、ユーザが Twitter のアカウントを削除した場合や Twitter によりアカウントが凍結された場合、Twitter 上のユーザのページに飛べなくなっていた。このような状態にある場合、そのユーザと Twitter で相談することは不可能であるため、推薦を行うユーザから除去する必要がある。

- 位置データという面で、Twitter ではハッシュタグが無い場合に各ツイートの正確な観光の特定や検索が難しいため、Instagram の方が良いのではないか。

今回はコミュニケーションをとる場として優れている Twitter を用いたが、奥らの研究 [10] で、Twitter と Foursquare、Panoramio のデータを利用していったように、今後は Twitter 以外のサービスも利用したデータ収集方法についても検討したい。

- 良いユーザを見つけるための重要な情報を探すのに時間がかかる場合があった。

実験 2 を行う際にもインタフェースの改良を行ったが、相談者を探す際に、より分かりやすく効果的にユーザの情報を提示できるように、インタフェースの改良を行っていきたい。

第6章 おわりに

6.1 まとめ

日本を訪れる外国人観光客の増加と、日本人の英語力の低さを背景に、本研究では、英語を話す外国人観光客が日本を観光する際の相談相手となりうる Twitter ユーザの推薦方法について提案を行った。具体的には、日本についてあまり知らない外国人観光客に相談役として、英語を使用する日本の観光に役立つユーザを地域別に推薦する方法を提案し、その実装と評価を行った。これにより、外国人観光客は、推薦された Twitter ユーザの中から相談者を選び、Twitter 上でコミュニケーションをとることで、自由な話題で気軽に情報収集や問題の解決を行うことが可能となった。

提案手法は、以下の処理によって実現した。

1. 日本でつぶやいた Twitter ユーザの収集
2. 英語を用いるユーザの選別
3. 推薦に適さないユーザの除去
4. 観光に役立つユーザの推薦

Twitter ユーザの収集では、日本が含まれる範囲の位置情報付きツイートを取得し、英語を用いるユーザには、Twitter によるユーザの言語が英語と判定されていて、英語によるツイートを N 件 ($N = 10$) 以上しているユーザとした。推薦に適さないユーザの除去では、在日米軍関係者と思われるユーザと Bot の除去を行った。ここで、在日米軍関係者と思われるユーザは、ツイートの大部分が在日米軍に関わる地域からであることと、それらのユーザのツイートは、ほとんどが仕事や日常についてのつぶやきであり、観光に有用なツイートが非常に少ないことから、推薦に適さないユーザとした。ユーザの推薦では、まず、どのようなユーザが外国人の観光に有用なのかを、日本人の第一著者と外国人の留学生の 2 名を被験者として予備実験を行ない、明らかにした。そして、この

有用性判定の基準をもとにした以下の4つの手がかりと、ユーザが主に活動する地域情報を利用してユーザの推薦を行うシステムを実現した。

- A) 写真，チェックイン付きツイートの頻度
- B) ツイートの平均語数
- C) ツイートを行なった場所の数
- D) 日本の滞在期間

ユーザが主に活動する地域情報としては，“北海道・東北地方”，“関東地方”，“中部地方”，“近畿地方”，“中国・四国地方”，“九州・沖縄地方”，“該当なし”の7つのいずれかを付与した。

構築したシステムの評価を行なった結果，写真，チェックイン付きツイートの頻度（A）やツイートを行なった場所の数（C）という手がかりが，有用なユーザを判定するのに特に重要であることが分かった。また，ツイート量が多い場合に限り，滞在期間（D）という手がかりは有効であること，ツイートの平均語数（B）という手がかりは有効ではないことが分かった。

システムの有効性を検証するための被験者実験では，既存のサービス（ツイート検索）に基づいたユーザ検索手法と提案手法との比較を実験1，ツイート内容すべてを結合し，1つの文書とした従来の文書検索技術によるユーザ検索手法と提案手法との比較を実験2として，比較を行った。また，この実験に際してインタフェースを実現し，提案手法自体にもいくつかの変更を加えた。具体的には，システムの評価で意味が無いとされたツイートの平均語数（B）という手がかりを削除，入力としてフリーワードの検索クエリを設け，この検索クエリに適合するツイート数に応じて各ユーザに評価値を加算した。実験1の結果，検索の有効性や使いやすさにおいて，提案手法が既存手法を上回った。また，実験2の結果，提案手法で推薦されるユーザの質が高いことを明らかにした。

6.2 今後の課題

現在の提案手法では，様々な観光地を訪れ，日本の観光に詳しいユーザだとしても，データの収集期間内に検索クエリを含んだツイートをしていなければ，推薦されないか，相談者に選ばれないことが多い。そこで，今後の課題としては，データの収集期間を延ばしてデータ量を増やしたり，フリーワードの検索クエリを完全一致だけではなく語形

変化にも対応した照合を行うことで、検索クエリを含むツイートを行なっているユーザを増やし、推薦精度の向上を行うことが挙げられる。

また、システムの評価で意味が無いとされたツイートの平均語数(B)という手がかりの代わりに、ユーザのツイートの質を評価する項目を検討していきたい。インタフェースについては、ユーザの推薦を行う際に、地域だけでなくユーザが良く話す話題などを付与し、よりシステム利用者の好みに合わせた推薦を行うことも考えられる。

さらに、名前が短くあまり有名でない観光スポットの場合に誤ったユーザが推薦される問題についても検討を進めていきたい。

謝辞

本研究を行うにあたり，全過程において丁寧かつ適切なご指導を賜った筑波大学図書館情報メディア系の関洋平准教授に心から感謝申し上げます．研究の進め方から論文の執筆まで，迷走していた本研究が研究として形になったのは，関先生のご指導あってこそでした．本当にありがとうございました．

また，助言を下さった筑波大学図書館情報メディア系の佐藤哲司教授，審査を引き受けて下さった筑波大学図書館情報メディア系の高久雅生准教授，被験者実験を引き受けてくださった筑波大学の留学生の方々に深く感謝いたします．

さらに，研究を進めるにあたり様々な面でお世話になった，コミュニケーション理解研究室とコンテンツ工学研究室の皆様に感謝いたします．

参考文献

- [1] 中嶋勇人, 新妻弘崇, 太田学. 位置情報付きツイートを利用した観光ルート推薦. 情報処理学会研究報告 . データベースシステム (DBS) , Vol. 158, No. 28, 2013.
- [2] 新井晃平, 新妻弘崇, 太田学. Twitter を利用した観光ルート推薦の一手法. 第7回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015) , G7-6, 2015.
- [3] 佐伯圭介, 村山敬祐, 遠藤雅樹, 横山昌平, 石川博. 外国人向け観光プロモーション支援のためのソーシャルメディア分析手法の研究. 第6回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2014) , B4-2, 2014.
- [4] 奥村学. マイクロブログマイニングの現在. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 111, No. 427, pp. 19–24, 2012.
- [5] 原辰徳. 顧客経験と設計生産活動の解明による顧客参加型のサービス構成支援法～観光サービスにおけるツアー設計プロセスの高度化を例として～. 戦略的創造研究推進事業 平成23年度研究開発実施報告書, 2011.
- [6] Semiocast. L 'Asie devient leader sur Twitter Hausse de 20% des tweets en France. http://semiocast.com/en/publications/2010_07_02_Lasie_devient_leader_sur_twitter (accessed 2016-1-10).
- [7] 橋本康弘, 岡瑞起. 都市におけるジオタグ付きツイートの統計. 人工知能学会誌, Vol. 27, No. 4, pp. 424–431, 2012.
- [8] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You Are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*, pp. 759–768, New York, NY, USA, 2010.

- [9] 上村卓也, 新田直子, 馬場口登. 時空間的出現特性の違いを考慮した位置を表す語の抽出によるツイートの発信位置推定. 第7回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), C4-2, 2015.
- [10] 奥健太, 橋本拓也, 上野弘毅, 服部文夫. 位置情報付きツイート対応付けに基づく観光スポット推薦システムの開発. 第2回 Web インテリジェンスとインタラクション研究会, pp. 7–12, 2013.
- [11] 泰斗高畠, 正人香田. 1クラス svm と近傍サポートによる領域判別. オペレーションズ・リサーチ : 経営の科学, Vol. 51, No. 11, pp. 677–682, 2006.
- [12] 前田高志ニコラス, 吉田光男, 鳥海不二夫, 大橋弘忠. 重力モデルと TF-IDF を用いたジオタグ付き Twitter データからの観光地抽出と魅力の評価. 第7回 Web インテリジェンスとインタラクション研究会, pp. 61–65, 2015.
- [13] Takeshi Kurashima, Tomoharu Iwata, Takahide Hoshide, Noriko Takaya, and Ko Fujimura. Geo Topic Model: Joint Modeling of User’s Activity Area and Interests for Location Recommendation. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM ’13)*, pp. 375–384, New York, NY, USA, 2013.
- [14] Chen Hongbo, Chen Zhiming, Mohammad Shamsul Arefin, and Yasuhiko Morimoto. Place Recommendation from Check-in Spots on Location-Based Online Social Networks. In *Proceedings of the 2012 Third International Conference on Networking and Computing (ICNC ’12)*, pp. 143–148, Washington, DC, USA, 2012. IEEE Computer Society.
- [15] 村上嘉代子, 川村秀憲. 外国人から見た日本旅行 : 英語ブログからの観光イメージ分析. 人工知能学会誌, Vol. 26, No. 3, pp. 286–293, 2011.
- [16] 佐伯圭介, 遠藤雅樹, 廣田雅春, 倉田陽平, 横山昌平, 石川博. 外国人 Twitter ユーザの観光訪問先の属性別分析. 第7回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), C4-3, 2015.
- [17] 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. 情報処理学会研究報告 . データベースシステム (DBS), Vol. 162, No. 15, 2015.

- [18] 田島真悟, 牛尼剛聡. 実世界イベント情報共有を目的としたアドホックな twitter フォローネットワーク構成手法. 電子情報通信学会技術研究報告 . データ工学 (DE), Vol. 113, No. 105, pp. 41–46, 2013.
- [19] 稲垣陽一, 中島伸介, 張建偉, 中本レン, 桑原雄. ブログの体験熟知度に基づくブログランキングシステムの開発および評価. 情報処理学会論文誌 データベース, Vol. 3, No. 3, pp. 123–134, 2010.
- [20] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46, 1960.
- [21] J. Richard Landis and Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, Vol. 33, No. 1, pp. 159–174, 1977.

発表論文

1. 田中匠，関洋平：外国人観光客に向けた Twitter ユーザの地域別推薦の評価，ARG 第7回 Web インテリジェンスとインタラクション研究会, pp.67-72, 2015年11月．
2. 田中匠，関洋平：外国人観光客に向けた英語を使用する Twitter ユーザの推薦，観光情報学会 第11回研究発表会，pp.21-24，2015年8月．
3. 田中匠，関洋平：マイクロブログユーザのクラスタに着目したイベント手がかり語の抽出，第6回データ工学と情報マネジメントに関するフォーラム（DEIM2014），B6-2, 2014年3月．