

外国人観光客に向けた Twitter ユーザの地域別推薦の評価

田中 匠^{†,a} 関 洋平^{‡,b}

[†]筑波大学大学院 図書館情報メディア研究科 [‡]筑波大学 図書館情報メディア系

a) humi@slis.tsukuba.ac.jp b) yohei@slis.tsukuba.ac.jp

概要 2020年に東京オリンピックを控え、日本は外国人観光客の呼びこみに力を入れている。本研究では言語の壁に着目し、英語を用い日本について相談できる人を、日本についてあまり知らない外国人観光客に推薦することで、観光中に発生したトラブルの解決や観光意欲の増加、観光の充実を図る。英語を用い日本について相談できる人としては、日本において英語を用いる Twitter ユーザを対象とし、言語判定、推薦に適さないユーザの除去、有用性判定、地域属性付与を経て推薦を行う。推薦に適さないユーザとしては、在日米軍関係者と bot について調査を行い、除去を行う。また、有用性判定の基準は、日本を訪問する外国人の観光に役立つユーザについての調査結果を参考に決定する。ユーザの推薦は、有用性判定の基準をもとにした4つの手がかりと地域情報を利用することにより実現し、求める地域に詳しいユーザを推薦する。評価の結果、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が分かった。

キーワード マイクロブログ, Twitter, 観光, ユーザ推薦

1 はじめに

近年、日本を訪れる外国人は増加傾向にあり¹、2020年に開催予定の東京オリンピックでは今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できる。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によると、日本は世界で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²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。

本研究では、以上のような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。観光についての相談というと、旅行先おすすめサイトや観光レビューサイトなどが考えられるが、そういったサービスでは、決まった観光地についての相談しかできない、リアルタイムな相談ができないといったような問題がある。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 を用いる。そして、英語を使うことができ、外国人の観光に有用な Twitter ユーザを相談役として提供することで、外国人観光客の観光中に発生したトラブルの解決や観光意欲の増加、観光の充実を図る。

以下、2章で関連研究を述べる。3章では、ユーザ推薦の手順と日本を訪れた経験のある英語ユーザの収集に

ついて述べる。4章では、外国人の観光に有用なユーザかどうかという観点で予備調査を行った結果を述べ、5章では、地域別ユーザ推薦システムの実装とその評価について述べる。最後に、6章でまとめを述べる。

2 関連研究

日本に外国人観光客をさらに呼びこむことを目的とした研究には次のようなものが挙げられる。

原の研究 [1] 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いる。さらに、旅行情報サイト TripAdvisor³の 口コミ分析により、外国人観光客の満足、不満足ポイントの調査を行い、口コミの否定的意見の多くに「日本語の必要性」があったという結果を得ている。

本研究と同様に、Twitter の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 [2][3] がある。佐伯らは、外国語で行われたツイートをを用いて、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。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、ツイート頻度を利用して、外国人ユーザを在日外国人と訪日外国人に分類している。そして、分類した3属性について特定の観光スポットを訪れた訪問者数を調査し、それぞれの特徴について考察を行っている。ここで、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。

本研究は、言語の問題を解決することで日本に外国人観光客をさらに呼びこむことを目的とする。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

Copyright is held by the author(s).

The article has been published without reviewing.

¹<http://www.jnto.go.jp/jpn/>

²<http://www.efjapan.co.jp/epi/>

³<http://www.tripadvisor.jp/>

ため、日本人、在日外国人、訪日外国人の分類は行わないこととする。

3 日本を訪れた経験のある英語ユーザの収集

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 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収集する。
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。

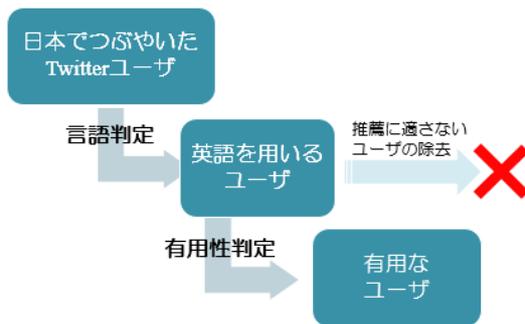


図 1 ユーザ推薦の手順

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 5 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 Twitter4j⁴を用いて、Twitter Streaming API⁵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収集

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収集する。英語を用いるユーザの条件は、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⁶以上行っているユーザとする。

⁴<http://twitter4j.org/>

⁵<https://dev.twitter.com/streaming/overview>

⁶調査に基づき N=10 とした。

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。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 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⁷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。

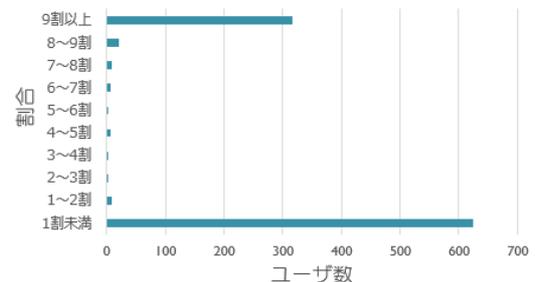


図 2 在日米軍に関わる地域でつぶやかれたツイートの割合とそのユーザ数

図 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。

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

⁷<http://www.mod.go.jp/j/publication/wp/>

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

4 有用なユーザについての予備調査

4.1 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。

4.2 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 1 著者と外国人の留学生の 2 名である。

4.3 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、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）

4.4 結果

4 段階評価における Cohen の 値 [4] は 0.34 となり、Landis らの基準 [5] 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。

4.5 考察

4 段階評価に差が出た理由として、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。

4.6 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

5 ユーザ推薦システムの構築

5.1 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、Instagram⁸の写真や twitter からのアップロード写真、swarm⁹によるチェックインを行っているツイートの頻度の割合が大きいほど高評価となる。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。平均ツイート語数が多いほど高評価となる。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。D では、データの収集期間を 12 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。

⁸<https://instagram.com/>

⁹<https://www.swarmapp.com/>

5.2 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。

5.2.1 地域の分割

付与する地域は、図3のような日本で一般的な地域区分¹⁰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3 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
- 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

5.2.2 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¹¹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1のような分布となった。

表1 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

5.3 システムの評価

A~Dの4つの手がかりを用いて作成したランキング（提案手法）と、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（比較手法）、合計5つのランキングを用いて、システムの評価を行う。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。作成した5つのランキングについて、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人手で判定を行い、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2に示す。

5.4 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、これらの地域の推薦ユーザを増やす方法について今後検討を行う。

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

¹⁰<https://ja.wikipedia.org/wiki/都道府県>

¹¹<http://www.finds.jp/wsdocs/rgeocode/index.html.ja>

表2 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

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ということがある。短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。このような、日本の観光に関して詳しくない可能性のあるユーザを正解として評価してしまったことが問題であり、評価方法の改善を行う必要がある。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、段階的に評価を行う必要がある。

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だろう。

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。

6 おわりに

本研究では、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。推薦は、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収集、推薦に適さないユーザの除去、地域情報の付与、観光に役立つユーザの推薦という処理に基づき、実現した。ユーザの推薦は、有用性判定の基準をもとに4つの手がかりを利用した。評価の結果、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が分かった。

今後は、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、システム利用者の

好みに合わせて推薦するための、提示の工夫を行う予定である。

謝辞

本研究の一部は、筑波大学研究基盤支援プログラム（Bタイプ）、科学研究費補助金基盤研究B（課題番号25280110）、萌芽研究（課題番号25540159）の助成を受けて遂行された。

参考文献

- [1] 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，http://www.ristex.jp/examin/service/pdf/kenkyu_h23_8-2.pdf, 2011.
- [2] 佐伯圭介，村山敬祐，遠藤雅樹ほか：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，DEIM Forum 2014, B4-2, 2014.
- [3] 佐伯圭介，遠藤雅樹，廣田雅春ほか：外国人 Twitter ユーザの観光訪問先の属性別分析，DEIM Forum 2015, C4-3, 2015.
- [4] Cohen, J. : A Coefficient of Agreement for Nominal Scales, Educational and Psychological Measurement, Vol. 20, No. 1, pp. 37-46, 1960.
- [5] J. R. Landis and G. G. Koch : The Measurement of Observer Agreement for Categorical Data, Biometrics, Vol. 33, pp. 159-174, 1977.