

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

佐伯 圭介[†] 村山 敬祐[†] 遠藤 雅樹^{††} 横山 昌平^{†††} 石川 博[†]

[†] 首都大学東京システムデザイン学部 〒191-0065 東京都日野市旭が丘 6-6

^{††} 首都大学東京大学院システムデザイン研究科 〒191-0065 東京都日野市旭が丘 6-6

^{†††} 静岡大学大学院情報学研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: †{saeki-keisuke@ed., murayama-keisuke@ed., ishikawa-hiroshi@}tmu.ac.jp,

††endo-masaki@ed.tmu.ac.jp, †††yokoyama@inf.shizuoka.ac.jp

あらまし 近年、日本を訪れる外国人観光客の数は増加傾向にある。今後観光客をさらに増やすためには、海外における観光宣伝の強化が必要と考えられる。しかし、外国人の興味、関心は日本人の考えとは異なる可能性がある。日本人が外国人の興味を知る方法のひとつにソーシャルメディアがある。特に Twitter はユーザの人数が多く、公開されていれば誰でも閲覧が可能である。そこで外国人の Twitter へのツイートを取得し、場所ごとの使用言語の違いや頻出する訪問順序に注目して、外国人の興味や関心を分析、可視化することにより、外国人に向けたプロモーションの支援を行うための手法を提案する。

キーワード Twitter, 観光, 情報推薦

1. はじめに

2013 年、1 年間の訪日外国人旅行者は 1 千万人を超え、過去最高を更新した^(注1)。「観光立国」を目指し、さらに外国人観光客を増やしていくためには、さらなるプロモーションや観光客の受け入れ体制の強化が必要と考えられる。外国人観光客へのプロモーションは日本に住み、受け入れる側である日本人が行うことが多いが、日本人が外国人の興味、関心を完全に理解することは大変難しい。世界各地、日本とは違うそれぞれの文化を持つ外国の人々の日本に対する興味、関心が日本人と同じとは限らないからである。もし、日本人には想像のつきにくい外国人特有の興味や関心、観光先などを知ることができれば、外国人に対し、さらに効果的なプロモーションができるだろう。

外国人の興味を知るには、世界的によく利用されている SNS (ソーシャル・ネットワーキング・サービス) の利用が効果的である。特に Twitter は世界中に多くのユーザがおり、公開されていれば誰でも閲覧が可能である。また、投稿が手軽でリアルタイム性が高いのが特徴である。日本に関する外国人のツイートを集めることができれば、外国人の興味を知ることができる。

プロモーションを行う日本人が知りたいことのひとつに「外国人観光客は日本でどこに行くのか」が挙げられる。観光客の目的地が分かれば、その周辺情報を調べることで興味の対象を推測でき、プロモーションの一助となる。また、観光客の移動順序が推測できれば、外国人向けのツアー企画や観光ルートの提案の手助けとなるだろう。

そこで本稿では、外国人に向けたプロモーションの支援を行

うための Twitter へのツイートの分析手法について提案する。

本稿の章構成を説明する。2 章では関連研究について述べる。次に、3 章では提案手法を述べる。4 章では実験方法を、5 章では実験結果と考察を述べる。6 章では、5 章にて評価の高い手法の汎用性を評価し、7 章では本稿のまとめと今後の課題を述べる。

2. 関連研究

Web 上の情報を利用し、観光情報を抽出する研究には次のようなものが挙げられる。

倉島らの研究 [1] では、日本語のブログから旅行の目的となる対象と個人の体験の抽出を行っており、村上らの研究 [2] では、英語で書かれたブログから外国人観光客の旅行先の感想の分析を行っている。特に Twitter を利用して観光情報を抽出しているものには、位置情報付きツイートを利用し、観光スポットの特徴化を行っている奥らの研究 [3] や、位置情報付きツイートを利用して観光ルートの推薦を行っている中嶋らの研究 [4] などがある。

本稿では投稿者の国籍が多様であり利用者も非常に多い Twitter を利用するが、Twitter を利用して観光情報を抽出している研究には位置情報付きのツイートを利用しているものが多い。橋本らの研究 [5] によると、ツイートに位置情報が付いている割合は日本国内で約 0.18% と非常に少ない。投稿者を外国人に限定した場合、十分なサンプルを集めるには非常に長い期間を要することが予想される。これは Twitter のリアルタイム性を生かそうとする場合に大きな弊害となる。

そこで本稿ではツイートの位置情報を利用せず、ツイート本文に対象の単語が含まれているか否かに注目し、観光情報の抽出を行う。

(注1) : http://www.jnto.go.jp/jpn/news/press_releases/pdf/pdf/131225-tenmillion.pdf, Accessed 2013

3. 提案手法

3.1 言語別観光先の推薦

ツイートを取得する際に同時に Twitter より取得できるユーザ情報の中に、ユーザの言語設定がある。ある言語圏からのツイートがほかに比べて多くなる地点が存在する場合、その地点はその言語圏の人たちにとって比較的興味深く、魅力的な地点であるといえるだろう。そのような地点が各言語ごとに推測できれば、外国人観光客に効率的に観光先の推薦やプロモーションが可能になる。

そこである地名を含んだツイートを言語ごとに集計し、日本語を除く全言語との比率を算出する。ツイート全体における言語の比率と比較することによって、その地域を訪れる外国人観光客の使用言語の傾向が推測できる。この情報と外国人の実際のツイート数を使用し、外国人観光客の言語別におすすめの観光先を提案する。

3.2 観光ルートの推薦

ツイートされた時刻を調べることで、観光地の時間帯別のツイートされる割合が分かる。他の場所に比べてある時間帯でのツイートが多くなる地名はその時間に訪れる観光客の割合が高いと言えるだろう。場所や近くにある施設によって、朝にツイートされる割合が高くなる地域もあれば、夕方からツイートされる割合が上昇していく地域もある。ツイートされる割合が高くなる時間帯は地域によって様々ではあるが、その時間帯が観光客にとって魅力的なものであると予想できる。また、外国人観光客が1日に複数の観光先を訪れ、その先々でツイートを行っている場合、異なる日本の地名が含まれたツイートを短時間に複数観測することができる。ツイートに含まれる地名をツイートされた時刻とともに取得することで、ある人物の移動、訪れた順番が推測できる。これらの情報を利用し、外国人におすすめな観光ルートを提案する。

外国人観光客にとって魅力的な観光ルートが推測できれば、宣伝における優先順位やツアールートの設定に活用することができる。

4. 実験方法

4.1 ツイートの取得

Twitter よりツイートを取得するため、Streaming API を使用する。キーワードを設定することにより、キーワードを含むツイートをリアルタイムに取得できる。キーワードの設定には、世界の各言語で展開されているインターネット百科事典である Wikipedia を利用する。Wikipedia には、カテゴリやテンプレートなどのある条件でまとめられたページが存在する。今回は東京 23 区内の主要地域についてまとめられたページ^(注2)をもとに、東京 23 区内の 56 地点を対象とする。この 56 地点について、Wikipedia の他言語版のページより該当地点を指す各言語での表記を取得し、キーワードとする。取得する情報は

表 1 言語一覧

言語名	略称
英語	en
中国語簡字体	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

1 件のツイートにつき、「ツイート時刻」「ユーザ名」「言語設定」「ツイート本文」の 4 つである。

4.2 言語別観光先の推薦

各地名ごとに地名を含む全てのツイート数と各言語でツイートされたツイート数の比を求め、言語ごとの比率を出す。ツイートの比率を出す言語は、インターネット上で多く利用されている 10 言語^(注3) から日本語を除いた 9 言語とした。なお、中国語を繁体字、簡体字を分けて扱っている。表 1 に言語とその略称の一覧を示す。

ある言語 l でツイートされた地名 a を含むツイートの全言語に対する比率を $A(a, l)$ とする。また、取得した日本語以外の全てのツイートにおける、言語 l でツイートされた割合を $B(l)$ 、地名 a を含む、言語 l でツイートされた総数を $C(a, l)$ とする。これらの値を使用し、言語別の地名のおすすめ度を数値化する。

数値化されたおすすめ度について、プロモーションを行う人にとって有用な情報となりうるか評価を行う。ある言語圏の観光客に対しプロモーションを行う際、やはり一番に推薦すべき場所は観光客が多く訪れる、人気のある地点だろう。また、その言語圏の観光客が訪れている割合がほかの言語圏の観光客よりも比較的大きい地点も優先的に推薦すべきだろう。その言語圏の観光客にとって魅力的な地点である可能性が高いためである。

よって、評価の基準に次の条件を設定する。

- ツイートの総数が多い地名が上位か
- 対象言語でのツイートの割合が多い地点のおすすめ度がツイートの総数が近いほかの地点より高いか

本稿では設定した条件に適するようにおすすめ度を数値化するため、以下のような複数の計算方法を提案する。対象地名が a 、対象言語が l であるときのおすすめ度を $S_1(a, l) \sim S_5(a, l)$ とし、算出の方法を次に示す。

$$S_1(a, l) = A(a, l) \cdot C(a, l) \quad (1)$$

$$S_2(a, l) = C(a, l) \cdot \cos \frac{\pi}{2\{A(a, l) + 1\}} \quad (2)$$

$$S_3(a, l) = \{A(a, l) - B(l) + 1\} \cdot C(a, l) \quad (3)$$

$$S_4(a, l) = C(a, l) \cdot \cos \frac{\pi}{2\{[A(a, l) - B(l) + 1] + 1\}} \quad (4)$$

(注2) : http://ja.wikipedia.org/wiki/Template:東京_23_区の地域, Accessed 2013

(注3) : <http://www.internetworldstats.com/stats7.htm>, Accessed 2013

表 2 時間帯番号対応表

時間帯番号 i	時間
0	6:00 ~ 10:00
1	10:00 ~ 14:00
2	14:00 ~ 18:00
3	18:00 ~ 22:00
4	22:00 ~ 翌 2:00
5	翌 2:00 ~ 翌 6:00

$$S_5(a, l) = C(a, l) \cdot \sin \frac{\pi \cdot \{A(a, l) - B(l) + 1\}}{4} \quad (5)$$

これらの計算式で算出したおすすめ度について、設定した条件に適合するかどうか評価を行い、より条件に適したものを採用する。

評価を行うにあたって、実際のツイート数の順位とおすすめ度の順位から算出した順位相関 ρ を算出する。また、ツイートの総数に 2 割以上の大きな差があり、かつおすすめ度の順位が入れ替わる回数 G_1 と、ツイートの総数の差が 1 割以下でおすすめ度の総数が入れ替わる回数 G_2 を計測する。なお、 ρ の計算には、スピアマンの順位相関係数を用いた。 ρ に正の相関があり、 G_1 の数が少なければ、ツイートの総数が多い地名がおすすめ度において上位であると言える、 G_2 の数が多ければ、対象言語でのツイートの割合が多い地点のおすすめ度がツイートの総数が近いほかの地点より高いと言える。

4.3 観光ルートの推薦

表 2 に時間帯番号対応表を示す。時間帯番号 i により 1 日を 4 時間ごとに 6 つの時間帯に分け、各時間帯でのツイート数の全体に対する割合を地名ごとに算出する。時間帯番号が i であるときの地名 a に対するツイートの割合を求め、 $M(a, i)$ とする。取得した全てのツイートに対する時間帯別のツイートの割合 $N(i)$ も同様に算出し、地名ごとに全体に対するツイートの割合で割る。この値を地名 a における時間帯 i でのおすすめ度 $O(a, i)$ とする。

$$O(a, i) = \frac{M(a, i)}{N(i)} \quad (6)$$

また、ある地名を含むツイートをしたユーザが 12 時間以内に別の地名を含むツイートをした場合に、ユーザが移動したとみなす。ある地名 a からある地名 b への移動の数を全ての移動の総数で割り、全移動に対する移動頻度 $P(a, b)$ とする。

O, P について、それぞれの最大値が 1、最小値が 0 となるように正規化したものを O', P' とした。これらを利用して、外国人観光客に対する観光ルートのおすすめ度として数値化する。

なお、観光ルートについて、次のルールを設定している。

- 観光を開始する時間帯は表 2 に示す時間帯番号で $i = 0, 1, 2$ のいずれかである。
- 必ず各時間帯で異なる地点へと移動する。
- 3 つ目の地点を訪れたところでルートを終了する。

数値化されたおすすめ度について、プロモーションを行う人にとって有用な情報となりうるか評価を行う。ルートを推薦するにあたって、最も推薦すべきルートは、移動頻度が高くなる

地点同士を結ぶルート、つまり実際に移動が多いルートであろう。そのようなルートは、外国人観光客にとって魅力的な地点同士を結んでおり、なおかつ交通の便が良いなどの理由で地点間の移動が比較的容易と考えられる。また、観光地には見学できる時間が決まっているなど、訪れるべき時間帯が存在する場合がある。外国人が多くツイートしている時間帯にその土地を訪れられるよう、ルートだけではなくスタートの時刻も提案する必要があるだろう。

よって、評価の基準に次の条件を設定する。

- 全移動に対する移動頻度が高い地点を結ぶルートが上位か
- 同ルートでも観光開始の時間帯によっておすすめ度に違いが出るか

本稿では設定した条件に適するようにルートのおすすめ度を数値化するために、以下の 4 つの計算方法を提案する。観光を開始する時間帯を i 、行き先をそれぞれ a, b, c とした場合に、各計算方法で算出したおすすめ度を $R_1(i, a, b, c)$, $R_2(i, a, b, c)$, $R_3(i, a, b, c)$, $R_4(i, a, b, c)$ として、算出の方法を次に示す。

$$R_1(i, a, b, c) = \{O(a, i) + O(b, i + 1) + O(c, i + 2)\} \cdot \{P(a, b) + P(b, c)\} \quad (7)$$

$$R_2(i, a, b, c) = O(a, i) \cdot O(b, i + 1) \cdot O(c, i + 2) + P(a, b) \cdot P(b, c) \quad (8)$$

$$R_3(i, a, b, c) = \{O'(a, i) + O'(b, i + 1) + O'(c, i + 2)\} \cdot \{P'(a, b) + P'(b, c)\} \quad (9)$$

$$R_4(i, a, b, c) = O'(a, i) \cdot O'(b, i + 1) \cdot O'(c, i + 2) + P'(a, b) \cdot P'(b, c) \quad (10)$$

これらの計算式で算出したルートのおすすめ度について、設定した条件に適合するかどうか評価を行い、より条件に適したものを採用する。

評価を行うにあたって、それぞれのルートのおすすめ度の順位と、 a から b 、 b から c への実際の移動総数の順位から順位相関 ρ を算出する。また、スタート時間が異なる場合の 1 ルートあたりのおすすめ度の順位の差 H を計測する。なお、 ρ の計算には、前節と同様にスピアマンの順位相関係数を用いた。 ρ に正の相関があれば、全移動に対する移動頻度が高い地点を結ぶルートのおすすめ度が上位であると言える、 H の値が大きければ、同ルートでも観光開始の時間帯によっておすすめ度に差が出ていると言える。

5. 実験結果

5.1 ツイートの取得

ツイートを取得する期間は、2013 年 12 月 3 日 12 時 40 分 06 秒から、12 月 24 日 12 時 40 分 06 秒までの 3 週間とした。ツイート取得に使用するキーワードは 56 地点で 536 語、日本語を除くと 480 語であった。取得したツイートの総数は全言語で 1,010,709 件、言語設定が日本語であったツイートを除くと 137,654 件であった。

表 5 各言語でのおすすめ順位 (東京)

順位	en	zh-cn	zh-tw	es	pt	de	ar	fr	ru	ko
1	渋谷	渋谷	神田	上野	神田	原宿	お台場	原宿	原宿	神田
2	原宿	銀座	渋谷	渋谷	原宿	渋谷	秋葉原	渋谷	八重洲	日比谷
3	神田	新宿	新宿	原宿	渋谷	神田	渋谷	神田	銀座	渋谷
4	新宿	池袋	秋葉原	神田	永田町	青山	神田	新宿	神田	新宿
5	銀座	青山	原宿	新宿	青山	新宿	赤坂	秋葉原	大手町	吉原
6	秋葉原	恵比寿	赤坂	秋葉原	秋葉原	六本木	新宿	銀座	渋谷	池袋
7	青山	日本橋	表参道	青山	新宿	銀座	原宿	六本木	新宿	中野
8	赤坂	下北沢	銀座	日比谷	上野	築地	築地	青山	六本木	青山
9	六本木	丸の内	池袋	六本木	六本木	秋葉原	新橋	上野	青山	秋葉原
10	表参道	表参道	六本木	お台場	池袋	吉原	池袋	永田町	秋葉原	原宿

表 3 順位相関 ρ と順位が入れ替わる回数

	順位相関 ρ	G_1	G_2
$S_1(a, l)$	0.96400	501	61
$S_2(a, l)$	0.97423	466	58
$S_3(a, l)$	0.99800	17	32
$S_4(a, l)$	0.99877	0	15
$S_5(a, l)$	0.99838	10	25

表 6 順位相関 ρ と順位の差 H

	順位相関 ρ	順位の差 H
$R_1(i, a, b, c)$	0.87860	532.29
$R_2(i, a, b, c)$	0.46147	1488.52
$R_3(i, a, b, c)$	0.93124	1381.44
$R_4(i, a, b, c)$	-0.03580	3889.42

表 4 おすすめ度 $S_4(a, l)$, $l = \text{en}$

地名 a	言語比率	ツイート数	おすすめ度 $S_4(a, l)$
渋谷	0.80004	13,503	10192.55
原宿	0.64143	13,064	9368.85
神田	0.28564	9,465	5626.09
新宿	0.77167	5,431	4065.98
銀座	0.83420	4,000	3047.92
秋葉原	0.76124	4,062	3031.61
青山	0.75265	2,982	2219.78
赤坂	0.83807	2,800	2135.76
六本木	0.80417	2,353	1778.20
表参道	0.84682	1,769	1352.47

表 7 おすすめ度 $R_3(i, a, b, c)$

時間帯 i	行き先 a	行き先 b	行き先 c	おすすめ度 $R_3(i, a, b, c)$
0	原宿	秋葉原	表参道	0.11399
2	表参道	原宿	秋葉原	0.09947
2	原宿	渋谷	新宿	0.08893
0	原宿	渋谷	新宿	0.08703
2	表参道	渋谷	新宿	0.08484
1	原宿	秋葉原	渋谷	0.08309
0	原宿	渋谷	表参道	0.08237
2	原宿	渋谷	六本木	0.08006
0	原宿	秋葉原	新宿	0.07896
0	上野	原宿	秋葉原	0.07597

5.2 言語別観光先の推薦

(1) ~ (5) 式それぞれで算出したおすすめ度 $S_1(a, l)$ ~ $S_5(a, l)$ における, ρ の平均値と G_1, G_2 を表 3 に示す.

表 3 によると, ρ はすべて 1 に近い値となっており, 大きな差はなかった. $S_1(a, l), S_2(a, l)$ と $S_3(a, l), S_5(a, l)$ を比較すると, $S_3(a, l), S_5(a, l)$ での G_2 は $S_1(a, l), S_2(a, l)$ での G_2 に比べ, 半分ほどになったが, G_1 の数には大きな差ができていた. また, $S_4(a, l)$ では, G_1 が 0 となった.

よって, 設定した条件に適合させるには, (4) 式を使用したおすすめ度の算出が適している. ただし, 言語比率の影響力は弱まるため, 言語比率の影響力を優先する場合には, (3), (5) 式を使用しても良いだろう.

(4) 式を使用した場合, 英語でのおすすめ度 $S_4(a, l)$ を表 4 に, 各言語でのおすすめ順位を表 5 に上位各 10 地名ずつ示す.

5.3 観光ルートの推薦

(7) ~ (10) 式で算出した観光ルートのおすすめ度 $R_1(i, a, b, c)$ ~ $R_4(i, a, b, c)$ における, ρ と H を表 6 に示す.

全移動に対する移動頻度が高い地点を結ぶルートが上位であ

るほど, 表 6 における ρ が 1 に近くなり, 観光を開始する時間帯によって同一ルートのおすすめ度の順位に違いが出るほど, 順位の差 H が大きくなる.

表 6 から, 最も ρ が 1 に近かったのは, (9) 式によりおすすめ度を算出した場合である. また, 順位の差 H は (10) 式でおすすめ度を算出した場合がほかに比べて大きくなったが, ρ が負の値となった.

よって, 設定した条件に合うようにおすすめの観光ルートを推測するには, (9) 式を使用するのが良いだろう.

(9) 式を使用して算出したおすすめ度 $R_3(i, a, b, c)$ を表 7 に示す.

6. 汎用性の評価

5 章で高評価を得たおすすめ度の算出式の汎用性を, 東京と同じく交通の便が良く選択枝の多い都市である大阪について同様の実験を行うことによって評価する.

東京の場合と同様に, Wikipedia の大阪市内の主要地域につ

表 10 各言語でのおすすめ順位 (大阪)

順位	en	zh-cn	zh-tw	es	pt	de	ar	fr	ru	ko
1	難波	難波	難波	難波	梅田	福島	難波	新世界	難波	難波
2	福島	日本橋	心斎橋	梅田	難波	難波	新世界	福島	新世界	新世界
3	梅田	梅田	平野	新世界	新世界	新世界	梅田	難波	福島	梅田
4	新世界	平野	梅田	福島	日本橋	心斎橋	中之島	梅田	心斎橋	道頓堀
5	心斎橋	本町	道頓堀	十三	心斎橋	梅田	福島	中之島	梅田	心斎橋
6	平野	天満	新世界	心斎橋	福島	中之島	放出	心斎橋	アメリカ村	日本橋
7	道頓堀	新世界	福島	中之島	アメリカ村	—	—	道頓堀	日本橋	平野
8	日本橋	福島	日本橋	アメリカ村	淡路	—	—	アメリカ村	茶屋町	中之島
9	アメリカ村	淡路	茶屋町	道頓堀	中之島	—	—	日本橋	千日前	千日前
10	十三	放出	十三	放出	長居	—	—	十三	中之島	福島

表 8 順位相関 ρ と順位が入れ替わる回数

	順位相関 ρ	G_1	G_2
$S_1(a, l)$	0.79095	54	3
$S_2(a, l)$	0.99846	52	3
$S_3(a, l)$	0.99979	1	2
$S_4(a, l)$	0.99980	0	2
$S_5(a, l)$	0.99980	0	2

表 9 おすすめ度 $S_4(a, l)$, $l = \text{en}$

地名 a	言語比率	ツイート数	おすすめ度 $S_4(a, l)$
難波	0.80671	2,429	1748.19
福島	0.89462	1,163	862.12
梅田	0.73113	1,172	819.21
新世界	0.45165	383	229.45
心斎橋	0.79240	271	194.02
平野	0.84426	103	75.11
道頓堀	0.55621	94	60.30
日本橋	0.64835	59	39.74
アメリカ村	0.89286	50	37.04
十三	0.76087	35	10.48

いてまとめられたページ^(注4)を使用し、大阪市内の主な地域とされる 70 地点についてすべての他言語版のページから該当地点を指す表記を取得し、ツイートを取得するためのキーワードとした。ツイートを取得する期間は、2014 年 1 月 16 日 14 時 30 分 48 秒から、2 月 6 日 14 時 30 分 48 秒までの 3 週間とした。ツイート取得に使用するキーワードは 70 地点で 101 語、日本語を除くと 31 語であった。取得したツイートの総数は全言語で 192,199 件、言語設定が日本語であったツイートを除くと 8,569 件であった。

ツイートの総数は東京で取得した場合と比べて約 19%、言語設定が日本語であったものを除くと約 6%と、全体の数も少なくなっているが、特に日本語以外でのツイート数が東京と比べ少なくなった。これは大阪の主要地域における Wikipedia の他言語版のページの数が東京の主要地域における数よりも少なく、ツイートを取得する際のキーワードがあまり設定できなかったことが原因として挙げられる。全体のツイート数についてはツ

イートの取得期間を延ばすことによって増やすことができるが、他言語でのキーワード設定については、解決策を考える必要があるだろう。

大阪における日本語以外でのツイートが少なかったため、4 章で提案した (7) ~ (10) 式によって外国人観光客向け観光ルートのおすすめ度を算出することはできなかったが、言語別観光先のおすすめ度については (1) ~ (5) 式によって算出することができた。4 章で提案した (1) ~ (5) 式によって算出された言語別観光先のおすすめ度 $S_1(a, l) \sim S_5(a, l)$ について算出した ρ , G_1 , G_2 を表 8 に示す。

表 8 より、大阪についておすすめ度を算出した場合も、東京と同じく (4) 式での算出が適合していた。また、東京にて次に条件に適していた (3), (5) 式についても条件に適合していた。

よって、東京、大阪のような主要な地域が密集している都市圏については、言語別観光先のおすすめ度の算出式について汎用性があると言える。

(4) 式を使用した場合の英語でのおすすめ度 $S_4(a, l)$ を表 9 に、各言語でのおすすめ順位を表 10 に上位各 10 地名ずつ示す。なお、上下のセルが結合されている箇所は同順位を示す。ドイツ語とアラビア語では 6 位までしか順位が算出されなかった。

7. おわりに

本稿では、外国人のツイートに出現する地名に注目し、外国人に向けたプロモーションの支援を行うための手法を提案した。その結果、東京 23 区内における言語圏別のおすすめの観光先と、外国人観光客向けのおすすめの観光ルートを推薦することができた。この結果は、外国人向けの観光プロモーションの一助となるだろう。また、算出方法の汎用性を確認するために同様の実験・評価を大阪を対象に行った。言語圏別のおすすめの観光先の推薦については汎用性の確認ができたものの、観光ルートの推薦についてはおすすめ度の算出ができず、汎用性の確認ができなかった。これは主に取得したツイート数が東京に比べ少ないことが原因であり、取得期間の延長や取得用のキーワードの設定方法の変更によって対応する必要がある。

なお、本稿では地名通りの場所を指している確率や、同一地名で複数地域を指す場合について考慮していない。これらを改善することによって、さらなる精度の向上や、あらたな推薦手

(注4) : <http://ja.wikipedia.org/wiki/Template:大阪市の地域>, Accessed 2014

法の提案が期待できるだろう。

以上を踏まえて、今後は今回出来なかったルート推薦手法の評価や推薦手法の改良，トレンドの発見，情報提供に向けた可視化などを検討する。

謝 辞

本研究の一部は，首都大学東京傾斜的研究費「関心に誘導されるデータマイニングの研究」によって行われた。

文 献

- [1] 倉島健，手塚太郎，田中克己，Blog からの街の話題抽出手法の提案，DEWS 2005.
- [2] 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 26 巻 3 号，pp.286–293.
- [3] 奥健太，橋本拓也，上野弘毅，服部文夫，地理情報推薦のための観光スポットと位置情報付きユーザ生成コンテンツの対応付け手法の提案，DEIM Forum 2013.
- [4] 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 Vol.2013-DBS-158 No.28.
- [5] 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 27 巻 4 号，pp.424–431.