

ソーシャルメディアを用いた訪日外国人へのパーソナライズされたイベント情報提供の最適化ルートの推薦

理学専攻・情報科学コース 1840648 今井 美希 (指導教員：小口 正人)

1 はじめに

近年日本を訪れる外国人観光客は急激に増加しており、2020年に開催される東京オリンピック・パラリンピックを考慮すると更なる増加が見込まれる。観光客の増加に伴い、有名な観光スポットなどの情報を載せたガイドブックやWebサイトは見受けられるようになってきた。しかしながら、それらの媒体に載っていないようなローカルな情報や今まさに開催されているイベントを知り得ることは、現状難しい。また興味のあるイベント情報を自身の手によって見つけるのは手間がかかる。そこで我々はSNSにある情報に着目した。本研究では位置情報とソーシャルストリームに基づき推定した訪日外国人の趣向によりイベント情報の順位付けを行い、その時その場で使える情報として最適なルートの提案を行う。

2 提案システム

2.1 提案システムの概要

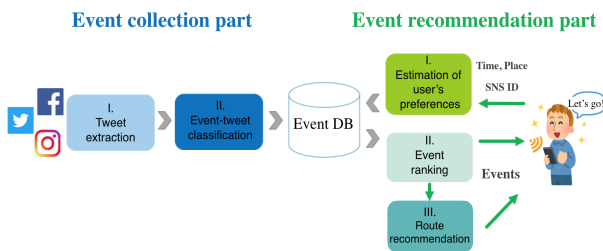


図1 提案システムの概要

観光者などに有用な情報をタイムリーにインバウンド対応で提示するために、本研究で提案するシステムの概要を図1に示す。システムは大きく分けて二つのパートからなる。

- **Event collection part**
 1. SNS から観光客にとって有用なイベント情報を抽出する
 2. イベントをカテゴリごとに分類し、保存する
- **Event recommendation part**
 1. ユーザの過去のツイートをカテゴリに分類し、分類結果によりユーザの趣向を推定する
 2. 推定した趣向、時間、位置情報によりイベントを順位づけする
 3. 順位づけしたイベントをその時、その場で使える情報としてユーザの趣向や状況にあわせてルートの提供を行う

Event collection part は工藤ら [1] によって研究が進められている。本研究では Event recommendation part の部分に焦点をあて、先行研究で抽出したイベント情報をユーザの SNS から推定した趣向に合わせて時間と場所を考慮したルートの提案を行う。

3 趣向の解析のための分類モデル

3.1 分類モデル作成

Twitter API のキーワード検索で各カテゴリごとのキーワードを設定し各カテゴリ 500 件のツイートを取得する。

カテゴリは、イベント情報の収集を行なっている先行研究に合わせて展示、映画、ライブ、舞台の4カテゴリとした。都市情報雑誌の代表格である「東京 Walker」[2] のイベントカテゴリを参考にし、予備実験で地名と日時を含むツイートの種類を分析した際に、一定数のツイートが得られると判明したカテゴリを設定してある。

取得したツイートを使い、ユーザのツイートを分類するための学習モデルを作成した。先行研究においてイベント分類で代表的な手法である SVM とランダムフォレストを比較したところ、精度、処理速度の観点からみてランダムフォレストの方が研究では優れていた為、本研究においてもランダムフォレストを使用して学習を行う。

英単語には過去分詞 (ed)、現在分詞 (ing)、複数形があり、分類の妨げとなると考えた。そこで、Tree Tagger[3] と呼ばれるパッケージを用いて形態素解析を行った。

パラメタを調整することで、よりランダムフォレストで正確な分類をできるようにする。そこでグリッドサーチという自動的な最適化ツールを使い、与えたパラメタの中で最も精度の良いものを選ぶ。学習データ用いてランダムフォレストによって学習を行った。

3.2 実験結果 (分類器の作成)

適合率 (Precision)、検出率 (Recall)、F 値 (f-score) などから見ても、高い結果が得られた。

表1 Classification accuracy

	precision	recall	f-score
Exhibition	0.79	0.83	0.81
Movie	0.96	0.82	0.89
Music	0.87	0.84	0.86
Stage	0.89	0.82	0.86
avg/total	0.84	0.83	0.83

4 イベントの順位付け

4.1 順位付け方法

1. 位置情報のみを考慮した場合
地図上の標準地域メッシュを利用し、現在地を考慮してメッシュ毎に Event score(a) をつけ順位付けを行う。図2のピンが立っているところを現在地 x とすると、Event score (a) = mesh(x) で図の該当する値を返す。

0.3	0.5	0.3
0.5	📍	0.5
0.3	0.5	0.3

図2 地図メッシュ

2. 趣向を考慮した場合
Event score (b) = (分類されたジャンルのツイート数)/(全体のツイート数) × Event score (a) とすることで、位置情報に趣向を加えて順位付けを行う。
3. メジャーなイベント、マイナーなイベント follower 数によるスコアを (2) に掛け合わせた。Event score (c) = Event score (b) × follower-score. follower 数によ

* This work is supported by Core Research for Evolutionary Science and Technology (CREST).

るスコアは follower 数による閾値を定め、予め決めたものである。

4.2 検証

ユーザに対してイベントの推薦を行った結果を示す。

1. 訪日外国人の趣向の解析

実際に日本へ観光に来た訪日外国人のツイートを用いて検証を行った。

(a) 検証データ

ツイート取得の際、ツイートの位置情報を日本国内の緯度経度に合わせ、更に言語を英語とすることで訪日外国人のツイートを取得した。今回はツイートをした人が在日外国人か、観光目的での訪日外国人か目視で判断した。外国人観光客と思われる5人分の過去のツイートを取得した。この5人をユーザと見立て、ツイートの分類を行い趣向の解析をしていく。

(b) ツイート分類

取得したツイートを作成したユーザの趣向性を判定するための分類モデルによって分類した。分類の結果を以下に示す。

表2 訪日外国人のツイート分類

	ユーザ1	ユーザ2	ユーザ3	ユーザ4	ユーザ5
展示会	274件	2048件	663件	448件	807件
映画	40件	66件	53件	48件	76件
ライブ	78件	80件	60件	61件	135件
舞台	144件	490件	236件	63件	141件

実際にユーザのツイートを目視で確認する。ユーザ1について過去のツイートを確認すると、アートに関するツイートやリツイートが多く確認でき、またミュージアムへ行ったツイートも確認できたため展示には興味があるのではないかと推測できる。また、実際に日本に滞在し観光していた際のツイートを確認すると、日本のミュージアムを観光していたツイートを確認することができたため、情報の推薦に役立つとみなすことができる。

表2の結果から、訪日外国人の趣向の傾向が展示会にあることが見受けられる。この結果を用いてツイッターの情報が無い観光客にも、訪日外国人の傾向を反映してイベントの配信を行えると考える。

2. イベントの推薦

表2のユーザ1に対してイベントの推薦を行った結果。

表3 位置情報, 趣向, メジャー, マイナー

	(1):位置情報	(2):(1)+趣向	(3):(2)+メジャー	(4):(2)+マイナー
1	ライブ 1(1.0)	展示 1(0.51)	展示 1(0.0561)	展示 2(0.0561)
2	映画 1(1.0)	展示 2(0.51)	舞台 1(0.0286)	展示 3(0.0382)
3	ライブ 2(1.0)	舞台 1(0.26)	展示 2(0.0255)	展示 1(0.0255)
4	ライブ 3(1.0)	展示 3(0.255)	ライブ 2(0.0154)	展示 4(0.021)
5	ライブ 4(1.0)	ライブ 1(0.14)	映画 2(0.0151)	ライブ 5(0.021)
6	展示 1(1.0)	ライブ 2(0.14)	舞台 2(0.0065)	舞台 2(0.0169)
7	展示 2(1.0)	ライブ 3(0.14)	映画 3(0.00525)	舞台 1(0.013)
8	舞台 1(1.0)	ライブ 4(0.14)	展示 3(0.00255)	映画 1(0.0105)
9	舞台 2(0.5)	映画 2(0.13)	映画 4(0.00245)	舞台 5(0.0091)
10	展示 3(0.5)	映画 1(0.07)	舞台 3(0.00245)	映画 5(0.00315)

表3は、左から(1):位置情報のみを考慮した場合、(2):(1)に加え趣向を考慮した場合、(3):(2)に加えメジャーなイベントを抽出した場合、(4):(2)に加えマイナーなイベントを抽出した場合である。位置情報のみを考慮したものと、それに加え趣向を考慮したものを比

べると、位置情報のみだった場合は同じスコアのもの10件中8件を占め、その中で適切に順位付けが行われていなかったものが、趣向を考慮することで本人の趣向により合ったリストになっているので趣向の考慮が有効である。趣向を考慮することで、位置情報だけでは遠いために下位にあったイベントが中位に表示されるようになった。

趣向を考慮した上で、メジャーなイベント、マイナーなイベントを考慮したものは、趣向のみでは上位10件に推薦されなかったイベントも推薦可能となった。例えば、テーブル3の下線部のイベントである。

5 ルートの推薦

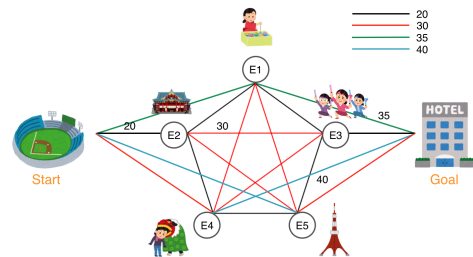


図3 Route recommendation

- 14:00 にスタート
 - 20:00 にゴール
 - イベントの滞在予定時刻はユーザが設定し、今回は60分とする
 - E1~E5はイベントであり、E1からE5の順に優先度、つまり4章での順位づけは低くなっている
1. はじめに以下のルートが最適なルートとして推薦
start - E2 - E4 - E1 - E3 - goal
しかしながらE2に予定より40分長く滞在した
 2. E4には行かないルートを推薦
E2 - E1 - E3 - goal
E1に予定より30分短い滞在であった
 3. 当初推薦されていたE4には行けませんが、E4より優先度は低いがE5には行けるルートを推薦
E1 - E5 - E3 - goal

6 まとめと今後の課題

イベントの順位付けにおいて位置情報のみと比べ、趣向を考慮した場合は、本人の趣向により合ったリストになっているので趣向の考慮が有効であると考えられる。ルート推薦においては、順位づけしたイベント情報をその時、その場で使える情報としてユーザの予定が変わった場合にも対応したダイナミックなルート推薦を行なった。

今後は、ルートの推薦における検証、検証手法を提案していきたい。

参考文献

- [1] 工藤 瑠璃子, 榎 美紀, 中尾 彰宏, 山本周, 山口 実靖, 小口 正人, "場所と時間を考慮した SNS データを用いた訪日外国人観光客へのタイムリーな情報配信", (DEIM2018), B1-3.
- [2] <http://www.walkerplus.com/spot-list/ar0313/>
- [3] <http://www.cis.uni-muenchen.de/schmid/tools/TreeTagger/>