

ソーシャルストリームを用いた訪日外国人の趣向を考慮したイベント情報の提供

今井 美希[†]工藤 瑠璃子[†]榎 美紀[‡]小口 正人[†][†]お茶の水女子大学[‡]IBM Research

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、位置情報を考慮するとともに、ソーシャルストリームに基づき訪日外国人の趣向を推定し、イベント情報を順位付けして配信を行う。

2. 提案システム

2.1 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。

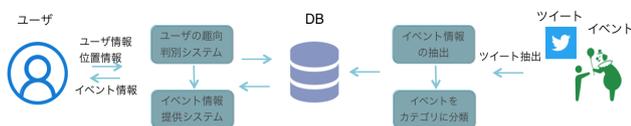


図1: 提案システムの概要

DBを介して図の右側が先行研究によって進められており、左側が本研究である。ソーシャルストリームを用いた場所と時間に対応するリアルタイムなイベント検索(DICOMO 2017) [1] という題で、工藤らがイベント情報の収集について研究を行っている。先行研究ではツイッターからイベント情報の抽出を行っている。

本研究では先行研究で抽出したイベント情報を、ツイッターに基づいて判別したユーザの趣向に合わせて、時間と場所を考慮した上で配信していく。

2.2 本研究

1. カテゴリ分類器の作成

- (a) あらかじめ、イベントのカテゴリに関する英語のツイートを取得

- (b) (1)-(a)のツイートを学習データとして、ユーザ(訪日外国人)のツイート(2)-(a)をカテゴリごとに分類する学習器を作成

2. ユーザ(訪日外国人)の情報取得

- (a) 過去のツイートを取得
(b) 現在の位置情報を取得

3. 情報提供

- (a) (2)-(a)で取得したツイートをカテゴリ分類器にかけ、趣向の解析
(b) 趣向に合ったカテゴリの情報を配信する
(c) 適合性フィードバックにより、更に個々のユーザにあったイベント情報の配信

3. 趣向の解析のための分類モデル

趣向の判定にはユーザの過去のツイートをを用いる。過去のツイートをイベントのカテゴリごとに分類し、分類結果をもとに各カテゴリの点数を求める。

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。

3.1 分類モデル作成のための学習データの収集

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、また言語を英語に設定することで英語のツイートを取得する。ツイートは各カテゴリ300ツイートずつ取得した。

3.2 分類モデル作成

3.1で取得したツイートをを使い、ユーザのツイートを分類するための学習器を作成した。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。

パラメー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なる。そこでグリッドサーチと

[†] Miki Imai, Ruriko Kudo, Masato Oguchi, [‡] Miki Enoki
Ochanomizu University ([†]), IBM Research - Tokyo([‡])

いう自動的な最適化ツールを使い、与えたパラメー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。

3.3 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。

	precision	recall	f1-score
museum	0.79	0.83	0.81
movie	0.96	0.82	0.89
concert	0.87	0.84	0.86
musical	0.89	0.82	0.86
other	0.70	0.85	0.76
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. インタラクティブに行った場合

より、ユーザそれぞれにあった情報を配信するため、適合性フィードバックを行う。

- 時間

時間を考慮してイベントを抽出。同時刻のものは更に (2) によって順位付け。

- 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 イベントの順位付け検証

あるユーザに対してイベントの推薦を行っていく。このユーザのその他に分類されたツイート以外の全ツイート数は 536 件であった。またユーザ 1 は渋谷駅にいと仮定して推薦を行う。

表 2: 場所, 趣向, メジャー, マイナー

	(1):場所	(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 2 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ものが 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 2 の下線部のイベントである。

5. 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。ただし趣向だけだと適切な場所と時間のイベントが得にくいので、場所と時間に趣向を加えたリストが、特定の時間と場所に特定の趣向を持ったユーザが居る場合の推薦リストとして最適そうである。

今後は現在地からイベント開催地までの移動時間も考慮した上で、「その時」「その場」で使えるイベント情報の配信を目指していきたい。また、アニメに関して学習データを再考すると共に、イベントデータの方もカテゴリ数に合わせて分類を行なっていく。

6. 謝辞

本研究は一部、JST CREST JPMJCR1503 の支援を受けたものである。

参考文献

- [1] "ソーシャルストリームを用いた場所と時間に対応するリアルタイムなイベント検索", DICOMO 2017 5C-5.
- [2] <http://www.walkerplus.com/spot-list/ar0313/>
- [3] <http://www.cis.uni-muenchen.de/schmid/tools/TreeTagger/>