

海外観光イベントに関する動画検索システムの開発

島田 恵輔[†] 難波 英嗣[†] 石野 亜耶[‡] 竹澤 寿幸[†]

[†] 広島市立大学大学院 情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

[‡] 広島経済大学 ビジネス情報学科 〒731-0138 広島県広島市安佐南区祇園 5-37-1

E-mail: [†] [‡] {shimada, nanba, ishino, takezawa}@ls.info.hiroshia-cu.ac.jp

あらまし 近年、海外旅行者は年々増加傾向にあり需要が高まってきている。海外旅行者が訪れようとしている場所で、滞在期間中にどのようなイベントが開催されるのかという情報は、旅行の計画を立てる時に非常に重要である。しかし、イベント関連の情報を人手で整備するのは非常にコストがかかる。また、日本国内のイベント情報と異なり、海外のイベント情報は知られていないことも多い。一般に、有名なイベントや伝統的なイベントの情報は、新聞や Web ポータルサイトや旅行ガイドブックなどから得られるものの、それらは旅行関連イベント全体のごく一部にすぎない。そこで本研究では、ある海外のイベントに関する動画を自動的に収集し、それらを地図上にマッピングすることで、ユーザが俯瞰的にイベント情報を閲覧できるイベント参加支援システムの構築を行う。

キーワード イベント情報、動画、マッピング、機械学習

1. はじめに

近年、海外旅行者は年々増加傾向にあり需要が高まってきている。海外旅行者が訪れようとしている場所で、滞在期間中にどのようなイベントが開催されるのかという情報は、旅行の計画を立てる時に非常に重要である。しかし、イベント関連の情報を人手で整備するのは非常にコストがかかる。また、日本国内のイベント情報と異なり、海外のイベント情報は知られていないことも多い。観光を支援する媒体として、旅行会社や政府の観光局などが運営する観光情報サイトや、旅行ガイドブック「るるぶ」などが挙げられる。観光情報の中でも、祭りやイルミネーション、マラソン大会などのイベントに関する情報は、旅行の計画を立てる観光客にとって重要な情報である。例えば、今から訪れようとしている場所で何が行われるのか、それがどんな様子なのか、といった情報が分かれば、旅先での行動が計画しやすくなる。しかし、Web サイトや旅行ガイドブックは基本的なイベント情報のみが提示されており、ユーザによるコメントの投稿数も少ないため、具体的にどのようなイベントなのかを把握することは困難である。

そこで本研究では、イベントに関する情報を含んだ動画に注目し、それらを地図上にマッピングすることで、ユーザが俯瞰的にイベント情報を閲覧できるイベント参加支援システムの構築を行う。動画は、観光客にとって未知なイベントでも、映像を見ることによって、どんなイベントなのかを視覚的に捉えることができる。このように、いつ、どこで、何が、どんな様子だったのかが得られるため、動画やブログエントリは

イベントに関して有益な情報源であると考えられる。本研究では、“祭り”や“イルミネーション”、“マラソン大会”などの観光客や一般の地域住民が参加できる行事や催しをイベントと定義し、イベントに関する情報を含む動画をイベント動画と定義する。

先行研究において島田ら[1]は、日本のイベント動画とイベントブログを自動的に収集し、地図上にマッピングすることでイベント参加支援システムの構築を行った。本研究では、対象を海外のイベントに広げ、海外のイベントに関する動画を自動的に収集し、それらを地図上にマッピングすることで、ユーザが俯瞰的にイベント情報を閲覧できるイベント参加支援システムの構築を行う。

本論文の構成は以下のとおりである。2 章では本研究で構築したシステムについて、その動作例を説明し、3 章で関連研究について述べる。4 章では観光イベントに関する動画の自動収集について、5 章は実験について述べる。また、結論と今後の課題については 6 章で述べる。

2. システムの動作例

本章では、構築するシステムについて、その動作例を説明する。本章で紹介するシステムは、先行研究[1]で作成したシステムである。本研究で対象とした海外のイベントを検索できるシステムは構築中である。

本システムでは、オンライン地図情報サービスである Google Maps¹を使用する。図 1 に動作例を示す。

¹ <http://maps.google.co.jp>



図 1: イベント参加支援システムの動作例

ユーザが検索語を入力することにより、検索語に関連した動画及びブログエントリが地図上にピン(図中①)として表示される。ピンをクリックすると吹き出し(図中②)が表示される。吹き出しの中には動画及びブログエントリの概要(図中③)が表示される。また、タイトルの部分をクリックすることで、動画(図中④)のリンク先にアクセスし、イベント情報を閲覧することができる。このシステムにより、あるイベントがどのぐらいの範囲で行われているかといったことがわかる。また、図の例にある花火のような「見る」ことに関するイベントの場合、地元の人にしかわからないような、いわゆる穴場がどこにあるのかといったこともわかるようになる。なお、地図上のピンはイベントごとに決められた掲載期間のみ表示させる。これによりユーザにとって不要な情報を排除することができる。

3. 関連研究

本研究は、観光情報の中でもイベントに注目し、観光支援システムの構築を目標としている。類似する研究として、Webを情報源としたイベント情報の抽出と、観光支援サービスが挙げられる。

3.1. Webを情報源としたイベント情報の抽出

Nanbaら[2]は、新聞記事からイベント情報を抽出する手法を提案している。現在より未来に開催される娯楽に関する催しや行事が記載されている新聞記事を「イベント記事」とし、新聞記事がイベント記事かを自動検出する。そして検出したイベント記事より、手がかり語の有無を素性とした機械学習を用いてイベン

ト名、開催日時、開催地、開催施設名といったイベント情報を抽出している。

吉田ら[3]は、ブログエントリとWebページからイベント情報を抽出する手法を提案している。いくつかのイベント名に対してブログエントリを収集し、そこからイベント名の前後に接続しやすいパターンを抽出する。そして、抽出したパターンを用いてWebからイベント名を収集している。この手法により、地域に根付いた「ふくろ祭り」というイベント名の抽出にも成功している。

岡本ら[4]は、ブログエントリからイベントを抽出する手法を提案している。地名を検索クエリとし、ブログ検索エンジンにより収集されたブログエントリを対象に、話題とその推移を抽出し、イベント情報の抽出を行っている。実際に、音楽グループの告知されていないイベントや、スポーツイベント、美術や写真の個展などのイベント情報が抽出されている。

金子ら[5]は、Twitterからイベント情報を抽出する手法を提案している。日本国内を分割し、それぞれの地域に対して、前日と比べ多くのユーザがツイートしたキーワードがイベントに関わるとし、キーワードを抽出する。そのキーワードを用いて収集した画像をクラスタリングし、代表画像を地図上にマッピングすることで、イベントの様子を視覚的に捉えやすくしている。台風や虹といった自然現象から、花火大会、イルミネーションなどといったイベントの抽出に成功している。

これらの研究はイベント情報を扱う点では本研究と似ている。しかし、本研究では、実際にイベントに参加した人の生の声を提示することを目的としており、

イベント情報そのものより、イベントに関する感想や意見が記述された動画とブログエントリに注目する点で異なる。

3.2. 観光支援サービス

安村ら[6]は、イベント情報をイベントの発生時間、および発生場所と定義している。イベントを処理できる地理情報システムの構築を目標に、Web からイベント情報を抽出する手法を提案している。

石野ら[7]は、旅行者が旅行記を記述した旅行ブログが観光情報を得るための有益な情報源であると考え、ブログデータベースより旅行ブログエントリを検出する手法を提案している。さらに Ishino ら[8]は、広島県の観光を支援するために、電停に関する旅行ブログエントリを検出する手法も提案している。本研究で扱うブログエントリは、この手法により検出された旅行ブログエントリを対象とする。

この他の手法を使用した観光支援サービスとして、藤井ら[9]の研究がある。広島 P2 ウォーカーで公開されている「ぶらり広島電停散歩 MAP²」では、収集した旅行ブログエントリを地図上にマッピングすることで、どこに関する旅行ブログエントリなのかが視覚的に分かるようになっている。しかし、一点に大量の旅行ブログエントリがマッピングされると、得たい情報を探ることが困難になる。そこで藤井らは、旅行者が知りたい情報を効率的に閲覧することができるよう、旅行ブログエントリを「買う」、「食べる」、「体験する」、「泊まる」、「見る」、「その他」の6つのタイプに分類する手法を提案している。イベントに関して情報を得たい場合、「見る」や「経験・体験」、または「グルメ」をクリックすると、イベントに関する旅行ブログエントリを閲覧することができる。しかし、そのイベントが現在行われているという保証は必ずしもない。そこで本研究では、提示させるイベント動画やイベントブログに掲載期間を付与し、旅行者が知りたい時期の情報を提示することを目的とする。

4. 観光イベントに関する動画の自動収集

観光イベントに関する動画とブログエントリの自動収集は、以下の3つのステップに分けられる。この3つのステップについて、4.1 節、4.2 節、4.3 節でそれぞれ説明する。

1. イベントデータベースの作成
2. 動画の収集

² <http://p2walker.jp/peace/ja/blog/>

³ https://en.wikipedia.org/wiki/Lists_of_festivals

⁴

<http://wikitravel.org/ja/%E3%83%A1%E3%82%A4%E3%83%B3%E3%83%9A%E3%83%BC%E3%82%B8>

3. イベント動画の自動判定

4.1. イベントデータベースの作成

各イベント動画に、イベント開催場所やその緯度、経度、掲載期間などのイベント情報が付与できれば、期間を決めて地図上に提示できる。そこで本研究では、イベント情報を収集したイベントデータベースを作成する。イベントデータベースに登録するイベントは Wikipedia の祭り一覧³や Wikitravel⁴から収集した。本研究では、イベント名、開催期間、掲載期間、開催場所、住所、緯度、経度をイベント情報と定義し、イベント名、開催期間、開催場所、緯度、経度を抽出した。掲載期間は開催期間1ヵ月前から終了日までとし、イベントデータベースを作成した。イベントデータベースの例を表1に示す。

表1: イベントデータベースの例

- New York Film Festival
- Independence Day Celebration Parade
- Montreal International Fireworks
- Kagurasai Ceremony
- Fiesta Carnival
- Cold Food Festival
- Dongzhi Festival
- Duanwu Festival
- Freespace Fest
- Fu Yang Festival

4.2. 動画とブログエントリの収集

本節では、イベント動画の候補となる動画の収集手法についての説明を行う。イベント動画の収集のためには、イベントについて書かれた動画が必要となる。そこで、4.1 節で作成したイベントデータベースよりイベント名を抽出し、それぞれの収集に使用した。動画の収集には、YouTubeDataAPI⁵を用い、イベント名を検索クエリとしてイベント動画候補を収集した。

4.3. イベント動画の自動判定

4.2 節では、イベント名をクエリとして動画を収集した。しかし、収集された動画には、クエリとしたイベント名が動画のタイトルや説明文に含まれてはいるが、実際にイベント参加者が撮影した動画ではない場合も含まれる。先行研究[1]では、イベント動画かどうか判定するために以下の2つの方法を利用した。

- (1) テキスト情報を用いた判定

⁵

https://developers.google.com/youtube/2.0/developers_guide_protocol?hl=ja

(2) 地理情報を用いた判定

(1) テキスト情報を用いた判定

日本語のイベント動画のタイトルや説明文に“開催”や“参加しました”など、実際にイベントへ参加した際に、よく使用される単語が含まれる。よって、以下に示すテキスト情報を使用してイベント動画の自動判定を行った。

- 動画のタイトルにイベント名が含まれているかどうか
- 動画の説明文にイベント名が含まれているかどうか
- “開催”や“会場”などのイベント開催時によく使用される単語の有無
- “参加しました”や“見に行った”などのイベントの参加者がイベントに参加した際によく使用される単語の有無

先行研究で利用した(1)テキスト情報を用いた判定に用いた手掛かり語は、日本語の動画を対象に収集したものである。

本研究では、英語でテキストが記述された動画を対象としている。人手でイベント動画であると判定された海外のイベント動画を図2に示す。図2の動画のテキストには、“Eve”というイベント動画であることがわかる単語が含まれている。このような、英語のイベント動画を検出するための手掛かり語を収集し、機械学習の素性に用いた。



図2: 人手でイベント動画と判定された動画の例

(2) 地理情報を用いた判定

動画内にある緯度経度とイベント範囲を照らし合わせるにより、地理情報からイベント動画の判定を行う。判定には上記と同じく機械学習を用いて自動

判定を行う。

図3は、「阿波おどり」というイベント名から収集した動画を地図上にマッピングした例になる。図3のように、イベント範囲外に表示された「外れ値」が存在することが確認できる。よって本研究では、この外れ値を One Class SVM を用いて検出を行う。

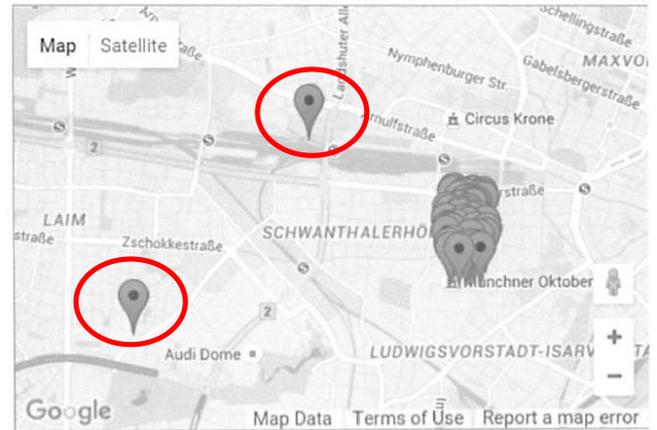


図3: イベント範囲外に存在する外れ値の例

(2)地理情報を用いた判定は、緯度経度の情報のみを利用しているため、言語に依存せず利用することが可能であるため、英語のイベント動画の収集に利用することができる。

5. 実験

本研究で行った実験とその結果について述べる。

5.1. イベント動画の自動判定

5.1.1. テキスト情報を用いた判定

データセット

4.2節で収集した動画2000件に対し、人手でイベント動画かどうかの判定を行った結果を用いた。2000件のうち、1239件がイベント動画であった。

比較手法

提案手法の有効性を確認するため、4.2節で収集した動画を、全てイベント動画として判定した場合を比較手法とした。

機械学習

イベント動画の判定の機械学習には TinySVM を用いた。2次の多項式カーネルを使用し、2分割交差検定を行った。評価尺度として、精度・再現率・F値を用いた。

実験結果

実験結果を表2に示す。表より、比較手法に比べ、提案手法では精度が0.238向上し、0.838と高い精度を得ることができた。

表 2: イベント動画の自動判定結果

手法	精度	再現率	F 値
提案手法	0.838	0.744	0.788
比較手法	0.620	1.000	0.764

5.1.2. 地理情報を用いた判定

データセット

実験データには、5 件のイベント名から収集した緯度経度付きの動画 226 件に対し、人手でイベント動画かどうかの判定を行った結果を用いた。人手でイベント動画の判定を行った結果を表 3 に示す。

表 3: イベント動画の人手での判定結果

イベント名	件数	イベント動画件数
Seoul International Fireworks Festival	46	37
Octoberfest	48	27
Hilton Arts Festival	36	16
Hollywood Film Festival	49	39
Winter Magic Festival	47	29

評価方法

提案手法の有効性を確認するため、収集した緯度経度付きの動画 226 件を、全てイベント動画として判定した場合を比較手法とした。イベント動画の判定の機械学習には One Class SVM を用い、外れ値の検出を行った。評価尺度として、正解率を用いた。

実験結果

実験結果を表 4 に示す。表より、比較手法に比べ、全てのイベントについて提案手法では高い正解率を得ている。よって、提案手法の有効性を確認できたといえる。

表 4: イベント動画の自動判定結果

イベント名	比較手法	提案手法
Seoul International Fireworks Festival	0.787	0.852
Octoberfest	0.563	0.745
Hilton Arts Festival	0.444	0.687
Hollywood Film Festival	0.796	0.821
Winter Magic Festival	0.617	0.778

考察

実験結果より、システムによってイベント動画と判定された 1161 件のうち、1038 件のイベント動画が、人手で判定した結果と一致した。しかし、人手によりイベント動画でないと判定された 638 件のうち、207 件がシステムによって誤って判定された。以下に検出誤りと再現誤りについて述べる。

- システムが誤ってイベント動画と判定した場合
人手ではイベント動画ではないと判定したが、システ

ムではイベント動画と判定した 123 件の検出誤りは、大きく次の 2 種類に分類できる。

- [i] タイトルや説明文にイベント名を含むにも関わらず、イベント動画でない動画(75%)
- [ii] イベント動画ではないにも関わらず、素性が付与されている動画(25%)

[i] タイトルや説明文にイベント名を含むにも関わらず、イベント動画でない動画(75%)

123 件のうち、93 件がタイトルや説明文にイベント名を含むにも関わらず、イベント動画出なかった。例えば、“Seoul International Fireworks Festival”を説明文に含んだ動画について、実際に動画を調べたところ、動画は花火大会後の電車の様子を撮影したものであった。

[ii] イベント動画ではないにも関わらず、素性が付与されている動画(25%)

システムがイベント動画と判定した 1161 件のうち、イベント動画と正しく判定した 1038 件の中に“fest”、“fair”、“festival”のいずれかの単語が存在する動画が 578 件、イベント動画でないと判定した 123 件の中には 30 件存在した。このような識別能力の低い語は、例えば情報利得などを用いて除外する必要があると考えられる。

- システムが誤ってイベント動画でないと判定した場合

これらの 201 件の誤りは、大きく次の 3 種類に分類できる。

- [i] 英語で表記されていない動画(80%)
- [ii] 説明文自体がない動画(16%)
- [iii] その他(4%)

[i] 英語で表記されていない動画(80%)

201 件のうち、161 件がハングルなど、英語以外の言語で表記された動画であった。

[ii] 説明文自体がない動画(16%)

201 件のうち、動画の説明文が 1 単語あるいは説明文がない動画は 21 件あった。

6. おわりに

本研究では、海外のイベントに関する動画を自動で収集する手法を提案した。また、それらを地図上にマッピングすることで、ユーザが俯瞰的にイベント情報を閲覧できるイベント参加支援システムの構築を行った。

謝辞

本研究の一部は総務省による戦略的情報通信研究開発推進制度（SCOPE）の支援を受けて行われた。

参考文献

- [1] 島田恵輔, 難波英嗣, 竹澤寿幸, 石野亜耶, “観光イベントに関する動画およびブログ検索システムの開発”, 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(DEIM2015), 2015.
- [2] Nanba, H., Saito, R., Ishino, A., and Takezawa, T. “Automatic Extraction of Event Information from Newspaper Articles and Web Pages” ICADL 2013, LNCS 8279, 171-175, 2013.
- [3] 吉田将人, 福原知宏, 増田英考, “ブログ記事とWebページを用いたイベント情報抽出手法の提案”, 情報処理学会研究報告, デジタルドキュメント2009(35), pp.37-44, 2009.
- [4] 岡本昌之, 菊池匡晃, “ブログからの地域イベント情報抽出”, 情報処理, Vol.51, No.1, pp.14-17, 2010.
- [5] 金子昂夢, 柳井啓司, “位置情報付き画像ツイートを利用した視覚的なイベント検出”, 第5回データ工学と情報マネジメントに関するフォーラム(DEIM2013), 2013.
- [6] 安村祥子, 池崎正和, 渡邊豊英, 牛尼剛聡, “blogマッピングを用いたイベント情報検出”(DEWS2007), 2007.
- [7] 石野亜耶, 難波英嗣, 竹澤寿幸, “旅行ブログエントリーからの観光情報の自動抽出”, 日本知能情報ファジィ学会誌, Vol.22, No.6, pp.667-679, 2010.
- [8] Aya Ishino, Hidetsugu Nanba, Toshiyuki Takezawa, “Construction of a System for Providing Travel Information along Hiroden Streetcar Lines”, Proceedings of the 3rd IIAI International Conference on e-Services and Knowledge Management (IIAI ESKM 2012), 2012.
- [9] 藤井一輝, 石野亜耶, 藤原泰士, 前田剛, 難波英嗣, 竹澤寿幸, “多言語旅行ブログエントリーを用いた観光情報提示システム”, 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), 2014.