

観光イベントに関する動画とブログの自動収集

島田 恵輔[†] 山本 夏生[†] 石野 亜耶[‡] 難波 英嗣[‡] 竹澤 寿幸[‡]

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

[‡] 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

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

あらまし 旅行の計画を立てる時に、「今から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」という情報は非常に重要である。しかし、開催期間が限定的な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ため、例えば、旅行会社や政府の観光局などが提供する雑誌や Web 上のポータルサイトなどでは、知名度の高いもののうちの一部を簡単に紹介するにとどまっている。そこで本研究では、あるイベントに関する動画とブログを自動的に収集し、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。動画とブログの自動収集において、動画では精度 0.855、再現率 0.834、ブログでは精度 0.862、再現率 0.794 の割合で正しく収集できた。

キーワード イベント情報、動画、ブログ、マッピング

1. はじめに

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。こういったイベントについて検索できる Web サイトとして、2006 年に、ユーザが自由にイベント情報を投稿・共有することができるソーシャルイベントサイト「eventcast¹」が、2008 年には、総合イベント情報サイト「ことさが²」が公開されている。しかし、このような Web サイト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とブログエントリに注目し、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。また、ブログエントリは、個人が自由に意見や感想を記述しており、実際にイベントに参加した人にしか知り得ない情報が記述されていることもある。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、ブログエントリをイベントブログと定義する。

また、すべてのイベント動画、イベントブログを地図上にマッピングすると、夏に行われるイベントと、冬に行われるイベントの情報が同時に提示される問題が起こる。そこで本研究では、各イベント動画やイベントブログに、“いつからいつまで地図上に掲載するか”という情報を付与する。本論文では、これを掲載期間と呼ぶ。各々のイベント動画やイベントブログに掲載期間を付与することで、観光客に「滞在期間中にどのようなイベントが開催されるのか」を効率的に提示することができる。

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するシステムについて、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とブログエントリの自動収集について、5 章は実験について述べ、6 章で考察を行う。また、結論と今後の課題については 7 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³について、その動作例を説明する。本研究では、オンライン地図情報サービスである Google Maps を使用する。図 1 に動作例を示す。

¹ <http://clip.eventcast.jp/>

² <http://cotosaga.com/>

³ <http://165.242.101.30/cgi-bin/travel/world/map.cgi>



図 1: システムの動作例

地図上に表示されたピンをクリックすることで吹き出しが表示される。吹き出しの中には、イベント動画やイベントブログのタイトル、その地点の緯度、経度が表示される。タイトルの部分をクリックすることで、動画もしくはブログエントリ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。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

齊藤ら[1]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[2]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[3]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[4]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

石野ら[5]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。さらに石野ら[6]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[7]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 MAP⁴」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、得たい情報を探すことが困難になる。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

⁴ <http://p2walker.jp/peace/ja/blog/>

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。さらに、本研究でも、藤井らのタイプ分類に合わせることにより、ユーザの要求に合ったイベント動画やイベントブログを推薦することが可能となる。

4. 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

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. 動画とブログの収集
3. イベント動画とイベントブログの自動判定

4.1. イベントデータベースの作成

各イベント動画やイベントブログに、イベント開催場所やその緯度、経度、掲載期間などといった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。イベント情報の収集には、日本全国で約43,000件のイベント情報を得ることができるYahoo!ロコ⁵を使用した。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Yahoo!ロコから抽出した。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。その例を表1に示す。

表 1: イベントデータベースの例

イベント名	2013 広島みなと夢 花火大会	神戸ルミナリエ
開催期間	2013年07月27日	2013年12月05日 ～2013年12月16日
掲載期間	20130627-20130727	20131205-20131216
開催場所	広島港 1万トンバ ース	旧居留地界限およ び東遊園地
住所	広島県広島市南区 宇品海岸3丁目	兵庫県神戸市 中 央区海岸通
緯度	34.353177623	34.6881337053942
経度	132.470301389	135.196380615234

4.2. 動画とブログエントリの収集

本節では、イベント動画やイベントブログの候補となる動画やブログエントリの収集手法についての説明を行う。イベント動画やイベントブログの収集のためには、イベントについて書かれた動画やブログエントリが必要となる。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、YouTubeDataAPI⁶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。ブログエントリの収集には、石野らの手法によって検出された旅行ブログ1,073件を対象とし、イベント名をタイトルか本文に含むブログエントリをイベントブログ候補として収集した。

4.3. イベント動画とイベントブログの自動判定

本節では、4.2節で収集した動画とブログエントリから、イベント動画とイベントブログの自動判定を行う手法について説明を行う。イベント動画の自動判定については4.3.1節、イベントブログの自動判定については4.3.2節で説明する。

4.3.1. 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。そこで本研究では、4.2節で収集された動画に対し、イベント動画かどうかを、機械学習を用いて自動判定する。

図2は、人手でイベント動画であると判定された動画の一例である。図2に示すイベント動画のように、“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。よって本研究では、機械学習に以下の素性を使用することで、イベント動画の自動判定を行う。

- 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

6

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

⁵ <http://loco.yahoo.co.jp/>



図 2: 人手でイベント動画と判定された動画の例

4.3.2. イベントブログの自動判定

イベント動画と同様に、4.2 節で収集したブログエントリーの中には実際にイベントに関係しないものも存在する。イベントブログかどうか判定するためには、以下の 2 つの方法が有用であると考える。

- (1) 日付情報を用いる判定
- (2) イベント動画との類似度を用いた判定

(1) ブログエントリーには書かれた日付が存在するため、これに注目する。イベントに参加したブログユーザは、2 週間以内にブログを書くと考え、イベント開催後 2 週間に書かれたブログエントリーをイベントブログと判定する。また、「昨年こんなイベントに参加し、今年も参加するんだ!」といったような内容の場合、イベント開催前にブログを書くと考えられるので、イベント開催前 2 週間に書かれたブログエントリーもイベントブログと判定する。このとき、「もみじまつり」のような日本各地で行われるイベントの場合、イベント開催期間を一意に決めることができない。そのため今回は、「宮島水中花火大会」のような一意に決まるイベントのみを対象とする。

(2) ブログエントリーの本文に注目し、4.3.1 節でイベント動画と判定された動画の説明文との類似度が高いブログエントリーをイベントブログと判定する。

図 3 に人手でイベントブログであると判定されたブログエントリーの例を示す。

8 月 14 日、夏の夜を 5 千発の花火で彩る今年 34 回目の「宮島水中花火大会」が行われ、宮島の島に 4 万 6 千人、対岸に 21 万人が観賞したそうです。私は交通渋滞が嫌で、今まで見に行くことさえ躊躇していましたが、フェリーに乗って海上から見物をするという企画を知り、生まれて初めて見ることができました。

図 3: 人手でイベントブログと判定されたブログエントリーの例

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

データセット

実験用データには、4.2 節で収集した動画 1,905 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 2 に示す。

表 2: イベント動画の人手での判定結果

イベント動画 (件)	その他 (件)	合計 (件)
1,398	507	1,905

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 1,905 件を、全てイベント動画として判定した場合を比較実験とした。

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・再現率を用いた。

実験結果

実験結果を表 3 に示す。表 3 の実験結果より、比較手法に比べ、提案手法では高い精度を得ることができた。イベント動画は日々撮影され、インターネット上にアップロードされるため、本研究では再現率よりも精度を重要視している。よって、提案手法の有効性を確認できたといえる。

表 3: イベント動画の自動判定結果

手法	精度	再現率
提案手法	0.855	0.834
比較手法	0.734	1.000

5.2. イベントブログ

データセット

実験用データには、4.2 節で収集したブログエントリー 533 件より、一意に決まるイベント名で収集されたブログエントリー 178 件に対し、人手でイベントブログかどうかの判定を行った結果を用いた。人手でイベントブログの判定を行った結果を表 4 に示す。

表 4: イベントブログの人手での判定結果

イベント動画 (件)	その他 (件)	合計 (件)
126	52	178

比較手法

提案手法の有効性を確認するため、4.2 節で収集したブログエントリー 178 件を、全てイベントブログとして判定した場合を比較実験とした。

判定方法

今回、イベントブログの判定には、4.3.2 節で述べた手法(1)を適用した。評価尺度として、精度・再現率を用いた。

実験結果

実験結果を表 5 に示す。表 5 の実験結果より、比較手法に比べ、提案手法では高い精度を得ることができた。イベント動画と同様に、イベントブログは日々更新されるため、本研究では再現率より精度を重要視している。よって、提案手法の有効性を確認できたといえる。

表 5: イベントブログの自動判定結果

手法	精度	再現率
提案手法	0.862	0.794
比較手法	0.708	1.000

6. 考察

本章では、イベント動画の自動抽出、イベントブログの自動抽出についてそれぞれ考察を行う。

6.1. イベント動画

イベント動画の判定において、機械学習に用いた素性はテキスト情報のみ使用しているが、人手の判定ではテキスト情報を考慮しつつ、実際に動画を見て判断している。実際に動画を見ると、テキスト情報にイベント名が含まれているがイベント動画ではない動画が存在する。その動画の例を図 4 に示す。



図 4: テキスト情報にイベント名を含むが人手でイベント動画でないと判定された例

図 4 の動画はタイトルや説明文にイベント名を含んでいる。しかし、動画の内容は舞台上上がる前の練習風景を撮影しており、イベント動画としては不適切といえる。このような動画は実際に閲覧してみないとイベント動画かどうかの判定をすることができない。そこで今後の課題として、テキスト情報だけでなく視覚情報も考慮してイベント動画の判定を行う必要があると考えられる。

6.2. イベントブログ

イベントブログの自動判定の結果について、(1)提案手法により誤って抽出した例と、(2)提案手法により抽出できなかった例について分析を行う。

(1) 提案手法により誤って抽出した例

人手ではイベントブログでないと判定したが、提案手法によってイベントブログと判定された例を図 5 に示す。

長崎空港に着くと長崎名物「ちゃんぽん」と「かすてら」がお出迎えです(^^)ランタンフェスティバルの期間だからからランタンも飾られていましたよ～空港からはバスで長崎市内へと向かいます。約 35 分で市内に到着して、まずホテルに荷物を置きに行きます
*****略*****
次の日記からランタンフェスティバルについて書きたいと思います(^^)

図 5: 誤ってイベントブログとされた例

図5は、そのブログエントリーにはイベントの様子は書かれていないが、「ランタンフェスティバル」というイベント名を含んでいるため、誤って判定されたと考えられる。

(2) 提案手法により抽出できなかった例

人手ではイベントブログと判定したが、提案手法によってイベントブログでないとして判定された例を図6、図7に示す。

去年の2月のことになりますが、家族で毎年恒例の宮島かき祭りに行ってきました。お目当ては、かきの浜焼き体験です。場所を借りて、殻付きのかきを自分で焼いて食べるものです。15個くらい入っているものが一パックで1000円でした。うちは、家族3人で三パックも食べました。つまり1人平均15個食べたこととなります(汗)。

*****略*****

図6: イベントブログと判定されなかった例1

待ちに待った9月です。何を待っていたの?.....夏休み終了を待ちました。やっと静かになった我が家、ゆっくり花火写真を整理♪8月14日開催の宮島の水中花火をご紹介～。宮島水中花火大会は関門海峡花火の翌日なんです。私、地元なので午後ゆっくり撮影場所に移動!

*****略*****

図7: イベントブログと判定されなかった例2

図6、図7は、どちらも「宮島かき祭り」、「宮島水中花火大会」というイベントに関するブログエントリーである。しかし、ブログエントリーが書かれた日付がイベント開催時と離れてしまっているため、今回の手法ではイベントブログと判定されなかった。このようなブログエントリーに対しては、本文の内容に注目し、イベント動画の説明文との類似度を測ることで、正しく判定を行うことができると考える。

7. おわりに

本研究では、イベント参加支援システムを構築することを目標に、イベントに関する動画とブログの自動収集を行った。機械学習や日付情報を用いた自動判別方法を提案し、動画については精度0.855、ブログについては精度0.862でイベント動画、イベントブログを検出できた。今後の課題として、イベント動画の判定に関しては、テキスト情報だけでなく視覚情報も考慮してイベント動画の判定を行い、イベントブログの判定に関しては、日付情報だけでなく本文の内容を用

いて判定を行うことが挙げられる。

また、全国各地で行われる同じイベント名のイベントに対しても、開催場所や住所などといった他のイベント情報を用い、イベント動画やイベントブログを適切なイベントと紐付ける必要がある。そして、イベント動画やイベントブログに「体験する」や「見る」といったタイプを付与することにより、実際に身体を動かして楽しみたい人には「トライアスロン」や「マラソン大会」など、身体は動かさず、何かを見て楽しみたい人には「花火大会」や「イルミネーション」など、ユーザの要求に合ったイベント動画やイベントブログを推薦することが可能になる。

参考文献

- [1] 齊藤隆太, 石野亜耶, 難波英嗣, 竹澤寿幸, “新聞記事と Web からのイベント情報の自動抽出”, 第5回 Web とデータベースに関するフォーラム (WebDB Forum), 2012.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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”, 日本知能情報フレンジ学会誌, Vol.22, No.6, pp.667-679, 2010.
- [6] 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.
- [7] 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.