

画像共有サイトを利用した実世界イベントの検出

組橋 祐亮[†] 新田 直子[†] 馬場口 登[†]

[†] 大阪大学大学院工学研究科 〒 565-0871 大阪府吹田市山田丘 2-1

E-mail: †kumihashi@nanase.comm.eng.osaka-u.ac.jp, ††{naoko,babaguchi}@comm.eng.osaka-u.ac.jp

あらまし 本稿は、画像共有サイトに公開された画像に付与された、撮影時間・場所・画像の意味内容を表すテキストタグを用いた、実世界に発生するイベントの検出手法を提案する。イベントの中には、意味的に関連が深く、同一時区間に異なる場所で発生するものが多く存在する。本研究では、これらのイベント集合をイベントクラスとみなし、まず画像共有サイト上の画像から、撮影時間・テキストタグの類似度に基づきイベントクラスを検出する。次に、イベントクラスを表す画像集合から、撮影場所の類似度に基づき、個々のイベントを検出する。このような2段階の処理により、撮影画像の少ないイベントの検出を目指す。

キーワード イベント検出, 画像共有サイト, イベントクラス

Real-World Event Detection Using Image Sharing Site

Yusuke KUMIHASHI[†], Naoko NITTA[†], and Noboru BABAGUCHI[†]

[†] Graduate School of Engineering, Osaka University
2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan

E-mail: †kumihashi@nanase.comm.eng.osaka-u.ac.jp, ††{naoko,babaguchi}@comm.eng.osaka-u.ac.jp

Abstract This paper proposes a real-world event detection method by using the time and location information and text tags attached to the images in the photo sharing site. Semantically related events often occur around the same time at different locations. Considering a group of these events as an *event class*, the proposed method firstly detects event classes from all images in the photo sharing site based on the similarity of their captured time and text tags. Then, from the images consisting each event class, events are detected based on the similarity of their captured locations. Such two-step approach enables us to detect events where a small number of images were captured.

Key words Event Detection, Photo Sharing Site, Event Class

1. はじめに

実世界においては、さまざまな場所、時間に多様なイベントが発生する。実世界をセンサにより観測し、これらのイベントを自動的に検出する研究が数多く進められている。例えば、道路で発生する渋滞や速度違反は、道路に設置した監視カメラやレーダーを用いた、走行している車の台数、スピードの計測や、定常状態から逸脱した異常状態の検知により検出することができる。また、地震は、全国各地に設置された地震計を用いた地面の揺れの大きさの計測により検出される。このように、実世界イベントの検出には、イベントが発生すると考えられる実世界の特定の地点に、検出対象のイベントに応じたセンサを設置することが不可欠であり、多様なイベントを多くの場所で検出するためには多大なコストが発生する。

そこで近年、人間がマイクロブログや画像共有サイトをはじ

めとするソーシャルメディアを通して、各自が実世界を観測して得たさまざまな情報を観測時間、場所と共に公開することに着目し、ソーシャルメディア上の情報から実世界の情報を獲得する研究が注目されている。人間は実世界の多くの場所に存在し、五感を用いて感知したデータの意味を解釈する機能を持つため、このように人間をセンサ (Citizen Sensor) [6] とみなして利用することは、センサ設置のコストを抑えた上で、多様な情報を取得できるという利点がある。

ソーシャルメディアを用いた実世界情報獲得に関する主な既存研究として、特定の場所に特徴的な情報を取得するものが挙げられる。例えば、Kennedy ら [3] や Zheng ら [9] は、多くの人が観光などの際に有名なランドマークを撮影して画像共有サイトに公開することに基づき、ランドマーク検出手法を提案した。ランドマーク画像は時間に依存せず特定の場所で撮影され、ランドマークを表すテキストタグが付与されることが多いため、

各テキストタグが付与された画像の撮影場所の分布に基づき、ランドマークを表すテキストタグ、画像を検出することができる。また、Yinら[8]は、多くの人が、車や料理、アクティビティなどに関して各地で撮影した画像を画像共有サイトに投稿することに着目し、各地に特徴的な車や料理、アクティビティの種類などを検出する手法を提案した。

さらに、特定の場所・時間に特徴的な情報を取得することにより、イベント検出を試みる研究も進められている。例えば、Sakakiら[5]は、多くの人間が地震を経験した際に地震が起こったことをTwitterに投稿することに着目し、地震の震源地の検出手法を提案した。これは、検出イベントを予め地震と特定し、Twitterに投稿された地震に関連したキーワードを用いたツイートの投稿場所を追跡することにより実現できる。これに対し、Rattenburyら[4]やChenら[2]は、多くの人が集まり画像を撮影するようなイベントでは、ランドマークと同様に、イベントが発生した時間・場所で撮影された画像にイベントを表すテキストタグが付与されて画像共有サイトに投稿されることから、各テキストタグが付与された画像の撮影場所及び撮影時間の分布に基づき、イベントを表すテキストタグ、画像を検出する手法を提案した。この手法では、あらかじめ検出対象となるイベントを特定する必要はないが、検出できるイベントは多くの人が参加し撮影するようなものに限られる。

ここで、イベントの中には、意味的に関連の深く、同時期にさまざまな場所で発生するものが多く存在する。例えば、春には桜や卒入学、夏には花火や海水浴、秋には運動会や紅葉、冬にはクリスマスや正月に関連したイベントが各地で発生する。各イベントを撮影した画像は少ない場合でも、これらの関連した複数のイベントが、類似した時間に撮影され、共通するテキストタグが付与されることが期待される。そこで本研究では、このようなイベントの関連性に着目した2段階の処理により構成されるイベント検出手法を提案する。提案手法は、意味的に関連するイベント集合をイベントクラスと捉え、まず撮影時間とテキストタグの類似した画像集合を抽出することによりイベントクラスを検出する。さらに抽出した各画像集合から、撮影場所が類似した画像集合を抽出することにより個々のイベントを検出する。複数の関連イベントにより構成されるイベントクラスを抽出する段階を踏むことにより、撮影画像の少ないイベントについても検出可能と期待される。

2. 2段階クラスタリングによるイベント検出

本研究は、画像共有サイトに投稿された N 枚の画像 $I_n (n = 1, \dots, N)$ からなる画像集合 P から、実世界で発生したイベント $e_i (i = 1, \dots)$ 、及び意味的に関連するイベントから構成されるイベントクラス $ec_j = \{e_{j1}, e_{j2}, \dots\} (j = 1, \dots)$ を検出することを目的とする。

イベント e_i は、特定の時間に特定の場所で発生する出来事と定義でき、発生時間 e_i^t 、発生場所 e_i^l 、意味内容を表すキーワード集合 e_i^W で表すことができる。ここで、イベントの中には前述のように、花見、入学式、海水浴、花火、紅葉狩り、成人式など同時期にさまざまな場所で発生するものが多い。そこ

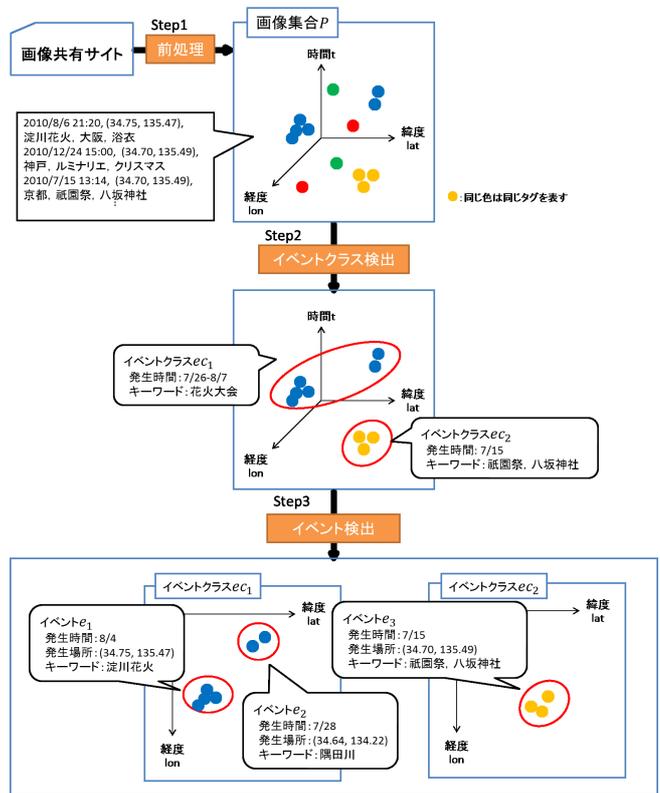


図 1 提案手法の概要

で、このような同時期に異なる場所で発生する意味的に関連するイベント集合をイベントクラスと定義すると、イベントクラス ec_j は、発生時間 ec_j^t 、及び意味内容を表すキーワード集合 ec_j^W で表すことができる。よって、本研究では、イベント $e_i = (e_i^t, e_i^l, e_i^W)$ 、イベントクラス $ec_j = (ec_j^t, ec_j^W)$ を抽出することを、イベント、及びイベントクラス検出とみなす。

画像共有サイトには、画像 I_n が、撮影時間 t_n 、撮影場所の緯度経度 $l_n = (lat_n, lon_n)$ 、投稿者名 u_n と共に、画像の内容を表すテキストタグの集合 $W_n = \{w_1, w_2, \dots\}$ がメタデータとして付与されて投稿される。ここで、意味的に関連するイベントが複数の場所で発生した場合、投稿者、撮影場所共に異なるが、撮影時間、テキストタグは類似した画像が投稿されると推察される。さらにこれらの画像のうち、各イベントで撮影された画像は、投稿者が異なっても、撮影場所は類似すると考えられる。そこで提案手法は、まず画像集合 P から、撮影時間、テキストタグが類似し、かつ投稿者が異なる画像集合 $P_{ec_j} (j = 1, \dots)$ をイベントクラス ec_j に属するイベントにおいて撮影された画像として抽出し、 P_{ec_j} を構成する画像に付与された撮影時間、テキストタグから、イベントクラス ec_j の発生した時間 ec_j^t 、及び意味内容を表すキーワード集合 ec_j^W を抽出する。次に P_{ec_j} から撮影場所の類似した画像集合 $P_{e_i} (i = 1, \dots)$ を抽出し、 P_{e_i} を構成する画像に付与された撮影時間、撮影場所、テキストタグから、イベント e_i の発生した時間 e_i^t 、場所 e_i^l 、及び意味内容を表すキーワード集合 ec_i^W を抽出する。

提案手法は図1に示すように、以下のステップにより構成される。

Step1) 前処理 :

画像共有サイトから取得した画像集合から、予め不要と考えられる画像やテキストタグを削除し、残ったテキストタグのついた画像 $I_n = (t_n, l_n, u_n, W_n) (n = 1, \dots, N)$ を画像集合 P とする。

Step2) イベントクラス検出 :

画像集合 P から、撮影時間、テキストタグの類似度に基づき、イベントクラス $ec_j (j = 1, \dots)$ に属するイベントにおいて撮影された画像集合 P_{ec_j} を抽出し、 $ec_j = (ec_j^t, ec_j^W)$ を決定する。

Step3) イベント検出 :

Step2) で抽出した画像集合 P_{ec_j} から、撮影場所の類似度に基づき、イベント $e_i (i = 1, \dots)$ において撮影された画像集合 P_{e_i} を抽出し、 $e_i = (e_i^t, e_i^l, e_i^W)$ を決定する。

各ステップの詳細を次節以降で述べる。

2.1 前処理

類似した撮影時間と撮影場所、テキストタグが付与された同一投稿者により投稿された画像は、イベントクラス検出において誤検出の要因となるため、予め削除する必要がある。そこで各画像に対し、以下の条件に該当する2枚の画像 I_α と I_β を同一の画像とみなし、画像 I_n として統合する。

$$u_\alpha = u_\beta \quad (1)$$

$$|t_\alpha - t_\beta| < TH_t \quad (2)$$

$$|l_\alpha - l_\beta| < TH_l \quad (3)$$

$$\frac{|W_\alpha \cap W_\beta|}{\min(|W_\alpha|, |W_\beta|)} > TH_W \quad (4)$$

ただし、画像 I_n の撮影時間、撮影場所は画像 I_α 、 I_β の平均とし、 I_n のテキストタグ集合は、 I_α と I_β のテキストタグ集合の和集合とする。

また、画像に付与されているテキストタグのうち、特定の時間にもみ発生するイベントクラスの内容を表すものとして不適切と考えられるほど多くの画像に付与されたものは、イベントクラス検出の精度及び処理速度向上のために、予め不要な情報として削除しておくことが望ましい。そこで、 $r\%$ 以上の画像に付与されているテキストタグを一般語として除去する。

以上の二つの処理後残った N 枚の画像 $I_n = (t_n, l_n, u_n, W_n) (n = 1, \dots, N)$ を画像集合 P とする。

2.2 イベントクラス検出

同時期に異なる場所で発生する意味的に関連するイベント集合をイベントクラスと定義すると、同一のイベントクラスに属するイベントにおいて撮影された画像は、投稿者、撮影場所共に異なるが、類似した撮影時間、テキストタグが付与されると考えられる。よって、画像集合 P から、撮影時間、テキストタグが共に類似した画像集合を抽出し、各画像集合内の画像に付与された撮影時間、テキストタグから、イベントクラスの発生時間、内容を表すキーワード集合を決定する。次に、ここまでの処理の結果、イベントにおいて撮影されたものでない画像集合が誤抽出される可能性があるため、画像集合のフィルタリングを行う。画像集合抽出、フィルタリングの2つの処理について以降の節で述べる。

2.2.1 画像集合抽出

まず、画像集合 P から類似した撮影時間、テキストタグを持つ画像集合 $P_{ec_j} (j = 1, \dots)$ を抽出するが、 P にはイベントクラスに属さない画像も多く存在し、抽出する画像集合の数は未知である。そこでSCAN [7] と呼ばれるグラフクラスタリング手法を P に適用する。SCAN は予めクラスタ数を与える必要がなく、どのクラスタにも属さないデータ（以下、孤立点と呼ぶ）の存在を許しながら、データをノード、データ間の関係をエッジで表したグラフから関係の深いノード集合をクラスタとして抽出するクラスタリング手法である。

まず画像 I_n をノード、画像間の類似度を対応するノード間のエッジの重みとすることにより、画像集合 P をグラフで表す。ここで、同一イベントクラスに属するイベントにおいて撮影された画像は、撮影された人物の名前、撮影したカメラの種類など、投稿者ごとに異なるテキストタグと共に、イベントクラスを表すキーワードのように多くの投稿者間で共通するテキストタグが付与されていると考えられる。また、イベントクラスに属するイベントは特定の時間に発生するため、投稿者が異なっても画像の撮影時間は互いに近い。よって、画像間の類似度は、撮影時間が非常に近く、かつ共通するテキストタグを持つ2枚の画像に対し高くなればよい。そこでまず撮影時間の類似度を、2枚の画像の撮影時間の差が非常に小さい時のみ高く、大きくなるほど低くなるよう指数関数により設定する。次にテキストタグの類似度を、2枚の画像に共通するテキストタグの有無とする。最後に、画像間の類似度をテキストタグの類似度と撮影時間の類似度の積とすることにより、共通したテキストタグを持つ画像間に対してのみ、撮影時間の近さに応じた類似度が設定される。つまり、画像 I_n 、 I_m 間のエッジの重みは次式で表される。

$$E(I_n, I_m) = \begin{cases} \exp\left(-\frac{|ct_n - ct_m|}{\frac{1}{N^2} \sum_n \sum_m |ct_n - ct_m|}\right) & (W_n \cap W_m \neq \emptyset) \\ 0 & (W_n \cap W_m = \emptyset) \end{cases} \quad (5)$$

次に得られたグラフに対しSCANを適用する。ここでSCANは、エッジに対する重みがないグラフを対象としたクラスタリング手法で、以下の式で定義されるノード a 、 b 間の構造的な類似度を用いる。

$$\sigma_{unweighted}(a, b) = \frac{|\Gamma(a) \cap \Gamma(b)|}{\sqrt{|\Gamma(a)| |\Gamma(b)|}} \quad (6)$$

ただし $\Gamma(a)$ は、ノード a の自身を含む隣接ノード集合であり、 $\sigma_{unweighted}(a, b)$ はノード a 、 b がどれだけ多くの隣接ノードを共有するかを表す。SCANは、 $\sigma_{unweighted}(a, b) \geq \epsilon$ となる隣接ノード b を μ 個以上持つノード a を中心に、共有する隣接ノードが多いノード同士を同じクラスタとしてまとめる。しかし、エッジに対する重みが存在するグラフには、式 (7) をこのまま用いることができない。そこで、ノード a 、 b 間の構造的な類似度を、次式に示すようにエッジの重みに基づき設定し、SCANを重み付きグラフに適用できるよう拡張する。

$$\begin{aligned} & \sigma_{weighted}(a, b) \\ &= \frac{2E(a, b) + \sum_{c \in \{\Gamma(a) \cap \Gamma(b)\}} \min(E(a, c), E(b, c))}{\sqrt{\sum_{x \in \Gamma(a)} E(a, x) \sum_{y \in \Gamma(b)} E(b, y)}} \quad (7) \end{aligned}$$

ただし $E(a, b)$ は、ノード a とノード b 間のエッジの重みである。 $E(a, b)$ が 0 もしくは 1 のみとなる、つまり、エッジに対する重みが存在しない場合、式 (6) と式 (7) は等価となる。画像集合 P から $\sigma_{weighted}(I_n, I_m)$ を用いた拡張 SCAN により得られたクラスタをイベントクラス候補 \hat{e}_{c_k} の画像集合 $P_{\hat{e}_{c_k}} (k = 1, \dots)$ とする。

次に、画像集合 $P_{\hat{e}_{c_k}}$ 内の画像に付与された撮影時間、テキストタグから、イベントクラス \hat{e}_{c_k} の発生時間 $\hat{e}_{c_k}^t$ 、内容を表すキーワード集合 $\hat{e}_{c_k}^W$ を決定する。ここでは、画像の撮影時間の最小値と最大値で表される範囲をイベントクラスの発生時間、画像集合の $s\%$ 以上の画像に付与されたテキストタグ集合をキーワード集合とする。

2.2.2 フィルタリング

2.2.1 節で得られた画像集合には、以下の要因からイベントにおいて撮影されていないものが含まれる可能性がある。

まず、2.1 節で述べたように、同一投稿者が投稿した類似した撮影時間、テキストタグを持つ画像集合はイベントにおいて撮影されたものでない可能性が高い。このような画像は前処理のような閾値を用いた処理では完全に除かれず、画像集合として誤検出される可能性がある。よって、投稿者が 1 人の画像集合が得られた場合、候補から削除する。

次に、画像共有サイトへの投稿数は時期によって異なり、特定の時期を撮影時間とする画像が他の時期と比較して特によくある場合がある。この時、さまざまな時間に撮影された画像に付与されたテキストタグであっても、画像の投稿数が多い時期を撮影時間とする多くの画像に付与されていると判断され、これらの画像集合が誤抽出されやすくなる。このようなテキストタグは、2.1 節で説明した一般語に相当すると考えられるが、前処理で行った単純な削除方法ではすべて削除することはできない。ただし、画像の投稿数が他より多い時期は複数回存在し得るため、同じテキストタグが多くの画像に付与された画像集合が複数個抽出されると考えられる。これに対し、各イベントクラスに属するイベントにおいて撮影された画像集合内で多くの画像に付与されているテキストタグは、該当するイベントクラスの意味内容を表すものであり、他の画像集合内の画像には付与されない可能性が高い。そこで、キーワード集合 $\hat{e}_{c_k}^W$ を用いて、次式で表すように、各画像集合候補 $P_{\hat{e}_{c_k}}$ の適切さを表すスコアを算出する。

$$score_{\hat{e}_{c_k}} = \max_{w \in \hat{e}_{c_k}^W} \left(\log \frac{K}{K_w} \right) \quad (8)$$

ただし、 K と K_w は、それぞれ画像集合数、タグ w が付与された画像を含む画像集合数である。つまり、 $score_{\hat{e}_{c_k}}$ は、画像集合候補 $P_{\hat{e}_{c_k}}$ が固有なキーワードを持たないとき低くなる。よって、 $score_{\hat{e}_{c_k}} \leq Th_s$ となる画像集合を削除する。

以上の二つの処理後、得られた画像集合をイベントクラス

$ec_j (j = 1, \dots)$ に属するイベントにおいて撮影された画像集合 $P_{ec_j} (j = 1, \dots)$ とする。

2.3 イベント検出

イベントには、祇園祭のように、他のイベントと意味的に関連せず、単独で発生するものも存在する。そのため、2.2 節の処理により検出されたイベントクラス ec_j には、複数地点で発生した複数のイベントを含むものと、単一地点で発生した単一イベントに相当するものがある。さらに、複数地点で発生するイベントの中には、花見のように、各イベントにおいて多くの人に画像が撮影されるものと、運動会のように、各イベントにおいてはそれほど多くの人に画像が撮影されないものが存在する。そこで、画像集合 P_{ec_j} に対し、撮影場所の類似度に基づきクラスターリングした結果により、3 種類の方法でイベントを検出する。

まず画像集合 P_{ec_j} を 2.2.1 節で説明したようにグラフで表す。各イベントにおいて複数の人が画像を撮影した場合、その撮影場所は類似すると考えられるため、以下のように計算される、画像 I_n と I_m の撮影場所の類似度をエッジの重みとする。

$$\begin{aligned} E(n, m) &= \exp\left(-\frac{\sqrt{(\text{lat}_n - \text{lat}_m)^2 + (\text{lon}_n - \text{lon}_m)^2}}{\frac{1}{N^2} \sum_n^N \sum_m^N \sqrt{(\text{lat}_n - \text{lat}_m)^2 + (\text{lon}_n - \text{lon}_m)^2}}\right) \quad (9) \end{aligned}$$

得られたグラフに拡張 SCAN を適用する。得られたクラスタが 1 つの場合、そのクラスタを P_{e_i} とする。複数のクラスタが得られた場合、各クラスタを P_{e_i} とする。また、クラスターリングの結果、孤立点が検出された場合、各画像を P_{e_i} とする。

次に、画像集合 P_{e_i} を構成する画像の撮影時間、場所、テキストタグ集合からイベント e_i の発生時間 e_i^t 、発生場所 e_i^l 、意味内容を表すキーワード集合 e_i^W を決定する。ここでは 2.2.1 節と同様に、画像の撮影時間の最小値と最大値で表される範囲をイベントクラスの発生時間、画像の撮影場所の平均値を発生場所とし、キーワード集合は、 P_{e_i} が画像集合の場合、 $s\%$ 以上の画像に付与されたテキストタグ集合、孤立点であるとき、画像に付与されたテキストタグ集合とする。ただし、各イベントに特有なテキストタグをキーワードとするため、イベントクラス ec_j のキーワード集合 ec_j^W は除く。なお、 $P_{e_i} = P_{ec_j}$ の場合は、撮影場所のみを設定すればよい。

3. 評価実験

画像共有サイトとして広く普及している Flickr [1] から、2011 年 1 月 1 日から 2011 年 12 月 31 日に、緯度が 33.5 度から 35.5 度、経度が 134.5 度から 136.5 度、つまり、図 2 に示すエリアで撮影され、少なくとも 1 つのテキストタグが付与された画像を収集した。撮影時間と撮影場所は画像に付与されたメタデータを参照した。収集した画像集合は、3,287 人が投稿した 108,477 枚の画像から成り、画像に付与されたタグの総数は 808,586 個、固有なタグ数は 35,521 個であった。

この画像集合に、 $Th_t = 3$ 日、 $TH_l = 0.1$ 度、 $TH_w = 0.5$ 、 $r = 10\%$ とし、2.1 節で記述した前処理を行った。削除された



図 2 対象エリア

タグと付与された画像の割合を表 1 に示す. これら 8 個のタグの削除により, 画像に付与されたタグの総数は 126,954 個となり, 約 84%のタグが削除された. また画像の総数は 15,110 枚となり, 約 86%の画像が削除された.

イベントクラス, イベント検出の有効性の評価を行うため, イベントクラス及びイベントを手で設定し, 上記の画像集合から各イベントに関連すると考えられる画像を選別し, 評価用データベースを作成した. ここでイベントには, 花見のように発生期間があいまいで比較的長期間のものと, 花火大会のように発生期間が明確に定まっているものが存在する. また, 紅葉狩りのように複数地点で発生しイベントクラスを構成するものと, 祇園祭のように単一地点で発生し, 他のイベントとは独立したものが存在する. そこで, 表 2 に示すように, 発生時間の期間, 発生場所の地点数が多様なイベントを設定し, 各イベントの発生時間と発生場所で撮影された画像, 及び各イベントを表すものと考え設定したテキストタグが付与された画像を抽出し, この和集合を画像集合 P とした. このように画像を抽出することにより, 撮影時間と場所は同じであるがイベントに関連しない画像, イベントに関連するがイベントを表すテキストタグが付与されていない画像, イベントを表すテキストタグが付与されているが, 撮影時間と場所がイベントの発生時間と場所と一致しない画像などのノイズが多く存在する画像集合が作成される. 表 2 に, 各イベントに対して抽出された画像数及びイベントを表すテキストタグが付与された画像数を示す. 例えば, 造幣局の花見に対しては, 4/14~4/20 の間に緯度 34.697 度, 経度 135.524 度を中心とした半径 5km の領域で撮影された画像, 及び, 花見, 桜, hamami, sakura, cherry のいずれかと造幣局というテキストタグが同時に付与された画像もしくは通り抜けというテキストタグが付与された画像を抽出し, 計 95 枚の画像を得た. このうち, 上記のテキストタグが付与され

表 1 削除されたテキストタグ

テキストタグ	割合	テキストタグ	割合
japan	35.44%	square	19.08%
iphoneography	19.02%	squareformat	18.70%
instagramapp	18.63%	uploaded:by=instagram	18.57%
kyoto	16.74%	osaka	11.89%

た画像は 6 枚であった.

画像集合 P に対し, イベントクラス, イベント検出を行い, 表 2 で設定したイベントクラス, イベントが正しく検出できるか検証する. 以下では, 設定したイベントクラス, イベントを正解イベントクラス, 正解イベントと呼ぶ. 検出結果は, 以下に示す適合率 $Precision$ と再現率 $Recall$ で評価する.

$$Precision = \frac{N_{r1}}{N_a} \quad (10)$$

$$Recall = \frac{N_{r2}}{N_c} \quad (11)$$

ただし, N_{r1} , N_{r2} は正しく検出されたイベントクラス, もしくはイベントの数を表す. イベントクラス検出においては, 検出された各クラス内の 10%以上の画像に正解イベントクラスに含まれる各イベントのテキストタグが付与されているとき, このイベントクラスを正しく検出したものとみなす. また, イベント検出においては, 検出された各クラス内の 10%以上の画像, もしくは検出された孤立点に, 正解イベントのテキストタグが付与されているとき, このイベントを正しく検出したとみなす. ここで, 1つの正解イベントクラス, もしくは正解イベントに対して, 2つ以上のクラスが検出される場合がある. この場合, 適合率では検出された複数のクラスすべて正解とみなした. よって, N_{r1} は重複を許し, 正しく検出されたイベントクラス, もしくはイベント数である. また, 再現率は正解イベントクラス, もしくは正解イベントのうち正しく検出されたイベントクラス, もしくはイベント数を表すため, 複数のクラスが検出された場合でも正解数は 1つとなる. よって, N_{r2} は N_{r1} から重複して検出されたイベントクラス, イベントを除去した数を表す. また, N_a は検出されたイベントクラス, もしくはイベントの数, N_c は正解イベントクラス, もしくは正解イベントの数である.

3.1 イベントクラス検出結果

SCAN は ϵ と μ の 2つのパラメータにより, 得られるクラスが影響される. ϵ を大きくすると, 高い類似度を持つノードのみが隣接するとみなされるため, 孤立点が増える傾向にある. 逆に小さくすると, 類似度がそれほど高くないノードも隣接するとみなされ, 多くのノードが一つのクラスにまとめられる. 本研究では, 多数の孤立点を含むと考えられるデータから類似度の高い小さなクラスを見つけることが目的であるため, ϵ は大きく設定する必要がある. また, μ を大きくするほど, クラスタを形成するために, 類似度が高いノードが多く必要となる. 従って, 投稿者数が少ないイベントでも検出できるようにするためには, μ はある程度小さいことが望ましい. そこで, 今回の実験では, $\epsilon = 0.7$, $\mu = 4$ に設定した.

画像集合 P に対するクラスタリング結果から, 投稿者が 1 人

表 2 設定したイベントとその画像集合

イベント クラス名	イベント名	抽出画像数	テキストタグ が付与された 画像数	発生時間	発生場所	テキストタグ
花見	造幣局の花見	95	6	4/14 - 4/20	(34.697, 135.521)	{ { 花見 or 桜 or hanami or sakura or cherry } and 造幣局 } or 通り抜け
	清水寺の花見	202	20	4/1 - 4/14	(34.995, 135.785)	{ 花見 or 桜 or hanami or sakura or cherry } and { 清水 or kiyomizu }
	奈良公園の花見	35	4	4/1 - 4/14	(34.685, 135.844)	{ 花見 or 桜 or hanami or sakura or cherry } and { 奈良公園 or narapark }
	姫路城の花見	20	5	4/1 - 4/14	(34.839, 134.694)	{ 花見 or 桜 or hanami or sakura or cherry } and { 姫路 or himeji }
	万博記念公園の花見	28	4	4/1 - 4/14	(34.811, 135.529)	{ 花見 or 桜 or hanami or sakura or cherry } and { 万博 or banpaku }
	大阪城公園の花見	153	6	4/1 - 4/14	(34.686, 135.523)	{ 花見 or 桜 or hanami or sakura or cherry } and { 大阪城 or osakajo }
	八坂神社の花見	206	13	4/1 - 4/14	(35.004, 135.779)	{ 花見 or 桜 or hanami or sakura or cherry } and { 八坂 or yasaka }
紅葉狩り	清水寺の紅葉狩り	298	27	11/11 - 12/4	(34.995, 135.785)	{ 紅葉 or momiji or maple or koyo or leaf } and { 清水 or kiyomizu }
	嵐山の紅葉狩り	97	33	11/12 - 12/4	(35.01, 135.682)	{ 紅葉 or momiji or maple or koyo or leaf } and { 嵐山 or arashiyama }
	比叡山の紅葉狩り	14	1	11/3 - 11/25	(35.068, 135.839)	{ 紅葉 or momiji or maple or koyo or leaf } and { 比叡 or hiei }
	勝尾寺の紅葉狩り	19	1	11/1 - 11/30	(34.865, 135.491)	{ 紅葉 or momiji or maple or koyo or leaf } and { 勝尾寺 or katsuoji }
花火大会	天神祭奉納花火	9	1	7/25	(34.694, 135.501)	{ 花火 or hanabi or fireworks } and { 天神 or tenjin }
	淀川花火大会	24	4	8/6	(34.716, 135.485)	{ 花火 or hanabi or fireworks } and { 淀川 or yodogawa or なにわ or naniwa }
	みなとこうべ海上 花火大会	11	5	8/6	(34.679, 135.188)	{ 花火 or hanabi or fireworks } and { 神戸 or こうべ or kobe }
	びわ湖大花火大会	2	1	8/8	(35.011, 135.863)	{ 花火 or hanabi or fireworks } and { 琵琶湖 or びわ湖 or biwako }
	熊野花火大会	1	1	8/17	(34.888, 135.102)	{ 花火 or hanabi or fireworks } and { 熊野 or kumano }
クリスマス	神戸ルミナリエ 中之島	45	12	12/1 - 12/12	(34.690, 135.187)	ルミナリエ or luminarie
	ルネッサンス	115	8	12/14 - 12/25	(34.6917, 135.491)	ルネッサンス or ルネッサンス or renaissance
祇園祭	祇園祭	230	33	7/1 - 7/31	(35.004, 135.779)	祇園祭 or gionmatsuri
花灯路	花灯路	24	6	12/9 - 12/18	(35.01, 135.682)	{ 花灯路 or hanatouro } and { 嵐山 or arashiyama }
時代祭	時代祭	21	8	10/23 - 10/23	(35.017, 135.782)	時代祭 or jidaimatsuri
万博公園の 梅まつり	万博公園の 梅まつり	19	3	2/16 - 3/10	(34.8114, 135.5294)	{ { 梅 or ume } and { 万博 or banpaku } } or { 梅まつり or umematsuri }

の画像集合を削除した結果、64個の画像集合 P_{eck} を獲得した。このうち、正しく検出されたイベントクラスは13個、重複を除くと8個であり、適合率と再現率はそれぞれ20.3%(13/64)、100(8/8)%となった。誤検出された画像集合は、nikon や canon などの画像を撮影したカメラに関連したテキストタグや、eyefi

のように、画像を画像共有サイトに投稿するためのサービスにより自動的に付与されるテキストタグをキーワードとして持つものが多い。このクラスタリング結果に対し、各画像集合内の $s = 50\%$ 以上の画像に付与されたテキストタグをキーワードとし、 $TH_s = 4$ としたフィルタリングを行った結果、誤検出

表 3 検出された正解イベントクラス

id	イベントクラス名	投稿者数	画像数	発生時間	キーワード集合
1	花見	110	153	4/1 - 4/18	sakura
2	花見	6	6	4/5 - 4/10	cherry
3	花見	3	3	4/7 - 4/17	2011
4	花見	3	7	4/1 - 4/14	nikon
5	紅葉狩り	118	165	11/10 - 12/4	京都
6	花火大会	13	13	7/25 - 08/08	fireworks
7	クリスマス	10	13	12/1 - 12/11	kobe
8	クリスマス	3	3	12/22 - 12/25	winter sky river
9	クリスマス	23	25	12/14 - 12/24	大阪
10	花灯路	3	3	12/15 - 12/18	嵐山
11	時代祭	6	6	10/23	京都
12	万博公園の梅まつり	4	4	2/27 - 3/06	万博記念公園 万博公園 梅 日本 大阪 吹田 plum expo 万博 岡本太郎 太陽の塔
13	祇園祭	43	60	7/1 - 7/31	京都

されていた画像集合を 29 個削除できた。前述のテキストタグは撮影時間に関連なく付与されるため、同じテキストタグが付与された画像集合が多く検出され、これらの画像集合に対する $score_{ec_k}$ が低くなったためである。

一方、正しいイベントクラスも 5 個削除された。これにより、フィルタリング後正しく検出されたイベントクラスは 8 個、重複を除くと 5 個であり、適合率と再現率は、それぞれ 22.9%(8/35), 62.5%(5/8) となった。表 3 に、フィルタリング前に検出されたイベントクラス $ec_j = (ec_j^t, ec_j^w)$ を各画像集合を構成する画像数、その投稿者数と共に示す。ただし網掛けは、フィルタリングにより、誤って削除されたイベントクラスである。単一地点で発生したイベントに加え、花火大会のように複数地点で発生したイベントがイベントクラスとしてまとめて抽出され、発生時間もおおむね表 2 に示すものと合致していることが分かる。しかし、花見やクリスマスのように、各イベントの発生期間が長期間でかつずれが生じる場合や、イベントを表す頻出テキストタグが複数ある場合などは、一つのイベントクラスに対して複数のクラスタが検出された。また、フィルタリング後に誤って削除されたイベントクラスは場所を表すテキストタグをキーワードとして持つものが多かった。例えば、紅葉狩り、祇園祭、時代祭はすべて京都のイベントであり、イベントクラスのキーワードが”京都”となった。このように場所を表すキーワードは、同じ場所で発生する複数のイベントに付与されることが多い。この影響による誤削除を避けるために、画像集合候補のフィルタリングの際、同じキーワードが付与されたクラスタの数だけでなく、クラスタ内の画像の撮影場所のばらつきも考慮する必要があると考えられる。

3.2 イベント検出結果

イベント検出においては孤立点が検出されすぎないように ϵ は低く設定し、 $\epsilon = 0.3$, $\mu = 6$ として、前節でフィルタリング前に検出された 13 個のイベントクラスの各画像集合に対してイベント検出を行った。その結果、16 個のクラスタと 77 個の孤立点の合計 93 個のイベントが得られた。このうち正しいイベントは、15 個のクラスタ、29 個の孤立点の計 44 個であり、重複を除くと 14 個であった。よって、適合率と再現率は、それぞれ 47.3%(44/93), 63.6%(14/22) となった。また、正しくクラスタとして検出された 15 個のイベント $e_i = (e_i^t, e_i^l, e_i^w)$ を表 4 に示す。奈良公園や造幣局の花見、みなと神戸海上花火大会のように撮影画像の少ないイベントも検出できていることが分かる。

例として、表 3 に示す $id = 1$ のイベントクラスに対して行ったイベント検出の結果を図 3 に示す。また、表 2 で示す花見の発生地点を図 4 に図示する。花見のイベントクラスから 4 つのイベントが得られ、うち 3 つのイベントが大きく、大阪、京都、奈良の 3 都市に分類されていることが分かる。さらに、図 3 より、検出されたイベントは花見の発生地点とほぼ一致することが確認できる。このように距離的に近いイベントはまとめて検出されるものの、おおむね正しい位置でイベントが検出された。また、発生時間もおおむね正しく検出されたが、淀川花火大会のように、実際より長い期間が検出された例もみられた。

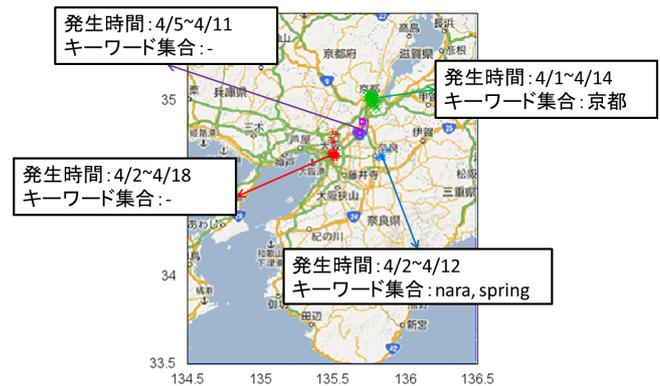


図 3 検出されたイベント例

これは、淀川花火大会が開催された時間と場所と類似した時間と場所で発生した別のイベントが、一つのイベントとして検出されたためである。

4. まとめ

本稿では、画像共有サイトに投稿される画像に付与された撮影場所・時間・意味内容を表すテキストタグを用いて、実世界のイベントを検出する手法を提案した。提案手法は、意味的に関連するイベントが同時期にさまざまな地点で発生することに着目し、まず、投稿された画像全体から、このようなイベントの集合で構成されるイベントクラスを検出した後、イベントクラスを表す画像集合から個々のイベント検出を行うことにより、投稿者の少ないイベントの検出を試みた。提案手法により、

表 4 検出された正解イベント

対応するイベントクラス id	イベント名	画像数	発生時間	発生場所	キーワード集合
1	清水寺の花見	91	4/1 - 4/14	(35.009, 135.773)	京都
	清水寺の花見	10	4/5 - 4/11	(34.894, 135.696)	-
	造幣局の花見	45	4/2 - 4/18	(34.709, 135.513)	-
	奈良公園の花見	6	4/2 - 4/12	(34.688, 135.840)	-
3	造幣局の花見	3	4/7 - 4/17	(35.018, 135.676)	2011
5	嵐山の紅葉狩り	157	11/10 - 12/4	(35.014, 135.750)	-
	勝尾寺の紅葉狩り	6	11/18 - 11/29	(34.841, 135.500)	autumn, 日本, 関西地方, japanese, olympusep1, kansai, fall, 大阪市, mino, minoh
6	淀川花火大会	8	7/25 - 8/6	(34.691, 135.491)	大阪
	みなとこうべ海上花火大会	4	8/6	(34.683, 135.186)	花火, kobe, hanabi
7	神戸ルミナリエ	13	12/1 - 12/11	(34.687, 135.189)	kobe
8	中之島ルネッサンス	3	12/22 - 12/25	(34.687, 135.506)	winter, sky, river
9	中之島ルネッサンス	24	12/14 - 12/24	(34.683, 135.502)	大阪
10	花灯路	3	12/15 - 12/18	(34.018, 135.676)	嵐山
11	時代祭	6	10/23	(35.023, 135.776)	temple
12	万博公園の梅まつり	4	2/27 - 3/6	(34.805, 135.519)	万博記念公園, 万博公園, 梅, 日本, 大阪, 吹田, plum, expo, 万博, 岡本太郎, 太陽の塔



図 4 花見の発生場所

2011 年に関西地方で撮影された画像集合を用いて作成したデータベースから、適合率 20.3%, 再現率 100% の精度でイベントクラスの検出に、適合率 47.3%, 再現率 63.6% の精度でイベントの検出を実現した。撮影画像が数枚のイベントについても正しく検出できることを確認した。今後の課題として、検出されたクラスを構成する画像の撮影場所を用いたフィルタリングの改善が挙げられる。また、より有効性の高い手法であることを検証するために、2 段階のクラスタリングを用いない手法との比較やあらかじめイベントを設定せずに、より広範囲で撮影された画像を用いた実験を行う必要がある。

文 献

- [1] “Flickr”, <http://www.flickr.com/>
- [2] Ling Chen and Abhishek Roy, “Event Detection from Flickr Data through Wavelet-based Spatial Analysis”, Proceedings of the 18th ACM Conference on Information and Knowledge Management(CIKM2009), pp. 523-532, 2009.
- [3] Lyndon S. Kennedy and Mor Naaman, “Generating Diverse and Representative Image Search Results for Landmarks”,

- Proceedings of the 17th International Conference on World Wide Web(WWW2008), pp. 297-306, 2008.
- [4] Tye Rattenbury, Nathaniel Good, and Mor Naaman, “Towards Automatic Extraction of Event and Place Semantics from Flickr Tags”, Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR2007), pp. 103-110, 2007.
- [5] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, “Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors”, Proceedings of the 19th International Conference on World Wide Web(WWW2010), pp. 851-860, 2010.
- [6] Amit Sheth, “Citizen Sensing, Social Signals, and Enriching Human Experience”, IEEE Internet Computing, Volume 38, Issue 4, pp 87-92, 2009.
- [7] Xiaowei Xu, Nurcan Yuruk, and Thomas A. J. Schweiger, “SCAN: A Structure Clustering Algorithm for Networks”, Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD2007), pp. 824-833, 2007.
- [8] Zhijun Yin, Liangliang Cao, Jiawei Han, Chengxiang Zhai, and Thomas Huang, “Geographical Topic Discovery and Comparison”, Proceedings of the 20th International Conference on World Wide Web(WWW2011), pp. 247-256, 2011.
- [9] Yan-Tao Zheng, Ming Zhao, Yang Song, Hartwig A. U. Buddemeier, Alessandro Bissacco, Fernando Brucher, Tat-Seng Chua, and Hartmut Neven, “Tour the World: Building a Web-scale Landmark Recognition Engine”, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR2009), pp. 1085-1092, 2009.