

# Twitter からのジオタグ画像収集による視覚的イベント検出

金子 昂夢<sup>†</sup> 柳井 啓司<sup>††</sup>

<sup>†</sup> 電気通信大学 電気通信学部 情報工学科 〒182-8585 東京都調布市調布ヶ丘 1-5-1

<sup>††</sup> 電気通信大学 大学院情報理工学専攻 総合情報学専攻 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: <sup>†</sup>kaneko-t@mm.inf.uec.ac.jp, <sup>††</sup>yanai@cs.uec.ac.jp

あらまし 近年, スマートフォンのような位置情報を扱えるデバイスの普及に伴い, 撮影した画像に位置情報を付加することが容易となった. また, ツイートと呼ばれる短文を投稿できるサービス, Twitter の普及により, 自分が今何をしているのか, どこにいるのかといったことを投稿する人が増加してきた. ツイートには本文の他に画像や位置情報を添付でき, 外出先で撮影した位置情報付き画像を即座に投稿することができる. 本研究では, これらの Twitter に投稿された位置情報付き画像ツイートから視覚的なイベントの検出を試みる. まず, 対象をより小さな地域に分割し, 各地域において前日の状態と比較して得られるスコアが一定以上となるキーワードを検出した. そして, 検出されたキーワードを含むツイートの各画像から画像特徴量を抽出し, クラスタリングを行った. クラスタリング結果からイベントの代表的な画像を選出し, 中心座標を計算して地図上に表示させた. 実験では, 日本とアメリカで投稿されたツイートをを用いて, それぞれのデータセットを作成した. 検出の結果, 台風や虹, 花火大会, スポーツの試合といったイベントが検出され, 画像からその様子を視覚的に捉えることができた.

キーワード Twitter, イベント検出, 位置情報付き画像

## 1. はじめに

近年, スマートフォンやタブレット端末といった位置情報を扱えるデバイスの普及に伴い, 撮影した画像に位置情報を付加することが容易となった. また, 140 文字以内のツイートと呼ばれる短文を投稿できるマイクロブログサービス Twitter の普及により, 自分が今何をしているのか, どこにいるのかといったことを投稿する人が増加してきた. ツイートには本文の他に画像や位置情報を添付でき, スマートフォンを利用して外出先で撮影した位置情報付き画像を即座に投稿することができる. このため Twitter には多くの位置情報付き画像がテキストデータと共に投稿され, これらのデータはリアルタイム性という重要な性質を持つ. Twitter に投稿された位置情報付き画像ツイートの例を図 1 に示す.

本研究では, Twitter に投稿された位置情報付き画像ツイートから視覚的なイベントの検出を試みる. ここでイベントとは, 台風や虹といった自然現象や, 花火大会のような局所的な行事のことを指す. これらのイベントを検出し, その情報を画像と共に地図上に表示することで, 何が・いつ・どこで起き, どんな様子だったのかを視覚的に捉えることを目標とする. さらに日本とアメリカで投稿されたツイートをを用いて, それぞれのデータセットを作成し, 互いの地域で検出されるイベントの違いを検証する.

## 2. 関連研究

Twitter は Web からのテキストマイニング対象として広く研究されており, Twitter に投稿されたツイートからイベント



図 1 位置情報付き画像ツイートの例

の検出を試みた研究も多くある.

地震や台風などの大規模で社会的な影響を持つイベントの検出を試みた研究として, 榊らの研究 [1] がある. Twitter のユーザを実世界のイベントを監視するソーシャルセンサと捉えたモデルを作り, イベントの検出とその位置の推定を行うシステムを作成した.

より局所的なイベントの検出を試みた研究として, Lee らの研究 [2] がある. 彼らは対象地域をいくつかのより小さな地域

に分割し、各地域の通常の状態を推測する。そして、通常の状態とは異なるツイートが多く投稿されたとき、イベントが発生したとみなしている。

画像を利用したイベント検出の研究として中地らの研究 [3] がある。あらかじめ「台風」、「正月」といったキーワードや 2011 年 3 月 11 日から 2011 年 3 月 12 日といった期間を設けて Twitter から位置情報付き画像ツイートを収集し、クエリに対するツイートを解析することで画像付きのイベント検出を試みた。異なる時間や地域で代表的な画像や類似する画像群の抽出を行うことで、抽出される画像の違いが示された。

さらに、位置情報付き画像ツイートを利用した研究として、柳井らの World Seer [4] がある。Twitter に投稿された位置情報付き画像をリアルタイムに監視し、地図上にマッピングすることでどのような写真がどこで撮られたのかを瞬時に視覚化され閲覧することができるシステムであった。

### 3. 手法の概要

中地らの研究 [3] では位置情報付き画像ツイートの収集に用いるクエリを手動で設定しているため、イベントの自動的な検出が行えず、ユーザが知らないイベントは検出できなかった。そこで本研究では、まず、イベントのキーワードをシステムにより自動的に検出することで、より多くのイベントを検出し、未知のイベントのキーワードも得られるようにする。さらに、キーワードを用いて収集した画像を解析することで、ユーザが知らないイベントでも画像により視覚的に捉えることができるようにする。本研究の手法の流れを以下に示す。

- (1) 特定の地域で頻出するキーワードの抽出
- (2) 抽出されたキーワードの統合・補完
- (3) キーワードに対する位置情報付き画像ツイートの収集
- (4) 画像から Bag-of-Features とカラーヒストグラムを生成
- (5) 2 つの画像特徴量を用いてクラスタリング
- (6) クラスタリング結果から代表画像選出・中心座標計算
- (7) 抽出されたイベントを地図上に表示

まず、対象領域をより小さな地域に分割し、各地域で前日の状態と比較して得られるスコアが一定以上となるキーワードを抽出する。抽出されたキーワードは、同日・同地域で抽出された同じイベントに関するものを統合し、キーワードを含むツイートから前後の文字の補完を行う。

次に、抽出されたイベントのキーワードを用いて位置情報付き画像ツイートを収集し、各画像から SURF による Bag-of-Features と RGB カラーヒストグラムを生成する。

そして、画像を Ward 法によりクラスタリングし、その結果から得られるスコアによりクラスタをランキングする。最もスコアが高いクラスタからイベントの代表的な画像を選出し、スコアが一定以上のクラスタに含まれる位置情報付き画像ツイートからイベントの中心座標を計算して地図上に表示する。またスコアが一定未満のクラスタはノイズ画像群とし、イベントから除外する。

## 4. 手法の詳細

### 4.1 キーワードの検出

視覚的なイベントの検出を行うために、位置情報付き画像ツイートの収集に用いるイベントのキーワードを検出した。

#### 4.1.1 タグの抽出

イベントのキーワードを検出するために、ツイートの本文に対して形態素解析を行い、特定の品詞の単語をタグとして抽出した。日本語の形態素解析にはオープンソースの形態素解析エンジン MeCab [5] を利用し、英語の形態素解析には TweetNLP [6] を利用した。それぞれの出力に従い、切り出された単語から品詞が「名詞」であるものを抽出し、重複がないように各ツイートのタグとして登録した。

#### 4.1.2 キーワードの抽出

抽出された各ツイートのタグのうち、特定の地域で前日の状態と比較して得られるスコアが一定以上となるものを 1 日ごとに探索し、イベントのキーワードとして抽出した。

まず、対象地域の様々な地域から抽出されるように対象を緯度・経度それぞれ 1 度ずつのグリッドにより分割し、各地域に重みを付けた。分割された各地域について、各データセットの全ての位置情報付き画像ツイートから、そのグリッド内でツイートしたユーザ数をカウントし、最も出現したユーザ数が多い地域が 1 となるように以下の式 1 により重みを付けた。

$$W_{i,j} = \frac{M + s}{N_{i,j} + s} \quad (1)$$

式 1 において、 $i, j$  はそれぞれ緯度・経度によるグリッドを表し、 $M$  は最も出現するユーザ数が多い地域のユーザ数、 $N_{i,j}$  はグリッド  $i, j$  に出現するユーザ数、 $s$  は標準偏差である。この重みによりスコアを補正することでユーザが少ない地域からでもキーワードが抽出されるようにした。

そして、タグに登録された各単語について、それぞれの地域でその単語をタグに含むツイートを投稿したユーザ数と前日のそれとの差に注目している地域の重みを乗じた値をスコアとし、スコアが一定以上となるものを抽出した。

$$S_{k,d,i,j} = (N_{k,d,i,j} - N_{k,d-1,i,j})W_{i,j} \quad (2)$$

式 2 において、 $k$  は各単語、 $d$  は注目している日、 $N$  はそれぞれの条件で単語  $k$  をタグに含むツイートを投稿したユーザ数である。今回の実験では、日本データセットでは 50、アメリカデータセットでは 200 以上となる単語をイベントのキーワードとして抽出し、キーワードによる位置情報付き画像ツイートの収集は、キーワードが抽出された日と地域に注目して行った。

#### 4.1.3 キーワードの統合・補完

形態素解析では単語を最小単位まで分割するため、例えば「花火大会」のような単語と単語が連結して構成されたものは、「花火」と「大会」のように別々に抽出されることがある。そこで、同日・同地域で抽出された各キーワードについて、同じイベントに関するキーワードの統合を行った。一方のキーワードをタグに含むツイートに、もう一方のキーワードが 50% 以上含

まれていれば、同じイベントに関するキーワードとして統合した。統合されたキーワード群に対し新たなキーワードを統合する場合、いずれかの組で条件を満たせば同じイベントに関するキーワードとして統合した。統合されたキーワードのうち、そのイベントのキーワードは最も多くのツイートに含まれるものとした。

また、形態素解析に用いる辞書により単語は切り取られるため、例えば「スカイツリー」のような辞書に登録されていない単語は、「スカイ」と「ツリー」のように必要以上に分割されてしまうことがある。そこで、抽出されたキーワードを含むツイートを利用して、キーワードの補完を行った。元のキーワードの一つ前の文字が80%以上同じ文字ならば、この文字でキーワードを補完し、さらに補完されたキーワードの一つ前の文字を調べ、再帰的に補完を行った。キーワードの後ろの文字についても同様に補完を行った。

#### 4.2 画像特徴量

キーワードを用いて収集した画像から類似する画像群を抽出するため、各画像から SURF [7] による Bag-of-Features と RGB カラーヒストグラムを生成した。

SURF は縦横共に 10 ピクセルごとを特徴点とし、特徴点ごとにスケールを 5, 10, 15 として 128 次元の特徴ベクトルを抽出した。そして、1000 次元の codebook を作成し、画像を Bag-of-Features で表現した。

また、RGB 色空間の各チャンネルを減色することで 64 次元のカラーヒストグラムを作成した。

#### 4.3 クラスタリング

各イベントを視覚的に表す画像群を抽出するため、キーワードを用いて収集した画像を画像特徴量を用いてクラスタリングを行った。クラスタリングには Ward 法を用い、閾値を設定することで任意のクラスタ数に画像が分類されるようにした。

Ward 法はクラスタ内の各要素からその中心までの距離を最小化する手法であり、凝集型階層的クラスタリングの一種である。全ての要素が独立したクラスタ状態から始め、以下の式 3 により定義されるエラーの増加量が最小となるクラスタの組を逐次的に併合する。

$$d(C_1, C_2) = E(C_1 \cup C_2) - E(C_1) - E(C_2) \quad (3)$$

ただし、通常  $E(C)$  はクラスタ  $C$  の全ての要素から中心までの距離の二乗の総和であるが、今回使用する画像特徴量は二つあるので、以下の式 4 のように  $E(C)$  を定義した。

$$E(C) = \sum_{x \in C} ((x_{BoF} - \bar{x}_{BoF})^2 w_{BoF} + (x_{Color} - \bar{x}_{Color})^2 w_{Color}) \quad (4)$$

式 4 において、 $x_{BoF}$  は画像  $x$  の Bag-of-Features、 $x_{Color}$  は画像  $x$  のカラーヒストグラム、 $\bar{x}$  はクラスタ  $C$  における各特徴量の中心、 $w$  は各特徴量の重みでそれぞれの次元数である。

本研究では任意のクラスタ数に分類されるように、エラーの増加量  $d(C_1, C_2)$  に対して閾値を設定し、 $d(C_1, C_2)$  が閾値よりも大きくなるか、クラスタ数が 1 となることをクラスタリン

グの終了条件とした。今回の実験では閾値を両データセットで 300 とした。

#### 4.4 視覚的な検出

クラスタリング結果により、視覚的に統一性があり画像枚数が多いクラスタに限定することでイベントを視覚的に表すことができる画像群を抽出し、地図上にイベントの検出結果を表示した。

各クラスタについて、より画像の類似度が高く、分類された画像枚数が多いクラスタが高いスコアとなるように、以下の式 5 によりスコアを付けた。

$$V_C = \frac{n_C^2}{E(C)} W_{i,j} \quad (5)$$

式 5 において、 $n_C$  はクラスタ  $C$  に分類された画像枚数、 $W_{i,j}$  は式 1 によって得られた、キーワードが検出された地域の重みである。この値が大きいほどそのクラスタのイベントに対する視覚的関連性が高く、小さいほどノイズである可能性が高いと言える。

このスコアが閾値以上のクラスタをイベントを視覚的に表すことができる画像群とし、閾値未満のクラスタはノイズ画像群としてイベントから除去した。また、全てのクラスタのスコアが閾値未満となったイベントは視覚的に統一性がないイベントとして一覧から除去した。さらに、各イベントにおいて、スコアが最も高いクラスタの中でクラスタ中心に最も近い画像をイベントの代表的な画像として選出し、閾値以上のクラスタに含まれる位置情報付き画像ツイートからイベントの中心座標を求めた。そして、得られた中心座標を用いて地図上にマーカーを設置し、イベントの情報と代表画像をマーカーに対応付け表示させた。今回の実験では、閾値を日本データセットでは 5、アメリカデータセットでは 20 とした。

## 5. 実験

### 5.1 データセット

今回の実験を行うにあたって、Twitter に投稿された位置情報付き画像ツイートから 2 つのデータセットを作成した。1 つは日本で 2011 年 2 月 10 日から 2012 年 9 月 30 日までに投稿されたものであり、約 3,000,000 件の位置情報付き画像ツイートである。もう 1 つはアメリカで 2012 年 1 月 1 日から 2012 年 12 月 31 日までに投稿されたものであり、約 17,000,000 件の位置情報付き画像ツイートである。

### 5.2 キーワードの抽出

それぞれのデータセットで抽出されたキーワードの一部を表 2、表 3 に示す。表中の「地域」は左からグリッドにより分割された各地域の南端、北端、西端、東端の緯度・経度を表し、「重み」は各地域の重みである。

### 5.3 キーワードの統合・補完

抽出されたキーワードについて、キーワードを含むツイートが 50%以上同じものだった場合に統合を行った。さらに、統合後のキーワードについて、キーワードを含むツイートの本文中でキーワードの前後の文字が 80%以上同じものだった場合に補完を行った。それぞれの処理を行った結果について、日本デー

タセットの例を表 4 に、アメリカデータセットの例を表 5 に示す。

#### 5.4 クラスタリング

検出されたキーワードを用いて収集した画像を画像特徴量を用いて、Ward 法によりクラスタリングを行った。

クラスタリング結果の例それぞれ図 2, 図 3, 図 4 に示す。図中の各クラスタの上部に記載されている“num”はそのクラスタに分類された画像枚数であり, “b\_score” と “c\_score” はそれぞれ Bag-of-Features とカラーヒストグラムの各画像からクラスタ中心までの距離の二乗に特徴量の次元を乗じた値の総和で, この値が小さいほどその特徴量についての類似性の密度が高いことを表している。また, “weight” はキーワードが検出された地域の重みを表し, “score” がこれらの値から計算されたクラスタのスコアである。

クラスタリング結果により, スコアが閾値未満のクラスタをノイズ画像群としてイベントから除去された。

図 2 の「花火」の例は, 日本データセットで 2011 年 8 月 6 日に東京で検出されたイベントであり, 最も高いスコアを持つ最上部のクラスタを見ると花火が写った画像が多く含まれている。

図 3 の「桜」の例は, スコアが閾値以上である上 2 つのクラスタを見ると, 両方とも桜が写り, このイベントに関する画像であるが, 背景の色の違いにより日中に撮影されたものと夜に撮影されたものに分類され, 時間によるイベントの様子の違いが示された。

図 4 の「Stanford Stadium」の例は, 2 つのクラスタはそれぞれ別の日に検出された同キーワードであるが別のイベントのクラスタである。この例では上部のクラスタを見るとサッカーの試合を写した画像が多く含まれているのに対して, 下部のクラスタがフットボールの試合を写した画像が多く含まれている。よって同じキーワードでも画像により, イベントの違いを視覚的に知ることができた。

#### 5.5 検出結果の表示

検出されたイベントを地図上に表示するため, スコアが最も高いクラスタの中でクラスタ中心に最も近い画像を代表的な画像として選出し, スコアが閾値以上のクラスタに含まれる位置情報付き画像ツイートから中心座標を求めた。

検出されたイベントについて, 地図上にマーカーを設置し, いくつかのマーカーをクリックした例を図 5, 図 6 に示す。Web 上のシステムでは, 表示された代表画像をクリックすることで, そのイベントのクラスタリング結果が表示され, イベントに関するその他の画像を見ることができる。代表画像をクリックした例を図 7 に示す。

最終的なイベントの検出結果は表 1 のようになった。データ

表 1 それぞれのデータセットの検出結果

	日本	アメリカ
イベント数	258	1676
代表画像の適合率 (%)	65.5	72.5

数や期間が異なるため, 検出されたイベントを単純に比較する

ことはできないが, 日本では主に祭り等の地域の行事や虹や台風などの自然現象が多く検出されたのに対して, アメリカでは野球やフットボール, バasketボールのようなスポーツの試合が非常に多く検出された。

## 6. ま と め

本研究では, Twitter に投稿された位置情報付き画像ツイートから視覚的なイベントの検出を行った。まず, 対象をより小さな地域に分割し, 各地域において前日の状態と比較して得られるスコアが一定以上となるキーワードを抽出した。抽出されたキーワードは同じイベントに関するものを統合し, キーワードを含むツイートから前後の文字の補完を行った。次に, 検出されたキーワードを用いて位置情報付き画像ツイートを収集し, SURF による Bag-of-Features と RGB カラーヒストグラムを用いて Ward 法によりクラスタリングした。そして, 各クラスタのイベントに対する視覚的関連性を表すスコアを計算し, クラスタのスコアに基づいて, 代表的な画像の選出と中心座標を計算し, 地図上に表示した。実験では, 日本とアメリカの 2 つのデータセットを作成し, 互いのデータセットで検出されるイベントの違いがあることが示された。

今後の課題として, まず, 検出するグリッドの大きさと時間の刻みを可変にし, より柔軟にイベントの検出が行えるようにする。また, 代表的な画像の選出方法を改良し, その精度を上げると共に, 同キーワードのイベントについて, 時間や地域における画像の違いを抽出することが考えられる。さらに, システムを自動化し, Twitter のタイムラインを監視することでリアルタイムなイベントの検出を行えるようにする。

## 文 献

- [1] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proc. of the International World Wide Web Conference*, pp. 851–860, 2010.
- [2] R. Lee and K. Sumiya. Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In *Proc. of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pp. 1–10, 2010.
- [3] Y. Nakaji and K. Yanai. Visualization of real world events with geotagged tweet photos. In *Proc. of IEEE ICME Workshop on Social Media Computing (SMC)*, 2012.
- [4] K. Yanai. World seer: a realtime geo-tweet photo mapping system. In *Proc. of ACM International Conference on Multimedia Retrieval*, 2012.
- [5] Mecab. <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>.
- [6] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Milles, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. Part-of-speech tagging for twitter: annotation, features, and experiments. In *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers*, pp. 42–47, 2011.
- [7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346–359, 2008.

表 2 日本データセットでの抽出結果の一部

キーワード	日付	地域 (lat,lng)	重み	スコア
雪	2011/2/11	34 ~ 35,135 ~ 136	1.96	135.5
地震	2011/3/11	35 ~ 36,139 ~ 140	1	55
花火	2011/8/6	34 ~ 35,135 ~ 136	1.96	149.2
大会	2011/8/6	34 ~ 35,135 ~ 136	1.96	68.7
台風	2011/9/21	35 ~ 36,139 ~ 140	1	62
アップル	2011/10/6	35 ~ 36,139 ~ 140	1	70
月食	2011/12/10	34 ~ 35,135 ~ 136	1.96	84.4
クリスマス	2011/12/24	35 ~ 36,136 ~ 137	2.9	55.2
大晦日	2011/12/31	35 ~ 36,139 ~ 140	1	68
初日の出	2012/1/1	35 ~ 36,139 ~ 140	1	84
マラソン	2012/2/26	35 ~ 36,139 ~ 140	1	77
六本木	2012/3/24	35 ~ 36,139 ~ 140	1	88
桜	2012/4/28	37 ~ 38,140 ~ 141	4.18	121.4
スーパー	2012/5/5	35 ~ 36,139 ~ 140	1	93
ムーン	2012/5/5	35 ~ 36,139 ~ 140	1	96
日食	2012/5/21	35 ~ 36,139 ~ 140	1	314
金環	2012/5/21	35 ~ 36,140 ~ 141	3.18	60.5
七夕	2012/7/7	34 ~ 35,135 ~ 136	1.96	56.9
祇園祭	2012/7/14	35 ~ 36,135 ~ 136	3.46	104
幕張メッセ	2012/8/11	35 ~ 36,140 ~ 141	3.18	168.9
阿波	2012/8/12	34 ~ 35,134 ~ 135	3.91	54.8
精霊流し	2012/8/15	32 ~ 33,129 ~ 130	4.69	51.6
送り火	2012/8/16	35 ~ 36,135 ~ 136	3.46	104
大文字	2012/8/16	35 ~ 36,135 ~ 136	3.46	83.2

表 3 アメリカデータセットでの抽出結果の一部

キーワード	日付	地域 (lat,lng)	重み	スコア
snow	2012/1/9	38 ~ 39,-78 ~ -77	8.02	248.7
sunset	2012/1/13	47 ~ 48,-123 ~ -122	10.7	290.4
Super	2012/2/5	37 ~ 38,-123 ~ -122	10.9	251.8
Bowl	2012/2/5	37 ~ 38,-123 ~ -122	10.9	251.8
Grammy	2012/12/12	34 ~ 35,-119 ~ -118	6.52	208.7
Valentines	2012/2/14	37 ~ 38,-123 ~ -122	10.9	438.1
SXSW	2012/3/8	30 ~ 31,-98 ~ -97	9.67	464.4
Festival	2012/3/24	25 ~ 26,-81 ~ -80	10.4	282.9
Music	2012/3/24	25 ~ 26,-81 ~ -80	10.4	272.4
Easter	2012/4/8	33 ~ 34,-85 ~ -84	9.13	703.1
shuttle	2012/4/17	38 ~ 39,-78 ~ -77	8.02	577.7
Jazz	2012/4/27	29 ~ 30,-91 ~ -90	10.3	228.2
eclipse	2012/5/20	37 ~ 38,-123 ~ -122	10.9	1544.2
WWDC	2012/6/10	37 ~ 38,-123 ~ -122	10.9	514.7
America	2012/7/4	33 ~ 34,-119 ~ -118	10.9	373.5
hurricane	2012/8/26	25 ~ 26,-81 ~ -80	10.4	241.0
rainbow	2012/9/5	37 ~ 38,-123 ~ 122	10.9	1423.7
49ers	2012/10/18	37 ~ 38,-123 ~ 122	10.9	262.8
Halloween	2012/10/31	40 ~ 41,-74 ~ -73	1.45	375.2
vote	2012/11/6	34 ~ 35,-119 ~ -128	6.57	345.6
Thanksgiving	2012/11/22	39 ~ 40,-85 ~ -84	8.69	573.6
Cristmas	2012/12/24	40 ~ 41,-75 ~ -74	3.94	1009.9
blizzard	2012/12/26	39 ~ 40,-87 ~ -86	9.93	208.5
NYE	2012/12/31	34 ~ 35,-119 ~ -128	6.52	436.9

表 4 日本データセットの統合・補完結果

抽出されたキーワード群	統合後	補完後
花火, 大会	花火	花火
森	森	こだまの森
台風	台風	台風一過
明治	明治	明治神宮
花火, 大会, 昼, 全国, 観覧席	花火	花火
ドーム	ドーム	札幌ドーム
リエ	リエ	ルミナリエ
アップル, 銀座	アップル	アップル
月食, 皆既	月食	皆既月食
六本木, ヒルズ	六本木	六本木
ホテル	ホテル	東京ホテル
マラソン	マラソン	京都マラソン
桜, 滝	桜	桜
スーパー, ムーン	スーパー	スーパームーン
ブルー, ムーン	ブルー	ブルームーン
国際, 幕張メッセ	幕張メッセ	幕張メッセ
海, 海浜, 中道	海	海の中海浜公園
日食, 金環	日食	金環日食
幕張, 県立, 海浜	幕張	幕張メッセ
母	母	母の日
スカイ, ツリー	スカイ	スカイツリー
鈴鹿, サーキット	サーキット	鈴鹿サーキット
送り火, 五山	送り火	送り火
風, 区, 湖畔, 杜, 国営, 北地	風	風の草原

表 5 アメリカデータセットの統合・補完結果

抽出されたキーワード群	統合後	補完後
Golden,Gate	Golden	Golden Gate
SXSW,Convention,sxsw	SXSW	SXSW
Auritorium,Shores	Auditorium	Auditorium Shores
Rangers,Ballpark	Rengers	Rengers Ballpark
Eclipse,eclipse	Eclipse	Eclipse
Summer,Fest,Press	Summer	Free Press Summer Fest
E3,expo	E3	E3
West,WWDC,Apple	WWDC	WWDC
Festival,North	North	North Beach Festival
Dodger,stadium	Dodger	Dodger Stadium
Bowl	Bowl	The Hollywood Bowl
rainbow,Double,Rainbow	rainbow	rainbow
Theater,Greek	Theater	Greek Theater
Theatre,Fox	Theatre	Fox Theatre
Apple,Store,iPhone	Apple	Apple Store
shuttle,Space,Endeavor	shuttle	shuttle
Hotel,Casino	Hotel	Hotel & Casino
Halloween,costume	Halloween	Halloween
Bowl,Rose,UCLA	Rose	The Rose Bowl
Square,Union	Union	Union Square
Color,Candlestick	Color	The Color Run
Oracle,Arena	Arena	Oracle Arena
Field,LP	Field	LP Field
Square,Times	Times	Times Square

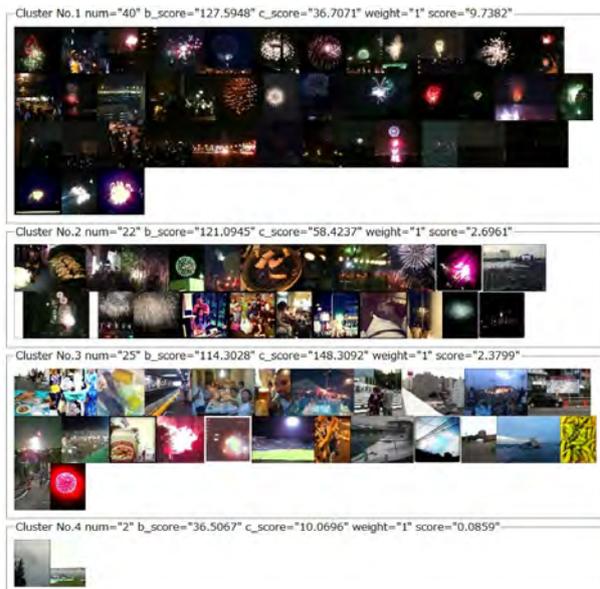


図 2 「花火」のクラスタリング結果



図 5 日本データセットにおける検出結果



図 3 「桜」のクラスタリング結果



図 6 アメリカデータセットにおける検出結果



図 4 「Stanford Stadium」のクラスタリング結果



図 7 代表画像をクリックした例