

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

金子 昂夢[†] 柳井 啓司^{††}

[†] 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

^{††} 電気通信大学 大学院情報理工学研究科 総合情報学専攻 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: [†]kaneko-t@mm.inf.uec.ac.jp, ^{††}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。検出の結果、台風や虹、花火大会、イルミネーション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。

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

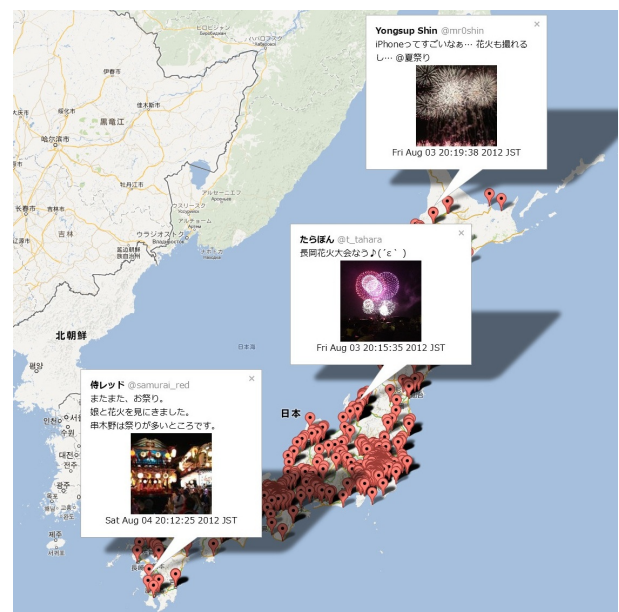


図 1 位置情報付き画像ツイートの例

を作成した。タイムラインを監視することでリアルタイムな検出を行い、高精度かつ高速な検出が行われた。

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。まず、対象地域をいくつかのより小さな地域に分割し、ツイート数、ユーザ数、ユーザの移動状況の 3 点を考慮して、その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。

また、Hong ら [3] は、地域によって投稿されるキーワードの違いや、ユーザによる傾向の違いを考慮してイベントの検出

を行った。各ユーザについて、ユーザが興味を示し話題に挙げるイベントの傾向や、普段活動している地域からユーザのプロファイリングを行い、地域によるキーワードの傾向と組み合わせることでツイートの分類を行った。実験では、位置情報が未知のツイートについて、その位置を高い精度で推定することができ、特定の地域で多く投稿されるキーワードが抽出された。

Liら[4]は、Twitterに投稿されたツイートをいくつかのセグメントに分割することでイベントの検出を行った。出現頻度が急上昇したセグメントをイベントとして抽出し、Wikipedia内のアンカーテキストからイベントのニュース性を計算し、ニュース性が一定以上だったものに限定することで、高い精度でイベントが検出された。また、キーワードの抽出にマルチグラムを用いているため、検出されたイベントのテキスト情報が多く、よりイベントを捉えやすい結果が得られた。

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[5]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。ツイートの解析には、画像から画像特徴量を抽出し、代表的な画像の選出や類似する画像群の抽出を行うことで視覚的な情報を得ている。異なる位置や時間で得られる画像の比較により、同じ時間において、台風が現在来ている地域と台風が通り過ぎた地域で多く投稿される画像が異なることや、正月の初日の出から初詣、夕食といったような一日の流れが画像により示された。

3. 手法の概要

中地らの研究[5]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、日本国内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を補完する。

次に、検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。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[6]を利用した。MeCab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。ただし、品詞が「名詞」の単語のうち、顔文字や特定の記号は品詞細分類が「サ変接続」として切り出されたため、これらと品詞細分類が「数」として切り出されたものはタグから除外した。MeCabによるタグ抽出の例を図2に示す。

長岡の花火大会です♪花火綺麗(´▽`) #長岡まつり2012

↓ 形態素解析

長岡	名詞,固有名詞,一般,*,*,*,長岡,ナガオカ,ナガオカ,,
の	助詞,格助詞,一般,*,*,*,の,ノ,ノ,,
花火	名詞,一般,*,*,*,花火,ハナビ,ハナビ,,
大会	名詞,一般,*,*,*,大会,タイカイ,タイカイ,,
です	助動詞,*,*,*,特殊・デス,基本形,です,デス,デス,,
♪	名詞,サ変接続,*,*,*,*,
花火	名詞,一般,*,*,*,花火,ハナビ,ハナビ,,
綺麗	名詞,形容動詞語幹,*,*,*,綺麗,キレイ,キレイ,,
(´▽`)	名詞,サ変接続,*,*,*,*,
#	記号,一般,*,*,*,#,#,#,
長岡	名詞,固有名詞,一般,*,*,*,長岡,ナガオカ,ナガオカ,,
ま	フイラー,*,*,*,*,ま,マ,マ,,
つ	助動詞,*,*,*,下二・タ行,基本形,つ,ツ,ツ,,
り	助動詞,*,*,*,文語・リ,基本形,り,リ,リ,,
2	名詞,数,*,*,*,2,二,二,,
0	名詞,数,*,*,*,0,ゼロ,ゼロ,,
1	名詞,数,*,*,*,1,イチ,イチ,,
2	名詞,数,*,*,*,2,二,二,,

↓ タグ(サ変接続と数以外の名詞)を抽出

長岡,花火,大会,綺麗

図2 MeCabによるタグ抽出の例

4.1.2 キーワードの抽出

MeCab により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、国内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。グリッドにより分割された地域を図 3 に示す。赤線が緯度・経度それぞれ 1 度ずつのラインであり、枠内が分割された各地域となる。分割された各地域について、今回の実験で収集した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、このスコアが 50 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

MeCab による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・同地域で検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

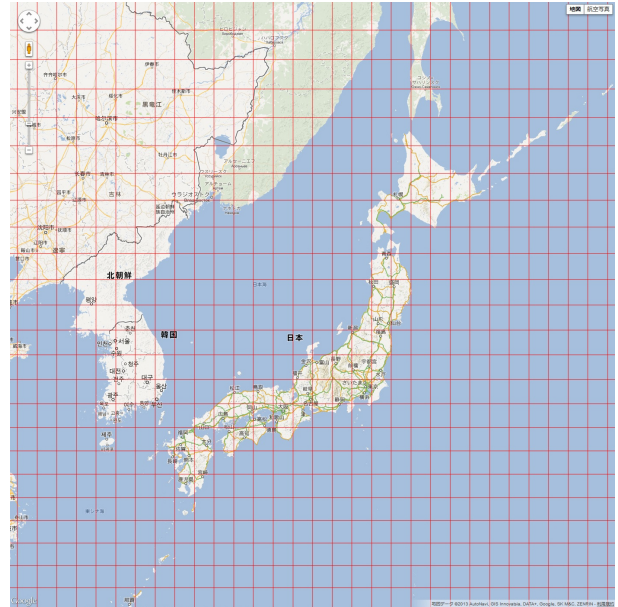


図 3 グリッドにより分割された地域

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 による Bag-of-Features と RGB カラーヒストグラムを生成した。

4.2.1 SURF

SURF (Speeded-Up Robust Features) [7] は H. Bay らによって考案された、画像の局所特徴量を特徴ベクトルで表現する手法である。画像の回転やスケールの変化、照明の変化に頑健であり、類似する他の手法よりも高速に特徴量を抽出することができる。SURF は通常 64 次元の特徴ベクトルで表現されるが、本研究では拡張ディスクリプタを使用し、128 次元の特徴ベクトルを抽出した。また特徴点は縦横共に 10 ピクセルごとの点とし、特徴点ごとにスケールを 5, 10, 15 とし SURF を抽出した。

4.2.2 Bag-of-Features

Bag-of-Features (BoF) は、Bag-of-Keypoints や Bag-of-Visual-Words とも言われ、自然言語処理の手法である Bag-of-Words を画像処理に応用したものである。Bag-of-Words では、文書を単語の集合とみなし、単語の語順を無視した各単語の出現頻度により文書表現する。Bag-of-Features では、文書を画像に、単語を特徴量に置き換え、特徴点の位置を無視した各特徴量の出現頻度により画像を表現する。

Bag-of-Features の生成のために、まず、局所特徴量を量子化することによって visual word と呼ばれる代表的な特徴ベクトルの集合 (codebook) を作成する。Codebook は、画像から抽出された局所特徴量を k-means 法によって k 個にクラスタリ

ングし、各クラスタの中心を visual word とすることで得られる。そして、画像から抽出された局所特徴量を codebook 中の visual word のうち最も近いものに割り当てることで、k 個の特徴ベクトルの出現頻度からなるヒストグラムが得られる。このヒストグラムを正規化することにより、画像を Bag-of-Features で表現することができる。

本研究では各イベントについて、検出されたキーワードを用いて収集した画像から SURF を抽出し、抽出された全ての局所特徴量を $k=1000$ としてクラスタリングを行い、codebook を作成した。従って、各イベントはそれぞれ別の codebook を持ち、それぞれの codebook により画像を Bag-of-Features で表現した。

4.2.3 カラーヒストグラム

カラーヒストグラムは、画像の各ピクセルの色情報を調べ、色の出現頻度をヒストグラムで表現したものである。色空間には RGB 色空間、HSV 色空間、Lu*v*色空間などがあるが、本研究では各ピクセルを Red, Green, Blue の 3 チャンネルの濃度で表す RGB 色空間を利用した。各チャンネルは通常それぞれ 256 段階の値を取り、16,777,216 色を表現することができる。本研究では、各チャンネルをそれぞれ 4 分割することで 64 色に減色し、各色の出現頻度からなる 64 次元のヒストグラムを正規化してカラーヒストグラムを得た。

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 Bag-of-Features とカラーヒストグラム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 100 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。

今回の実験ではこのスコアが 5 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、スコアが 5 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 5 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、スコアが 5 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。

5. 実験

5.1 データセット

今回の実験を行うにあたって、あらかじめ Twitter に投稿された位置情報付き画像ツイートを約 3,000,000 件収集した。これらのツイートは日本国内で 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、画像はサムネイルをダウンロードして利用した。

5.2 キーワードの抽出

閾値を 50 とした場合のキーワードの抽出結果の一部を表 1 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。2011 年 2 月 10 日から 2012 年 9 月 30 日までの期間で抽出されたキーワードの数は 306 となった。

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結果を表 2 に示す。また、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結果を表 3 に示す。

表1 キーワードの抽出結果の一部

キーワード	日付	地域	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
淀川	2011/8/6	34,35,135,136	1.96	72.6
ドーム	2011/8/10	43,44,141,142	3.96	51.5
雨	2011/8/19	35,36,139,140	1	60
台風	2011/9/21	35,36,139,140	1	62
富士山	2011/9/24	35,36,138,139	3.35	67
アップル	2011/10/6	35,36,139,140	1	70
銀座	2011/10/6	35,36,139,140	1	51
鈴鹿	2011/10/9	34,35,136,137	3.94	78.8
サーキット	2011/10/9	34,35,136,137	3.94	67
時代	2011/10/23	35,36,135,136	3.46	55.4
月食	2011/12/10	34,35,135,136	1.96	84.4
皆既	2011/12/10	34,35,135,136	1.96	58.9
リエ	2011/12/10	34,35,135,136	1.96	51
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
元旦	2012/1/1	35,36,139,140	1	69
明治	2012/1/1	35,36,139,140	1	50
恵方	2012/2/3	35,36,139,140	1	63
スキー	2012/2/11	36,37,138,139	3.69	77.5
バレンタイン	2012/2/14	35,36,139,140	1	58
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
ホテル	2012/5/6	35,36,139,140	1	59
タイ	2012/5/12	35,36,139,140	1	73
母	2012/5/13	35,36,139,140	1	63
日食	2012/5/21	35,36,139,140	1	314
金環	2012/5/21	35,36,140,141	3.18	60.5
一過	2012/6/20	35,36,139,140	1	61
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
東北電力	2012/7/14	37,38,139,140	4.4	79.2
スカイ	2012/7/28	35,36,139,140	1	54
ツリー	2012/7/28	35,36,139,140	1	53
平和	2012/8/6	34,35,132,133	4.08	77.5
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
沖縄	2012/9/8	26,27,127,128	4.17	66.8
サー	2012/9/8	26,27,127,128	4.17	50.1
風	2012/9/15	38,39,140,141	3.9	58.5
湖畔	2012/9/15	38,39,140,141	3.9	66.3
日本平	2012/9/15	34,35,138,139	4.15	78.8

表2 キーワードの統合結果

統合されたキーワード群	統合後のキーワード
花火, 大会	花火
花火, 大会, 淀川	花火
花火, 大会, 江戸川	花火
花火, 大会, 多摩川	花火
花火, 大会, 昼, 全国, 観覧席	花火
花火, 隅田川	花火
花火, 諏訪湖	花火
アップル, 銀座	アップル
月食, 皆既	月食
六本木, ヒルズ	六本木
風, 雨	風
桜, 綺麗	桜
桜, 滝	桜
スーパー, ムーン	スーパー
ブルー, ムーン	ブルー
国際, 幕張メッセ	国際
海, 海浜, 中道	海
日食, 金環	日食
幕張, 県立, 海浜	幕張
花火, 大会, 隅田川	花火
スカイ, ツリー	スカイ
鈴鹿, サーキット	サーキット
送り火, 五山	送り火
風, 区, 湖畔, 杜, 国営, 北地	風

表3 キーワードの補完結果

補完前	補完後
ドーム	札幌ドーム
アップル	アップルストア
リエ	ルミナリエ
明治	明治神宮
恵方	恵方巻
スキー	スキー場
マラソン	東京マラソン
マラソン	京都マラソン
スーパー	スーパームーン
ブルー	ブルームーン
ホテル	東京ホテル
母	母の日
海	海の中道海浜公園
森	こだまの森
一過	台風一過
東北電力	東北電力ビッグスワンスタジアム
恋	つま恋
スカイ	スカイツリー
サーキット	鈴鹿サーキット
平和	平和記念公園
マリン	QVC マリンフィールド
サー	エイサー
風	風の草原
日本平	アウトソーシングスタジアム日本平

5.4 クラスタリング

抽出されたキーワードを用いて収集した画像から SURF による Bag-of-Features と RGB カラーヒストグラムを生成し、エラーの増加量に対する閾値を 100 として 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は以下の表 4 のキーワードを例にそれぞれ図 4、図 5、図 6 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、赤枠で強調表示されている“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。各クラスタはこのスコアによりソートされ、上方によりスコアが高いものが表示されている。さらに、各クラスタ内の画像はクラスタ中心との距離によりソートされ、クラスタ中心に最も近い画像が左上に、次に近い画像がその右隣に順に表示されている。

クラスタリング結果により、スコアが 5 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し、全てのクラスタのスコアが 5 未満のイベントは一覧から除去した。従って、図 4、図 5、図 6 では赤枠のスコアが 5 未満のクラスタはそのイベントから除去された。

キーワード「ツリー」では、クリスマスツリーが多く分類されたクラスタのスコアが最も高く、スカイツリーが分類されたクラスタが次に高いスコアとなった。このキーワードが検出された日付は 12 月 23 日であるため、クリスマスに関するイベントの可能性が高く、クラスタリング結果においてもクリスマスツリーを多く含むクラスタの方が、よりイベントに対する視覚的関連性が高いという結果となった。

キーワード「桜」では、スコアが 5 以上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。

また、キーワード「虹」によって収集される画像は虹の特徴が小さく、写りこむ建物や背景の色、雲の状態が多様であるため、上位 2 つのクラスタを見ると、両方とも虹が写っている画像が多いが、下位のクラスタのスコアは 5 未満となった。

表 4 クラスタリング結果例のキーワード

キーワード	日付	地域	枚数
ツリー	2011/12/23	35,36,139,140	91
桜	2012/4/21	34,35,135,136	57
虹	2012/5/4	35,36,139,140	93

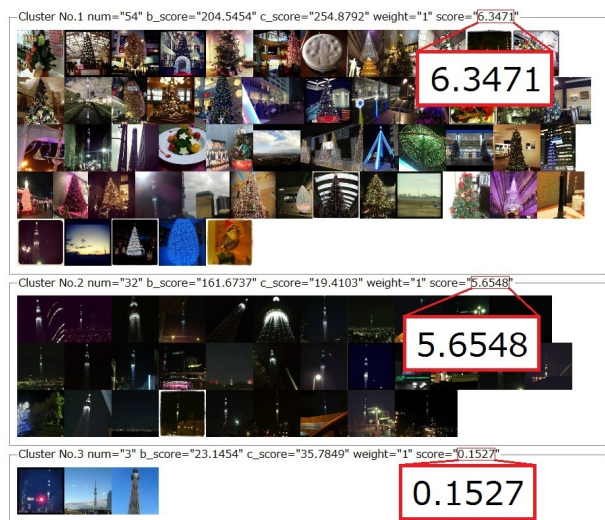


図 4 「ツリー」のクラスタリング結果



図 5 「桜」のクラスタリング結果



図 6 「虹」のクラスタリング結果



図 7 検出されたイベントのマーカ-をクリックした例

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5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-を設置し、いくつかのマーカ-をクリックした例を図7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。

また、代表的な画像の選出結果について、正しい画像が選出された例を図8に、間違った画像が選出された例を図9に示す。図8の上段左の「時代」とは「時代祭」という歴史上の様々な時代を再現した衣装や道具を身につけた人々が行列を作る祭りであり、鎧を着て馬に乗った人が画像に写っている。また、上段右の「東京ホテル」とは、この日東京スカイツリー周辺で開催されたイルミネーションイベント^(注1)であり、ライトアップされたスカイツリーが代表的な画像として選出された。下段右の「幕張メッセ」は、幕張メッセで行われた音楽イベントであり、実際のイベントの様子が表されている。最終的に検出されたイベントの数は258となり、代表画像の適合率は65.5%となった。

6. 考 察

6.1 キーワードについて

本研究は視覚的なイベントの検出として大きく2つのステップからなる。2つのステップとは、イベントのキーワードの検出と、イベントを視覚的に表す画像の抽出であるが、各イベントの画像収集には検出されたキーワードを用いるため、キーワードの検出は次のステップに大きく影響する。さらに、キ



図 8 正しい代表画像が選出された例



図 9 間違った代表画像が選出された例

ワードの検出は分割された地域ごとに独立して行われるため、各地域のグリッドの大きさによって検出されるキーワードが異なってくる。

今回の実験では、局所的なイベントと大規模なイベントの両方が検出されるように、グリッドの大きさは緯度・経度1度とした。しかし、検出されたイベントを見ると、両種のイベントは検出されているが、例えば局所的なイベント「花火」では、グリッド内の2つの場所で花火大会が行われているものがあり、それらが1つのイベントとして検出されてしまった。また、例えば大規模なイベント「桜」では、桜の開花状況は緯度・経度が1度変わってもあまり大きく変わらないため、同日に隣り合う地域で検出されることがあったが、これらは別々のイベントとは考えられない。このように、グリッドの大きさを固定して局所的なイベントと大規模なイベントの両方を対象に検出を行うと、イベントやその情報を誤って検出してしまう。従って、グリッドの大きさを可変にすることで非常に近い場所で行われた同キーワードのイベントでも、それぞれのより小さなグリッド内で一定以上のスコアが得られれば別々のイベントとして検出することができ、スコアの閾値をグリッドの大きさに比例させれば、大規模なイベントを1つのイベントとして検出することが可能であると考えられる。また、グリッドの大きさを可変にすることでイベントのスケールを推測することができ、検出されたイベントの情報量が増えるという利点もある。

キーワード検出のもう一つの欠点として、前日の状態と比較してスコアを計算している点がある。前日の状態と比較してスコアを計算する利点としては、ツイートのデータが2日間あれば検出を行うことができ、大きな計算コストを伴わないことが

(注1): <http://tokyo-hotaru.jp/index.html>

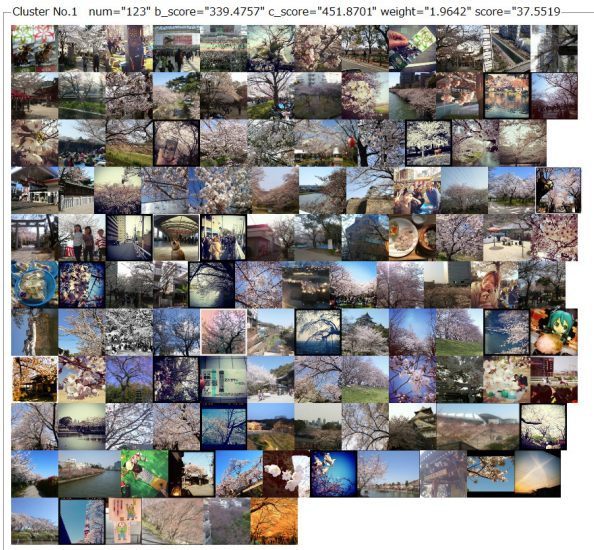


図 10 間違った画像が選出された最高スコアのクラスター

挙げられるが、例えば 2 日間連続で開催されるイベントは、1 日目が検出されても翌日のハードルが上がり、2 日目が検出されない。従って、単純に 1 日ごとに検出を行わず、検出する時間の刻みを可変にし、検出されるイベントの継続時間を求める必要がある。

6.2 画像について

視覚的なイベント検出の結果として、画像を用いることによって得られるイベントの情報量がテキストのみと比べて非常に大きいことは明らかである。さらに、キーワードだけではあまり想像できないイベントや「幕張メッセ」のように場所の名前がキーワードとして検出された場合でも、画像によりイベントの様子を捉えることができた。

代表的な画像は、正しい画像が選出されなかったイベントがいくつかあった。しかし、これらの間違った画像が選出されたイベントの最高スコアのクラスターを見ると、他に正しいと思われる画像が多く含まれている場合があった。例として、図 9 の「桜」の最高スコアのクラスターを図 10 に示す。図中のクラスターには桜の画像が多くあるが、選出された代表画像は最も左上の画像であり、「桜」の代表画像とは言えない。代表的な画像は単純に最高スコアのクラスターの中心に最も近い画像を選出するのではなく、他の選出方法に改良することで、代表画像の適合率の上昇が期待できる。

本研究で用いたデータは位置情報と画像の両方が付加されたツイートであった。人は視覚的に何か珍しいものやことが起こったからこそ、写真を撮り、他の人と共有するため Twitter に投稿する。従って、これらのツイートは視覚的な珍しさとその土地との関連性を持つデータであり、何らかのイベントに関連している可能性が高い。そのため位置情報付き画像ツイートを用いることによって、誤検出は最小に抑えることができるが、投稿される数が少ないため、特に地方におけるイベントの検出は難しいと考えられる。

7. おわりに

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、各クラスター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、最もスコアが高いクラスター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ーからイベントの中心座標を計算して地図上に表示した。結果として、国内の様々な地域から台風や虹といった自然現象や、花火大会やイルミネーションのような局所的なイベントが検出され、未知のイベントやキーワードだけでは想像できないイベントでも、画像によりその様子を視覚的に捉えられることを確認した。

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、システムを完全に自動化し、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。本研究ではリアルタイムな検出を想定し、画像はサムネイルを、局所特徴量は高速な SURF を利用しているが、画像のダウンロードや特徴ベクトルの生成にはやはり時間が掛かるため、高速に検出できるようシステムを最適化する必要がある。

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] L. Hong, A. Ahmed, S. Gurumurthy, A. J. Smola, and K. Tsioutsoulis. Discovering geographical topics in the twitter stream. In *Proc. of the ACM International World Wide Web Conference*, pp. 769–778, 2012.
- [4] C. Li, A. Sun, and A. Datta. Twevent: Segment-based event detection from tweets. In *Proc. of the Conference on Information and Knowledge Management*, 2012.
- [5] 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.
- [6] Mecab. <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>.
- [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.