

画像の位置推定を用いたマイクロブログからの視覚的なイベント検出

金子 昂夢¹ 松尾 真¹ 柳井 啓司¹

1. はじめに

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、Twitter の普及により自分の現在の状況を投稿する人が増加してきた。Twitter にはテキストの他に画像や位置情報を添付でき、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、Twitter に投稿された位置情報付き画像ツイートからリアルタイムなイベント画像の検出を試みる。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、視覚的に捉えることを目標とする。

2. 関連研究

Twitter は Web マイニング対象として広く研究されており、イベントの検出を試みた研究も多々ある。

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。あらかじめイベントに関連する位置情報付き画像ツイートを収集し、画像付きのイベント検出を行うことで、時間や地域による代表画像の違いが示された。

また、我々が以前提案したシステム [1] では、利用したデータは位置情報と画像の両方を持ったツイートのみであったため、検出されないイベントも多く、収集できた関連画像も少なかった。従って、本研究では位置情報付き画像ツイートに加えて、位置情報無し画像ツイートおよび画像なし位置情報付きツイートも利用するイベント検出手法を提案し、さらにイベントキーワード検出手法を改良することによって、これらの問題点を解決する。

3. 提案手法

3.1 システムの流れ

本研究では、任意のイベントキーワードを自動的に抽出し、キーワードを用いて収集した画像を解析することで、イベントを視覚的に捉えることができるようにする。本システムの流れを以下に示す。

- (1) 位置情報か画像を持ったツイートの収集
- (2) イベントキーワードの検出
- (3) 画像ツイートの位置推定
- (4) 収集した画像のクラスタリング
- (5) 代表画像選出・イベントのマッピング

3.2 対象領域の分割

地域によるユーザ人口の偏りを考慮するため、対象領域を緯度・経度それぞれ 0.5 度ずつのグリッドにより分割し、以下の式 1 により各地域の重み付けを行った。

$$W_{i,j} = \frac{\#users_{max} + s}{\#users_{i,j} + s} \quad (1)$$

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、 $\#users_{max}$ は最も出現するユーザ数が多い地域のユーザ数、 $\#users_{i,j}$ 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。 $W_{i,j}$ により最も出現ユーザ数の多い地域の重みは 1、それ以外では 1 よりも大きい値となり各地域ごとのキーワードのスコアを平滑化するものである。

3.3 文字列の非重要度の計算

本研究では、ユニグラムを用いて日本語では 1 文字、英語では 1 word を最小単位のユニットとして、前後に出現するユニットを順に結合し、新たなユニットとして解析した。各文字列のユニットについて 1 日を時間の区切りとし、ある 1 日に対象の文字列を投稿したユーザ数から、以下の式 2 により非重要度を定義した。

$$Dic(w) = \sum_{i,j} \frac{E(\#users_{w,i,j})^2}{V(\#users_{w,i,j}) + 1} W_{i,j} \quad (2)$$

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、 $\#users$ はある 1 日にそれぞれの地域で文字列 w を含む投稿をしたユーザ数であり、 $E(\#users_{w,i,j})$ は対象の投稿をした 1 日の平均ユーザ数、 $V(\#users_{w,i,j})$ はその分散である。日常的に広い地域で投稿される文字列は非重要度が大きくなり、イベントのキーワードとして検出されにくく、日によって投稿数に大きな差があるもの、日常的にはあまり出現しないものは非重要度が小さくなり、ある 1 日に集中した投稿があった場合イベントとして検出されることになる。

3.4 キーワードの検出

位置情報付きツイートをを用いて、文字列のバーストスコアを計算し、キーワードの検出を行った。3.3 と同様にテキスト解析を行い、文字列のバーストスコアを以下の式 3 により定義した。

$$S_{w,i,j} = \frac{\#users_{w,i,j}}{Dic(w)} W_{i,j} \quad (3)$$

式 3 において、 w は各文字列ユニット、 $\#users$ は対象となるツイートを投稿したユーザ数である。 $S_{w,i,j}$ は文字列の非重要度と地域の重みを考慮した各地域における文字列 w のバーストスコアとなり、今回はこのスコアが 200 以上となる文字列をイベントのキーワードとして検出した。

ここで、検出には各グリッドのうち隣接する 4 つのグリッドを 1 つのウィンドウとして同時に調査し、検出されたキーワードは、隣接するウィンドウで検出されたキーワードと比較し、互いのキーワードに関連するツイートが 50% 以上同じツイートの場合、同じイベントのキーワードとして統合した。

¹ 電気通信大学大学院 情報理工学研究所 総合情報学専攻

3.5 位置推定による画像データ拡張

本手法では、イベントに関連するツイートであるかそうでないかの二値問題として画像ツイートの位置推定を行った。

各キーワードの検出に貢献したツイートをポジティブ、いずれかのキーワードを持ち、別の地域で投稿されたツイートをネガティブなデータとし、テキストによる Naive Bayes 法と画像による Naive Bayes Nearest Neighbor (NBNN) 法を組み合わせ判定した。NBNN 法では画像から SIFT 特徴を抽出し、テキストの結果と組み合わせるためにコサイン類似度を用いた。クエリが分類されるクラス c はテキストと画像それぞれの手法を結合し、以下の式 4 により決定した。

$$\hat{c} = \arg \max_c P(c) \prod_{i=1}^n P(x_i|c) \sum_{j=1}^v \frac{d_j \cdot NN_c}{\|d_j\| \|NN_c\|} \quad (4)$$

式 4 において、 x はキーワードの文書ベクトルであり、 n はクエリが持つキーワード数、 d はクエリ画像の局所特徴ベクトル、 v は検出された局所特徴数、 NN_c は各クラスの学習サンプルで最も類似した局所特徴ベクトルである。

3.6 代表画像選出

それぞれのイベントを視覚的に表すことができる代表的な画像を抽出するため、画像のクラスタリングを行った。画像特徴量として、Deep Convolutional Neural Network[2] の第 7 層の出力である 4096 次元のベクトルを L2 正規化して利用した。

クラスタリングには、まず、以下の式 5 によりクラスタのスコアを定義した。

$$V_C = \frac{\#images_C}{\sum_{x \in C} \|x - \bar{x}\| + 1} \quad (5)$$

式 5 において、 $\#images_C$ はクラスタ C に分類された画像数、 x は各画像の DCNN 特徴ベクトル、 \bar{x} はクラスタ中心のベクトルである。クラスタスコアはクラスタ内の画像数が多く、クラスタ内の類似度がより高ければ高い値となる。クラスタリングは Ward 法を改良し、クラスタからノイズの除去も行うように以下のステップにより行った。

- (1) 全ての要素が独立したクラスタ状態から開始
- (2) 各クラスタにおいて各要素を取り除いた場合のクラスタスコアを計算
- (3) 各クラスタにおいて他クラスタの要素を追加した場合のクラスタスコアを計算、ただし移動元のクラスタスコアを下回る場合は 0 とする
- (4) 2., 3. のうち最大スコアが得られる処理を実行
- (5) 実行する処理がない場合は終了、いずれかの処理を実行した場合 2. に戻る

クラスタリング結果により、最も高いスコアを持つクラスタを代表クラスタとし、さらに、代表クラスタに含まれる画像のうち、最もクラスタ中心に近い画像をイベントの代表的な画像として選出した。

4. 実験

今回の実験を行うに当たって、2012 年を対象にあらかじめ位置情報が画像が付加されたツイートを収集し、2012 年全体のツイートをを用いて地域の重みと文字列の非重要度を学習した。

イベント検出は 2012 年 8 月を対象に行った。検出されたキーワードの一部を表 1 に、収集された画像のクラスタリング結果を 8 月 6 日に検出されたイベント「ねぶた」を



図 1 2012 年 8 月 6 日の「ねぶた」画像。

例に図 1 に示す。図中の各クラスタはスコアによりソートされ、各画像はクラスタ中心との距離によりソートされている。また、各画像は元々位置情報を持っていたものは赤枠、位置推定により追加された画像は黄枠で表示されている。本手法で 2012 年の 8 月に検出されたイベントの総数と検出精度、各代表画像の適合率を以下の表 2 にまとめた。

表 1 キーワードの検出結果の一部

キーワード	日付	地域 (lat,lng)	スコア
花火大会	20120801	33,129.5	297.7
虹	20120801	34,134.5	229.1
ROCK IN JAPAN	20120803	36,140	430.3
鮎まつり	20120804	34.5,138.5	265.1
ねぶた	20120806	40.5,140	255.7
阿波踊り	20120814	34,134	589.8
落雷	20120818	34,135	367.5
ブルームーン	20120831	34.5,136	269.7

表 2 イベントの検出結果

	提案システム	従来システム [1]
イベント数	310	35
イベントの精度 (%)	81.3	77.1
代表画像の適合率 (%)	88.7	65.5

5. おわりに

本研究では、Twitter に投稿されたツイートのうち位置情報が画像のいずれかを持ったツイートから視覚的なイベント検出を行った。従来のシステムよりも検出イベント数が大幅に増加し、イベント検出精度も向上した。

今回の実験では表記揺れによって正しく統合されないイベントがあったため、今後は、キーワードの共起ではなく、その位置情報を利用した同イベントのキーワードの同定を行うことでこの問題が解決する予定である。また、代表画像だけでなく他の画像の利用価値を再検討し、より視覚的に多様な出力を可能なシステムを実現することを考えている。

参考文献

- [1] Kaneko, T. and Yanai, K.: Visual Event Mining from Geotweet Photos, *Proc. of IEEE ICME Workshop on Social Multimedia Research (SMMR)* (2013).
- [2] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems* (2012).
- [3] Nakaji, Y. and Yanai, K.: Visualization of Real World Events with Geotagged Tweet Photos, *Proc. of IEEE ICME Workshop on Social Media Computing (SMC)* (2012).