

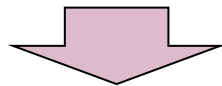


位置情報付き画像ツイートを利用した 視覚的なイベント検出

電気通信大学 電気通信学部 情報工学科
金子 昂夢, 柳井 啓司

背景

- スマートフォンの普及
-位置情報付き画像
- Twitterの普及
-リアルタイムな投稿



位置情報付き画像ツイートの増加

目的

- ツイートからイベントの検出
 - 台風や虹のような自然現象
 - 花火大会のような局所的な行事
- イベントの情報量の向上
 - 何が・いつ・どこで
 - どんな様子だったのか





関連研究

- テキストと位置情報を利用した研究は多い
- 榊らの研究[WWW 2010]
 - Twitterのユーザ=ソーシャルセンサ
 - 高速・高精度に地震を検出
- Leeらの研究[ACM SIGSPATIAL WS 2010]
 - 対象をより小さな領域に分割
 - 通常の状態との違いにより検出



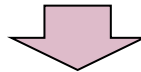
関連研究

- 画像を用いた研究は非常に少ない
- 中地らの研究[ICME WS 2012]
 - あらかじめ特定のキーワードや日時を与える
 - 位置情報付き画像ツイートから代表画像を選出
 - 時間や地域の違いを検出

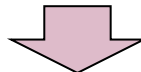


方針

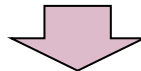
1. 特定の地域で頻出するキーワードの抽出



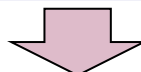
2. 抽出されたキーワードの統合・補完



3. キーワードに対する画像の解析



4. イベントの代表画像選出・中心座標計算



5. 検出されたイベントを地図上に表示



タグの生成

- ツイート本文からタグを生成
 - 形態素解析エンジンMeCabを利用
 - 品詞が「名詞」の単語
 - 品詞細分類が「サ変接続」、「数」は除外

長岡の花火大会です♪花火綺麗(´▽`)#長岡まつり2012

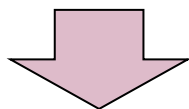
| | |
|-------|--------------------------------|
| 長岡 | 名詞,固有名詞,一般,*,*,*長岡,ナガオカ,ナガオカ,, |
| の | 助詞,格助詞,一般,*,*,*の,ノ,ノ,, |
| 花火 | 名詞,一般,*,*,*,*花火,ハナビ,ハナビ,, |
| 大会 | 名詞,一般,*,*,*,*大会,タイカイ,タイカイ,, |
| です | 助動詞,*,*,*特殊・デス,基本形,です,デス,デス,, |
| ♪ | 名詞,サ変接続,*,*,*,*,* |
| 花火 | 名詞,一般,*,*,*,*花火,ハナビ,ハナビ,, |
| 綺麗 | 名詞,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 | フィラー,*,*,*,*,*ま,マ,マ,, |
| つ | 助動詞,*,*,*下二・夕行,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,, |



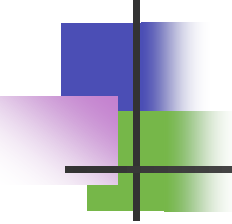
タグの生成

- ツイート本文からタグを生成
 - 形態素解析エンジンMeCabを利用
 - 品詞が「名詞」の単語
 - 品詞細分類が「サ変接続」、「数」は除外

長岡の花火大会です♪花火綺麗(´▽`)#長岡まつり2012



長岡、花火、大会、綺麗



キーワードの抽出

- 緯度・経度1度ごとのグリッドに分割
- 出現するユーザ数から各地域を重み付け

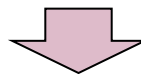
$$W_{i,j} = \frac{M + s}{N_{i,j} + s}$$

- スコアが50以上となる単語を抽出

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

キーワードの抽出

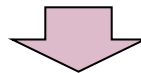
キーワード「花火」



地域の重み
1.9642

2012/8/3
16ユーザ

2012/8/4
134ユーザ



キーワード「花火」の
この地域における
2012/8/4のスコアは
 $(134 - 16) \times 1.9642$
= 231.7756



キーワードの統合・補完

- キーワードの統合

- ツイートが50%以上同じなら統合

「花火」、「大会」 → 「花火」

- キーワードの補完

- 前後の文字が80%以上同じならその文字で補完

「スカイ」 → 「スカイツリー」



クラスタリング

- 画像特徴量

- SURFによるBag-of-Features

- RGBカラーヒストグラム

- Ward法

- エラーの増加量が最少の組を逐次的に併合

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

$$E(C) = \sum_{x \in C} ((x_{BoF} - \overline{x_{BoF}})^2 w_{BoF} + (x_{RGB} - \overline{x_{RGB}})^2 w_{RGB})$$



視覚的な検出

- 各クラスタのスコアを計算

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

- スコアが高いほど視覚的関連性が高い
 - 最高スコアのクラスタから代表画像選出
 - 5以上から中心座標計算
 - 5未満はノイズ画像群



実験

- データセット

- 日本国内

- 2011年2月10日から2012年9月30日

- 約3,000,000件

- 画像はサムネイル



キーワードの抽出

| キーワード | 日付 | 地域 | キーワード | 日付 | 地域 |
|-------|------------|----|-------|-----------|----|
| 雪 | 2011/2/11 | 大阪 | 初日の出 | 2012/1/1 | 東京 |
| 地震 | 2011/3/11 | 東京 | マラソン | 2012/2/26 | 東京 |
| 花火 | 2011/8/6 | 大阪 | 六本木 | 2012/3/24 | 東京 |
| 大会 | 2011/8/6 | 大阪 | 桜 | 2012/4/28 | 福島 |
| 台風 | 2011/9/21 | 東京 | 幕張メッセ | 2012/4/28 | 千葉 |
| 富士山 | 2011/9/24 | 山梨 | ホテル | 2012/5/6 | 東京 |
| アップル | 2011/10/6 | 東京 | 日食 | 2012/5/21 | 東京 |
| 月食 | 2011/12/10 | 東京 | 金環 | 2012/5/21 | 茨城 |
| リ工 | 2011/12/10 | 兵庫 | 一過 | 2012/6/20 | 東京 |
| クリスマス | 2011/12/24 | 東京 | 送り火 | 2012/8/16 | 京都 |
| 大晦日 | 2011/12/31 | 東京 | 大文字 | 2012/8/16 | 京都 |

キーワードの統合

| 統合前 | 統合後 | 統合前 | 統合後 |
|----------|------|------------|-------|
| 花火,大会,淀川 | 花火 | 幕張,県立,海浜 | 幕張 |
| アップル,銀座 | アップル | 花火,大会,隅田川 | 花火 |
| 鈴鹿,サーキット | 鈴鹿 | スカイ,ツリー | スカイ |
| 月食,皆既,月 | 月食 | サーキット,鈴鹿 | サーキット |
| 六本木,ヒルズ | 六本木 | 花火,大会,江戸川 | 花火 |
| 風,雨 | 風 | 花火,大会 | 花火 |
| 桜,綺麗 | 桜 | 花火,諏訪湖 | 花火 |
| スーパー,ムーン | スーパー | 送り火,五山 | 送り火 |
| 国際,幕張メッセ | 国際 | 花火,大会,昼,全国 | 花火 |
| 海,海浜,中道 | 海 | ブルー,ムーン | ブルー |
| 日食,金環 | 日食 | 風,湖畔,北地 | 風 |

キーワードの補完

| 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 |
| スーパー | スーパームーン |
| ホテル | 東京ホテル |
| 母 | 母の日 |
| 海 | 海の中道浜公園 |

| 補完前 | 補完後 |
|-------|-----------------|
| 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 |
| ブルー | ブルームーン |
| サー | エイサー |
| 風 | 風の草原 |



クラスタリング

- 以下のクラスタリング結果

| キーワード | 日付 | 地域 | 枚数 |
|-------|------------|----|----|
| 花火 | 2011/8/6 | 東京 | 89 |
| ツリー | 2011/12/23 | 東京 | 91 |
| 桜 | 2012/4/21 | 大阪 | 57 |
| 虹 | 2012/5/4 | 東京 | 93 |
| 東京ホテル | 2012/5/6 | 東京 | 85 |

Cluster No.1 num="40" b_score="127.5948" c_score="36.7071" weight="1" score="9.7382"



9.7382

Cluster No.2 num="22" b_score="121.0945" c_score="58.4237" weight="1" score="2.6961"



2.6961

Cluster No.3 num="25" b_score="114.3028" c_score="148.3092" weight="1" score="2.3799"



2.3799

Cluster No.4 num="2" b_score="36.5067" c_score="10.0696" weight="1" score="0.0859"



0.0859

「花火」のクラスタリング結果

Cluster No.1 num="54" b_score="204.5454" c_score="254.8792" weight="1" score="6.3471"



Cluster No.2 num="32" b_score="161.6737" c_score="19.4103" weight="1" score="5.6548"



Cluster No.3 num="3" b_score="23.1454" c_score="35.7849" weight="1" score="0.1527"



0.1527

Cluster No.4 num="2" b_score="8.2787" c_score="39.7969" weight="1" score="0.0832"



0.0832

「ツリー」のクラスタリング結果

Cluster No.1 num="32" b_score="89.4698" c_score="127.6658" weight="1.9642" score="9.2631"



9.2631

Cluster No.2 num="24" b_score="77.7001" c_score="90.9009" weight="1.9642" score="6.7104"



6.7104

Cluster No.3 num="1" b_score="0" c_score="0" weight="1.9642" score="0.0002"



0.0002

「桜」のクラスタリング結果

Cluster No.1 num="37" b_score="106.2934" c_score="141.4798" weight="1" score="5.5252"



Cluster No.2 num="51" b_score="163.8098" c_score="398.7042" weight="1" score="4.6239"



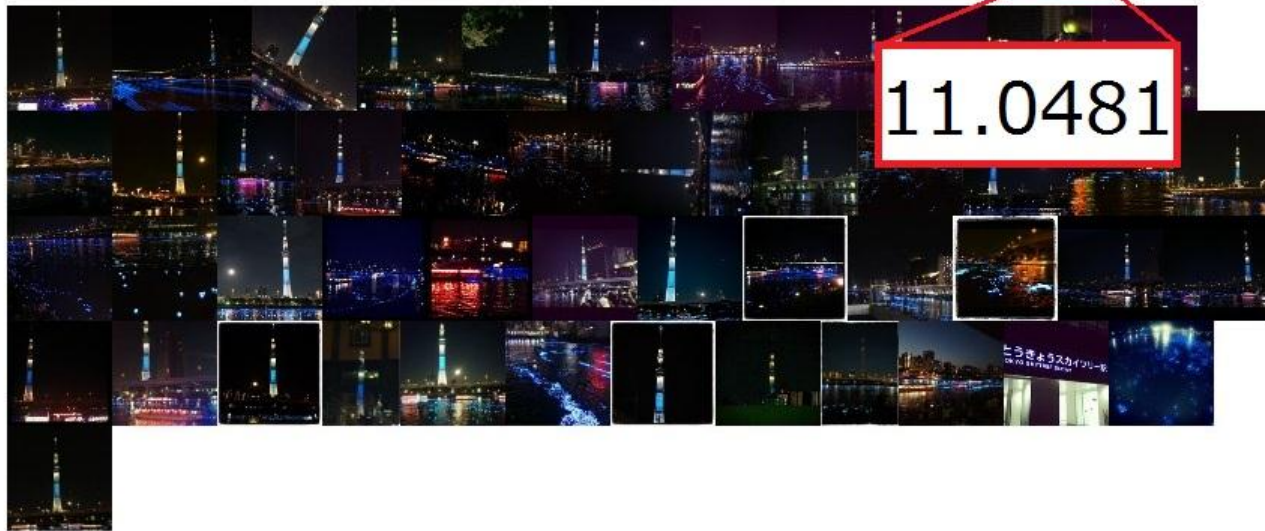
Cluster No.3 num="5" b_score="11.4403" c_score="99.1252" weight="1" score="0.2261"



0.2261

「虹」のクラスタリング結果

Cluster No.1 num="48" b_score="164.1649" c_score="44.3774" weight="1" score="11.0481"



11.0481

Cluster No.2 num="27" b_score="107.3085" c_score="152.9358" weight="1" score="2.8012"



2.8012

Cluster No.3 num="5" b_score="53.0887" c_score="2.3569" weight="1" score="0.4509"



0.4509

Cluster No.4 num="5" b_score="18.4136" c_score="113.2231" weight="1" score="0.1899"



0.1899

「東京ホテル」のクラスタリング結果



イベントの検出結果

- 検出されたイベントを地図上に表示
 - 中心座標からマーカールを設置
 - イベントの情報と代表画像を対応付け
- 検出されたイベント数は258
- 代表画像の適合率は65.5%



台風
 2012/9/15
 緯度:26.30841
 経度:127.75061





まとめ

- 画像を用いた視覚的なイベント検出
 - キーワードの検出
 - イベントを視覚的に表す画像を抽出
- 画像の持つイベントの情報量は多い
- グリッドの大きさに依存
 - 同グリッド内では1つのイベント
 - 隣接するグリッドでは別のイベント



今後の課題

- 柔軟な検出
 - グリッドの大きさを可変にする
 - 時間の刻みを可変にする
- リアルタイムな検出
- 代表画像選出方法の改善