

位置情報付き画像ツイートを利用した リアルタイムイベント画像検出

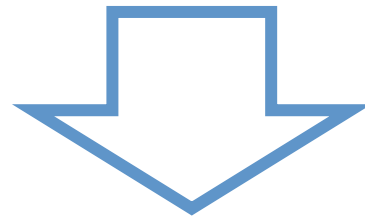
金子 昂夢、柳井 啓司

電気通信大学 大学院情報理工学研究科

総合情報学専攻

背景

- スマートフォンの普及
-位置情報付き画像
- Twitter
-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

- Twitterストリームからイベントの検出
 - 天災、自然現象
 - 祭り、スポーツの試合
- 視覚的な検出
 - 代表画像の選択
 - 地図上に表示
- リアルタイム検出



画像と共に地図上に配置

関連研究: Twitterイベントマイニング

- 多くがテキスト情報のみを利用
- 榊らの研究 [WWW 2010]
 - ユーザをソーシャルセンサと捉えたモデルを構築
 - 地震の発生を位置と共に高速に検出
- Leeらの研究 [ACM SIGSPATIAL WS 2010]
 - 対象地域をより小さな領域に分割
 - ツイートの頻度をモニタリング

関連研究: Twitter画像マイニング

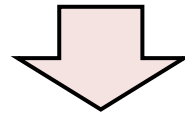
- Twitterに投稿された画像を利用
- 柳井の研究 [ACM ICMR 2012]
 - リアルタイムに投稿された画像をモニタリング
 - 地図上に画像をマッピング
- 中地らの研究 [ICME WS 2012]
 - キーワードに関連する代表画像を抽出
 - 位置や時間による違いを比較

システムの概要

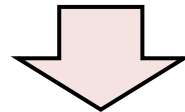
- バッチ処理システムの構築
 - テキスト解析によりイベント検出
 - 画像解析により代表画像抽出
- リアルタイム検出
 - タイムラインをリアルタイムに監視
 - 検出処理を定期的に実行

システムの概要

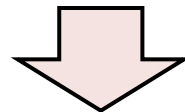
1. イベントキーワードの検出



2. キーワードの統合・補完



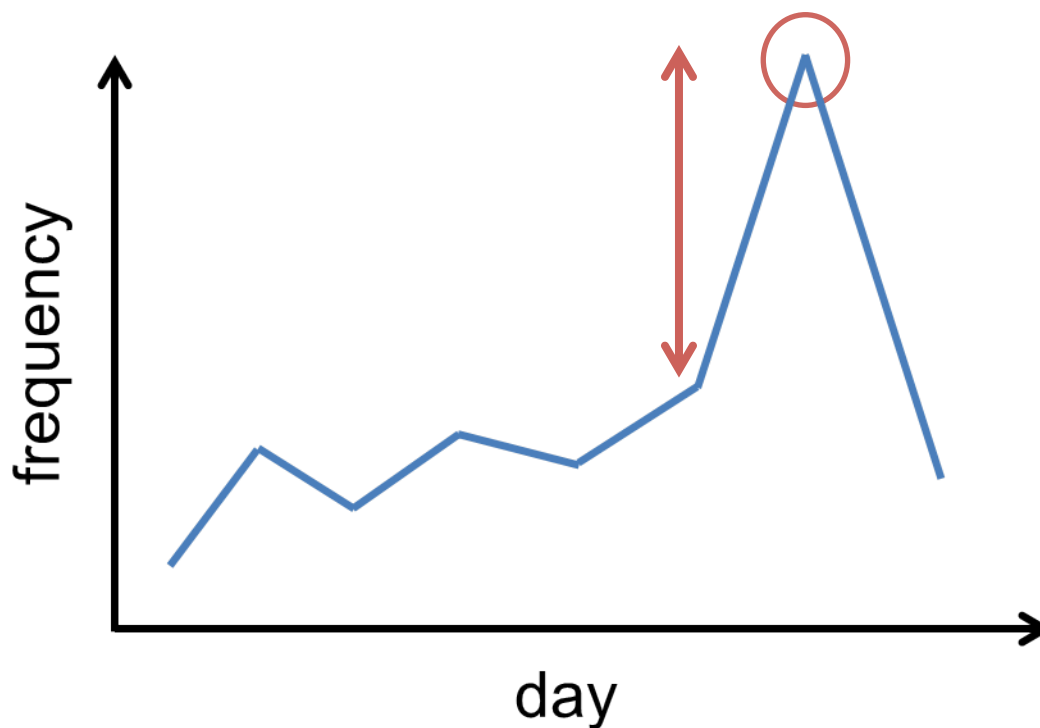
3. 画像のクラスタリング



4. 画像と共に地図上に表示

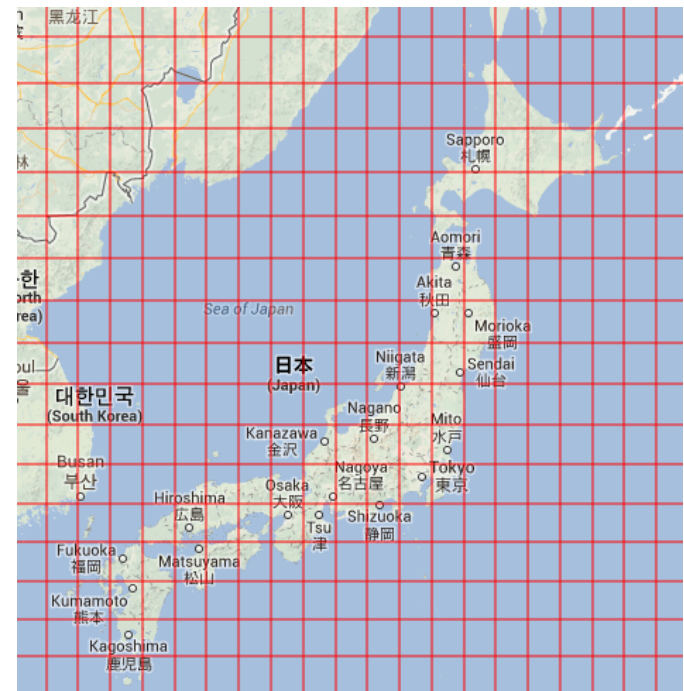
イベントキーワードの検出

- 単語の出現頻度からバースト検出



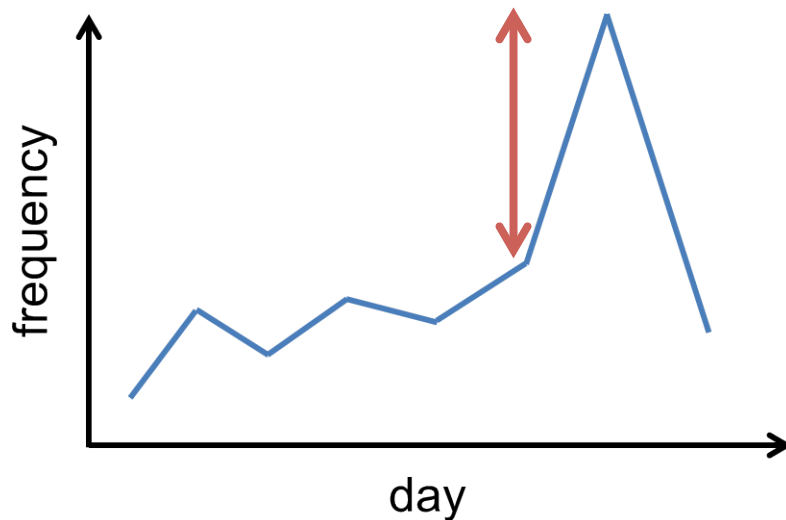
イベントキーワードの検出

- 形態素解析
 - MeCab(<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>)
- 対象地域をより小さな領域に分割
 - 緯度・経度1度ごとのグリッド

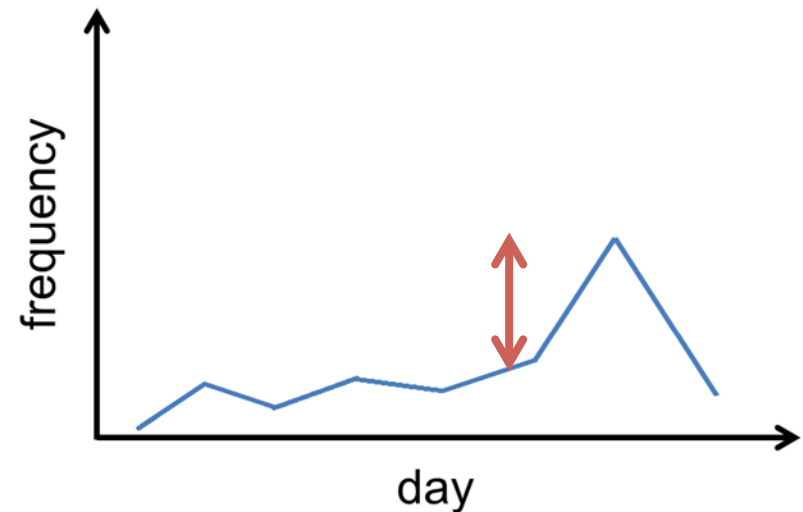


イベントキーワードの検出

- ユーザ人口の違い
 - バーストスケールの違い



人口の多い地域



人口の少ない地域

イベントキーワードの検出

- 地域の重みを計算

$$W_{area} = \frac{\#users_{\max} + sd}{\#users_{area} + sd}$$

- キーワードのスコアを計算
 - ユーザ数の前日との差
 - 50以上

$$S_{w,date,area} = (N_{w,date,area} - N_{w,date-1,area})W_{area}$$

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
 - 最もスコアの高いキーワードを利用

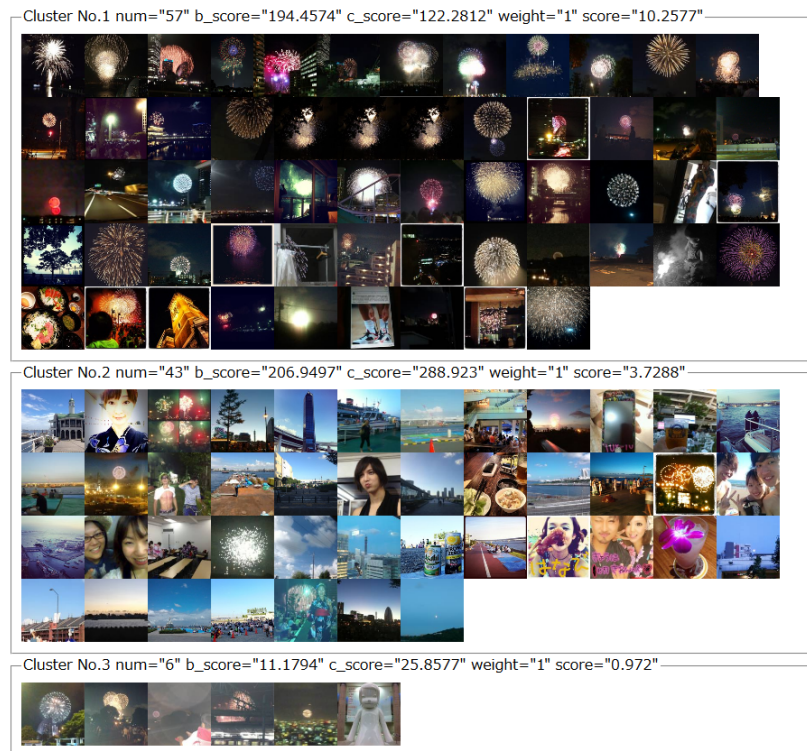
「花火」、「大会」 → 「花火」

- キーワードの補完
 - 文字が80%以上同じ
 - 前後の文字を再帰的に補完

「スカイ」 → 「スカイツリー」

画像のクラスタリング

- 画像特徴量
 - SURFのBag-of-Features
 - カラーヒストグラム
- Ward法
 - 階層型クラスタリング手法
 - 閾値で終了



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

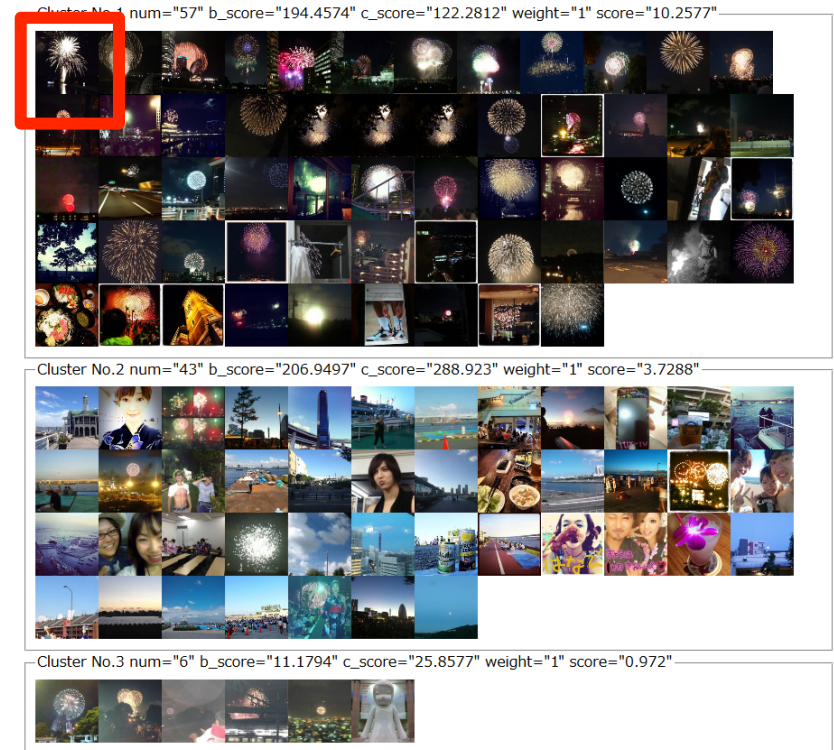
代表画像の選択

- 代表クラスタの選択
-クラスタのスコアを計算

$$V_C = \frac{\# photos_C}{E(C)} W_{area}$$

- 代表画像の選択
-最大スコアのクラスタ中で
中心に最も近いもの

- 低いスコアを持つクラスタを除外



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300万件の位置情報付き画像ツイート
- リアルタイム処理実験
 - 2014年3月10日
 - プールしながら100秒ごとに検出処理を実行

キーワードの検出結果

キーワード	日付	地域
雪	2011/2/11	大阪
地震	2011/3/11	東京
花火	2011/8/6	大阪
大会	2011/8/6	大阪
台風	2011/9/21	東京
富士山	2011/9/24	山梨
アップル	2011/10/6	東京
月食	2011/12/10	東京
リエ	2011/12/10	兵庫
クリスマス	2011/12/24	東京
大晦日	2011/12/31	東京

キーワード	日付	地域
初日の出	2012/1/1	東京
マラソン	2012/2/26	東京
六本木	2012/3/24	東京
桜	2012/4/28	福島
幕張メッセ	2012/4/28	千葉
ホテル	2012/5/6	東京
日食	2012/5/21	東京
金環	2012/5/21	茨城
一過	2012/6/20	東京
送り火	2012/8/16	京都
大文字	2012/8/16	京都

キーワードの統合・補完結果

統合・補完前	統合結果	補完結果
スカイ、ツリー	スカイ	スカイツリー
花火、大会	花火	花火
日食、皆既	日食	皆既日食
月食、皆既、月	月食	皆既月食
マラソン	マラソン	東京マラソン
リエ	リエ	ルミナリエ
スーパー、ムーン	スーパー	スーパームーン
送り火、五山	送り火	送り火
鈴鹿、サーキット	鈴鹿	鈴鹿サーキット
台風、一過	一過	台風一過

「花火」のクラスタリング結果

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



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



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



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



0.0859

「桜」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"



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



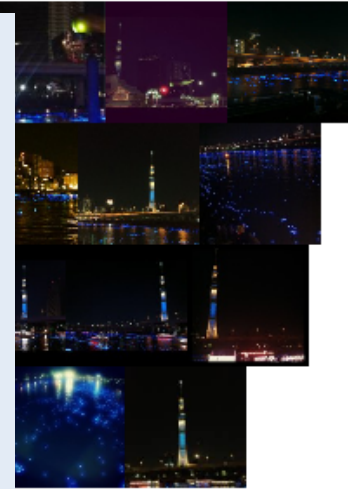
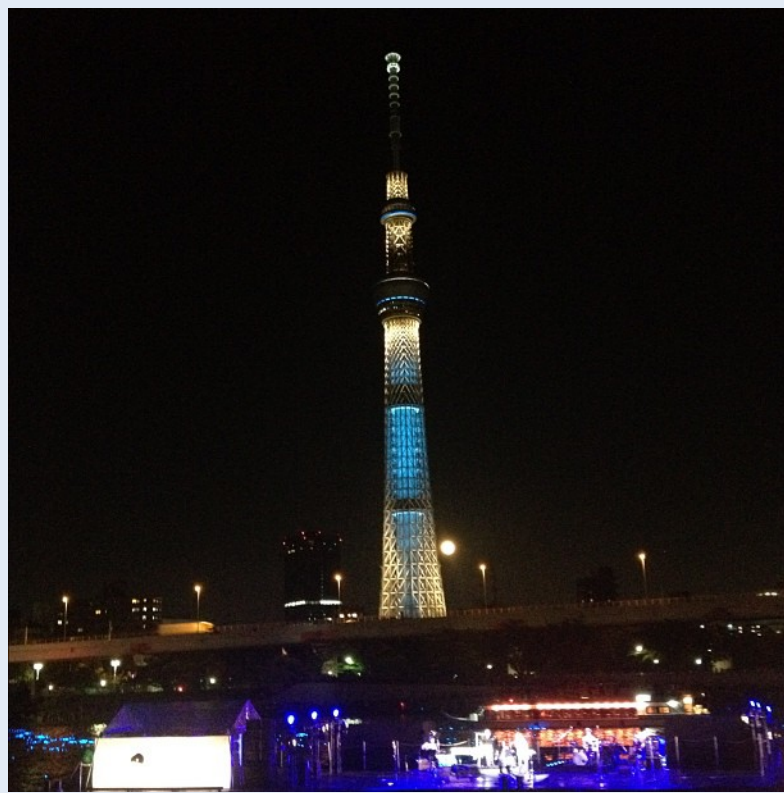
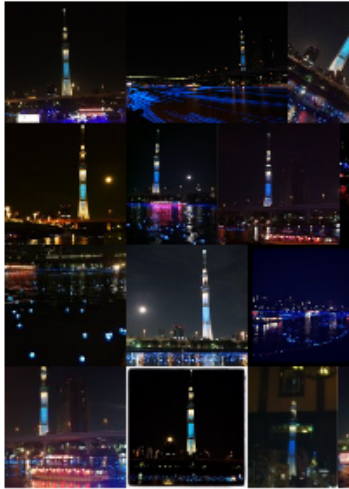
Cluster No.3 num="1" b_score="0" c_score="0" weight="1.9642" score="0.0002"



代表画像の選択

東京ホテル 2012/5/6

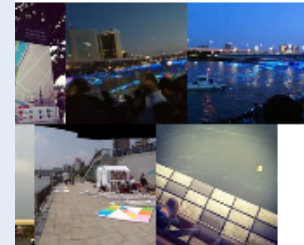
Cluster No.1 num="48" b_score="164.1649" c_score="44.3774" weight="1" score="11.0481"



Cluster No.2 num="27" b



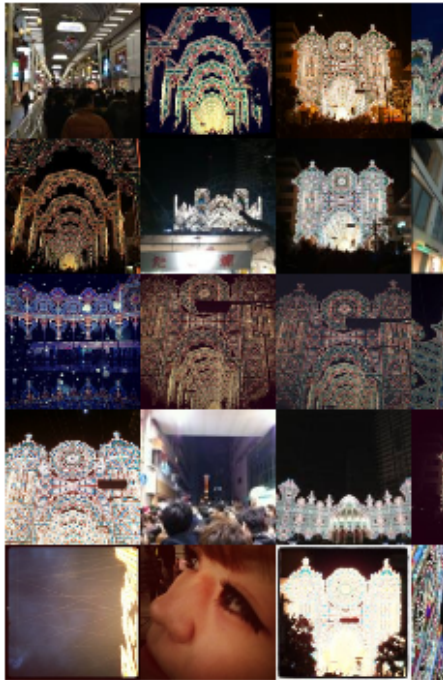
= "2.8012"



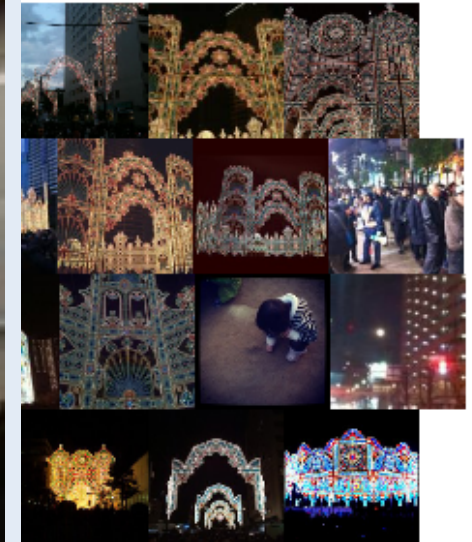
代表画像の選択

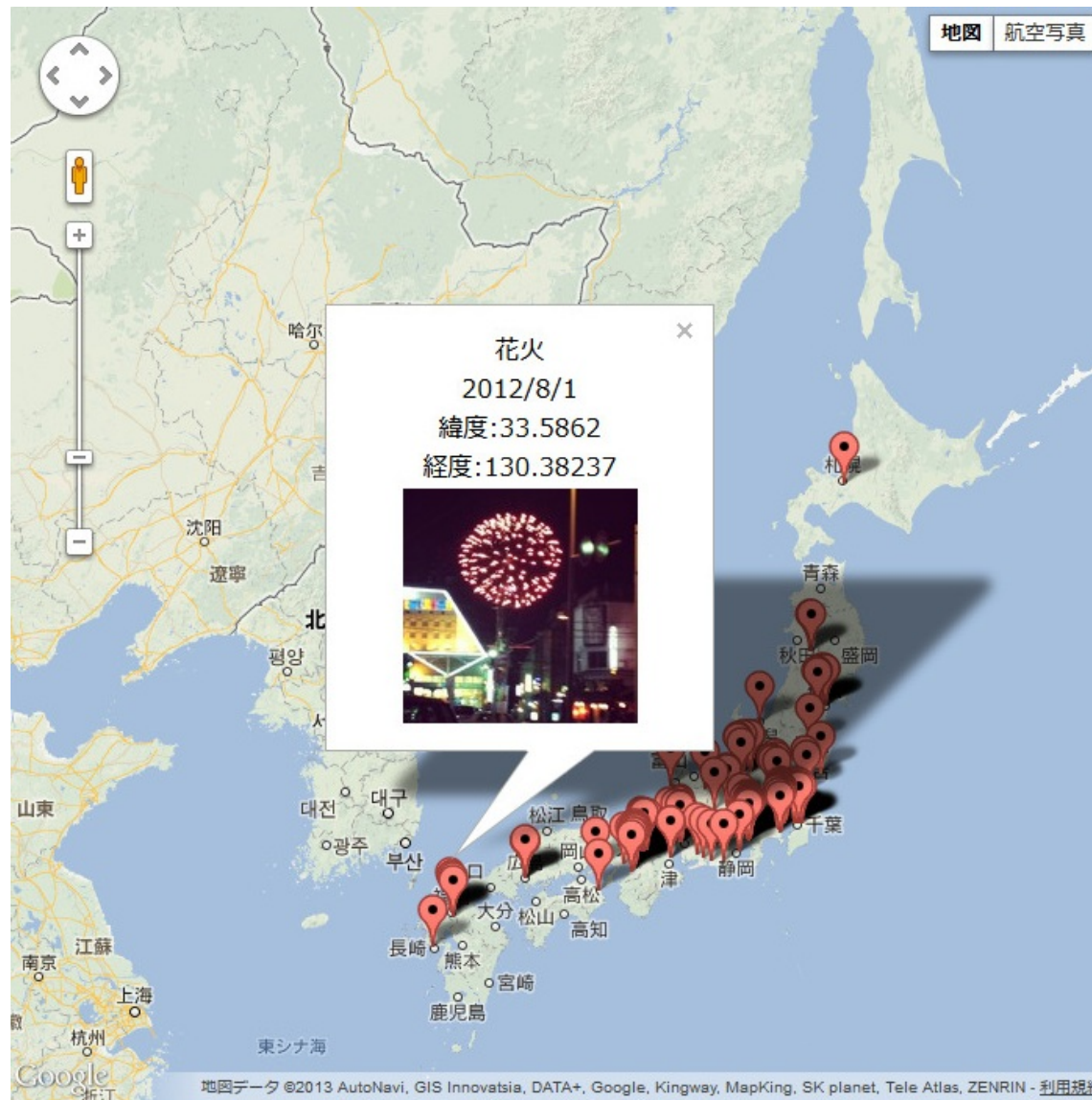
ルミナリエ 2011/12/10

Cluster No.1 num="53" b_sc



42" score="12.5526"





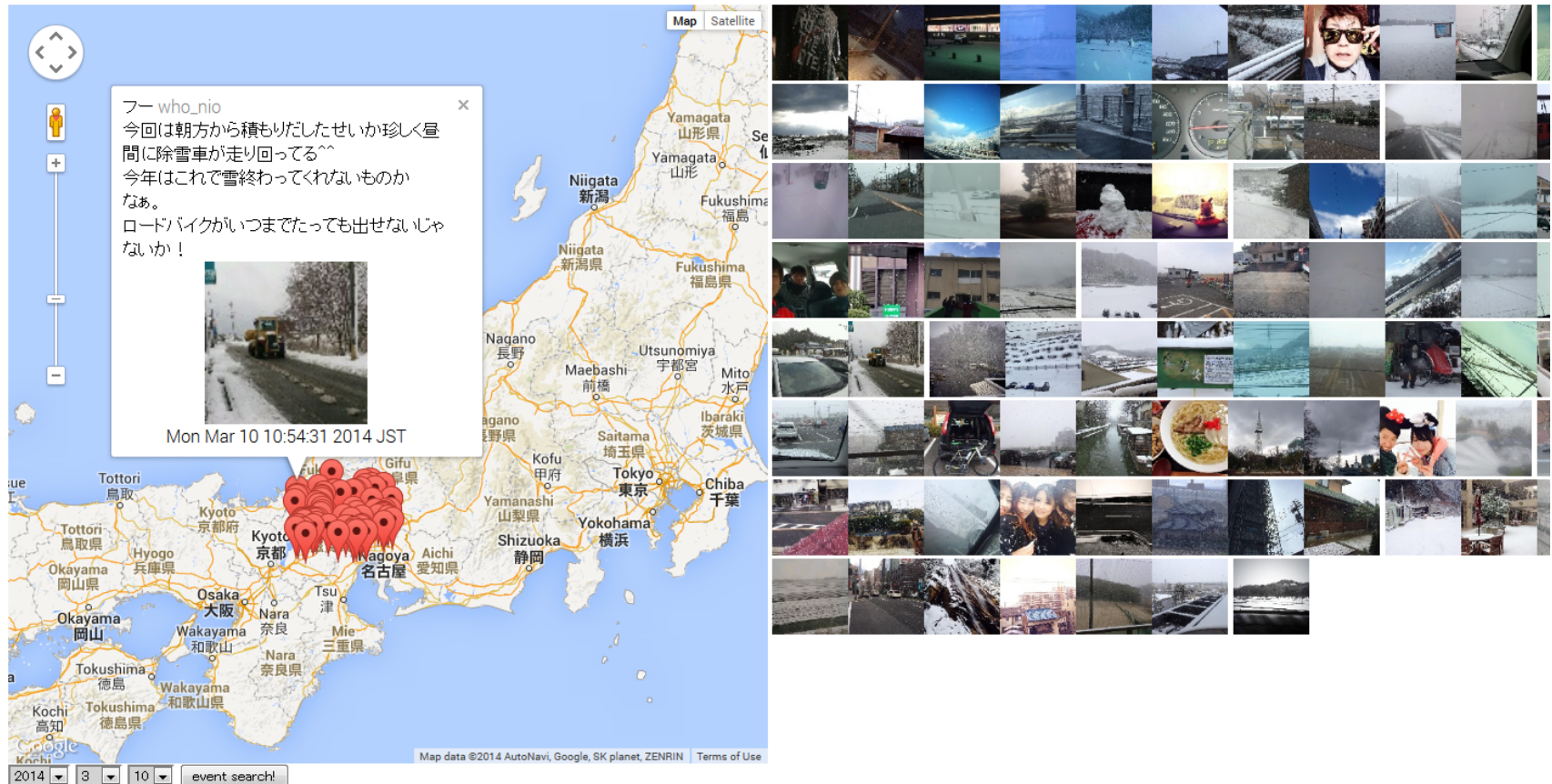
データ選別結果

リアルタイム処理実験

キーワード	位置(lat,lng)	時刻(h:m:s)
雪	43.0,141.4	8:41:23
雪	35.0,136.1	8:51:25
富士山	35.3,138.9	12:9:41
雪	36.6,136.6	12:11:57
雪	35.0,135.7	12:30:20
雪	35.7,137.9	12:35:19
雪	34.5,136.7	12:38:40
雪	34.9,135.8	12:48:45

2014年3月10日の検出結果

検出された「雪」の画像群



まとめ

- ジオタグ画像ツイートを用いたイベント検出
 - キーワードの検出
 - 代表画像の選択
 - リアルタイム検出
- 検出結果
 - イベント検出数: 258件
 - 代表画像の適合率: 65.5%

今後の予定

- 検出アルゴリズムの改良
 - 空間規模の推定
 - 期間の推定
 - n-gram
- データの拡張
 - 画像か位置のみのデータの利用
- 代表画像の選択手法の改良