



ソーシャルセンサによるオンデマンド流行検知

岡村, 直人
関, 和広
上原, 邦昭

(Citation)

神戸大学都市安全研究センター研究報告, 17:177-184

(Issue Date)

2013-03

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81011421>

(URL)

<https://hdl.handle.net/20.500.14094/81011421>



ソーシャルセンサによるオンデマンド流行検知

On-demand Epidemic Surveillance Using Social Sensors

岡村 直人¹⁾
Naoto Okamura
関 和広²⁾
Kazuhiro Seki
上原 邦昭³⁾
Kuniaki Uehara

概要: 本論文では, Twitter における位置情報付き発言を利用し, 「特定的话题が盛り上がっている地域の自動検出 (バースト地域検出)」と「特定の地域で盛り上がっている話題の自動検出 (バースト単語検出)」を可視化するシステムを提案する. また, ユーザが適切なキーワードを表現できるとは限らないため, キーワードに意味的に関連する単語をオンデマンドに抽出し, 検索語として提示する. 本システムの有用性を評価するため, 実世界データを用いた実験では, 検出地域, 検出単語について Web ニュースなどの他の情報源と比較・考察する.

キーワード: マイクロブログ, バースト検出, 地域検出, 関連語

1. はじめに

近年, スマートフォンなど, GPS を内蔵したモバイル端末の普及に伴い, 位置情報付きのサービスが増加している. 例えば, foursquare では, 駅前のスターバックスに行くと, その場所に来たことを知らせるチェックインと呼ばれる仕組みによってメッセージを残すことができる. また, 過去にチェックインした人を知ることや, チェックインの回数によってお店からサービスを受ける場合もある. 他にも, スマートフォンで写真を撮影し共有する Instagram と呼ばれるソーシャルネットワークサービスがある. このサービスでは, Twitter のようにフォローした人の写真がタイムラインに現れ, 写真にキャプションとジオタグを付けることで, どこで何を撮影したのかを友人と共有することができる. このようなサービスの登場により, ユーザが興味を持つ発言を位置情報付きで投稿する機会が増加している.

位置情報付きの投稿の中には, 投稿場所に関する内容であったり, その地域で盛り上がっている話題についての内容がある. これらの位置情報付きの投稿を解析すれば, 実世界で起きているイベントを検出することができる. 従来, Twitter ではユーザが任意のキーワードで検索した際に, キーワードを含む発言 (ツイート) が抽出される. 抽出された発言には位置情報が付いていないため, ユーザはそのキーワードがどの地域で盛り上がっているかを知ることはできない. また, ユーザは, 必ずしも自分の検索意図に適切なキーワードを表現できるとは限らない. この問題に対して, ユーザの入力したキーワードに意味的に関連した単語 (関連語) を提示すれば, ユーザの情報要求をより効果的に満たすことができる. なお, Twitter では, 現在 Twitter 上を流れている発言の中から頻繁に話題に上っているキーワードをリアルタイムに抽出して表示するトレンドと呼ばれる機能がある. しかし, この機能でも, 東京・名古屋・大阪などの大都市を除き, 地域毎の話題を検出することはできない.

そこで本研究では, 位置情報付き発言を収集し, ユーザが任意のキーワードで検索した際に, キーワードの関連語を提示し, キーワードを含む発言がどこで盛り上がっているかを検出し, 可視化するシステムを提

案する。また、地域をグリッドで分け、それぞれの地域で盛り上がっているキーワードを検出する。検出地域、検出単語について Web ニュースなどの他の情報源と比較を行い、本研究でのシステムの有用性を評価する。

2. 関連研究

マイクロブログはブログの一種であり、一般的なブログよりも短い文章をウェブ上に公開したり、ユーザー同士でコミュニケーションを行うことができる。米国で 2006 年にサービスが開始された Twitter の流行を皮切りに、Facebook, Google+, Tumbler などが公開され、多くのユーザーに利用されている。このマイクロブログサービスの急速な普及とともにマイクロブログを対象とした研究が盛んに行われている。中でも本研究が対象としている Twitter は、2012 年 3 月現在、世界で 1 億 4000 万人以上¹のユーザーが利用する代表的なマイクロブログサービスである。Twitter では、ユーザーがツイートと呼ばれる 140 字以内の短い文書を投稿したり、他のユーザーが投稿した文書を閲覧したりすることができる。

(1) Twitter を用いたサーベイランスに関する研究

Twitter を用いたサーベイランスに関する研究として、花粉症観測システム¹⁾がある。このシステムの目的は、花粉症の症状を訴える人々の地域毎の分布を天気予報の地図のように作成し、花粉症の広がりを可視化することである。システムは初めに「花粉症」というキーワードを含む発言を取得し、そこから花粉症の症状を分類し、その結果を都道府県毎に集約し可視化する。可視化することにより、どの都道府県に花粉症の症状を含む発言が多いかを容易に把握することが出来る。また、取得した全ての発言に対して位置情報が付加されているわけではないため、ユーザーのプロフィールの情報をを用いて位置情報を取得する。Twitter ではユーザープロフィールの中に「位置情報」という項目があり、多くのユーザーがここに居住地を書いている。

他にも、Twitter を用いた感染症を観測・予測する研究は、主に対象となる感染症をキーワードとして、Twitter 検索を行い、キーワードを含む発言を収集する。例として、感染症の中で毎年世界中で流行している「インフルエンザ」に注目した研究を挙げる。インフルエンザ流行の予測を行う研究²⁾では、アメリカの疾病管理センター (CDC) のデータと Twitter データの両方を用いる自動回帰モデルを提案している。CDC は ILI 症例と呼ばれるインフルエンザの症例を収集しているが、公開までに 1~2 週間の遅延が出る。そのため、インフルエンザ流行の早期検知のタイムリーな方法が必要となる。そこでこの研究では、予測する週の 1~2 週間前の ILI 症例データと最新のインフルエンザを含む発言を自動回帰モデルに適用し、ILI 症例数を予測した。また、retweet と呼ばれるオリジナルの発言を転送したものや、同一ユーザーが ILI に感染してから 1 週間以内に ILI について言及した発言を無視することで予測精度を向上させた。

Culotta^{3, 4)} は、Twitter の発言からインフルエンザの割合とアルコールの販売量を推定を行う研究をしており、インフルエンザに関連するキーワード (flu, cough, headache など) の検索によって収集された発言を用いて、インフルエンザの割合を推定する線形回帰モデルを提案している。キーワードによってはインフルエンザの症状がない発言が多く収集される場合があるため、この研究では ILI 症例データと相関の低いキーワードを誤ったキーワードとして除外して回帰モデルに適用している。さらに、収集データの時系列を逆順にして学習することで、インフルエンザのピークが推定できることを示した。また、drunk (酔っ払っている) というキーワードを含む発言を回帰モデルに適用することで、アルコールの販売量を推定でき、従来の CDC による電話健康調査システムの代わりになることを示した。

これらの研究は、あらかじめキーワードを設定してサーベイランスが行われているものの、ユーザーが任意の話題について検索する際、適切なキーワードを表現できるとは限らない。そこで次節では、キーワードに意味的に関連した単語を提示する研究について説明する。

(2) キーワードの関連語抽出に関する研究

キーワードの関連語抽出は、ユーザーがより適切な検索クエリを考える際に有効である。そこで、ユーザーの入力したキーワードに意味的に関連した単語 (関連語) を提案するシステムがいくつか研究されている。この関連語を提案するためには、ユーザーが入力したキーワードに加えて、さらに他の情報源が必要となる。その情報源としてユーザーからのキーワードを一旦サーチエンジンに入力し、そこから得られた検索結果を用いる手法がある。堀ら⁵⁾は、Web 上で最大の百科事典である点、多数のユーザーが頻繁に更新することで最新の

¹<http://blog.twitter.com/2012/03/twitter-turns-six.html>

内容が維持される点から関連語抽出の情報源として Wikipedia を使った関連語提案システムを提案した。このシステムでは、ユーザに入力させた検索キーワードに関連した文書を Wikipedia 内の、検索キーワードに関連する記事から段落を抜き出すことで収集する。ここから抽出した名詞群に対し、単語間の距離に着目して単語の関連度を算出し、関連度の高い単語から関連語としてユーザに提示している。

3. 提案システム

(1) 位置情報付きの発言の収集

本研究では、地域の話題検出を行うため、位置情報付きの発言を収集する必要がある。そこで、Twitter 社が提供する Streaming API を用いる。Streaming API では、収集範囲の緯度経度を指定することでその範囲に含まれる位置情報付き発言を収集することができる。また、キーワードを指定して収集する必要がないため、位置情報付き発言全体を収集することができ、任意のキーワードでの話題検出を行う本研究に適している。日本全体を収集範囲とした場合、1日に11~14万件の発言が収集できる。

(2) 関連語の推薦

ユーザが入力したキーワードの関連語候補を抽出する情報源として、本研究では Twitter の発言を用いる。これは、キーワード毎に話題の時間変化のパターンが異なっていること、キーワードを含む発言の中には話題に関連する単語が含まれ易いという直感に基づく。例えば、キーワード「地震」とそれに関連する単語である「震源」、「震度」は同時期に複数の発言に現れやすくなると考えられる。すなわち、いつ話題が活発に言及されるかという話題の時間変化を推定できれば、話題に関連する関連語をより正確に推薦できると考えられる⁶⁾。この考えを基に、次の手順で関連語の推薦を行う。

1. 検索キーワードの検索結果（上位5件）の発言から形態素解析を用いて候補単語をピックアップする。
2. 検索結果の発言から、検索キーワードの少なくとも1つと候補単語を含む発言のスコアが高い上位10件を抽出する。
3. 抽出した発言の時間情報と検索スコアをもとに時間モデル（拡張キーワード時間モデル）を作成する。
4. 検索結果の上位10件の検索キーワード時間モデルと拡張キーワード時間モデルとの KL 擬似距離を比較する。
5. KL 擬似距離の小さい拡張キーワード時間モデルを作る候補単語の上位10件を関連語として抽出する。

時間分布のモデル化と時間変化の類似度については、宮西らの論文⁶⁾を参照されたい。

(3) バースト地域検出

本研究では、日本全体を緯度経度共に 1.0° と 0.5° 間隔のグリッドによって各地域に分ける。また、各地域の代表点、人口比を求めるために以下の手順で処理を行う。

1. 発言の緯度・経度を用いて、対象期間で収集された全発言を各地域に分類する。
2. 地域 r における中心点（緯度・経度）を仮の代表点 r_k とする。
3. r に分類された発言 r_i の緯度・経度と r_k との距離を全ての r_i について計算し、平均 ave_{r_i} をとる。
4. r_k から ave_{r_i} 離れた点を r の代表点として更新する。
5. r における発言数を対象期間の日数で割り、1日の平均発言数を計算する。この平均発言数を人口比とみなす。

あるキーワードの検索結果を上記と同様に各地域に分類し、それぞれの発言数を地域の観測数とする。各地域における観測数と日にちを用い、次節で述べる時空間集積性の検定法の考えを基にバースト地域の検出を行う。

a) 空間・時間集積性の検定法

ある地点において継続して観測されたデータから、その発生の集積性（クラスター）を検出するためには、時間集積性の検定法が利用できる。また地域的な集積性も同時に検出するためには空間・時間集積性の検定法が利用できる。Kulldorff^{7,8)}は、解析時点を含んだ「新鮮なクラスター（alive cluster）」を同定する前向き（prospective）の方法を提案している。この解析によって、解析時点において、まさに起こっている突発的な事象の発生が「いつから起きていたのか」を検出することができる。

ところで、一般的な統計的検定の有意性の判定基準としては、 p 値が0.05や0.01などの値を用いることが多い。しかし、日々のサーベイランスにおいて5%の確率で起こるということは、 $1/0.05 = 20$ でほぼ20日に1回の頻度で起こっても不思議がないということになる。このように「〇〇日に1回の頻度よりも稀である」という期間をRecurrence Interval (RI)という。その考えから、毎日行われるサーベイランスにおいては、その有意性の判定基準として、ほぼ1年に1回の頻度よりも稀な現象 (RI=365日) に対応する $p = 1/365 = 0.0027$ がひとつの基準として用いられる。

話題検出を目的とした場合、その発生時点とともに発生地域の同定が重要となる。そこで本研究では、Kulldorffが提案した平面上での疾病集積性の検定である spatial circular scan statistic に時間のデータを導入した cylindrical scan を用いる。

(4) バースト単語検出

従来のバースト解析手法は、計算量がやや大きいことと、ツイートデータのような大規模データに不向きであることなどの欠点がある。そこで本研究では、HeとParker⁹⁾のアイデアを基にバースト単語検出を行う。具体的には、検索対象期間の発言に対して形態素解析を行い、名詞句のみを抽出する。この抽出した単語それぞれに、元の発言から得られた位置情報を付ける。そして、tf-idf法の文書を地域に置き換えることで、閾値を越えた単語をバースト単語候補として決定する。このバースト候補単語それぞれにMACDを用いてバースト分析を行う。MACDとは、EMA・指数平滑移動平均を使用したテクニカル指標である。MACDとそれを移動平均化したシグナルとの2本の移動平均線を用いることにより、相場の周期とタイミングを捉える指標である。時刻 n_1 から n_2 までのMACD値は以下で表される。

$$MACD(n_1, n_2) = EMA(n_1) - EMA(n_2) \quad (1)$$

離散的時系列 $X = \{x_t \mid t = 0, 1, \dots\}$ における変数 $x = x(t)$ のある時刻 n のEMA値を以下の式で示す。

$$EMA(n)[x]_t = \alpha x_t + (1 - \alpha)EMA(n-1)[x]_{t-1} \quad (2)$$

$$= \sum_{k=0}^n \alpha(1 - \alpha)^k x_{t-k} \quad (3)$$

ここで、 $\alpha = 2/(n+1)$ とする。

4. 評価実験

(1) 実験設定

実験データとして、2012年6月12日～2012年12月31日に収集した16,256,140件の位置情報付き発言を用いた。緯度24.84～46.80、経度122.87～153.01の範囲（日本全体）を対象の地域とした。また、実世界データとしてMSN産経ニュースの地方ニュースから地域毎にニュースを収集し、計239件を取得した。

(2) 検出地域に関する実験

検出された地域とMSN産経ニュースを比較し、その関連を評価する。MSN産経ニュースの各ニュースから、ニュースの日付を検索日、そのニュースを表す単語をキーワードと考え、発言（ツイート）を検索する。評価のため、検索された発言を用いて検出されたバースト地域が、ニュースに記載されている地域（MSN産経ニュースでは都道府県名）と一致する場合に成功事例、しない場合に失敗事例とする。ここでニュースに記載されている地域に発言が存在しない、もしくは極端に少ない（1, 2件など）場合は地域の特性が現れない（地域性なし）ニュースと考える。本研究では、このようなニュースはバースト検出地域と比較を行っても判断することができないため、あらかじめ評価用のニュースから除外する必要がある。そこで、239件のニュースそれぞれに対して上記の検索設定で検索を行い、地域の話検出システムのGoogle Mapsへのマーカー表示による可視化により、地域性の有無を判断した。その結果、地域性ありと分類されたニュース45件に対して、バースト地域検出を行った。

この実験では、最大クラスター数を5、最大時間長を5に固定し、2週間分の発言データを対象としているため、有意性評価の p 値は $p = 1/(52/2) = 0.038$ とした。各地域の中心点の緯度経度を代表点とし、グリッ

ド単位を 1.0° と 0.5°, 関連語を含める場合と含めない場合の 4 種類でバースト地域検出を行った。ここでは、グリッド単位を 1.0° にした場合を one-grid (OG), グリッド単位を 0.5° にした場合を half-grid-exword (HG) とした。また、関連語を含める場合をそれぞれ OG-exword (OGEX), HG-exword (HGEX) とした。ニュースの記事に合った関連語がない事例を除き、関連語を含めた検索は 31 事例に対して行う。1 件のニュース事例に対して、日別で複数のバースト地域が検出される場合は、1 つでも地域が一致すれば成功事例とした。表 1 にバースト地域検出の実験結果を示す。

表 1: バースト地域検出の実験結果.

検出設定	成功事例数	成功事例の割合
OG	37	0.822
HG	36	0.8
OGEX	19	0.612
HGEX	22	0.709

この結果より、関連語を含めた OGEX, HGEX よりも関連語を含めない OG, HG の方がニュースの地域を多く検出できていることが分かる。そこで、関連語を含めた場合のみ失敗した例を挙げ、地域を検出できなかった原因を調べた。表 2 に関連語を含めた場合の失敗例を挙げる。

表 2: 関連語を含めた場合の失敗例.

地域	日付	ニュース例	検索キーワード	関連語
青森	7-30	田んぼアートで町おこし	田んぼアート	田んぼ
宮城	10-10	仙台育英高校が東北高校野球優勝	仙台育英	優勝

この例から、一般語が関連語として提示され、関連語を含めて OR 検索をした場合に検索キーワードのみでは出てこなかった別の話題を検出してしまったものと考えられる。

(3) 検出単語に関する実験

本節では、検出された単語と地域が実世界のイベントと関連している例を挙げる。収集された単語に対して提案手法によりバースト単語、地域、日付を検出する。検出された単語に対して、単語+地域をキーワードとし日付によって web による期間検索を行い、検索結果に関連しているイベントがあれば、バースト単語検出が出来た事例として挙げる。表 3 にバースト単語の検出事例を挙げる。

表 3: バースト語の検出例.

検出地域	検出日付	検出単語	検索キーワード
岡山	11-9	注意報	11-8 に岡山県で「食中毒注意報」が発令された
茨城	11-9	震源	11-9 に茨城県北部で震度 4 の地震が発生

これらの例は、実世界のイベントに関連した単語と地域を検出することができたことを示す。また、このようなイベントは発生してから発言が増えるため、バーストする日付は実際にイベントが起こった日付と同日後になると考えられる。しかしながら、他の検出単語はほとんど地名という結果となり、関連している事例はあまり見られなかった。検出候補の単語は文書を地域と置き換えた tf-idf 法により決定しているため、ある地域でのみ出現しその地域での出現頻度が高い地名が検出単語として多く現れてしまう。地名以外の単語を重要語にするには、tf-idf 法を日付に注目させて、ある特定の日に多く出現し、他の日ではあまり見られないような単語を重要語とするなどの方法が考えられる。

(4) 関連語に関する実験

本節では、ユーザが入力したキーワードと抽出された関連語の関連を調べる。表 4 にキーワードと関連語の意味的に関連している例を示す。なお、関連語の上位 10 語は空白で区切り表示する。

表 4: Related words

検索キーワード	関連語上位 10 語
ボージョレ 解禁	ヌーヴォ 今日 本日 皆様 ワイン 今年 フランス お待ち 限定 みたい
不発弾 閉鎖	不発 空港 仙台 予定 全便 決定 ウェイ 発着 国内線 地区
宮城 地震	発生 時刻 松島 震源 速報 津波 心配 付近 震度 高城
立てこもり 信金	事件 豊川 状態 膠着 意図 想像 全員 確保 人質 無事

ボージョレ・ヌーヴォの解禁日で検索した例では、ボージョレに関する「ヌーヴォ」や「ワイン」などの意味的に関連した単語や、「本日」「今年」のように解禁に関連する単語が抽出された。また、仙台空港で不発弾が見つかり閉鎖されたニュースに対しては、閉鎖に関して「国内線」が「全便」欠航したことや、「仙台」空港で起きたことに関連する単語を抽出できた。このように、キーワードに意味的に関連した単語を抽出できることが確認できた。

5. おわりに

本研究では、Twitterにおける位置情報付き発言を利用して、ユーザが入力したキーワード、期間におけるバースト地域検出、バースト単語検出を行った。バースト地域検出では、地域をグリッド単位で分け、発言の位置情報を用いて分類、人口比を1日の発言数を用いて、cylindrical scanに適用することで実現した。バースト単語検出は、tf-idfにより地域特有の重要語を求め、重要語とその地域に対してMACDを計算することで実現した。また、キーワードと単語の時間変化に注目し、時間モデルのKL距離を計算することで、キーワードに意味的に関連した単語の提示を行った。さらに、ユーザが任意のキーワード、期間で検索した発言から、バースト地域・バースト単語を検出し、地図上へマッピングすることで可視化する地域の話検出システムを開発した。

提案手法のバースト地域検出が実世界のイベント検出に有効か調べるために、Webニュースが対象とする地域と検出たバースト地域が一致するか評価を行った。その結果、キーワードのみを含む発言を利用したバースト地域がWebニュース45件に対して、8割程度一致し、関連語を含めた発言を利用する場合でも7割の事例で地域が一致した。関連語を含めて一致数が減ったのは、一般語を関連語とすることで別の話題を検出してしまふことが考えられる。また、検出されたバースト単語・地域・日付をWeb検索に利用することで、実世界でのイベントと関連している事例を紹介し、ある地域でバーストしている単語をユーザに可視化させることができた。これより、本システムを利用することで、ユーザが興味のあるキーワードに対するバースト地域、ある地域でバーストしている単語を地図上へマッピングし、可視化することができる。

今後の課題としては、グリッドのさらなる細分化による検出数の変化、バースト候補単語の選定方法などが挙げられる。バースト候補単語は、本研究ではtf-idf法を用いて選択したものの、出現頻度が低くてもある特定の地域のみにも出現する単語が選択される傾向があるため、改善が必要である。また、本研究で用いたWebニュースだけでなく、別の実世界データを用いてバースト地域検出の正確さを検証していく予定である。

参考文献

- 1) 高橋哲朗, 野田雄也: 実世界のセンサーとしてのTwitterの可能性, 電子情報通信学会技術研究報告. 言語理解とコミュニケーション, Vol. 110, No. 400, pp. 43-48, 2011.
- 2) Harshavardhan, A. and Avinash, G. and Ross, L. and Ssu-Hsin, Y., Benyuan, L.: Predicting flu trends using Twitter data, *Proceedings of the 1st International Workshop on Cyber-Physical Networking Systems (CPNS2011)*, pp. 713-718, 2011.
- 3) Culotta, A.: Lightweight methods to estimate influenza rates and alcohol sales volume from Twitter messages, *Language Resources and Evaluation*, Vol. 47, No. 1, pp. 217-238, 2012.
- 4) Culotta, A.: Towards detecting influenza epidemics by analyzing Twitter messages, *Proceedings of the 1st Workshop on Social Media Analytics*, pp. 115-122, 2010.
- 5) 堀憲太郎, 大石哲也, 峯恒憲, 長谷川隆三, 藤田博, 越村三幸: Wikipediaからの拡張クエリ生成によるweb検索とその評価, 第20回セマンティックウェブとオントロジー研究会, pp. 769-778, 2009.

- 6) 宮西大樹, 関和広, 上原邦昭: マイクロブログ検索のための時間情報と非時間情報を統合したクエリ拡張, 情報処理学会論文誌, Vol. 54, No. 4, 2013.
- 7) Kulldorff, M.: Prospective time periodic geographical disease surveillance using a scan statistic, *Journal of the Royal Statistical Society*, Vol. 164, No. 1, pp. 61–72, 2001.
- 8) Takahashi, K. and Kulldorff, M. and Tango, T., Yih, K.: A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring, *International Journal of Health Geographics*, Vol. 7, No. 14, pp. 1–14, 2008.
- 9) He, D., Parker, D.: Topic dynamics: an alternative model of bursts in streams of topics, *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 443–452, 2010.

筆者: 1) 岡村直人, システム情報学研究科, 学生; 2) 関和広, システム情報学研究科, 講師; 3) 上原邦昭, システム情報学研究科, 教授

On-demand Epidemic Surveillance Using Social Sensors

Naoto Okamura
Kazuhiro Seki
Kuniaki Uehara

Abstract

Twitter, the most widely used micro-blogging service, provides a search functionality to find tweets containing user-specified keywords. However, the retrieved tweets are simply presented as a list, which does not allow the users to grasp the geographical distribution of the topic represented by the keywords. Also, the search is only concerned with keyword match, not considering temporal properties of the topic. These issues are critical for topics which spread over both time and geographical regions. To deal with such spatio-temporal information needs for microblog, this paper focuses on the use of geo-tagged microblog posts (tweets) and develops a system to visualize automatically detected regions in which a specific topic is actively mentioned and automatically detected topics which is actively mentioned in a specific region. The first process to detect such regions is referred to as “burst region detection”, the second process as “burst topic detection”. The burst region detection takes advantage of an approach from spatial epidemiology, specifically, the cylindrical scan technique, and adapts it to geo-tagged tweets retrieved for a given search query. Because the users may not be able to express their intended topic as a query, our system also suggests its related keywords automatically extracted from a Twitter corpus by considering the temporal characteristics of the query. For the burst topic detection, tweets in each region are processed by an existing burst detection algorithm to identify temporally rising topics in the region. The resulting detected regions or topics are displayed on a map using the Google Map API. For qualitative evaluation, the proposed system is tested on a large corpus of geo-tagged tweets collected for six months. Through the experiments, the effectiveness and usefulness of the system are demonstrated by comparing the detected topics and regions with other information sources including web news.