

# 災害発生検知へのビッグデータ複合分析の適用可能性について

小西 一也<sup>1,a)</sup> 坂野 鋭<sup>1</sup> 雨宮 俊一<sup>1</sup>

**概要:** 災害発生を迅速かつ網羅的に検知することは、減災のために非常に重要である。我々は、災害発生検知の実現に向け、種々のセンサデータ等を高速に統合処理する、ビッグデータ複合分析の適用可能性を調査している。調査では、ジャカルタ市で頻繁に発生している洪水による交通支障の発生検知を対象とした。そして、市内で十分に取得できる車両プローブデータとマイクロログに着目し、各データをストリーム処理して、位置情報を用いて統合する方針を立案した。そして、本方針の妥当性を明らかにするために、各データのストリーム処理結果を地図上に重畳表示する実験を行い、洪水による交通支障発生を迅速に目視把握できるか検証した。本稿では、実験結果から本方針の妥当性を明らかにできたことを報告し、ビッグデータ複合分析の技術課題を整理する。

**キーワード:** 防災、複合分析、ビッグデータ

## 1. はじめに

本稿において、我々は既存センサデータの統合による災害発生検知の可能性を調査したので報告する。

近年、大規模な自然災害の被害損失が世界で増加している。2010年代前半における全世界の災害被害額は、30年前のおよそ10倍に当たる8,571億ドルに達している。特にアジアの新興国では、人口増加や経済発展に伴い急速な都市化が進む一方、防災施策が間に合わず、経済的被害が頻繁に発生している。このような状況において、防災はもとより、減災により損失を減らすことが重要な課題となっている。

減災のためには、迅速に災害発生を検知して対策を打つことが重要である。一般に災害発生の検知には、何らかの検知用センサが用いられる。しかし、局所的な被害が同時多発的に発生する大規模災害を想定すると、センサを広域に渡り大量に設置しておく必要があり、コストが非常に高くなってしまふ。そこで我々は、必ずしも災害検知のために設置されたのではない種々の既存センサデータ等を統合処理することにより、災害発生を検知する可能性を検討している。

このように複数種類、大量のセンサデータを用いることにより、必ずしも本来のデータ収集目的ではない情報を明らかにする手法を我々は「ビッグデータ複合分析」と呼んでいる。単一種の既存センサデータからでは十分に災害検知ができなくても、複数種の既存センサデータで相互に情報を補完することで、十分な災害検知が可能になる場合がある。

本稿では、インドネシアのジャカルタ市の大規模洪水時に、物流等に影響を与えている交通支障（通行ができなくなった道路）の発生検知を対象とした。そして、実際の車両プローブデータとマイクロログをストリーム処理して、結果を地図上に表示する実験システムを構築し、洪水による交通支障の発生を迅速に目視把握できるか検証した。

## 2. 関連研究

複数種のデータを統合処理し、災害状況を明らかにする取り組みとして、金ら [1] はビッグデータ複合分析により災害状況をシミュレーションし、防災や減災の計画策定に適用する研究を行っている。金らは、災害シミュレーションデータや交通シミュレーションデータ等を時間軸と空間軸で整合をとってデータベース化し、各データを同一仮想空間上で表現している。この方法により、東日本大震災発生時の詳細な災害状況が一定精度で再現されている。我々が検討する災害発生検知とは異なるが、複数種のデータについて時間軸と空間軸で整合をとる点が共通している。

<sup>1</sup> (株) NTT データ技術開発本部,  
〒135-8671 江東区豊洲 3-3-9 豊洲センタービルアネックス,  
NTT Data Corp. Toyosu, Tokyo, 135-8671

<sup>a)</sup> konishiky@nttdata.co.jp

また、我々は災害発生検知のためのセンサデータの一つとしてマイクロブログに着目している。同様の研究として、榊ら [2] はマイクロブログから地震発生の検知を試みている。榊らは、記述内容が地震発生に関連するマイクロブログを機械学習を用いて選定し、選定したマイクロブログの発信位置から地震発生エリアを推定している。この方法により、高精度に地震を検知できている。ただし、検知対象としては比較的広域に渡る災害事象が想定されている。

マイクロブログ記述内容の局所的な位置推定に関しては、北本ら [3] がオープンに地名辞書を構築し、自然言語文に含まれる地名や住所を自動的に抽出するジオタギング API をフリーで公開している。同綴地名の解決等の課題はあるものの、局所的な降雨・降雪状況などを一定精度で検知できている。しかし、我々が最初のターゲットとしたジャカルタ市に関しては地名辞書が存在せず、ジオタギング API はインドネシア語に非対応である。

### 3. ジャカルタ市の洪水による交通支障の発生検知

我々は、ジャカルタ市の洪水による交通支障の発生検知に取り組んだ [4]。本取り組みは、2014 年の総務省公募案件である「アジア・オセアニア地域における準天頂衛星のメッセージ機能の活用等に関する調査」の一環として実施したものである。

ジャカルタ市は人口 1,000 万人を超える世界屈指のメガシティであり、成長著しいインドネシアにおける経済活動の中心地である。しかし都市機能は発展途上にあり、排水設備が十分でないため、雨季となる 1 月～2 月には毎年大規模な洪水災害が発生している。また、日常から交通渋滞が激しいこともあり、洪水により道路通行ができなくなると特に物流に影響があり、経済的に深刻な被害が発生する。このような背景を踏まえ、本取り組みは、洪水により発生する交通支障を迅速に検知することで、回避ルートによる物流を可能にし、経済への影響を最小限に留めることを目指す。

ジャカルタ市は車両登録台数が乗用車、貨物車、バスを合わせると 300 万台を超える。また、スマートフォンの個人保有率は 23 % に上り、マイクロブログに関しては全世界の発信件数のうちおよそ 2.4 % を占める。これは、都市ごとに比較したとき世界一の発信件数になる。このような実情を踏まえると、ジャカルタ市においては車両やスマートフォンを移動センサとして捉え、車両の位置・速度情報等を持つプローブデータや、スマートフォンから発信されるマイクロブログを分析することで、広域かつ詳細に、災害発生を検知できる可能性がある。例えば、洪水により交通支障が発生すると、発生箇所には車両が進入できなくなり、またその周辺で洪水の発生状況を記述したマイクロブログが発信されることが予想される。

そこで我々は、車両プローブ数の減少から交通支障の発生を推定し、マイクロブログから洪水の発生状況を記述したものを抽出するとともにその位置を推定する方法を検討した。そして、各推定結果を地図上に重畳表示する実験システムを構築した。図 1 に、実験システムの全体像を示す。実験システムの表示を確認することで、地図上で交通支障の発生箇所を迅速に目視把握でき、さらにその周辺で発信されたマイクロブログの内容から原因が洪水であることを判断できるようになることを見込む。

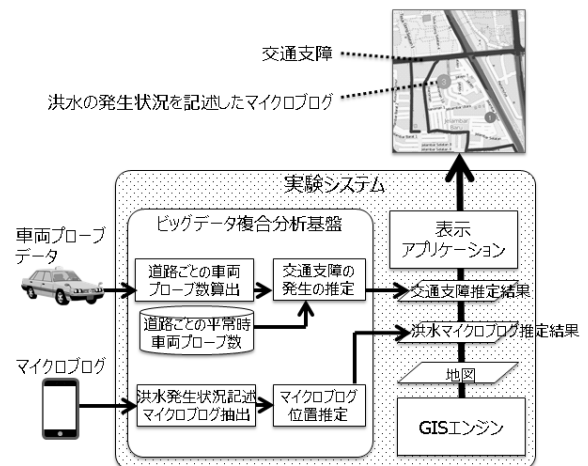


図 1 実験システムの全体像

そして、2014 年に洪水が発生した当時の実際の車両プローブデータとマイクロブログを実験システムで再生し、当時の洪水状況を記録した洪水マップと重ね合わせることで、目視把握により検討方針の妥当性を確認した。なお、洪水マップは最短 6 時間間隔で洪水の発生状況を数百平米程度の集落単位で人手で確認、集計したものである。

以下に、車両プローブデータ処理、およびマイクロブログ処理の概要を示す。また、時々刻々と発生するこれらのデータについて、各処理を遅延なく実行するためのビッグデータ複合分析基盤の概要も示す。

#### 3.1 車両プローブデータ処理

本取り組みでは、インドネシア大手タクシー会社の、ジャカルタ市で運行されているタクシーに搭載されたプローブデータを利用した。タクシー会社は各車両の走行中の位置情報をモバイルネットワーク経由で収集しており、予約電話があった際に顧客に最寄りのタクシーを配車できるよう、車両位置と空車情報を確認している。タクシープローブデータの諸元は以下の通りである。

- カバーエリア：ジャカルタ市
- タクシー台数：約 16,000 台 (実験対象期間中の最大値)
- プローブデータ間隔：30 秒～270 秒
- データ構成：タクシー ID、緯度、経度、スピード、空車・乗車フラグ、他

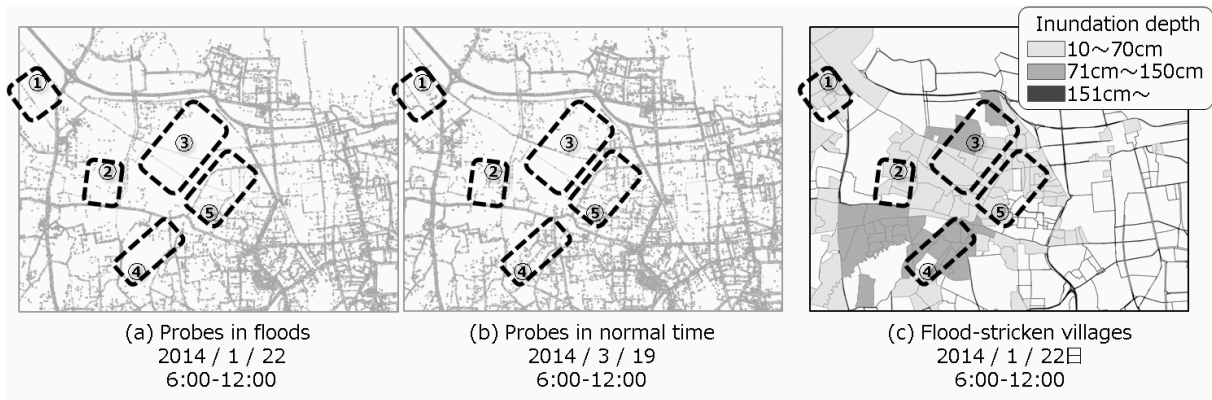


図 2 平常時と洪水発生時のタクシープローブ分布

タクシープローブデータからは、上記のデータ構成に示した情報の他に、道路情報との紐付けにより、どの道路を走行中であるのかといった情報が得られる。

ここで、平常時と洪水発生時の、観測対象道路におけるタクシー通行状況の想定を図 3 に示す。観測対象道路におけるタクシーの通行台数は時々刻々と変化するが、平常時であれば曜日や時間帯に応じて概ね一定となることが予想される。一方、洪水発生時は、道路浸水によりタクシーの流入数が少なくなるなど、この傾向が変化することが想定される。

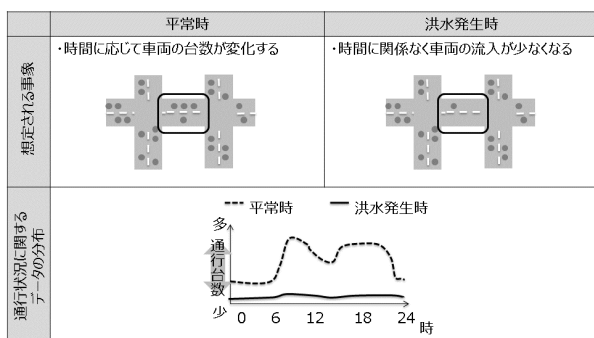


図 3 観測対象道路におけるタクシー通行状況の想定

実際の洪水発生時と平常時におけるタクシープローブの分布を、図 2(a) および (b) に示す。図中の点は、6 時間の間に観測されたタクシープローブの位置をプロットしたものである。破線枠で囲むエリアでは、(b) に示す平常時にはタクシープローブが存在し通行があるのに対し、(a) に示す洪水発生時にはタクシープローブが減少して通行がほとんどなかったことがわかる。また図 2(c) の洪水マップと重ね合わせると、洪水発生時にタクシープローブが減少したエリアの周辺の集落で浸水が確認されていたことが分かった。この結果から、タクシープローブの減少を捉えることで、洪水による交通支障の発生を推定できると考える。

そこで、交通支障検知方法として、道路ごとにタクシープローブ数を平常時の同曜日・同時間帯の平均タクシー

プローブ数と比較し、減少した道路を抽出する方法を検討した。実験システムでは、平常時のデータセットを約 2 ヶ月間のタクシープローブデータから作成した。そして、実際に洪水が発生した 2014 年 1 月中旬から 2 月中旬のタクシープローブデータを再生し、道路ごとに各 1 時間のタクシープローブ数を算出して、平常時との比較で 50 % 未満に減少している道路を抽出した。

洪水発生初期である 2014 年 1 月 18 日の 11 時から 12 時のタクシープローブデータを、上述の方法で処理し、地図上に表示した結果を図 4 に示す。洪水マップと重ね合わせたところ、破線枠で囲むエリアで、交通支障の発生と洪水の発生に対応関係があることを確認した。



図 4 交通支障推定結果と洪水マップの重畳表示

また、破線枠で囲むエリア内の道路におけるタクシープローブ数の時系列変化を図 5 に示す。この結果、洪水発生初期である 2014 年 1 月 17 日から 1 月 20 日において、著しくタクシープローブ数が減少していることも確認できた。

以上の結果から、平常時と比べてタクシープローブ数が著しく減少する道路を抽出することで、洪水による交通支障の発生を検知できる可能性を見出せた。

### 3.2 マイクロログ処理

本取り組みでは、マイクロログとしてツイッターを利用した。

まず、2014 年 1 月の実際に洪水が発生した日に、ジャカルタ市の BARAT 地区周辺の約 10km 四方のエリアでジオタグ付きで発信された、インドネシア語で洪水を意味する

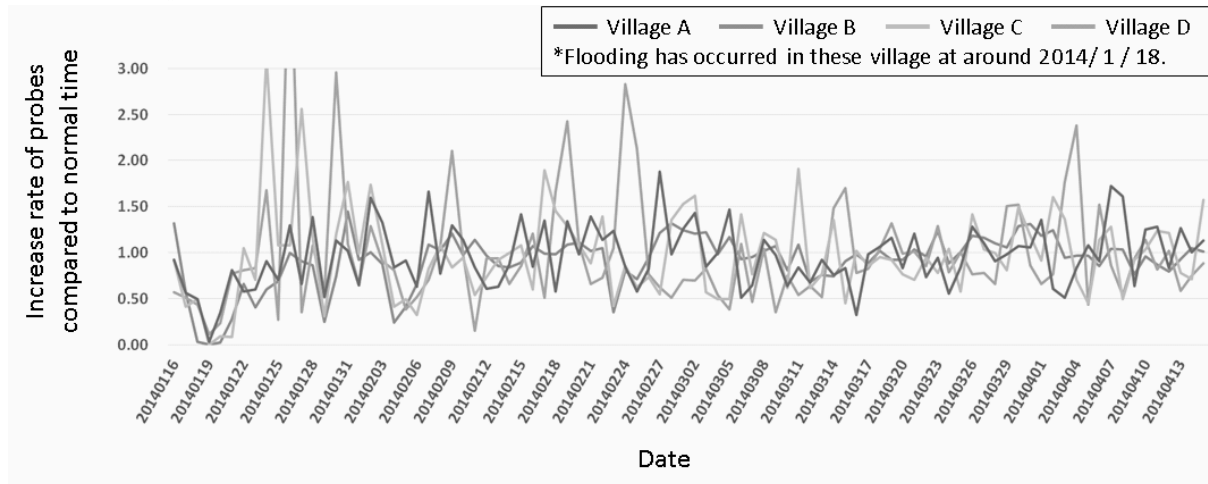


図 5 交通支障発生集落のタクシープローブ数の時系列変化

キーワード「banjir」が出現するツイッターの発信件数を調査した。1時間ごとのツイッター発信件数の変化を図6に示す。図示の通り、洪水発生の初期にツイッター発信件数が増加していたことがわかった。この結果から、洪水の状況を記述しているツイッター（以降では洪水状況ツイッターと呼ぶ）を抽出し、記述内容の位置を推定することで、洪水発生を検知できる可能性がある。

### 3.2.1 洪水状況ツイッターの抽出

洪水を意味するキーワード「banjir」が出現しても、洪水の状況を記述していないツイッターも多く存在する。例えば、一般的な洪水事象についての内容を記述しているツイッターや、比喩表現として「banjir」を用いて洪水と無関係の内容を記述しているツイッターがある。これらと洪水状況ツイッターの識別については、識別性能の優れた機械学習モデルであるSVM(Support Vector Machine)を採用することとし、具体的にはLIBSVM[5]を利用した。学習データは、「banjir」が出現する2,000件のインドネシア語ツイッターに対し、人手で洪水状況ツイッターか否かを判定して作成した。なお、現状ではSVMのチューニングに至っていないため、3-交差検定で適合率0.50、再現率0.53という結果になっている。

この識別の他にノイズフィルタリング処理を実装し、2014年1月の実際に洪水が発生した日にBARAT地区周辺エリアでジオタグ付きで発信された、「banjir」が出現するツイッター約121万件に適用したところ、約39万件の洪水状況ツイッターを抽出した。抽出結果の一部を表1に示す。抽出結果にはノイズも混在していたものの、洪水の状況を記述しているツイッターを抽出できていることが確認できた。

### 3.2.2 ツイッター記述内容からの位置推定

ツイッターの記述内容の位置推定には、洪水状況ツイッターがランドマーク名を用いて記述される場合があることに着目し、著名ロケーションSNSにユーザが自由に登録した膨大なランドマークリスト(ランドマーク名, カテゴリー,

表 1 洪水状況ツイッターの例

洪水状況ツイッター	意味
Selama banjir kami sewa getek dan perahu utk pulang pergi ke kantor. Di jalanan Kapuk Raya org ramai melewati air selutut. Ojek getek panen.	洪水の間、私たちは通勤のために(小さい)船とかだをレンタルしていた。Kapuk Raya 通りでは水がひざまでだった。「いかだタクシー」が大儲けだ。
@TMCPoldaMetro jl.raya kapuk indah terendam banjir 50-80cm	@TMCPoldaMetro 宛 Kapuk Indah 大通りが 50-80 cm の洪水になっている
Coba peruntungan menjar- ing ikan dikala sungai banjir di Jembatan Gantung, Daan Mogot	Daan Mogot、Jembatan Gantung にて、洪水に魚釣りをする。(あわよくば)
Condet-Daan Mogot ± 1,5 jam. Banjir dimana mana	Condet から Daan Mogot まで 1.5 時間。どこも洪水だ
Yang banjir mana suaranya?! *GayaVokalis	洪水になっているみなさまは??? (バンドのボーカル風で)
@ikarays yahhh iyalah ilang kan banjir , fisheyenya dibutuhkan sama ikan	洪水なんだからなくしちゃったよ。フィッシュアイは魚が必要としている

り、住所、緯度経度等の情報)を利用した。そして、洪水状況ツイッターに一致するランドマーク名が出現する場合に、そのランドマークの位置をツイッター記述内容の位置として推定した。BARAT 周辺エリアには 10 万件程度のランドマーク名が存在したが、少数のユーザにしか用いられないランドマーク名も多かったため、実験ではチェックインユーザ数の多い 1 万件程度を限定して使用した。また、チェーン店名など同一ランドマーク名が複数存在する場合は、洪水公共ツイッターに道路名も合わせて用いられる場合があったため、道路名も合わせて照合し、位置を一意に推定した。

なお、現状ではインドネシア語の自然言語処理を適用できてなく、ツイッター記述内容においてランドマーク名ではない部分であってもリストと照合してしまっており、目

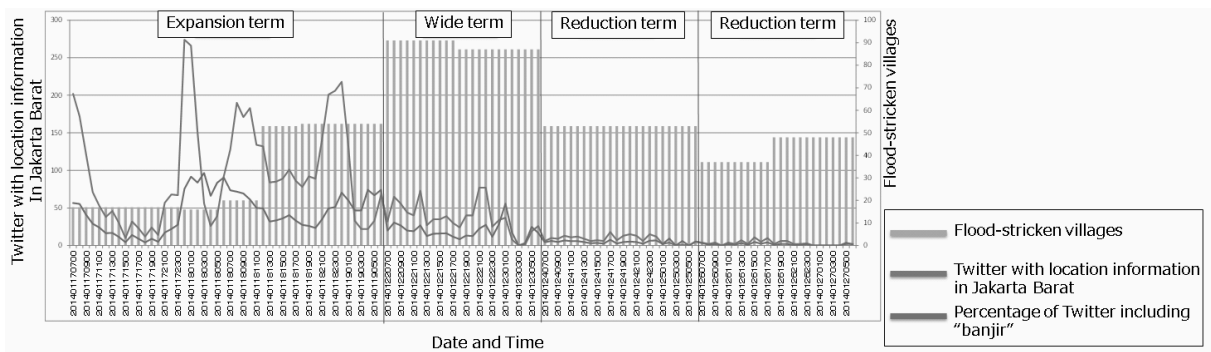


図 6 洪水発生時のツイッター発信件数の時系列変化

視でノイズが多く確認できる状況である。

このツイッター記述内容の位置推定方法を、洪水状況ツイッターとして抽出した約 39 万件に適用したところ、約 36,000 件について記述内容の位置を推定できた。推定結果の一部を図 7 に示す。洪水が発生した集落内に位置推定された洪水状況ツイッターの内容を確認したところ、適切に位置推定できた場合には、洪水発生を裏付けられることを確認できた。

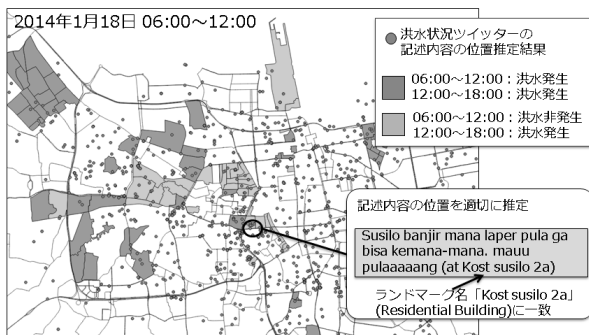


図 7 ツイッター記述内容の位置推定結果の例

### 3.3 ビッグデータ複合分析基盤

洪水による交通支障の発生箇所を迅速に把握可能にするためには、時々刻々と発生するタクシープローブデータとツイッターを、遅延なく処理し続ける必要がある。我々は、ストリームデータのリアルタイム分析基盤である CEP(Complex Event Processor) を用いたビッグデータ複合分析基盤を検討しており、これを用いて各処理を実行した。ビッグデータ複合処理基盤は、複数種のデータの時間軸と空間軸の整合をとる統一時間軸・空間軸マッピング処理、分析に適切な時間単位・空間単位にデータをまとめる時間・空間ウィンドウ処理、そしてまとめたデータを分析するデータ分析処理を高速に実行する基盤である。

タクシープローブ処理では、道路ごとに各 1 時間のタクシープローブ数を平常時と比較することにより、その時間における交通支障の発生を検知した。これにより、1 時間ごとに交通支障の発生を検出できる可能性がある。リアル

タイム性をより高めるためには観察時間を縮め、検知間隔を短くすることが考えられる。ビッグデータ複合分析基盤により 10 分間隔の処理も問題なく実行できることから、実際に 10 分間隔での実験も行ったが、タクシープローブ数の変動が大きくなってしまい、時間帯の通行状況を適切に表すことができなかった。現状では、検知間隔を 1 時間とすることにより、交通支障の発生を適切に把握できるものと考えている。

一方ツイッター処理については、20 分間隔で局所エリアごとの洪水状況ツイッター発信件数を算出して地図上に表示したところ、発信件数が滑らかに増減することを目視確認した。この結果から、現状ではツイッター処理は 20 分間隔程度が適切であると考えられる。

これらの結果は、洪水発生把握の時間間隔について、従来は人手で最短 6 時間だったものを、格段に短縮できることを示している。今後は、具体的なアプリケーションと、その要件に応じた時間間隔の短縮を検討していく。

## 4. 目視評価

洪水発生時のタクシープローブデータおよびツイッターを再生し、実験システムで収集、処理、結果表示して洪水マップと照合することで、交通支障の発生や洪水状況ツイッターが表示される場所と、実際に洪水が発生した集落の関係を目視確認した。

2014 年 1 月 18 日の 11 時から 12 時の各データの処理結果と洪水マップとの照合結果を図 8 に示す。図示の通り、洪水が発生した集落において交通支障の発生を検知でき、さらにその周辺に位置推定された洪水状況ツイッターの内容を確認することで、その位置で発生している洪水事象を大まかに捉えることができた。

## 5. ビッグデータ複合分析の技術課題

実データを用いた実験により、ビッグデータ複合分析による災害発生検知の可能性を見出すことができた。実際の検知に向けては、検知する災害や使用するデータの特性に応じて、それぞれに適切な処理を実行できるビッグデータ



図 8 各データ処理結果と洪水マップの照合結果の例

複合分析基盤を実現していく必要がある。今後解決に取り組む技術課題を、以下に整理する。

- (1) 統一時間軸・空間軸マッピング処理： 複数種のセンサデータを統合する際、多様な時間軸・空間軸に柔軟にマッピングし、各軸での統合を試行できる必要がある。ジャカルタ市の洪水による交通支障の発生検知では、車両プローブデータとマイクロログをマッピングする空間軸として、集落や道路が考えられる。それぞれにマッピングして統合を試行することで、適切な時間軸・空間軸を明らかにできる。
- (2) 時間・空間ウィンドウ処理： 各センサデータの処理を、任意の時間単位・空間単位に柔軟にまとめ、処理を試行できる必要がある。ジャカルタ市の洪水による交通支障の発生検知では、試行により車両プローブデータは1時間間隔、マイクロログは20分間隔の時間単位を明らかにした。データを補完してから処理する場合、統合してから処理する場合など、それぞれで任意の単位にまとめあげて処理を試行することで、適切な時間単位・空間単位を明らかにできる。
- (3) データ分析方法： 各センサデータを分析する際、典型的な分析手法は選択的に実行できるべきである。ジャカルタ市の洪水による交通支障の発生検知では、車両プローブ数について平常時との乖離の算出を行ったが、このような処理は他のセンサデータについても必要なことが想定されるため、選択的に実行できる環境を用意しておくべきである。

また、ジャカルタ市の洪水による交通支障の発生検知に関して、車両プローブデータ処理およびマイクロログ処理の高度化に向けての主な課題を、以下に整理する。

- (1) 車両プローブデータ処理における道路特性の考慮： 本取り組みでは、平常時との乖離を評価する閾値を、すべての道路で同一としたが、実際には高速道路や一般道路といった種別や道路長、平常時の混雑度などの特性があり、それに応じて閾値を設定することで、より適切に交通支障の発生を検知できるものと考えられる。
- (2) マイクロログ処理におけるインドネシア語解析の高度化： 本取り組みでは、洪水状況ツイッターの選定

にSVMを用いたが、学習データはわずか2,000件であり、チューニングもできていない。高精度化のために、学習データを増やしての再学習や精度向上のためのチューニングが必要である。また、記述内容の位置推定は、本文のすべてをランドマークリストと照合したが、ノイズを減らすために、構文解析により地名となる部分のみと照合するように修正する必要がある。

## 6. おわりに

ビッグデータ複合分析による、局所的な災害発生を迅速かつ網羅的に検知することを目標に、ジャカルタ市の洪水による交通支障の発生検知について、実データを用いて実験を行ったところ、災害発生検知の可能性を見出すことができた。

今後は、ビッグデータ複合分析を実現するための統一時間軸・空間軸マッピング処理等の数理モデル、またジャカルタ市の洪水による交通支障の発生検知について、高精度化するための各データ処理の数理モデルを構築していく。

## 参考文献

- [1] 金進栄, 花房比佐友, 桑原雅夫, 他: 災害時と平常時の交通マネジメントのためのデータ融合と解析, 生産研究 65(2), 195-200 (2013)
- [2] T. Sakaki, M. Okazaki, Y. Matsuo: *Earthquake shakes Twitter users: real-time event detection by social sensors*, Proceedings of the 19th international conference on World wide web, 851-860 (2010).
- [3] 北本朝展, 相良毅, 有川正俊: *GeoNLP: 自然言語文を対象とした高度なジオタキシングに向けて*, CSIS Days 2011, No.D10 (2011)
- [4] 小西一也, 田代裕和: *ビッグデータ分析の社会実験事例, Japio YEAR BOOK 2015 - ビッグデータの活用 -*, 44-51 (2015)
- [5] LIBSVM - A Library for Support Vector Machines, <<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>>