

位置情報ビッグデータによるデモ参加者の人流解析

○水野貴之（国立情報学研究所） 大西立顕 渡辺努（東京大学）

概要 欧州では、難民・移民等に対して排他的な社会に傾こうとしている。その傾向をとらえ、住民との交流を促進するためには、どのような社会システムが必要であるかを調査するために、スマートフォンの位置情報ビッグデータと SNS における位置情報付きの投稿を利活用することを提案する。特に、本論文では、特定の政治的な思想に共感する人々が集まるデモに注目し、DBSCAN を応用して、デモの時空間における変化が定量化できることを示す。また、デモ参加者の人流と SNS におけるフォローアカウントから、デモ後の報道事実と矛盾なく、デモ参加者の居住地域や応援政党の偏りを検出できることを示す。

キーワード: 位置情報ビッグデータ、クラスタリング、属性推定、社会調査、デモ活動

1 はじめに

近年、ヨーロッパ諸国では、民族独立、テロ反対、難民の受入反対に対する大規模デモが頻発している。背景は、地域格差の拡大、同時多発テロの発生、受け入れた難民による犯罪の発生による国民の不安の高まりである。ヨーロッパは、マイノリティに対して排他的なセグリゲートされた社会に傾こうとしている。高い解像度で時間的空間的なセグレーションの変化を検出することは、エビデンスにもとづいた文化や安全に関する政策作成するために重要である。

本研究では、政治的な主張を共有する人々が一同に集まるデモに注目する。スマートフォンのアプリケーションを提供するいくつかの業者は、同意のもとで、そのアプリケーションを利用するユーザーの位置情報の履歴を収集している。我々は、2015年以降のパリ、ブリュッセル、バルセロナ、ミュンヘンにおいて、ソフトバンク及びその提携企業によって収集された約4億ログの位置情報[1]と、位置情報が付加された約1250万ツイートを利用して、デモの規模や挙動、デモ参加者の地理的な属性や彼らのSNSにおける関心などの社会調査が可能であることを示す。

本論文では、本研究で調査をおこなった様々な政治的なデモの中から、デモ後の多くの報道により、分析結果の妥当性を容易に検証することが可能な、2015年1月11日にパリで起きた反テロデモを例にとりあげ、本提案手法を紹介する。

この論文の章立ては次のように構成される。2章では、DBSCANを利用して普段以上に人の集まるエリアを検出し、デモの時空間における変化を定量化する手法を提案する。3章では、2章で特定したデモの時空間クラスタに所属した人々のデモ前後の移動や、SNSにおける政党アカウントのフォロー状況から、デモ参加者の居住地域や応援する政党に関する情報を得ることが可能なことを、デモについて報道された事実と照らし合わせて示す。4章はまとめである。

2 デモの時空間における変化の定量化

我々は、DBSCANを利用して人の集まるエリアを検出する手法[2]を応用することによって、デモの時空間における変化を定量化する。

2015年1月11日に、パリのレピュブリック広場を中心として発生したフランス史上最大規模の反テロデモ

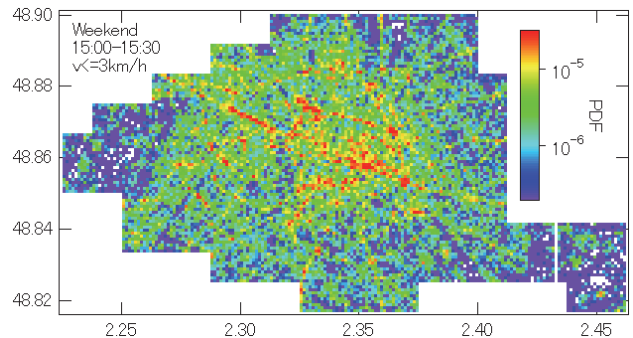


Fig.1 パリの週末における流動人口の時空間確率密度分布. 本図は確率密度分布の15:00-15:30の断面. 縦軸と横軸は緯度と経度.

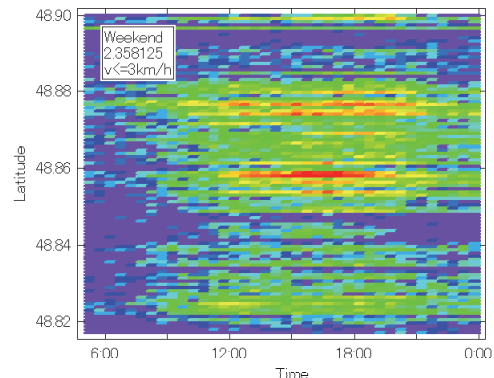


Fig.2 パリの週末における流動人口の時空間確率密度分布. 本図は確率密度分布の緯度×時間の空間の断面. 経度を東経2.358125度に固定している. 図中の色が示す確率密度は、Fig.1の基準と同じ.

[3, 4]を例に、本提案手法の詳細を説明する。本提案手法は(1)時空間の周期の抽出、(2)時空間における異常値の検出、(3)時空間クラスタリング、(4)要約統計量によるクラスタ変化の観測の4つのステップで構成される。

2.1 時空間の周期の抽出

駅など普段から人が多く集まっている場所に、普段と変わらず多くの人があいても、それはデモなどの人が集まるイベントが起きているわけではない。そこで、

まず、パリの与えられた全ての平日または休日をアグリゲイトして、平均的な日中の人数の、時間と空間の3次元空間における確率密度分布を算出する。ここで、デモ等のイベントにおいて参加者は、その開催場所で滞留しているため、時速3km/h以下の人限定して確率密度分布を算出する。Fig. 1は、週末における、流動人口の確率密度分布の15:00-15:30の断面を表す。シャンゼリゼ通り周辺に多くの人がいることが分かる。同じく、Fig. 2は、経度をレピュブリック広場付近の座標(東経2.358125度)に固定して、時間と緯度の2次元断面で確率密度分布を見たものもある。どの時間帯においてもお昼に人が多くいることがわかる。この平均的な時空間確率密度を基準として、各日の時空間確率密度の異常値(外れ値)を検出することで、デモの期間や範囲を特定する。

2.2 時空間における異常値の検出

各時空間座標に訪れる人数が、Fig. 1やFig. 2で見てきた普段の流動人口の確率密度分布に従ってポアソンの決まる状況をNullモデルとして仮定する。そして、与えられた日、ここでは、2015年1月11日における時空間座標ごとに、訪れた人数がポアソン分布における 3σ を超えた場合を異常値(人が普段に比べて非常に多くいた)として定義する。

Fig. 3は、大規模な反テロデモが発生している2015年1月11日の15:00から15:30までの異常値が現れた座標を表す。また、Fig. 4は、Fig. 2と同じ東経2.358125度で観測した異常値の緯度×時間の空間の断面である。パリのレピュブリック広場を中心に、異常値が多数発生していることが分かる。このデモが開催された午後では、多数の異常値が観測される。このことから、時間的にも、空間的にも固まっている異常値の時空間座標をクラスタリングすることにより、デモの時空間における変化をとらえることができる。

2.3 時空間クラスタリング

異常値の時空間の散布図に対してDBSCANによるクラスタリングをおこなう。ここでは、時空間上での1単位の距離を、時間軸では30分、空間軸では100mに設定した。また、DBSCANにおけるクラスタに座標を加える条件を、時空間上で距離3.6までの範囲に20メッシュ(1メッシュは100m×100m×30分の解像度)以上の異常値が発生しているとした。この条件下で、2015年1月11日の異常値をクラスタリングすると、2つのクラスタが抽出できる。Fig. 5は、2015年1月11日の15:00から15:30におけるクラスタリングの結果である。レピュブリック広場周辺に大きなクラスタと、その北西部に小さなクラスタが観測される。この大きなクラスタの半径は約1.2kmであった。

2.4 要約統計量によるクラスタ変化の観測

デモの時空間における変化を定量化するために、各時刻におけるクラスタの半径と中心座標に注目する。

2015年1月11日に発生した反テロデモには、フランスのオランド大統領の呼びかけで、イギリスのキャメロン首相、ドイツのメルケル首相、スペインのラホイ首相、イスラエルのネタニヤフ首相、そしてパレスチナ

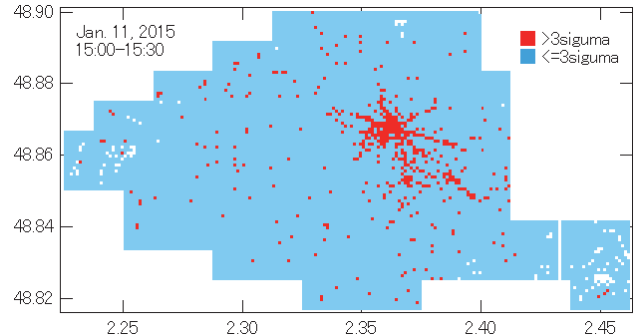


Fig. 3 パリにおいて2015年1月11日の15:00から15:30の間に 3σ を超える異常値が発生した座標。

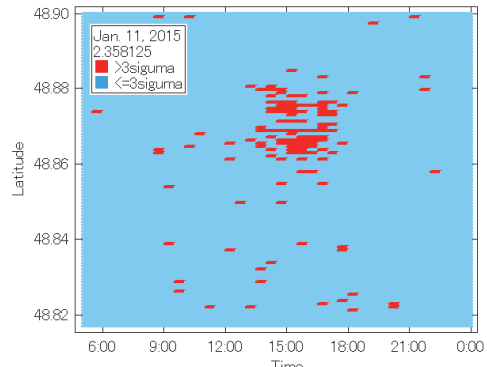


Fig. 4 パリにおいて2015年1月11日に 3σ を超える異常値が発生した座標。本図は緯度×時間の空間の断面。経度を東経2.358125度に固定している。

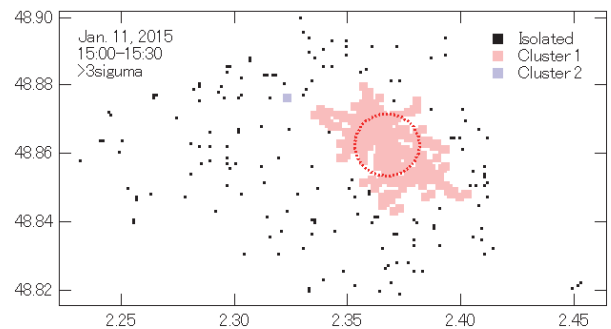


Fig. 5 パリで2015年1月11日に異常値が発生した時空間座標のクラスタにおける15:00-15:30の断面。黒はクラスタに属していない異常値の座標。赤と青は、それぞれ、クラスタに属する異常値の座標を表す。円は、赤のクラスタの半径を示す。

暫定自治政府のアッバス議長など、世界各国の指導者も参加し、15時からはレピュブリック広場から南東方向に腕を繋いで行進をおこなった。デモは日が落ちて後も続き、行進参加者は160万人以上にのぼった[3, 4]。

Fig. 6は、Fig. 5で観測された赤のプロットのクラスタ1と青のプロットのクラスタ2の中心座標の軌跡を表している。また、Fig. 7は各時刻における各クラスタの平均半径を表す。デモが発生したレピュブリック広場周辺で観測されたクラスタ1は、午前9時半ごろに半径100m程度の小さな規模で発生した、その後、成長し、その半径は12時から20時かけて1kmを越えた。20時以降は急激にクラスタの分解が進み、21時半にはクラスタ

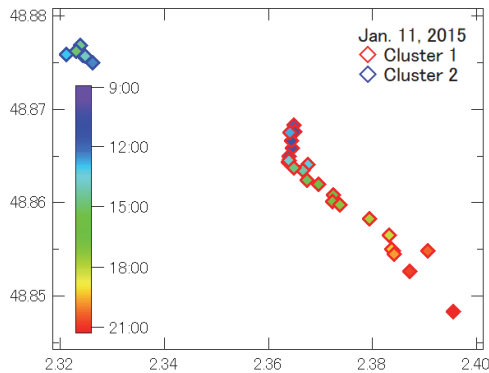


Fig. 6 パリにおける2015年1月11日のクラスタの中心座標の時間変化. 赤はクラスタ1を, 青はクラスタ2を表す. プロットの中の色の変化は時刻を表す.

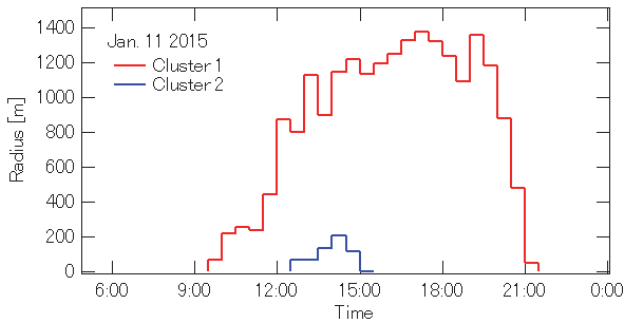


Fig. 7 パリにおける2015年1月11日のクラスタ半径の時系列. 赤はクラスタ1を, 青はクラスタ2を表す.

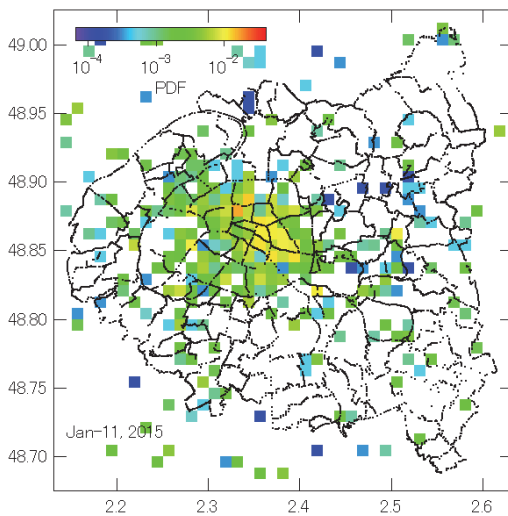


Fig. 8 デモを表す時空間クラスタ1に属した人々の自宅メッシュの空間的確率密度分布. 黒点はパリ, 及びパリに隣接するセーヌ＝サン＝ドニ, オー＝ド＝セーヌ及び, ヴァル＝ド＝マルヌの3県内のコミューンと呼ばれる行政区画を表している.

が消滅した. クラスタ1はサイズが変化するだけでなく, 移動もしている. Fig. 6が示すように, クラスタ1の中心は15時頃までは, それほど動かないが, 15時から南東方向へ移動した. このクラスタ1が示す動きは, まさに, 2015年1月11日にパリで発生した反テロデモについて報道された内容と同じである. つまり,

本提案手法により, デモの時空間における変化の定量化ができていけると言える.

2015年1月11日には, 反テロデモを表すクラスタ1以外に, もう一つ, 半径100m程度のクラスタ2が12時半から14時半に検出された. このクラスタ2の大半を, パリ近郊列車が発着するサン・ラザール駅が占めている. 従って, このクラスタ2は, パリ近郊から反テロデモに参加するために訪れた人々によって作られたと想像することができる. このクラスタ2発生の理由を明らかにすることは, 今後の課題である.

3 デモ参加者の地理的な特徴と政治的な関心

政治的なデモは, そのデモの主題となる思想に共感する人々が一斉に集まる. つまり, 参加者のデモ後, 及びデモ以前の移動を調べることで, そのような思想を持つ人々が多く住む地域, または, 多く働く地域, さらには逆に, そのような思想に共感しない人々が住む地域や働く地域を特定することができる. これは, 欧州におけるセグリゲーションを観測するうえで, 重要な情報である.

デモに参加した人々が, デモの主題となる思想に共感していること以外の特徴を, ここではTwitterにおける各政党へのフォロー情報から調査する. あくまで, Twitterユーザーに限定されてしまうが, デモ参加者のサイバー空間における政治的な関心を捉えることができる.

2015年1月11日の反テロデモにおいて, フランスのオランド大統領は, "every citizen can come...it is not controlled."と言いました. しかしながら, パリ郊外の西部の住民がデモに参加していないことが, 新聞紙Le Mondeで報じられている[5]. また, 大統領は世界各国の指導者に参加を呼びかけましたが, フランスの政党の1つFront NationalのMarine Le Pen党首には声をかけませんでした[6]. これらの事実は, デモ参加者の居住地域の推定と, デモに参加したTwitterユーザーによるFront Nationalが持つ公式Twitterアカウントへのフォローの低さから確認することができる.

3.1 デモ参加者の居住地域の特徴

空間を500m×500mメッシュに分割し, 人々の各メッシュにおける滞在時刻を観測すると, 各人は夜10時から朝8時まで0.7以上の高い確率で, 同じメッシュに滞在していることが分かる. そのメッシュを各人の自宅メッシュと定義する.

Fig. 8は, デモを表す時空間クラスタ1に属した人々の自宅メッシュの空間的確率密度分布である. 反テロデモがおこなわれた地域を中心にパリ市内の様々な地域の人々が参加していることが分かる. 一方で, パリ郊外に目を向けてみると, 北東部と南東部の住民の参加が少ないことが分かる.

我々は, デモを表す時空間クラスタ1に属した人々の自宅メッシュの空間的確率密度 $p(s)$ から, エントロピー $H_{demo} = -\sum_s p(s) \log p(s)$ を算出し, そして, このデモ参加者のエントロピー H_{demo} と, 2015年1月11日以外の休日に, この時空間クラスタ1と同じ座標に訪れた人々の自宅メッシュ分布から算出されるエントロピー H_{ord} を比較する. エントロピー H が高いほど, 様々な地域から一様に住民が集まっていることを意味

Table 1 デモを表すクラスタ 1 に訪れた人々の自宅メッシュ分布から得られる情報量エントロピー. エントロピーは, 東経 2.4 度以上のパリ郊外の東部と, 東経 2.3 度以下のパリ郊外の西部とで, それぞれ求めた.

地域	2015/1/11	2015/1/11 以外の休日
東部	5.97	6.70
西部	6.05	6.22

Table 2 デモを表すクラスタ 1 で Tweet をした人々がフォローしている政党の公式 Twitter アカウントの構成比

政党名	2015/1/11	2015/1/11 以外の休日
Les Republicains	0.24	0.28
Parti Socialiste	0.30	0.30
Front National	0.06	0.11
En Marche	0.40	0.31
	1.00	1.00

する. Table 1 は, パリ郊外の東部と西部のそれぞれで, エントロピーを比較した結果である. 西部では, $H_{demo} \approx H_{ord}$ であり, 普段と同じく, 様々なメッシュに住む人々がレピュブリック広場に集まっていることがわかる. 一方で, 東部では $E_{demo} < E_{ord}$ となり, 特定のメッシュに住む人々のデモへの参加が多い, または, 特定のメッシュに住む人々のデモへの参加が少なくなっていることを意味する. つまり, 普段よりも, レピュブリック広場に来る地域住民の構成が西部では偏っている. この解析結果は, 新聞紙 Le Monde で報じられた, 2015 年 1 月 11 日の反テロデモにパリ郊外の西部の一部地域の住民が参加していないという事実と矛盾しない.

3.2 デモ参加者が Twitter でフォローする政党

Twitter において, どの Twitter アカウントをフォローするかで, その Twitter ユーザーの属性推定が可能である [7]. ここでは, Tweet 時に位置情報を付加した Geo-tagged tweet を用いて, 2015 年 1 月 11 日のパリでの反テロデモを表す時空間クラスタ 1 の内部で Tweet をした Twitter ユーザーに着目する. それらのユーザーは, しばしば, フランスにある各政党の公式 Twitter アカウントをフォローしている. その構成比を表したのが Table 2 である. ここでは, 主要 4 政党: Les Republicains, Parti Socialiste, Front National, En Marche の公式 Twitter アカウントのみに着目した. デモが起きた日にクラスタ 1 の場所で, 位置情報を付けて Tweet した人のフォロー政党は, En Marche の 0.4 が最も高く, Front National の 0.06 が最も低い. これは, もともと Front National もフォロワー数が少ないことも影響している. 我々は同じく, デモ以外の休日においてクラスタ 1 と同じ場所で, 位置情報を付けて Tweet した人のフォロー政党を調査する. Table 2 が示すように, デモが発生した 2015 年 1 月 11 日以外でも, 構成比が Front National は 0.11 と低い. しかしながら, 構成比が, デモの日に最も下降したのは Front National である, この結果は, Front National の Marine Le Pen 党首に声がかからなかった事実や, 反 EU や移民排斥を掲げる Front National と連帯を求めるデモの主張とは相容れない点

が多いことから矛盾がない.

4 まとめと今後の展望

本論文では, 2015 年 1 月 11 日にパリで起きた反テロデモを例にとりあげて, 大量のスマートフォンの位置情報から, 普段に比べて人が多く集まる異常値を検出し, その異常値を DBSCAN によりクラスタリングすることで, デモの時空間における変化を定量化できることを示した. 夜 10 時から朝 8 時までに高い確率で滞在する 500m×500m の範囲を, その人の自宅メッシュと定義することにより, デモを表すクラスタの場所に, デモの日にいた人々と, デモを除く休日にいた人々とで, 自宅メッシュの空間分布の違いを情報量エントロピーで比較した. パリ郊外の西部では, 顕著な違いは見いだせない. 一方で, パリ郊外の東部では, デモ参加者の地域に偏りが発生していることが見いだせた. この結果は, デモ後の報道と矛盾がない. また, 我々は, デモの日とデモを除く休日のそれぞれで, 2015 年 1 月 11 日のデモの場所を表すクラスタ内で, 位置情報付きの投稿を Twitter に送った人々について, 彼らがフォローする政党アカウントを調査した. Front National をフォローする人々の割合が, デモ参加者では顕著に低いことが分かった. この結果も, デモ前後の報道事実と矛盾がない.

このように, 位置情報ビッグデータを使ってデモ参加者に対する社会調査が可能である. 2015 年 1 月 11 日のパリのデモは話題性が高く, 多くの関連する報道がなされたが, 数多起る通常のデモでは, 発生したという事実程度しか報道されない. しかしながら, 回数を重ねるに連れて巨大化するデモもあり, 早期に調査が必要な場合もある. 広い視点で客観的に, ビッグデータを利用してデモをナウキャストすることは, 社会の多様性を維持するうえで重要なシステムになり得る.

本会議では, これらの研究結果に加えて, 人々の出発地と到着地をエッジで繋いだネットワークから, 2nd order MapEquation [5] を用いて人々の生活コミュニティを検出する. パリ郊外の東部に位置するテロ直後のデモへの参加者が少ないコミュニティの住民は, 他のコミュニティへの移動が特に少ないことを示すことができる. また, これらのコミュニティの時間変化をとらえるカップリングネットワークについても提案する.

謝辞

本研究は, JST さきがけネットワーク「人流ビッグデータによる地球規模の課題解決のための情報学と社会科学の融合基盤構築」, JSPS 科研費 17H05123, 17KT0034, 16H02872 の助成を受けたものです.

参考文献

- 1) https://www.agoop.co.jp/en/news/detail/20170629_01.html (Accessed on Feb. 3, 2018).
- 2) J. Capdevila, G. Pericacho, J. Torres, and J. Cerquides: ICA3PP, pp. 356-373 (2016).
- 3) Paris attacks: Millions rally for unity in France. BBC News., 1/11 (2015).

<http://www.bbc.com/news/world-europe-30765824> (Accessed on 2017/2/15)

- 4) Massive crowds join march for solidarity in Paris. The Washington Post, 1/11 (2015).
https://www.washingtonpost.com/world/hundreds-of-thousands-are-expected-to-march-in-paris-solidarity-rally/2015/01/11/3befc7e2-996e-11e4-927a-4fa2638cd1b0_story.html (Accessed on 2017/2/15)
- 5) Le Monde, 1/15 (2015)
- 6) Marine Le Pen dénonce son "exclusion" de la "marche républicaine". Ledauphine.com, 1/8 (2015).
<http://www.ledauphine.com/france-monde/2015/01/08/marine-le-pen-denonce-son-exclusion-de-la-marche-republicaine> (Accessed on 2017/2/15)
- 7) A. Culotta, J. Cutler: Journal of Artificial Intelligence Reserch 55, pp. 389-408 (2016).