

# SNS 上のソーシャル・ネットワーク分析としての Condor を用いた Twitter 分析の手法と可能性

井 村 直 恵  
Ken RIOPELLE

## 要 約

インターネットの発展とともに経営学でもビッグデータ分析で知見を導くデータサイエンスが広まっている。しかし、実際には SNS データの中身を分析した研究は経営学分野ではあまりない。本稿では日本におけるビッグデータの利活用を概観した後、インターネット上のソーシャルネットワーク情報である Twitter 研究の現状を把握する。これにより日本では Twitter 研究が情報学分野に偏っていることを確認した。この理由として、Twitter 情報を分析する為には API を用いてデータを取得することが必要であるが、この手法や技術に関する情報が少ないことが明らかになった。この分析から経営学分野で Twitter 情報の分析が少ない原因の一端がデータ収集の手法であると捉え、本研究では TwitterAPI を利用したデータ収集の手法として MIT の Gloor 教授の開発した Condor を用いてデータの収集及びソーシャルネットワーク分析を行う手順を検討した。

## 1. はじめに

マーケティング分野を始めとして、経営学の分野でも Twitter や SNS の分析に対する関心は高い。企業の側でもこれらインターネット上の情報を自らの経営情報として利活用する重要性はますます高まっている。また、分析結果から予測をして、事前に対策を打つことで、世の中のサービスの仕組みが変わり、人々の生活が変わり、我々の仕事の仕方も変化する。かつてはコールセンターが蓄積してきたユーザーの声の中でも、特に製品・サービスに関するちょっとした不満や不具合が Facebook や Twitter などのソーシャルメディアにつぶやかれ、ビッグデータとなる。これらを収集・分析し、改善につなげることで、より良い製品・サービスの開発につながる。交通、エネルギー、流通、医療、農業、製造、金融、政府や自治体、教育等多くの分野でビッグデータが新たな知見を提案し続けている。日本 IBM ではビッグデータの特性を多様性(variety)、頻度(velocity)、量(volume)としており、これらデータの約 80%は非構造化データと呼ばれるデータベースの分析に適さない電子メールや、画像、動画等として蓄積されている。こうした従来の構造化データとは異なる特性を持つデータを利活用することが、企業の競争優位や社会の革新につながる。

本稿では、近年のビッグデータ分析を巡る環境を概観した後、SNS データの分析をする為に Twitter データを例に、API を用いて Twitter データの収集、分析を行う方法を議論する。本稿では MIT の Gloor 教授が開発した Condor というアプリケーションを用いて行う手順を説明する。Condor を利用すれば、プログラミング等の専門知識を要せずに、Twitter, Facebook, e-mail 等のソー

シャル・ネットワークデータが取得可能となる為、学生や院生、情報学以外の研究者にとって、インターネット上の情報を収集・分析することが容易になる。

## 2. 日本における Twitter 研究の現状

### 2-1. 日本におけるビッグデータ利活用

ビッグデータという用語が新聞紙上や CiNii に登場するのは 2011 年頃である。総務省は平成 29 年度情報通信白書において、2017 年を「ビッグデータ利活用概念の到来」と呼ぶ。具体的には、2016 年末から 2017 年にかけて、官民データ活用推進基本法の制定や改正個人情報保護法の全面施行といった法整備を進めた。こうしてデータの利活用についての基本的理念を定め、今後行政手続きや民間取引のオンライン化を目指すとする。その上で、2018 年を「ビッグデータ利活用元年」と位置づけた。情報通信白書（2017）では、ビッグデータを収集する手段が IoT（Internet of Things）であり、それを分析・活用する手段が AI（人工知能；Artificial Intelligence）である。世界では急速に IoT を活用してビッグデータ分析を企業や国の競争力とする方向が進んでおり、我が国も遅滞なくこうした環境を戦略的に活用し、競争力を確保する必要がある。携帯電話を保有する個人の GPS 情報（NTTDocomo の「モバイル空間統計」）、自動車の走行等の情報（トヨタ自動車のテレマティックスサービス、ソニー損保のテレマティックウェアラブル端末）、ヘルスケア情報（ドコモ・ヘルスケアの「ムーヴバンド3」、オムロン・ヘルスケアの「Wellness Link」等）、金融関連情報（日立製作所の「金融 API 連携サービス」）等、日本企業も様々な分野でビッグデータ利活用による新たな社会基盤の開拓に取り組んでいる。しかし総務庁調査は、日本企業によるデータ利活用は欧米企業と比較して遅れていることを示す（図 1）。

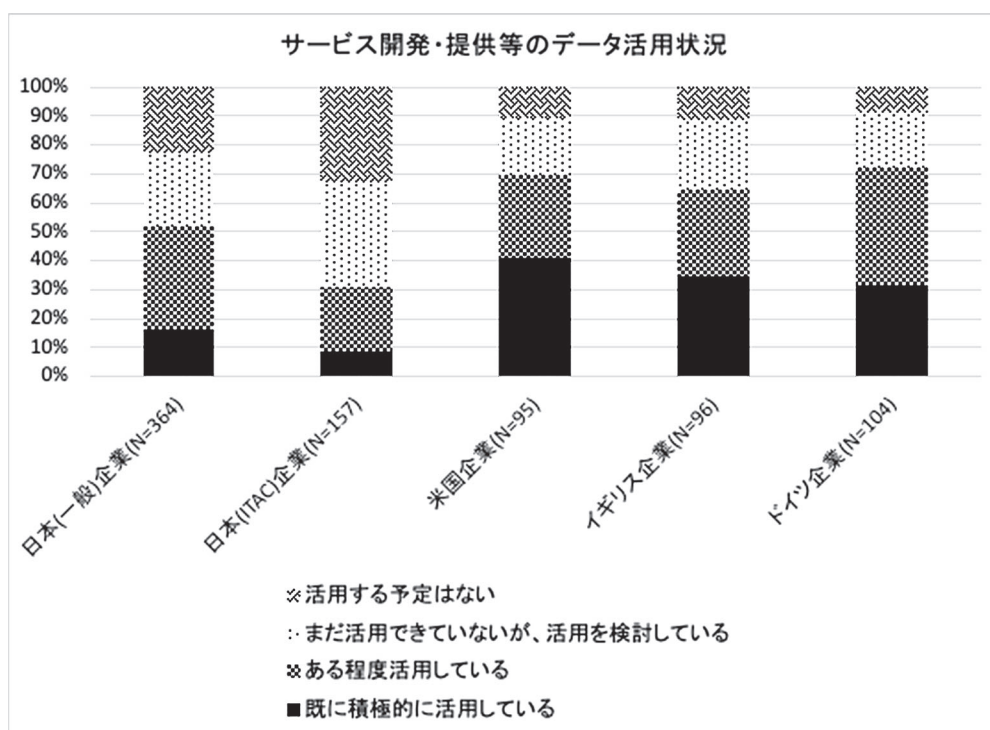


図1 サービス開発・提供等のデータ活用状況（産業データ）

出所：総務省「安心・安全なデータ流通・利活用に関する調査研究」（平成 29 年）

2016 年の第 46 回世界経済フォーラムの年次総会（通称「ダボス会議」）、続く 2017 年のダボス会議では、IoT や AI、ロボット技術等のもたらす社会変革を「第 4 次産業革命」として議論し、インターネットというインフラに対応した社会インフラ整備、サステイナビリティと産業革命等について議論した。ビッグデータを活かすデータサイエンティストの育成、企業の枠を超えた社会基盤の再構築は急務である。日本は、東日本大震災の後、ICT 利活用の重要性に関する認識が官民で高まり、2016 年 4 月 13 日に発生した熊本地震の際には、様々なツールを利用して情報発信・情報共有が行われた。テレビ、ラジオ等従来の情報発信インフラに加え、Twitter も民間レベルでの情報共有に大きな役割を果たした。井ノ口 & 今井（2018）はツイッター投稿画像を用いて屋内外被害の機械分類を提案する。曹（2018）や池田他（2018）はツイートを介したデマ拡散について分析する。横田 & 山田（2018）は、感情極性値の地理的・経時的分析を行った。Twitter 上のビッグデータを分析する為には、データの収集スキルは分析スキルと同様に重要である。曹（2018）は Twitter データの収集を外注してデータを購入している。池田他は、メンバーの 1 人である鳥海の研究室が Twitter 情報検索サービスを開発している。尚、横田・山田（2018）は分析プロセスについての記述はあるが、ツイートの収集手法については記載がなく不明である。

## 2-2. ビッグデータの構造

照井（2018）は、ビッグデータの構造をマーケティングを例に、図2のように整理する。横軸の構造化データとは、定量化されているデータ、非構造化データとは、文章などのテキストや、音声、画像などを指す。また、縦軸はデータの量を指し、上はデータの規模が大きいものを想定する。構造データとは、例えば、いつ、どの商品を、いくつ購入したかなどの消費者の購買行動を記録した定量化されたデータであり、非構造データとは、主に消費者の態度を表す定性的データである。小規模データは理論の検証や予測を目的として、推定や仮説検定を行う。小規模データでは計量経済学や統計学を適用する一方、大規模データの分析では、機械学習やデータマイニング、人工知能などのコンピュータを駆使したデータ分析を実施する。その際の分析は、パターン認識や発見、予測にあり、データの次元圧縮や分類を行う。

第Ⅳ象限は構造化された少数データであり、消費者行動、ロジット / プロビットモデルを用いた消費者のブランド選択行動が研究対象となる。分析対象の商品のカテゴリー内の競合関係や補完関係が分析の対象とされ、消費者は一部のサンプルパネルに限られる。

第Ⅲ象限は、消費者行動研究に基づいて設計された調査票を用いた、自由記述を含めた市場調査や顧客満足度調査などが代表的なデータとして使用される。第Ⅲ象限のデータの分析にあたっては、数量化理論が統計学上のツールとして利用されることが多い。

第Ⅰ象限は全数の顧客及び全数の商品の大規模構造データが対象となる。分析の規模拡張性（スケーラビリティ）が求められる。

第Ⅱ象限は、大規模非構造データで、ネット上のブログや SNS、Twitter、企業と顧客とのコミュニケーションによるインタラクションなどのテキストや、画像、音声などのデータが中心となる。近年はビッグデータの時代になり、この領域のデータが激増している。本研究では第Ⅱ象限のデータを分析する手法につき検討する。以下では Twitter に焦点を当てる。Twitter に着目するのは、Facebook 等とは異なりオープンソースであるため、利用者の側で公開設定の範囲を限定していなければ、API を活用することでデータ収集が可能になるためである。

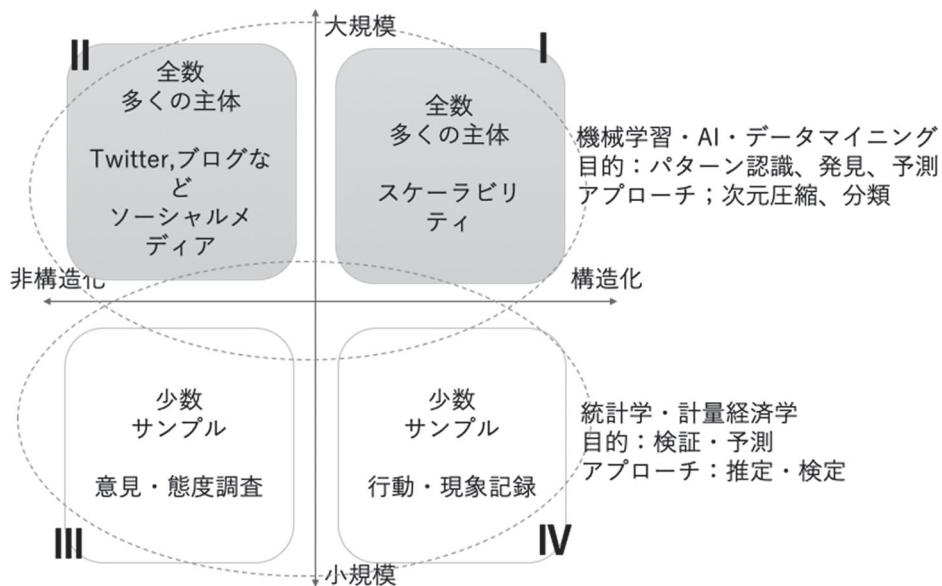


図2. 大規模データの構造

出所：照井（2018）. p13

### 2-3. Twitter 研究の現状

Twitter は、マイクロブログと呼ばれる 200 字程度まででつぶやきを公開もしくはグループを決めて公開範囲を限定して書き込みを行う各種のサービスの中で、最も一般的なツールである。

2018 年 12 月までの段階で、CiNii 上で Twitter に関する研究は、2,140 件存在する。中には、著者が不明な文章（日経 BP 社等のビジネス向け雑誌の記事が多い）や、同じタイトルで複数の学会で発表されたものも多い。その他、Twitter で検索した一覧には、Twitter が登場する以前の 1975 年や 1990 年代の論文もリストされる。明らかに Twitter に関する記事ではないものや著者が明らかでないもの、重複タイトルの研究をリストから削除すると、1,712 件抽出される。Twitter を抽出、分析する上での情報処理、人工知能、経営学、医学、メディア論、その他と分類したものが表 1 である。情報学が一番多いが、経営学が次に多く、ついで人工知能の論文である。人工知能学会の論文は、情報学関連の学会と重複して同タイトルで提出されていたものも多い。ついでメディア論、医学が続く。その他には、公共政策や言語学の論文が含まれる。経営学の論文の中には、観光学関連が最も多い。また、トピックとして多く出現するのは、東日本大震災や熊本地震等の災害に関するテーマであった。

表 1. CiNii の Twitter に関する研究の分野別分類

分野	情報学	人工知能	経営学	医学	メディア	その他
論文数	1,434	102	121	23	63	97

出所：筆者作成 2019 年 1 月 11 日アクセス 一部重複しているため合計は 1,712 件ではない。

Twitter を分析した海外の研究、特に英語圏の研究では、経営学関連は、LIWC というテキスト分析アプリケーションを用いて言語学的な分析に加え、感情を分析するものが多い。LIWC を用いれば、Tweet のテキストの内容から、品詞の使用頻度と肯定的な心理・否定的な心理、気分の変化（怒り、幸福感、不安など）、思考プロセス（確信、ためらいなど）、社会的関心事（家族、友人など）、時間感覚（過去や未来に関する表現）などの出現を分析可能である。例えば Dzogang et al. (2018) は英国の 54 都市から 2010 年から 2014 年の間に Twitter に投稿された 8 億件のツイートを分析し、英国の人々の心理状況の推移を分析した。ツイート内の言葉遣いを分析し、うつ状態や季節性衝動障害等の様々な精神的疾患のサインを検出した。LIWC を用いた分析では、気分の変化だけでなく、名詞や冠詞、前置詞の使用等、言語学的分析を用いて思考スタイルや性別、知的水準等も分析されている (Pennebaker, 2011)。近年では SNS 上の情報から個人の人格特性の予測 (Golbeck & Robles, 2011; Golbeck et al., 2016; Quercia et al., 2011b; Quercia et al., 2012; Sumner et al., 2012; 富永 & 土方, 2019) や行動との関係 (Kosinski et al, 2013) が注目されている。

Golder & Macy (2011) は世界 84 カ国 200 万人以上の Tweet を分析し、異文化比較を行った。LIWC を用いれば、テキストから品詞特性や用語の出現頻度が分析できるだけでなく、思考スタイルと表現方法（品詞の使用頻度、用語の選択）や認知プロセス、行動などとの相関を分析することが可能になる。LIWC は日本語に対応していない為、日本語で感情を分析する為には高額な商用ツールに頼らざるを得ないのが現状である。

投稿の内容のテキスト分析だけでなく Twitter や Facebook のプロフィール画像と個人特性との関係 (Celli et al, 2014; 富永他, 2014) を分析した研究もある。

他に、Twitter を用いた分析では、ツイート間の関係性、つぶやかれている場所などの情報を利用してソーシャル・ネットワークを析出する手法もある。Gloor (2017) では、Condor というインターネット情報収集アプリケーションを用いて、2015 年の米国大統領選のキャンペーンにおけるバーニー・サンダース、ドナルド・トランプ、ヒラリー・クリントン、ジェブ・ブッシュ 4 名に関するツイートをソーシャル・ネットワーク分析している。Gloor は、それぞれの候補者の名前に加え、#feelthebern, #Hillary2016, #makeAmericaGreatAgain, #jeb2016 などのハッシュタグでツイートを収集してマッピングした。その結果、クリントンは米国中心だが、世界中でツイートされていた。トランプに関するツイートは海外のコメントが肯定的な反面、米国内のコメントは否定的なものが多かった。これは、US のラテンアメリカマイノリティと違法移民に関する彼の攻撃に影響を受けていた。また、ブッシュはツイート数が少なく、肯定的なものではなかったことから、ブッシュが早期に離脱することは予言できた。その他、サンダースとトランプのリーダーシップの相違や機械学習を用いてファン層 (Gloor は tribe と呼ぶ) のコミュニケーション行動の相違を導き、選挙結果の予測に用いられることを示した。米国大統領選挙やブーム、災害、インフルエンザの流行のように、ソーシャルメディアの占める影響が大きい出来事については、ツイッターを分析して予測ツールとして用いたり影響力の強い人 (インフルエンサー) の発見を行うことが可能である。



## 2-4. Twitter を分析する際の課題

経営学分野の研究者による Twitter など SNS を用いた分析が限られているのは、Twitter 上の情報収集に関する技術的課題、社会情報やマーケティング、観光など、Twitter 上のつぶやきを用いた社会変動を研究対象とする分野の分析に向くという特徴があるためである。同時に、Twitter 上のつぶやきを収集する手法があまり知られておらず、データサイエンスもしくはプログラミングの知識が必要とされるためであると考えられる。

Twitter 上の情報を収集する方法として、鳥海（2015）はリアルタイム検索、API による収集、アプリケーションを用いる方法（大阪大学松村の開発した TTC、東京大学鳥海の開発した WTC）、データの購入などを提案するが、その後の 3 年間で状況は変化している。

近年、ビッグデータ分析の手法に関するデータサイエンティストの解説書も増加している。尾崎（2014）は、データマイニングの手法としての分析方法、プログラミングを R を使って行う方法を紹介する。R とは、S 言語を元に、1996 年にニュージーランド・オークランド大学の Ross Ihaka と Robert Clifford Gentleman が開発した統計分析パッケージである。オープンソースで無料で使用できることが最大の特徴である。

酒巻 & 里（2014）では、ビッグデータを活用した分析として、インターネット上での書き込み検索ワードのログ、アクセスログ等の他、位置情報ログ、通話や通信ログ、コンビニなどの購買履歴等の利用を指摘している。しかしいずれの書籍も、インターネット上の書き込みを収集する方法については説明していない。分析の前段階であるデータ収集については、コードを書けることが分析の前提となっている。プログラミングする必要があることが、情報学以外の分野の研究者による Twitter 分析の少なさの一因になっていると推察される。

## 3. ソーシャル・ネットワーク分析の近況

金光（2018）は、近年のソーシャル・キャピタル（社会関係資本）と経営についての論文集である。企業内ネットワーク、地域ネットワーク、企業／組織と社会の関わりという 3 種のソーシャル・キャピタルに分類し、経営学関連の研究者の最近の分析が纏められている。分析の手法に用いられているのは、理論レビュー 3 件、質問票調査 4 件、2 次データ分析 1 件、事例分析 3 件である。日本の最新のソーシャル・キャピタルの研究においても、ビッグデータを収集して分析された研究は含まれていない。金光は Adler&Kwon（2002）が指摘した 2012-2013 年以降論文数が減少傾向にあることを受け、「ソーシャル・キャピタル研究が転換点にあるサイン」、と述べるものの、論文集は方法的には従来からの方法論を踏襲するものが多い。ビッグデータ時代に対応したデータを収集・分析した研究は限られている。

米国では、例えば MIT Center for Collective Intelligence が過去 10 年以上にわたって社内外での e-mail のソーシャル・ネットワーク分析やテキスト分析等を実施し、ソーシャル・キャピタルとし

でのチーム・イノベーションや、リーダーシップの要件を導出するなど、ビッグデータ分析をソーシャル・キャピタルの研究に用いたものも見られる。例えば Gloor & Giacomelli (2014) では、グローバル組織におけるリーダーやグローバルに活躍する大企業が彼らのサービスに対する大口顧客の態度を e-mail の分析により明らかにした。その結果、コミュニケーションの態度と頻度を軸としてコミュニケーション・パターンを抽出している。彼らは、個人の e-mail の送受信を分析し、送信優位か受信優位かを指標化して「貢献指標」と呼ぶ。貢献指標の高い人は、情報発信志向が高く、コミュニケーションやコネクター、ゲートキーパーとしての役割を果たす。一方、各種専門家 (subject-matter experts) や知識ブローカーは情報発信志向が低いことを導いた。この時彼らが分析に用いた分析パッケージが Condor である。

#### 4. ネットワーク情報分析ツールとしての Condor

Condor は、MIT の Gloor 教授によって開発されたインターネット上の情報、e-mail や Web, Twitter などのソーシャル・ネットワークを可視化する分析ツールである。Condor では、前述の e-mail を用いた分析のほか、Internet 上の Web やブログ、Wikipedia、Twitter をクロールし、データベース上にデータの蓄積を行い、ソーシャル・ネットワークを可視化することが可能である。2019 年 3 月時点では英語のアプリケーションであるがデータの収集・分析は多言語に対応して可能である。以下では、Condor を用いて Twitter 上のつぶやきを収集し、ソーシャル・ネットワークを分析するプロセスを説明する。

##### 4-1. 前提条件

Condor は WindowsPC でも Apple 社の Mac でもインストール可能である。Condor の利用には、必要に応じて MySQL<sup>1)</sup> を事前にインストールしておくことと利便性が上がる。

Condor をインストールするためには、Condor3 をダウンロード<sup>2)</sup> し、ライセンスを取得する。学術目的での使用の場合は無償利用が可能である。分析者は、事前に自身の Twitter アカウントの取得<sup>3)</sup> が必要である。

1) 世界中で最も利用されているとされるオープンソースのデータベース。

2) ダウンロード先サイト <http://www.ickn.org/download.html>

学術利用ライセンスは無償で提供される。インストール後、GalaxyAdvisors か Prof.Peter Gloor にライセンス供与依頼のメールを送信する必要がある。送信先は GalaxyAdvisors の HP (<http://guardian.galaxyadvisors.com/>) 参照。左記サイト内で、マニュアルのダウンロードとチュートリアルビデオの参照も可能。

3) <https://twitter.com/login>



## 4-2. Twitter 分析のプロセス

### 4-2-1.API キーの取得

ツイートをダウンロードするには、API キーを取得する必要がある。API とは、Application Programming Interface の略でソフトウェアやアプリケーションの一部を外部のソフトウェアやアプリケーションと機能を共有したり操作するためのインターフェイスである。異なるアプリケーションやサービス間で認証機能やチャット機能を共有したり、一方から取り込んだデータを他方で解析したり出来るようになる。API キーはこの API を呼び出すアプリまたはプロジェクトの識別・承認に用いられる。以前は Twitter アカウントを取得していれば、同時に API キーの取得が出来た。しかし 2018 年 7 月以降、Twitter の API を取得するためにはデベロッパー向けアカウントの取得が必要となった。この審査には使用目的の説明（英語）等を経て、通常 3 日から 2 週間ほどを要すると言われている。だが、Condor ではこのプロセスを経ずに API キーの取得が可能である。通常の Twitter アカウントを有していれば、Condor は分析者の許可を得ることで、Twitter の探索に必要な API キーを 2 つ、一時的に発行する。それを使えば、Twitter の探索を、デベロッパー向けアカウントの申請・承認を待たず、即時に行うことが出来る。

### 4-2-2.Condor の起動からツイートの検索まで

データベースの導入を行う場合は、必要に応じて、予め MySQL をインストールしておく。分析に際しては、MySQL を立ち上げた後、Condor3 を立ち上げる。

File メニューから、create か switch to your database を選択し、create a new dataset でファイルの名前を決めた後、画面を閉じる。

次にメニューから、Fetch>Fetch Twitter を選び、dataset を選択する。（図 3）

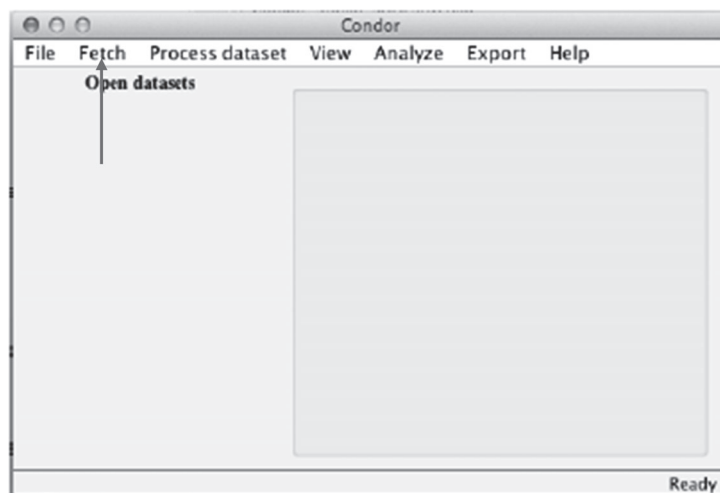


図 3 Condor 画面

選択したい単語やキーワードを query に入力後、Next をクリックする。2 語以上の場合は語句と語句の間に“,”（半角カンマ+半角スペース）を挿入。2 語をセットとして検索する場合には、“検索語句”のように、検索したい語句を“”で囲む。検索語は日本語でも英語及びその他の言語でも可能である。ここでは例として NAFTA を検索する（図4）。



図4 query の入力画面

Login with twitter を選択肢、Next をクリックする<sup>4)</sup>（図5）。



図5 Condor 内で Twitter ログイン画面

4) ここでは、TwitterAPI を所持していない場合のログイン方法について説明した。もしすでに TwitterAPI を所持しているか、Twitter のデベロッパーアカウントを取得している場合“Use the credentials from my app”を選択し、その後の指示に従う。この場合の画面は、以下のものとは異なる。Login with twitter を選択すると、以前に Condor で分析したことがある場合でも、consumer key と consumer secret key が消去されるので注意が必要である。

Login with Twitter をクリックする (図 6).

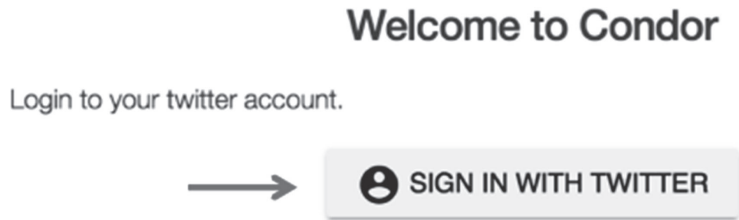


図 6 Condor 内で Twitter にサインイン画面

Sign in with Twitter をクリックして, 自身の Twitter アカウントにログインする (図 7).



図 7 TwitterAPI 1 次キーの取得画面

Authorize app をクリックする (図 8)。

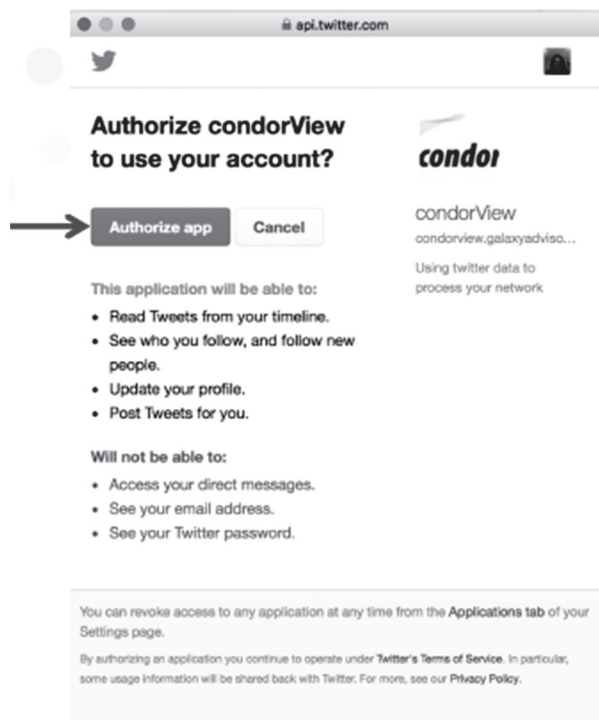


図 8 Authorize app 画面

Condor が access code と secret code の一時的キーを生成する(図 9)。それらを Condor Fetch ウィンドウの該当するコードにコピー＆ペーストする (図 10)。

**Welcome to Condor**

access code:

secret code:

図 9 取得した TwitterAPI 一時キーの access code と secret code 表示画面

入力が終われば Next をクリックする (図 10)。

**Twitter Fetcher**

Allow Condor accessing your Twitter data.  
Follow the instructions in your browser to obtain a access and secret code.

Login with Twitter

Copy&paste access code:

Copy&paste secret code:

< Prev
Next >
Cancel

図 10 取得した TwitterAPI 一時キーの入力画面

#### 4-2-3. 検索結果の表示

File メニューから、先程探索に使用した Twitter dataset を開く。MySQL をインストールしていない場合、検索結果は ".cdf" という拡張子で保存される。自身が保存した場所から File>Import dataset で.cdf ファイルをインポートする<sup>5)</sup>。

View > Create static view メニューで結果を提示する。今回の説明では、1 時点での検索を説明しているが、Condor の利点はダイナミックな変化を示すビジュアライゼーションが可能な点である。その場合、検索画面にて検索時期を設定した上で、View>Create dynamic view を選択することで、経時的变化を表示する。

以上のプロセスを経て、Twitter のネットワークが表示される (図 11)。

5) MySQL を導入していない場合、インポートに時間が掛かる点、自身がデータを保存した場所を記録もしくは記憶しておく必要がある点などの手間がかかるが、検索及び表示は導入していなくても可能である。

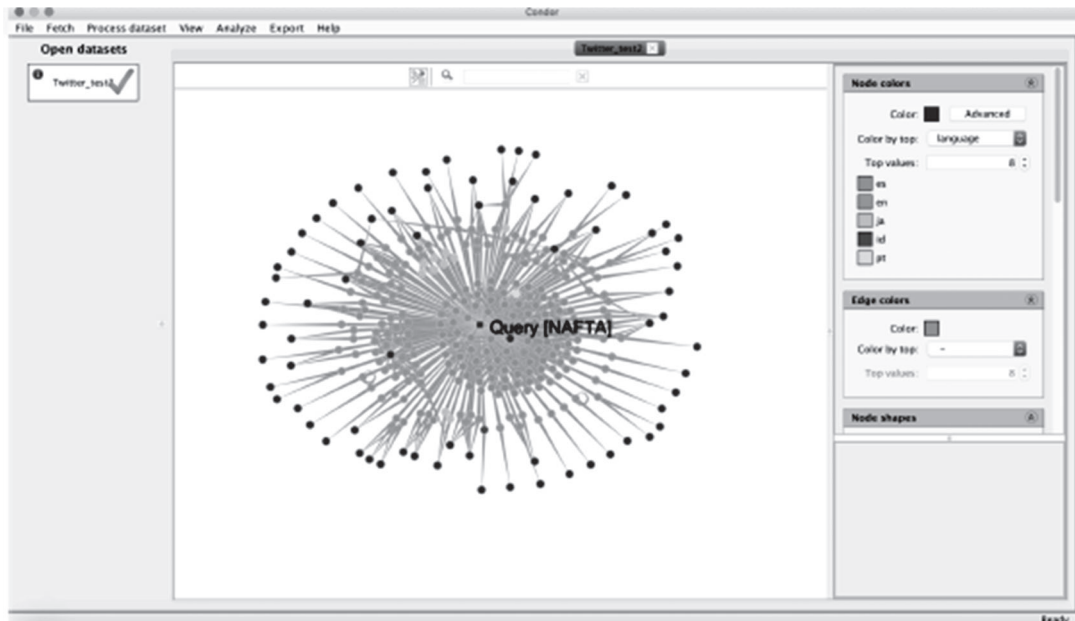


図 11 NAFTA に関するネットワーク

Process dataset>Remove specific actors で query を入力，実行すると，Query を除く，ツイート間のソーシャルネットワークが表示される。

例えば，2019 年 1 月 26 日テニスの全豪オープン優勝を果たした「大坂なおみ」選手に関する 2019 年 1 月 29 日のツイートの媒介中心性 (betweenness centrality) を表示したものが図 12，図 13 である。媒介中心性とは，あるノードを経由することが，他の 2 つのノードにとって最短経路になっているケースの高さを見る指標である。この指標が高いほど，ネットワークのハブになっている可能性が高い。検索は「大坂なおみ」及び「Naomi Osaka」で行い，2 つのファイルを合併した。Condor はこの他にもネットワーク分析の結果を表示可能である。また，Node colors の変数を変えれば，ツイートの言語や，どこでツイートされているか（ユーザーの側で GeoLocation の識別可能な位置情報を付与したツイート設定をしている場合），色分け表示される。図 12 はツイートされている場所を，図 13 はツイートされた言語を示す。ツイートされた場所の検出は，ツイッターの GeoLocation の設定に依存する為，日本でも日本，東京，東京日本，Japan，japan などが別の情報として検出されている。一方，ツイートされた言語は検出される言語の重なりはない。右上のツイートの塊の中心にはオバマ前米国大統領のツイートがあり，オバマ元大統領とつながって左のツイートが大坂なおみ選手本人のものである。日本語ツイートの多さから，大坂選手の全豪オープン優勝は日本での反応が大きかったことが示されている。Condor では，媒介中心性以外にもネットワーク分析の結果の析出ができるよう，細かいオプションが組み込まれている。アクターのプロフィールとツイートの内容も CSV ファイルか Excel ファイルとしてダウンロード可能であり，LIWC などで



ツイート内容のテキスト分析を行う際には、これらデータを Export>ExportCSV から抽出して分析する。

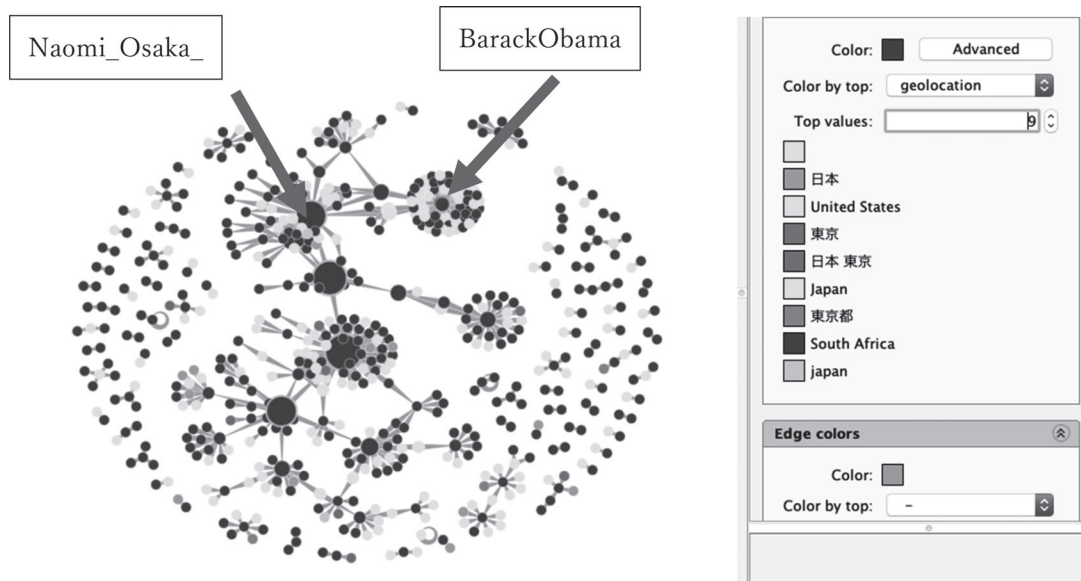


図 12 大坂なおみ +Naomi Osaka に関するツイート場所別表示 (2019 年 1 月 29 日取得)

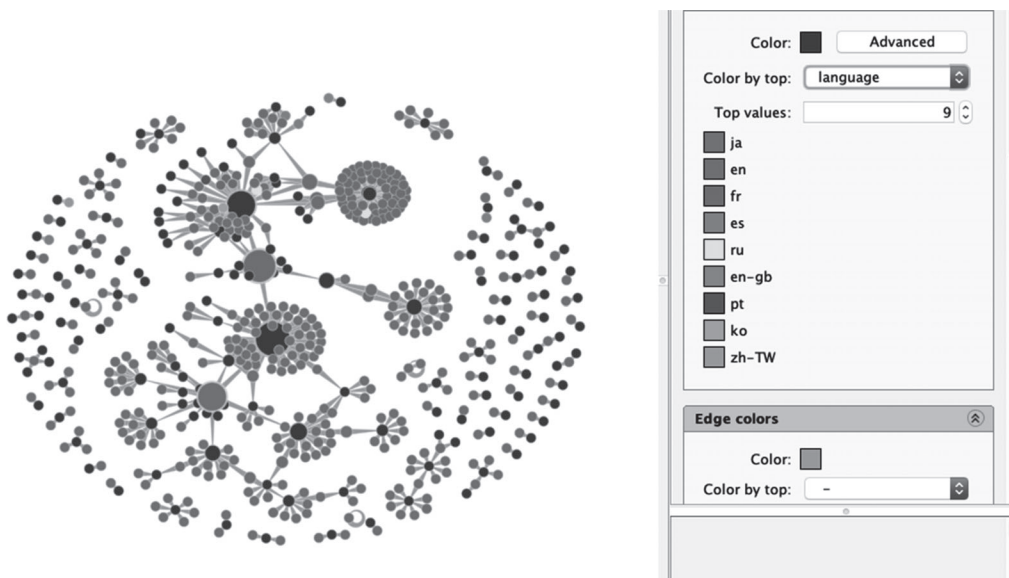


図 13 大坂なおみ +Naomi Osaka に関するツイート言語別表示 (2019 年 1 月 29 日取得)

## 5. まとめ

本稿では、経営学の分野におけるビッグデータ分析の可能性を探る上で、ビッグデータに関する近年の動きを考察し、MITの研究チームが開発した Condor というツールを用いて Twitter 情報を収集する方法について検討した。ここでは Twitter からデータを取得するための条件や、分析のプロセスを提示した。

Condor は Twitter の他にも、e-mail, Wikipedia, Facebook, Web, blog 等インターネット上の情報を収集し、ネットワーク分析を行うための多くのオプションを提供する。Condor 以外にもこれらの情報を収集する方法は多数あるが、Condor を用いる利点は、プログラムに関する専門知識を有さずとも GUI によって社会情報のビッグデータ収集が可能になる点である。

Condor を用いる限界もある。Tweet の内容をテキスト分析する上では、日本語だとダウンロードで文字化けが起きることがある。その場合データの修正が必要となる。また、Condor に起因しない条件として、ツイートは一定期間経つと消える。書き込みをした当事者が削除することもあるが、その場合データの修正が必要である。TwitterAPI を利用する場合、一回の検索で得られる情報量、1 人のユーザーに対して第三者が取得可能な情報量（富永・土方（2019）によれば最大 3200 件）遡れる期間など、API の制限を受ける。他には、今回は検索事例として、今回は「大坂なおみ」選手を用いたが、同時期に活動休止を発表したアイドルグループ「嵐」に関する記事を収集する場合、「嵐」という query では目的外の情報を収集してしまう。query は注意深く選択したり、分析目的に合う結果が得られる query が得られるまで、試行錯誤する必要がある。また、Condor に限ったことではないが、2018 年に Twitter が API 取得に対する戦略を変更したり、地理情報の取得方法を変更したように、技術的進化や変化の影響を受けて、データ収集の可否や範囲が変わる可能性がある。Twitter は公共のメディアであるが、プライバシーについては、一定の配慮が求められる。Condor を用いた具体的な分析については、別稿にて議論したい。

## 謝辞

本研究は、JSPS 科研費 15K13045, 18H00899, 18K18586 による助成を受けた。

## 参考文献

- Adler, P. S. & S.Kwon (2002) "Social capital: Prospects for a new concept." *Academy of Management Review*, Vol.27, pp17-40.
- 曹慶鎬 (2018) 「インターネット上の災害時「外国人犯罪」の流言に関する研究：熊本地震発生直後の Twitter の計量テキスト分析」『応用社会学研究』 pp. 79-89.
- Cilli F., Bruni E., Lepri B., (2014) "Automatic Personality and Interaction Style Recognition from Facebook Profile Pictures", *Proc. of the 22<sup>nd</sup> ACM Int. Conf. on Multimedia.*, ACM. pp.1101-1104.

- Dzogang F., Lightman S., Cristianini N. (2018) Diurnal variations of psychometric indicators in Twitter content. PLoS ONE 13 (6): e0197002. <https://doi.org/10.1371/journal.pone.0197002>
- Gloor P. (2017) *Sociometrics and human relationships*, Emerald Publishing.
- Gloor P. & Giacomelli G. (2014). Reading Global Clients' Signals. *MIT Sloan Management Review*. 55. 23+.
- Golbeck J., Robles C., Edmondson M. and Turner K. (2011a) "Predicting Personality from Twitter", *2011 IEEE Third Int. Conf. on Privacy, Security, Risk and Trust and 2011 IEEE Third Int. Conf. on Social Computing*, IEEE, pp.149-156.
- Golbeck J., Robles C. and Turner K., (2011b) "Predicting Personality with Social Media", CHI'11 Extended Abstracts on Human Factors in Computing Systems, ACM, pp. 253-252.
- Golder S. A. Macy M. W. (2011) "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures", *Science*, Vol.222, Issue 6051, pp1878-1881.
- Kosinski M., Stillwell D. and Graepel T. (2013) "Private Traits and Attributes are Predictable from Digital Records of Human Behavior", *Proc. of the National Academy of Sciences*, Vol.110, No.15, pp.5802-5805.
- 池田 圭佑, 榊 剛史, 鳥海 不二夫, 栗原 聡 (2018) 「口コミに着目した情報拡散モデルの提案およびデマ情報拡散抑制手法の検証」『情報処理学会論文誌数理モデル化と応用 (TOM)』 pp.21-36.
- 井ノ口宗成 & 今井淳 (2018) 「twitter 投稿画像を用いた屋内外被害の機械分類にかかる基礎研究：2016 年熊本地震を事例として」電子情報通信学会技術研究報告, pp.1-4.
- 金光淳・編著 (2018) 『ソーシャル・キャピタルと経営』 ミネルヴァ書房.
- Pennebaker J. W. (2011) *The secret life of pronouns; what our words say about us*, Bloomsbury Press.
- Quercia D., Lambiotte R., Stillwell D., Kosinski M. and Crowcroft J. (2012) "The Personality on Popular Facebook Users", *Proc. on the ACM 2012 Conf. on Computer Supported Cooperative Work*, ACM, pp.955-964.
- Quercia D., Kosinski M., Lambiotte R., Stillwell D. and Crowcroft J., (2011) "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter", *2011 IEEE Third Int. Conf. on Privacy, Security, Risk and Trust 2011 IEEE Third Int. Conf. on Social Computing*, IEEE, pp. 180-185.
- 酒巻隆治, 里洋平 (2014) 『ビッグデータを活かすデータサイエンスークロス集計から機械学習までのビジネス活用事例』東京図書.
- 総務省『平成 29 年度版情報通信白書』日経印刷.
- Sumner C., Byers A., Booschever R. and Park G.J. (2012) "Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistics Analysis of Tweets", *2012 11<sup>th</sup> Int. Conf. on Machine Learning and Applications*, IEEE, pp.386-393.
- 照井伸彦 (2018) 『ビッグデータ統計解析入門』日本評論社.
- 富永登夢, 土方嘉徳, 西田正吾 (2014) 「アイコン画像に着目したツイッター研究の提案」第 28 回人工知能学会全国大会論文集, 3M4-4in.pp.1-4.
- 富永登夢, 土方嘉徳 (2019) 「Twitter ユーザーの受け取るフィードバックと人格特性の変化に関する調査と分析」『知能と情報 (日本知能情報ファジィ学会誌)』 Vol.31, No.1, pp.516-525.
- 鳥海不二夫 (2015) 「Twitter 上のビッグデータ収集と分析」, 組織科学, Vol48.No4, pp47-59.

横田 尚己 & 山田圭二郎 (2017) 「熊本地震のつぶやきに見る感情極性値の時空間解析」『都市計画論文集』公益社団法人日本都市計画学会, pp1081-1087.

## Current status and issues on SNS Big data analysis: Methods for Twitter analysis by Condor

Naoe IMURA  
Ken RIOPELLE

### ABSTRACT

The development of the internet brings rapid spread of big data analysis and data science. However, not many Online Social Network analyses were found in management studies in Japan. We consider that one of the reasons is the lack of data collection method. In this research, we introduced data collection method and social network analysis using Twitter API and “Condor”.

