

第22回 ビッグデータ等の利活用推進に関する産官学協議のための連携会議 議事概要

【日 時】 令和6年8月7日（水）16:00～18:00

【場 所】 総務省第二庁舎6階特別会議室（Web会議併用）

【出席者】

構 成 員：高橋座長、和泉構成員、大西構成員、庄司構成員、田原構成員、水野構成員

審議協力者：株式会社ナイトレイ 石川代表取締役、総務省行政評価局政策評価課 菊永企画官

総 務 省：山田審議官、重里統計企画管理官、谷本統計員会担当室長、田村統計委員会担当室次長、植松参事官、河野企画官

【議 事】

- （１）人流データとしてのSNS投稿データ分析と活用事例
- （２）政策効果の把握・分析手法の実証的共同研究について
- （３）ビッグデータ等の利活用推進に係る政策評価について

【配付資料】

資料1 人流データとしてのSNS投稿データ分析と活用事例

資料2 政策効果の把握・分析手法の実証的共同研究について

資料3 統計品質管理推進室が実施するビッグデータ利活用推進事業について
～ビッグデータ等の利活用推進に係る政策評価について～

参考資料 第21回ビッグデータ等の利活用推進に関する産官学協議のための連携会議 議事概要

【（１）人流データとしてのSNS投稿データ分析と活用事例】

株式会社ナイトレイ石川代表取締役から、資料1に沿って説明。主なやり取りは以下のとおり。

- SNSデータが人流データとして捉えられるというお話について理解が深まった。SNSデータといっても、アプリケーションの種類によって、利用者や特徴を有していると思うが、SNSごとの特徴について、貴社としてどのように認識しているか御教示いただきたい。
- 当社としては、ビジネスとしてSNSデータを利用しているわけであるが、SNSごとの利用規約が定められており、SNSによっては、企業・事業者において勝手にクローリング・クリーニングしたデータの利用及び販売を認めていないものがある。また、かつて当社が解析していたあるアプリケーションについては、利用規約の変更が行われた結果、事業と

してデータを使いづらくなったという事情もある。こうした背景から、当社として問題なく解析できるデータは、最も主要なアプリケーションであるXに限定されているという現状がある。

なお、日本においては、ある地域でXの利用者が多い・少ないという傾向はなく、世界的に見ると、日本はXの利用者数の上位に位置している。その意味で、日本国内を解析する上で、Xの活用は適していると認識している。逆に、Xよりも別のSNS利用者が多いという国もあるため、世界中をXだけで解析するのは、日本のみを解析するよりも、難易度が高いと考えている。

→ SNS運営企業側の事情もあり、様々な御苦勞をされているのだと感じた。他方で、位置情報が取得でき、かつそれに付随して、投稿者の感情や写真等のデータも取得できるのであれば、解析対象としてはSNSではなくてもよいのかなとも思った。

○ 1点目、解析に当たっては、各投稿に付加されている位置情報を使うとのことであったが、そもそも位置情報はどのようにして付加されるのか御教示いただきたい。2点目として、投稿される膨大なテキストデータを解析するよりも、位置情報が付加されている投稿のみを分析することに大きなメリットがあるということか。3点目、スマートシティやまちづくり政策にも活用している事例を御紹介いただいたが、もう少し詳細をお聞きしたい。

→ 2点目について先に申し上げると、大前提のコンセプトとして、そもそも曖昧なデータは扱いたくないというのがある。SNSデータを人流データとして活用する以上は、時間・場所・施設名（緯度・経度）という3つの信頼性の高い情報が得られた投稿のみが当社の解析結果データベースに取り込まれるようにしており、これらの情報が欠けている投稿については、なるべくデータベースに入らないようにしている。

1点目の位置情報の付加の方法に関して申し上げますと、ユーザーがアプリ上で施設を選択する方法もあれば、ユーザーが投稿した「〇〇に来た」というテキストデータから情報を、我々独自のアルゴリズムで実際に来ている投稿を特定するという解析方法がある。解析に当たっては、何十通りもの条件を設けており、それらに合致する投稿を取り出して、解析しているため、結果として、分析対象は少なくなっている。

→ フィルタリングをよく工夫されていると理解した。

→ 大量のテキストデータ分析を行っている事業者もあるため、当社としては差別化を図るためにも、このような方法で解析を行っているところ。

3点目については、スマートシティ推進に当たり年代や性別による課題を解決する必要があるところ、SNSデータからそれらの情報を正確に分析することは難しいため、SNSデータだけを使ってスマートシティに活用するというケースはそこまで多くない。携帯キャリアの滞在人口データ等の正確な年代や性別を捉えやすいデータを使って、対象となる人々の移動や滞在の実態、困りごと等をデータとして読み解きつつ、最後に、人々は

ここでこういうことをしているなど、手触り感を出すためにSNSデータを組み合わせるといふが多い。

- 1点目、2011年頃から事業を開始されたとのことであるが、当時と今を比べると、SNSの使い方も変わってきているように思う。データの収集・解析を行っている御社として、SNSの使い方に変化があったとお考えか。2点目、御紹介いただいた観光事例以外の分野で、SNSデータと位置情報を掛け合わせたデータの使い道があるのかどうか。3点目、SNSの中には、動画など、言葉以外の内容が近年、増えてきている。そうした投稿からは、その当時、そこに何があったのかが見える形になっている。こうした使い方をする場合、マルチモーダルに、文章以外のデータを解析する場合、技術的な壁もしくは法的な壁はあるのかどうか、お聞きしたい。

→ 1点目については、解析するSNSの幅を広げると、他のアプリケーションなど圧倒的に動画の投稿が増加していると思われるが、当社では解析する対象のフォーカスを意識しており、ユーザー自身の発信が大きく変わったとは思っていない。現在分析しているXの投稿可能字数は初期に比べて増えているものの、比較的短文で、言いたいことをオープンに発信するというコンセプトは変わっていない。ただ、初期の頃は場所の発信をするのは物珍しく、チェックインという概念があったが、そのような発信は減っているように感じる。

2点目については、防災の分野がある。他のデータは匿名化処理を1か月かけて行うようなデータもあるが、SNSはリアルタイムな情報発信やコミュニケーション手段としての利用が多い特徴があるため、当社のSNS解析アルゴリズムは最短10分程度で最新データを取り込み、解析して見える化できることに関心をいただくこともある。例えば、先週発生した水害について、SNSでどのような発信がなされているかを解析し、過去データとして読み解いた上で、今後、同様の災害が起こったときに、SNSでリアルタイムに住民の状況を捉えられるようにしたいといった相談を受けるケースも増えてきている。

また、東京五輪の際は東京都から依頼を受け、五輪開幕直後の人流データを分析し、コロナ禍で多数の外国人が訪れる状況において、バブル方式（選手や運営関係者を外部と接触させない方式）の運営ができているかを把握し、何らかの予兆が確認できた場合は、東京都としても注意喚起をするといったリスク管理の観点からの活用事例もあったところ。

3点目について、当社として解析しやすいデータがXのデータであったということで、動画の解析には注力していない。なお、災害などニュース性のある事故等に関しては、SNS上に動画で発信する者も多いと思われるところ、他社においては、動画で発信された情報をAIの画像解析エンジン等を用いてタイムリーに解析し、ニュース速報や取材先の選定に活用できるような技術を作っている事業者もいる。

【(2) 政策効果の把握・分析手法の実証的共同研究について】

行政評価局政策評価課菊永企画官から資料2に沿って説明。主なやり取りは以下のとおり。

- 資料にも改善の方向性として書かれているが、因果的な関係の深掘りについては非常に重要である。その際、まずは関係者において、ロジックモデルをきちんと作成することが重要であり、イベントからターゲットとするアウトカムまで、どのような中間因子があり、その中間因子を測定するためには、どのようなデータを補助的に使うのか、そうしたことを整理すると、よりデータを使うための指針ができると考える。

加えて、波及効果として、意図せざる効果が発現する場合もあるだろう。例えば、ある施設・イベントに人は呼べたが、別のところから人を奪ってしまう結果になった際、トータルでみて、本当に施策の効果があったのかどうかは分からないだろう。したがって、ロジックモデルを作った上で、総合的に評価することが重要と考える。

- 事前にロジックモデルを作成すべきという点はおっしゃるとおりで、例えば、予算事業の効果検証に当たっては、行革事務局が主導している行政事業レビューシートを作成することになっているところで、レビューの際には、その事業の目的は何で、どうやって事業を進めていくのかを整理することが最も重要な点であり、この方針のもと、各省庁で取組が進められていると承知している。

また、波及効果の測定については、現場の担当者の声を十分に聞き取って進めていくことが重要と思うので、御指摘の点は引き続き留意していきたい。

- アドバイスとしていくつか申し上げたい。1点目として、施設利用者の属性・ニーズの把握が難しかったというお話だったが、新規に立ち上げるイベント・施策等については、一定の効果が見込めない・効果が不明確な場合は意思決定しないというのがEBPMの本質であり、企画段階で、こういう効果があるからこれを実施しますという順で意思決定すべきと考える。

2点目として、データは後から取ろうとすると非常に大変であるため、効果の捕捉を行う際に、あらかじめ、どのデータを使うか、どうやってデータを取るかを政策立案時に決めておかないと、関係者は苦勞すると思う。3点目としては、データをいかにして継続的に取得し続けるかという点も考えておく必要があるだろう。

- 当局としても、ロジックモデルを作る際には、どういったデータが必要なのかについては、セットで考える必要があると常日頃から関係者にはお伝えしてきている。例えば、新規の予算事業については、御指摘のとおり、あらかじめ測定データを考えておくことが必要ではあるものの、事業実施までに十分な時間・余裕がないというケースもある。複数年継続している事業については、既存のデータがあるため、そうしたデータを用いて効果検証し、事業の軌道修正をし、新規の事業については、余裕が出てきたら、政策立案段階から効果検証の点も踏まえて検討していくのだろうと考えている。

○ 御紹介いただいた調査研究事例において、データを取り扱う者や方法の適正性の確保について、どのように取り組んでいるのか。地方公共団体においては、例えば、観光にきた外国人旅行者をどこまでも追跡するといったデータ測定をしているという話も聞くところ、政策改善の目的は理解できるものの、データの取扱いに課題があると思われる事例も聞いている。

→ 総務省としては、データ全般の取扱いを所掌しているわけではないが、統計法上、都道府県及び政令市が統計調査を行う際は、総務省に届出を行うこととなっているが、市町村については届出が不要なので、例えば、独自の条例を設ける等により、一定の規律を設けて実施している例があるのではないかと思料する。国の統計調査の場合は、各省庁の統計調査を総務省が審査するプロセスが確立しているところ。データ全般について、規律・規制を設けることはなじまないと考える。

→ 統計調査の範囲の外で、政策効果の把握を行おうとする場合のデータの取扱いについて懸念があったので、質問させていただいたところ。

○ 今後、同じようなデータを活用したいと思った際に、データの所在を明らかにしておくことは重要と考えるが、調査研究で用いたデータは現在どこで保管されているのか。

→ 実証的共同研究では、原則、当局がデータを保管し、共同研究の相手方にも同様に共有している。府中市の案件は契約上、ローデータは保有していないが、集計データを当局及び府中市で保管・共有しており、姫路市の案件はローデータ及び集計データともに当局及び姫路市で保管・共有している。

御指摘のとおり、事例を横展開していく上では、データの取扱い・管理も必要と考えるが、現状は、実証的共同研究という形で、1つの事例を研究していくことに留めている段階であり、現状このような扱いとしているところ。

【(3) ビッグデータ等の利活用推進に係る政策評価について】

事務局から、資料3に沿って説明。主なやり取りは以下のとおり。

○ 今後の取組の方向性について、「ノウハウも含めた実践的な研究」というのが重要。実証研究により得られたノウハウとビッグデータ・ポータルがうまくリンクして、成功事例を生み出すことができれば、他省庁にもカスタマイズして、展開していくことができると思うので、ぜひ実証研究の成果とポータルが連携する形で事業を進めていただければと思う。

以上