

ビッグデータ連携会議におけるこれまでの事例整理

－ 公的統計作成への活用を中心に －

1 本とりまとめの目的

- ・本連携会議は、統計的分析や統計作成における各府省・地方公共団体・民間企業等でのデータ等の相互利活用を推進することを目的として開催
- ・これまで会議において紹介されたビッグデータ利活用の先進事例を通じて明らかとなった、利活用におけるメリットや課題及びその解決方法など、関係者において共有することが有用である内容を整理
- ・今後、本とりまとめの内容を参考にビッグデータ利活用の促進を図る

2 ビッグデータの活用法やメリット

(1) 既存統計の補完

- ・ビッグデータの特徴である高頻度・多種大量にデータが生成されることを生かし、既存統計では把握されていない時間的・空間的・統計分類的観点等から既存統計を補完
- ・既存調査の完全な代替ではなく、調査項目の中で対応可能となる一部代替や、新たな指標の作成などで活用
- ・これまでの連携会議において紹介された事例・意見は以下のとおり

①速報性

- ・公表の早期化や短期での動向把握を念頭に、POSデータ等の民間ビッグデータを活用し、従来よりも速報性に優れた指標を開発

②詳細化

- ・詳細なスケールで人の移動を捉えるニーズが広まっており、統計調査において設定・把握しているゾーンよりも詳細な動きを把握するため、ビッグデータを用いて、ゾーン内部を分割した目的手段別OD表（起終点表）を推計（「総量」を統計調査で担保した上で、ビッグデータの「比率」を活用して、より細かい内訳を把握）

③カバレッジの拡大

- ・ネット販売の増加等を背景に、ウェブスクレイピングにより、ネット販売価格の直接把握や膨大な量の価格データによる統計精度の向上が期待。ネット販売が主流であること等から「外国パック旅行費」、「国内航空運賃」「国内宿泊料」の価格動向をウェブスクレイピングにより把握。従前に比べ、旅行会社・プラン数や路線数の把握対象の拡大、価格収集期間や高頻度（毎日）の価格把握などカバレッジの拡大が可能となり、統計精度の向上に寄与

④新たな指標の作成

- ・Twitter から景気に関連するキーワードを含むツイートを抽出し、センチメント評価を行うことで景況感に関する指標を作成
- ・検索データによって意識化されていない国民の関心を把握

(2) 報告者負担や業務負担の軽減

- ・ウェブスクレイピングにより、報告者負担をほぼゼロにするとともに、収集自動化により業務を効率化。また、比較的安価にデータ収集
- ・調査をPOSデータで代替することで家電大型専門店（報告者）の負担を軽減

3 ビッグデータの活用における課題及びその解決方法

- ・ビッグデータを活用するためには、精度検証やデータの入手方法の確立、人材面等の課題が存在し、これらの解決を図ることが必須
- ・これまでの連携会議において紹介された課題及びその解決に向けた取り組みの例や意見は以下のとおり

(1) データの精度検証

- ・ビッグデータはそれぞれデータ生成状況等に伴う特徴を有しており、その特徴に留意した利活用が必要
 - －本連携会議「メッシュ型流動人口検証WG」における、都内地域でのアプリGPSによる流動人口データの検証では、アプリ利用者属性に起因する標本の偏りが存在することを確認。その上で、国勢調査や基地局データと相関を有することや、国勢調査の差率標準偏差基準では2000人以上のメッシュで一定レベルの信頼性を保有すること、大量の人の動線のハブとなる都心ターミナル駅や海岸沿い等隣接するメッシュ間の人口差が大きいエリアでの有効性を確認
 - －一般にウェブスクレイピング情報は店頭・ネット販売価格全体を代表していないことに留意し、ネット販売が主流であることやデータ特性等の費用対効果を見極め、価格動向をウェブスクレイピングにより把握する品目を選定。また、例えば「外国パック旅行」については、旅行会社から販売実態に関する詳細な情報提供が得られたことにより、膨大な数の商品情報からノイズを除去し、物価動向把握において前提となる同品質の商品の価格追跡が可能に
- ・データ提供側企業との連携・工夫により精度の維持向上を図る面あり
 - －POSデータについて一定の店舗数を保ちながらデータを収集し、仮にこれらの店舗の中から閉店になるものがあつた場合でも、データ提供側で同規模の店舗を割り当てることで対象店舗の状況を調整する場合あり

(2) ビッグデータの入手

- ・ビッグデータは民間の企業活動に伴うデータであるため、官側での利用にあたっては、データの入手方法・枠組の確立や、その持続可能性が必要
- ・民間企業等との信頼関係を構築
 - －産学官の連携による研究協議会を設立し設立趣旨に賛同する企業が参画。公的施策への貢献という企業側の社会貢献に対する意識に支えられている
 - －情報の扱いについても、各企業が提供可能と判断したデータを利用。データを取り扱う官の職員を限定するほか、大学研究者らも国家公務員として任用し守秘義務の下で情報管理及び保秘を徹底。参加企業間でのデータの共有は行わない。成果物の公表に際しては参加企業の個別データの状況が明らかとならないようにし、参加企業の了解を得た上で実施
 - －データホルダー企業と官側で連携協定を締結
 - －ウェブスクレイピングにおいても、データホルダーの理解・協力が必要であり、ネット情報の取集・統計作成に関して対象企業から承諾。また、サーバーへの負荷を考慮して、企業側から提示のあつたアクセス時間帯・頻度の制限、IPアドレスの事前登録などの対応を実施

- ・データ提供者へのインセンティブ付与
 - ーデータの分析過程で出てきた副次的な情報をデータホルダーにフィードバックするなど、Win-Win の関係となれる利活用モデルを構築する必要
 - ーデータホルダー単独ではできなかった精度向上に資する情報を提供するなど、先方がメリットを感じられるものになれば、例えばデータは購入ではなく、共同研究の枠組みなどより前向きな取組になり、社会的な価値の提供につながっていくと考えられる
- ・データの購入の場合、購入価格の上昇や、競争入札による入手データソースの変更などのリスクに備える必要

(3) ビッグデータを扱う人材・体制面

- ・ビッグデータを扱うスキルや知識を有する人材の育成・確保が社会全体として求められている中で、ビッグデータの活用を担う体制を整備していく必要
 - ー業務の内容・性質や、必要とされるハードウェア・ソフトウェア環境、及び人的リソースを考慮し、外部事業者との間で役割分担・委託を行い業務運営
 - ースキル・ノウハウが検討や開発業務を行った担当者に集中することがあるため、組織的な蓄積を図る必要
 - ーデータサイエンティストを職員として採用
 - ーデータホルダーと連携し、職員対象のセミナーの開催など人材育成に取り組み
 - ー精度検証の結果、精度が高く出せる部分とそうでない部分について整理し手引書を作成

(4) システム・技術面

- ・ビッグデータを適時適切に扱うためのシステム等を整備する必要
 - ーセキュリティを確保しつつ大量のデータを扱うための環境を構築
 - ーデータのフォーマット変換やエラーチェック、分類符号化の実施や、統計作成を実施するためのシステム、プログラム作成

(5) その他（情報保護、法令関係など）

- ・データホルダーにおいて提供可能な範囲でのデータを活用。データ利用に関する規約を策定

4 活用の推進に向けて

- ・新型コロナウイルス対応において、モバイルデータを用いた外出自粛の状況把握が注目されるなど、ビッグデータ活用の動きが進展
- ・一方で、プライバシーや情報保護、ガバナンス等の観点も重要性を増す
- ・社会経済のデジタル化が進む中で、ビッグデータの更なる活用が期待されており、活用における各種課題の解決に向け産官学で一層の連携を図る必要