

第17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
(第8回 統計技術・データソースの多様化等検討会と同時開催)
議事概要

(開催要領)

日時：令和4年3月18日(金) 16:00~18:00

場所：総務省第二庁舎7階中会議室 (Web会議)

(議事次第)

1 開 会

2 議 事

- (1) 自動車の走行データの活用について
- (2) 人流データを用いたテレワーク実施率の検討について
- (3) SDG11.3.1 (人口増加率と土地利用率の比率) の最終報告
- (4) ビッグデータ連携会議の取りまとめについて

3 閉 会

(配布資料)

- 資料1 Traffic Vision プローブデータ&サービス概要
- 資料2 人流データを用いたテレワーク実施率の検討
- 資料3 地球観測データを用いた SDG11.3.1 (人口増加率と土地利用率の比率) の試算・検証について (最終報告)
- 参考資料 第7回統計技術・データソースの多様化等検討会 (第16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 同時開催) 議事概要

(概要)

【(1) 自動車の走行データの活用について】

住友電工システムソリューション株式会社 コネクティッド事業部 事業推進2部 推進課の菅沼氏から、資料1に沿って説明。主なやり取りは以下のとおり。

- プローブデータは東日本大震災の際にも活用されており、いろいろな利用方法があることがよく分かった。例えば、車の所有者の年齢や性別、車のガソリン消費量、

EV 車であればバッテリー残量などは分かるのか。車の位置情報やジャイロセンサーのデータと組み合わせると、例えば OD 表を作成する際に属性別の集計などができるのではないか。

- 現在は、緯度・経度のデータのみ受領し、提供している。
- 他にも把握できる部分はあるが、使っていないということか。
- 今のところ、男女の別など個人に紐付く情報は保有していない。電力消費量、燃料消費量などの情報は分かるが、提供にまでは至っていない。
- EV 車の普及の際に、バッテリー残量などは有効に活用できるのではないか。

- データの網羅性を高めるため、他メーカーの車も含めたプローブデータをメーカー横断で集めることは考えているか。また、自治体でプローブデータを定常的に活用している事例があれば教えてほしい。
- 当社としては HONDA 車のみの扱いしか今のところ想定していないが、例えば大地震などの非常時には、各メーカーで収集したデータを持ち寄って、通行実績情報マップを公表したりしている。また、自治体によっては定常的に利用していただいているところもある。

- 自治体など行政機関の利用が多いとのことだが、民間の活用事例があれば教えてほしい。また、データの提供は自家用車に限定しているとのことだが、特に急ブレーキの話であれば自家用車に限らず営業車も入ってくる。営業車であれば、どこからどこまでどれほどの輸送量があって、どれほどの人が働いているかが分かるので、経済活動の分析にもつながると思う。営業車のプローブデータを使えないのはなぜか。また、自治体でデータを解析する場合、プローブデータならではの注意点等があれば教えてほしい。
- 一番利用していただいているのは自治体の道路の管理などで、民間企業からは問合せはあるが、なかなか利用にまでは至らないのが現状。また、当社のデータは一般の人が使う自家用車に焦点を当てており、営業車のデータは他のメーカーで集めているものがあるので、必要に応じてそれらをミックスして使うことを考えていくのが現実的ではないかと思う。統計に活用する際は、トリップの作り方など、細かい決めごとをしっかりと検討した上で使うのが重要かと思う。

【(2) 人流データを用いたテレワーク実施率の検討について】

厚生労働省 戸田政策企画官から、資料 2 に沿って説明。主なやり取りは以下のとおり。

- 例えば吉祥寺のグラフ (P7) だと、月によって人口自体にばらつきがあるが、深夜であっても月によって異なる。これはそもそもサンプルとしている人口が違うのか、

それとも他の要因があるのか。また、その差はテレワークの比率を出すときに調整しているのか。

→ 100メートル四方のメッシュなので、新しい住宅地ができたり、引っ越しをする人が多かったりということが反映されているのかもしれないが、検証の余地はある。また、月ごとのブレについては、大手町や霞が関などのオフィス街に関してはそれなりに安定している。目視で確認して、データとして問題なさそうだとすることで、特段処理はしていない。

○ 人流データは、コロナで非常に活躍しているデータ。このテレワーク率の分析は、地域間の比較ということが一つあるが、最後のアンケートなどを見ると、業種別でテレワーク率が違うということも考えられる。それを人流データで調べる場合、ある業種のビルが立っている土地を全てまとめた上で、それらの土地の通常の昼間人口と夜間人口の変動、またコロナ禍での変動を比べることによって、ある特定の業種が入っている土地でのテレワーク率は高くなっている、低くなっているという結果が出る。その際、データを建物を含んだ広い範囲のメッシュという単位ではなく、一つ一つそれらの土地に訪れた点で見るとは個人情報の問題もあり難しいが、例えば病院等の土地に訪れた人数を合計して見ることは可能である。データベンダーにデータをカスタマイズしてもらうのは難しいのか。公的統計で使っていくために、データをどの程度自由に扱えるのか。

→ ご指摘の点は重要なポイントだと思う。推測になるが、おそらく携帯電話のユーザー情報などにより、地域間の区切りをより正確にする技術はあると思う。ベンダーとコミュニケーションを取りながら進めていくことは今後のチャレンジングな課題と認識している。

○ 民間事業者からすると、データは存在しているがそれらがつながっていない状態になっていることが多い。携帯電話だと、キャリアが持っているユーザーアンケートの情報や、申込時の年齢や属性といった情報があるが、それらを関連付けて出すと、個人情報の問題なども発生する。あるいはベンダーにそういう情報を出せるのかというと、民間側が規制をかけて出さないという場合もある。

○ 例えば吉祥寺の方など、外出の目的が分かるとよい。その際の方法として、位置情報データを他の情報と突き合わせることで、分析ができるのではないかと。

【(3) SDG11.3.1 (人口増加率と土地利用率の比率) の最終報告】

国立研究開発法人宇宙航空研究開発機構 (JAXA) 第一宇宙技術部門衛星利用運用センタ

一 技術領域主幹 落合氏、同研究開発員 濱本氏、横浜市立大学大学院データサイエンス研究科 佐藤教授から、資料3に沿って説明。主なやり取りは以下のとおり。

- 2010年代の建物のデータの約98%が正のメッシュ内に入っていたという話について、98%というのはかなり高いが、どう理解すれば良いか。つまり、2010年代にできた建物の部分は元々都市に分類されていなかったが、建物が建つことで新たに都市になったというようなケースがほとんどなのか、分かれば教えてほしい。また、約1,400地点の情報を調べて、そのコンフュージョンマトリックスに基づきLGRを補正したという話があったが、その約1,400地点はどのように選ばれたのか。ここが偏っていると、全ての結果が偏るリスクがあるのではないか。
- 今回活用した火災保険業務データには、それぞれの建物ができた年代が入っている。それを郵便番号で集計して提供してもらっており、3次メッシュと重なっていない部分もあるが、経験上、郵便番号のデータは3次メッシュにほぼ完全に被覆できる。日本中全てというわけではないが、いくつかの都道府県においては、郵便番号のエリアは3次メッシュにほぼ包含されていると考えて特定している。解釈については、2010年代は建物がなかった場所に建物が新しくできたことで、2010年～2015年までのLGRに対して何らかの寄与していると考えた。LGRの評価値がプラスと識別されており、かつそこに建物ができたことをきちんと確認できる、という意味で妥当性があると判断した。実際にLGRを商業的に利用しようとする際は、LGRが正の値となっている場所は非常に商業活動が盛ん、あるいは新たな建物がその時期にできたことと判断できるため、誤った判別にはなりづらく、利用性が高いのではないか。
- 今回は表のみでコンフュージョンマトリックスを紹介しているが、JAXA 高解像度土地利用土地被覆図の検証データは基本的には日本全国からランダムサンプリングした地点になっている。都市部に限らず、都市以外の領域にも万遍なく分布しているデータを用いて検証すると、コンフュージョンマトリックスで示している分類結果になる。
- 新しく建った建物の有無を見ることで、LGRの増減が高い精度で分かることに驚いた。

【(4) ビッグデータ連携会議の取りまとめについて】

総務省統計改革実行推進室 前原参事官補佐から、取りまとめに向けて説明。主な発言は以下のとおり。

- ビッグデータで出てきた指標に対する再現性の確保と、そのための情報公開が重要である。

- データそのものに加え、データの処理方法も可能な限り公開して透明性を保つことが、再現性にもつながる。
- 民間事業者側など、回答者側の利便性・負担軽減も忘れてはならない観点。既存統計の作成においても、既存のデータを使う、なるべく人手を介さずにデータを生成する、ということが大事。
- 報告者負担、業務負担の軽減だけでなく、昨今問題となっている統計データの信頼性の回復のためにも、官民双方において、なるべく自動的にデータが連携していくことが望ましい。

以 上