

## 第6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 日 時 令和元年5月22(水) 15:00～17:00

2 場 所 総務省第二庁舎 6階特別会議室

### 3 出席者

- ・ 構成員 高橋座長、庄司構成員、田原構成員、水野構成員  
佐藤構成員(オブザーバー)
- ・ 審議協力者 経済産業省 大臣官房調査統計グループ 調査分析支援室
- ・ 事務局 総務省政策統括官(統計基準担当)付 統計委員会担当室

### 4 議 題

- (1) メッシュ型流動人口検証WGの結果報告について
- (2) ビッグデータを活用した新指標開発事業(経済産業省の取組み)について

### 5 配付資料

資料1 東京都における流動人口データの有効性の検証

～メッシュ型流動人口検証WG結果報告～

資料2 ビッグデータを活用した新指標開発事業(経済産業省の取組み)について

参考1 第5回BD連携会議 議事概要

### 6 議事概要

#### (1) メッシュ型流動人口検証WGの結果報告について

- 議題1に関して、ワーキンググループ主査である水野構成員から、資料1に基づき説明を行った。
- 携帯端末の位置情報を基に推計した民間企業が保有する流動人口データが国勢調査を補完するものとして有効であるという認識が構成員に共有されるとともに、ワーキングペーパーを総務省ホームページに公開することが承認された。

主な質問・意見は次のとおり。

- 国勢調査を補完したいという狙いであれば、24 時間平均データではなくて時間帯別のデータにして検証した方が良いと思うがどうか。
  - その通りであるが、比較対象としたモバイル空間統計データの時間帯別データが高額であり、十分な予算が確保できなかったため、24 時間平均データで実施した。
  
- GPS データを国勢調査と合わせているのであれば、総数も同じになり、バイアスがなくなるのではないかと考えるがどうか。
  - GPS データは、その夜間の人口を国勢調査で合わせて推計しているが、国勢調査と比較しているのは 24 時間平均であり昼間の人口も含んでいるため、その分の差異が散布図に現れている。また基地局データとの比較では、基地局データは 15 歳未満と 80 歳以上の人口を含んでいないので、GPS データとは総数が異なっていることもバイアスの要因となっている。
  
- 今後の展開として、年に 1 回とか定期的にこのデータの提供をうけて、継続的に取組を進めるような仕組みを考えているか。
  - 定期的な提供には費用がかかるので、データホルダーとの調整が必要になる。また一般に提供するためにはアプリケーションとセットで提供しないと利用してもらえないので、それらの課題を克服してから進めるものと考えている。今年度は地域を全国に拡大して検証を行う予定である。
  
- 全国に拡大して検証するにあたり、東京とどんなところが変わってくると想定しているか。
  - 東京での検証から人口の多いところは精度が良くなっているので、全国においても、都市部の人口が多いところは良いが、過疎地域は精度が落ちるのではないかと考えている。
  
- 基地局データは男女別、年齢別の属性を持っているが、GPS データは持っていないのか。持っていればサンプリングのコントロールが可能と思う。
  - 人口推計に使用できる属性は保有していない。
  
- 人口が少なくバイアスが発生しやすいエリアで、何らかのデータでバイアスを補完して精度をあげていくようなアイデアはないか。
  - 最初はバイアスがあることが分からなかったが、検証した結果、バイアスが場所の属性に依存していることがわかってきた。たとえば駅とか、それぞれの場所の属性情報を整理して、換算係数みたいなものを作って誤差を修正してやれば、より推定値というのはよくなるのではないかと考えている。

- 東京都市部のような大きなところでは、国勢調査を補完できるという結果になった場合に、こんな分析に有用であるというようなアイデアはあるか。
  - ▶ 現時点では具体的なものになっていないので、国勢調査を使用している地方公共団体等の行政機関などにヒアリングを行いたいと考えている。
- GPSデータが国勢調査と結構乖離しており、常住人口では捉えきれていない、ある時点・ある場所に実際にどれだけの人がいるかが分かることがビジネスの観点では価値があると思う。
- データを一般に活用できるようにするためには、データホルダーとなる企業とWin-Winの関係となれるビジネスモデルを構築する必要があるので、今回の取組をその呼び水にしたいと考えている。また今回の検証結果をデータホルダーに提供し、それぞれの流動人口データのブラッシュアップにつなげていってもらえるとありがたいと考えている。
- 今回の結果は総務省の立場で国勢調査を補完するものとして有効かという観点であるが、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良いと思う。そうすることで、例えばデータは購入ではなく、共同研究になったりして、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る。

## (2) ビッグデータを活用した新指標開発事業（経済産業省の取組み）について

- 議題2に関して、経済産業省大臣官房調査統計グループ調査分析支援室から、資料2に基づき説明を行った。

主な質問・意見は次のとおり。

- 基幹統計である商業動態統計調査の丁2調査との比較は実施するのか。また価格、数量に分けたデータの分析は行っているか。
  - ▶ 丁2調査との比較は現在、実施中であり5月24日（金）に公開を予定していますが、調査票の回収率は、全体で74%程度（23社中の17社）であるが、非常に相関が高く、相関係数は1に近いものとなっている。今回の試験調査は、「POS等ビッグデータを活用するといった新たな調査方法」の採用とその調査事務プロセスの検証や、「報告者負担の軽減化」、「統計業務の効率化」、「公表の早期化」の他、「景気動向把握の向上に資するための把握内容の詳細化」等の実現可能性などの精査に必要な基礎資料を得ることが主な目的であったため、価格、数量に関する分析は行っていない。

- 調査スキームの図では民間事業者から提出された調査票を報告者である家電大型専門店が日次でチェックすることになっているが、報告者の負担になっていないのか。
  - 報告者から「毎日調査票の内容を確認するのは作業量的に負担がある」という意見があるとともに、「民間事業者が POS データを受信した段階である程度の審査は行っているので、この確認自体が不要ではないか」という意見もあり、調査スキームを更に精査していく必要があると考えている。
  - 毎日の調査票の確認に効果があったかどうか、すなわち調査票の修正が発生していたかどうかを検証すべきと考えるが、どうか。
  - 2015 年 1 月 1 日まで遡って POS データを集めたが、報告者から民間事業者に送信された POS データが要因となった調査票情報の差し替えはなかった。
  
- 仲介している民間事業者は経済産業省から委託を受けているが、もともと報告者とのビジネス等での関係があるのか。
  - 民間事業者は、ビジネスとして、家電量販からの POS データを基にマーケティングデータを作成し、メーカーに提供している。
  - 中小企業庁では、支援団体が集めているデータをそのまま EBPM に使えるように統計化して、助成金や政策の精度向上につなげようとしている。国が何らかの形で関与するなどして、民間事業者から家電量販店に有用な情報をフィードバックすることにより、Win-Win の関係が構築できるようなスキームは描けないか。
  - 現在でも民間事業者が彼らのビジネスにおいて報告者である家電量販店に情報のフィードバックを行っているが、新しいテレビがいくらで何台売れたかなどミクロな情報であるので、統計調査の結果などマクロ情報の提供は喜ばれるかもしれない。一方、統計制度上、調査結果を一般公開するより前に、報告者に情報を返すことができない。また、報告者にとっては、家電量販店の動向が、IR を公開するより前に公表されては困るという事情があり、その整合をとるのは難しいと考えている。
  
- SNS への書き込みログを活用した新指標開発で、なぜ景気ウォッチャー調査を使ったかとか、AI を統計調査に活用するにあたって難しかったところとか、議論したことがあれば教えて欲しい。
  - 教師データとしては「景気ウォッチャー調査（内閣府）」の「景気判断理由集」の他に、「中小企業景況調査（中小企業基盤整備機構）」の「生声」がある。どちらの情報も、「判断」に対応する「その理由」が膨大なデータとして収集されています。これらの情報を機械学習させ、ツイッターからデータ抽出しポジネガ判定できる AI ロボットを作り、指標の作成を行った。当初の目的はそれぞれの指標を当てての事を目指したが、結果的に相関はあまり高くないが、山谷を見るのであれば使えるという結果となった。これとは別で鉱工業生産指数も残業というキーワードから予測を行っている

が、こちらはもう少し精度の良いものになっている。

- ▶ 鉱工業生産指数は企業の回答を基にした製造工業生産予測指数が既に経済産業省から出ているが、パフォーマンスはどうか。
- ▶ こちらは既に出ている予測指数に SNS 情報や為替、株価などの情報を追加したものであるため、精度はより高くなっていると思う。

○ 今回の試験調査票の回収率が 74%ということであるが、もともとの丁 2 調査と比べてどうか。

- ▶ 丁 2 調査の回収率は 100%。これに比べ、試験調査が 74%と低かった主な理由は、報告者側で POS データという詳細情報を外部に提供することの是非や、試験調査期間が短く、経産省側で取組の意義等を十分伝えられなかったことだと考えている。
- ▶ 民間事業者はビジネスとして POS データを集めているが、集めているデータの一部しか使用できなかったということか。
- ▶ ビジネスとしての民間事業者と家電量販店の契約では、データは第三者には開示できないものとなっている。今回の取組にあたり開示の許可が出たものが 74%であったということである。

○ 景況調査等の AI 分析で自然言語処理を使っていると思うが、毎年、より良い新しいモデルが出てきているが、モデルの更新は行っているか。

- ▶ 契約が切れており、フォローはできていないので推測になるが、「中小企業景況調査」では新しいデータは一切学習させられないのでモデルは古いままで、「景気ウォッチャー」の方は、毎月内閣府で公開されているので、学習し直そうと思えばできると思う。一方、POS 家電量販店動向指標に関しては、POS データのカバレッジが高く、日々収集した POS データを 6 分類に振り分け、集計するものなので、モデルを作る必要がなかった。

○ 2017 年度の POS 家電量販店動向指標では E C 販売が含まれないとなっているが、今後の方向として、モデルで補っていくものなのか、それとも E C を捉える方向で進めるのか。

- ▶ これは 2017 年度に行った試験的なプロトタイプの指標開発なので E C 販売を含めていなかったが、2018 年度の試験調査は E C も含めて集計をしており、E C 化率は勿論、E C 販売動向を捉えることも可能となっている。今後についても、E C の販売動向を正確に捉えることは大事なポイントだと考えている。

以上