

## 第12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3年3月10(水) 15:00～17:00

2 場所 総務省第二庁舎 6階特別会議室 (Web 会議)

### 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  
清家構成員 (オブザーバー)
- ・ 審議協力者 日本銀行 須合企画役  
横浜市立大学 佐藤教授  
東京大学 竹内教授  
JAXA 石田特任担当役
- ・ 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) 日本銀行によるビッグデータの利活用紹介
- (2) 観測データ利活用検証 WG の最終報告

### 5 配付資料

資料1 位置情報データによる経済活動のナウキャストニング

資料2 -1 地球観測データを用いた SDG15.4.2 (山地植生被覆指数) の試算について  
(最終報告)

-2 SDG15.4.2 (山地植生被覆指数) 評価結果報告

-3 SDG 指標 15.4.2 に関する今後の対応について (案)

参考資料 第11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

### 6 議事概要

- (1) 位置情報データによる経済活動のナウキャストニング

概要は以下の通り。

- 日本銀行 須合企画役より資料1の説明が行われた。

(参考:日本銀行HP掲載「位置情報データによる経済活動のナウキャストイング」

[https://www.boj.or.jp/research/wps\\_rev/wps\\_2021/data/wp21j02.pdf](https://www.boj.or.jp/research/wps_rev/wps_2021/data/wp21j02.pdf))

主な質問・意見は次のとおり。

- ショッピングセンターをみると、2020年3月は滞在人口が減った以上に売り上げが落ちているが、2020年3月以降も今までと同じ傾向なのか、それとも傾向が変わったのかどうかをお教えいただきたい。また工場でもテレワークが可能な場合には工場に人がいなくても生産はあまり落ちないと考えられるが、滞在人口と経済状況がどのくらい乖離してきたのか教えていただきたい。
  - ▶ 今回は2020年3月までのデータを使って分析をしており、それ以降のデータを用いた検証は出来ていない。仮に感染症下で個人の行動様式が変わって、オンライン消費にシフトした場合、これまで見られていたような相関関係は変わっている可能性はある。工場での生産活動も、リモートでの制御が出来るようになってくれば関係性が崩れると思うが、今はそこまで大きくシフトしてはいないという感覚であり、今後蓄積されるデータで検証していく必要があると思っている。
  
- 製造業と鉱工業生産の時間の絞り込みについて、例えば「食料品・たばこ工業」だと22時～23時となっていて、その他の業種だと遅い時間や早い時間や時間の長さも色々あるが、この時間帯(22時～23時)で相関が強いのであれば、もう少し早い時間帯から広め(ex:7時～23時)にとっても良いと思うが、その点をどのように考えたか。
  - ▶ 「食料品・たばこ工業」における22時～23時という絞り込みの結果は、前年比との相関が最も高い時間帯をピックアップしているものの、相関係数は他の業種と比べると低く、当てはまりがあまりよくないという印象を持っている。
  
- 労働投入を測るときは残業も捉えた方が良いので、24時間全体見るのがいいと思ったが、実際は時間を絞った方が精度は高くなると理解した。もっと情報を利用しようと思えば、時間帯毎の人流を生産の説明変数としてモデルを作るようなアプローチもあり得るかと思うが、そういう方向は考えることができるのか。
  - ▶ 夜間の残業時間帯についても時間帯別に全ての滞在人口を積み上げて検証を行った上で、当てはまりがあまり良くなかったため、今回のアプローチとなった。その要因を考えてみると、例えばある工場を考えたときに、昼は生産活動を行っているが、夜は設備メンテナンスを行っているケースなどがありうる。ナウキャストイングする際に、全ての情報が必ずしもベストな情報となるわ

けではなく、逆にノイズになる可能性があることも分かった。結局今回のような絞り込みをする事によって、統計的に見たときによりナウキャストイングに適しているものを抽出できていると考えている。

- 前半パートのサービスで、クラスタリングを行っていたと思うが、何の変数でクラスタリングをされたのか教えていただきたい。
  - ▶ ここでのクラスタリングは、時間帯ごとにみた滞在人口の時系列データを用いている。この滞在人口の推移が似ているメッシュが、同じクラスターに分類されることになる。
  
- 外食と居酒屋とありましたが、外食の相関が高いクラスター（クラスター1）と居酒屋と相関の高いクラスターは同じであったのか教えていただきたい。
  - ▶ クラスター1に分類されるメッシュが一番多く、全部（758メッシュ）のうち半分以上（435メッシュ）がクラスター1に含まれている。これは、居酒屋でも同じであり、クラスター1との相関が高い。
  
- 今回は位置情報データで消費や生産活動を見ているが、それ以外の経済活動で、位置情報データを用いて捉えることができるものはないか。
  - ▶ 論文で紹介した今回の手法を応用できるものとして、宿泊などが可能であると思うが、今回は検証を行っていない。
  
- 他のビッグデータと組み合わせて分析することに関する意見をお聞きしたい。
  - ▶ 1つのビッグデータだけでマクロの経済情勢を分析することはできないため、色々な組み合わせを行うことは有効だと考える。他方、ビッグデータを組み合わせれば行くほど、ノイズも多く含まれるため、そのバランスを考慮して分析していく必要があると思う。
  
- 欧米の中央銀行でこのようなビッグデータを活用した経済活動の把握はなされているのか教えていただきたい。
  - ▶ 欧米の中央銀行についても新型コロナウイルスの拡大以降、高頻度データ、または、ビッグデータを利用する分析は行われていると思うが、今回の我々の分析のような形で位置情報データを利用した分析は存在しない。米国では大型ハリケーンの直撃などにより消費が落ち込んだときの状況を、クレジットカードのデータを使ってナウキャストイングする取組みが既に行われてきたことが、位置情報データを利用していないことの背景かもしれない。

- 滞在人口の推移のグラフの表が幾つかあり、時々2種類の折れ線グラフが大きく乖離しているところがあるが、何がそうさせたのか、知見や仮説があれば伺いたい。
  - ▶ 例えばショッピングセンターの2019年9月の乖離は、同年10月の消費税率引き上げの影響を受けたものと推察している。一人当たりの消費額がコンスタントであれば来店客数と売上高は線形の関係になるが、9月は一人当たりの消費額が上がり、10月は下がったことにより、この乖離が生まれたものと考えられる。
  
- ショッピングセンターのグラフで、コロナ後の売上高の落ち込みが滞在人口の落ち込みよりも大きく、一人当たりの売上高が落ち込んでいることになるが、一般には家飲みや家庭内消費が増えたと言われており一人当たりの消費は増えていると思われる。この乖離についてどのように分析をしているのかお教えいただきたい。
  - ▶ 消費の行動様式がどう変わっているのかという質問であると認識しているが、本分析では2020年3月までのデータしか用いていないので確たることは申し上げられない。2020年4月以降のデータを見ながら個人の消費行動が変容したかどうかを確認する必要がある。

## (2) 観測データ利活用検証WGの最終報告について

概要は以下の通り。

- JAXA 特任担当役石田さまより資料 2-1、横浜市立大学佐藤教授より資料 2-2、総務省 政策統括官（統計基準担当）室国際統計管理官室津村管理官から資料 2-3 の説明が行われた。
- ワーキングペーパーに関してはワーキングの主査である水野構成員に一任されることが承認された。

主な質問・意見は次のとおり。

- 被覆分類の JAXA のデータや ESA のデータで分類のモデルというものはそもそもどのようなものか。また、ESA のデータよりも JAXA のデータを利用する方が、精度が高いとのことであるが、FAO の ESA のコンフュージョンマトリクスを書くと、精度が低いものになるということの意味するのか。

分類のモデルですが基本的には光学センサーの場合には違った波長域、すなわち可視、近赤外の領域から赤外域まで 10 バンドとか、ESA の場合ですと 20 バンド以上に波長帯域を持っており、各波長域での地表面からの反射の強さを 10 次元または 20 次元の空間のどこに位置するかでその分類をしている。そのため、使用している波長域やそのデータのビット数が多ければ多いほど細かい分類が出来ることになる。原理は、波長域における反射強度の違いから分類していることになる。分類精度としては、JAXA が 10 分類で 84% の分類精度、ESA は 20 数分類で 71% と若干低い分類精度になっている。しかし、植生・非植生という 2 分類にした場合は飛躍的に精度が上がるがその結果は確認していない。

以上