

第5回 統計技術・データソースの多様化等検討会  
(第14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と同時開催)  
議事概要

---

(開催要領)

日時：令和3年9月29日(水) 15:00~17:00

場所：オンライン開催

(議事次第)

1 開 会

2 議 事

- (1) 「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」及び「統計技術・データソースの多様化等検討会」の同時開催について
- (2) オルタナティブデータ活用の黎明期を超えるために：AI経済センシングセンター構想について
- (3) 農業統計における人工衛星データの利活用について
- (4) 第4回観測データ利活用検証WGの報告
- (5) ビッグデータ等の活用に係る今後の検討の方向性について

3 閉 会

(配布資料)

資料1 「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」及び「統計技術・データソースの多様化等検討会」の同時開催について

資料1の参考1 「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」の開催について

資料1の参考2 統計技術・データソースの多様化等検討会の開催について

資料2 オルタナティブデータ活用の黎明期を超えるために：AI経済センシングセンター構想

資料3 農業統計における人工衛星データの利活用について

資料4 第4回観測データ利活用検証WGの報告

資料5 ビッグデータ等の活用に係る今後の検討の方向性について

資料5の参考 公的統計へのビッグデータの更なる活用に向けて -ビッグデ

一タ連携会議におけるこれまでの事例整理-

参考資料 第13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

---

(概要)

【(1)「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」  
及び「統計技術・データソースの多様化等検討会」の同時開催について】

事務局から、資料1に沿って説明。

【(2) オルタナティブデータ活用の黎明期を超えるために：AI経済センシング  
センター構想について】

東京大学大学院の和泉構成員から、資料2に沿って説明。主な発言は以下の  
とおり。

- 自然言語処理分野では、評価指標やタスク設定に基づいて競い合っているため分野として精度が上がっていることが分かるが、金融業界は何をしているのかが分かりづらい。和泉構成員が研究している「因果」をどう評価するのかがこれから重要になると思っているが、分野の中での評価尺度、学会における共通タスクの設定などの動きは現在あるのか。
- 評価については、まさに立ち上げているところ。実際に経済的な因果が正しいのかどうかについては、ベンチマークのデータセットを2,000個ほど作り、そのうち何パーセントが正しいものだったのかを調べていくことを考えており、こうした取組が業界の発展に必要。ただ、これが自然言語処理の研究として高い精度を持つということと、実際に役に立つのかということは別であり、実際にどんなパフォーマンスが出たのかをプラスアルファで検証していく必要がある。また、現時点では自分たちの分析したデータでうまくいった、という個別のケースがあるだけで、共通のタスクなどはない。分野自体が黎明期ということでは理解してもらえればと思う。
- データセットは、先行研究を集めて作っているのか。また、データ間の因果関係は時期や場所などの条件によって異なってくると思うが、それはどう記述するのか。解析に横串を通すという話で言うと、基本的な解析方法を提示することはできるが、ローデータ（生のデータ）を解析できるように処理するところが大変ではないかと思った。また、レシピを公開・共有することで複数のサービス間におけるデータ構造のハーモ

ナイズにつながる可能性もあると感じた。

出来合いのものというのと、例えばGPSのデータでグーグルが提供しているモバイルインデックスなどが相当するイメージか。実はGPSではいろいろなことができるのだが、例えばこんなことができる、というのを提示するということか。

また、カタログという意味では、それぞれのデータがいつ、どのように作られたかをしっかり記録することには意義がある。データの正当性をある程度担保することで使いやすくすることも重要ではないか。

→ 評価用のデータセットは、特に英語の方は、既に因果関係があるかないかのタグ付けをされた公開データを使う手もある。それだけだと分野も広い上、目的にあったものかどうか分からないので、データ自体も手で作っているという状況。

因果関係については、例えばオリンピックの影響を調べる時、2016年のテキストデータと2020年のテキストデータを使った解析では全く違った結果が出る。むしろ、普遍的な因果は存在しないので、分析の目的によって関連性や構造は変わってもいいのではないかと思っている。また、解析手法の前段階のデータ処理が大変というのもそのとおりだが、情報ベンダーからデータを公開する際、一緒に分析コードも出そうとしている。ベンダーと契約し、前処理は終わった形で提供していくのが目標。

御指摘のとおり、GPSのモバイルインデックスなどもかなりイメージは近い。ただ、一番違うのは、例えば普段は混ぜないようなGPSとクレジットカードのデータを併せて使うことで、ある時期の経済指標との相関関係が分かるというような、データ融合の面が前面に出されているところ。データを複合させることでこういうことができる、という情報をカタログの中にどんどん出していきたい。

- 整理されていないオルタナティブデータも、今までよく使われているデータと組み合わせることでバラバラと思っていたものに見通しが立つということもあるかと思うが、今の共同研究グループでそういった議論はあるか。
- まさに共同しているグループで、欠損データの補完方法やデータの信頼性を統計的に研究している。マクロ経済の研究者に指標の作成グループに入ってもらい、時系列的に足りないデータを補ってもらうなど、体制も拡大している。また、信頼性の低いオルタナティブデータを別のオルタナテ

ィブデータで更に代替するといった技術も開発しているところ。

【(3) 農業統計における人工衛星データの利活用について】

農水省・木村統計企画管理官から、資料3に沿って説明。主な発言は以下のとおり。

- ポリゴンデータの提供には非常に手間がかかると思うが、どこまで有効に使われているのか教えてほしい。また、オープンデータでよくある話として、民間企業で自社のデータと重ね合わせる際にこのデータが合わない、足りないなどの要望が出てくるのではないかと思うが、どのようなものがあるのか教えてほしい。
- 農業関係団体等からは、農地管理、生産管理に活用しているという話は聞いている。昔ながらのやり方よりは進んでいるという認識はあり、オープンデータを提供した価値はあると思っている。ただ、具体的にどの程度まで活用しているのかは分からない。コストに見合うかどうかというよりは、できる限り有効に、より多くの方に利用してほしいと考えている。  
また、データ自体はHPに載せて公開する形であるため、明確にどのような要望があるかは認識していない。ただし、耳に届いていないだけだと思うので、利用者側の意見をどのような形で吸い上げていくかは考えたい。
- この件は特に日本全国カバーしなければならないこともあり、衛星データは使うのにかなりコストがかかるのではないかと思うが、予算的にはどれくらいの規模になるのか。
- 一般的な調査研究として業者に委託契約して行う程度の金額。実際に衛星データを利活用する際には、産総研やJAXAなどと協定を結び、協力関係を築いた上で進めているが、公的性格が強いので協力が得られやすいこともあり、調整コストという意味では民間企業より若干低いと思う。
- 筆ポリゴンデータは、農水省で実施している別の統計調査との関連はあるのか。人工衛星データを使って作柄を予測することが可能になれば、都道府県よりも細かい地域についても把握できると思う。
- 農地の区画情報は衛星画像のない時代から、母集団情報としていかに効率的に整備できるのか、これまで段階的に議論してきた。オープンデータは、民間を含む農業の振興にもつながるのではないかというところ

も含め、趣旨としては後からついてきたもの。作柄予測は、データ分析技術が進めばよりきめ細かい分析が可能になるのではないかと考えている。質の改善に向け、いただいた御意見を役立てていきたい。

【(4) 第4回観測データ利活用検証WGの報告】

国立情報学研究所 水野構成員から、資料4に沿って説明。

【(5) ビッグデータ等の活用に係る今後の検討の方向性について】

事務局から、資料5に沿って説明。

主な発言は以下のとおり。

- デジタル庁が発足したこととの関係は何かあるか。デジタル庁が設置したデータ戦略推進WG（これまでのデータ戦略タスクフォース）などの会議はこちらにも当然関係してくると思うが、現時点では特に密接に関係しているとは感じていない。タスクフォースでも産業界の分野をまたいだデータ連携の取組の話も出ているので、もっと絡むといいのではないか。
- IT室時代には情報共有をしていたが、御指摘のとおり、現時点でデジタル庁と本格的な連携はできていない。今後、デジタル庁の取組とも更に連携を深めていきたい。
- （タスクフォースの）メンバーとして、ビッグデータ連携会議は大変いいモデルになる議論をしていると思うので、是非アピールしてほしい。
- オルタナティブデータを活用した経済分析に関しては、本年度から「オルタナティブデータ推進協議会」や付随するサブの委員会が立ち上がり、レギュレーションや人材育成についてかなり近い議論を行っている。そちらは金融分析が中心になるが、海外の事例も含めて情報共有すると良いかと思う。
- 貴重な情報をいただき感謝。今後、情報収集していきたいと思う。
- モノの平均単価を出すような統計、スーパー業界の売上遷移等の統計

をそれぞれ別の省庁で作成しているが、民間から出すのは基本的に全部POSデータだと思う。一度データを出せば、それが各統計に利用される、ということになると、企業側の手間も減り、データを出すインセンティブにもつながる。省庁間の統計の連携も検討していくと良いのではないか。

→ 省庁間の連携については課題も多いが、御指摘を踏まえ何ができるか検討していく。

以 上