

## 第 8 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 11 月 13 (水) 10:00～12:00

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 総務省統計局 統計調査部消費統計課物価統計室
- ・ 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) 消費者物価指数(CPI)へのウェブスクレイピングの活用について
- (2) その他

### 5 配付資料

資料1 消費者物価指数(CPI)へのウェブスクレイピングの活用について

参考1 これまでにビッグデータ連携会議で取り上げた事例

参考2 第 7 回 B D 連携会議 議事概要

### 6 議事概要

#### (1) 消費者物価指数(CPI)へのウェブスクレイピングの活用について

- 議題 1 に関して、総務省統計局統計調査部消費統計課の山形物価統計室長より、資料 1 に基づき説明を行った。

概要は以下の通り。

- 近年のネット販売の増加や、ネット情報の収集技術の進展を背景に、ウェブスクレイピングの有用性が拡大しており、外国パック旅行費・運賃・宿泊料に関しては対応できる目処が立った。
- ウェブスクレイピングにより、ネット販売価格の把握拡大、膨大な量の価格デー

タによる統計精度の向上、報告者負担軽減・業務効率化が期待できるが、その活用拡大に当たっては、品目ごとのネット販売の進展状況や、ネット情報のデータ特性などに照らし、費用対効果を慎重に見極めていくことが必要

主な質問・意見は次のとおり。

- ウェブスクレイピングを採用することにより、業務負担はどうなると評価しているか。(構成員)
  - 検討段階では今までの統計調査と異なる新たな知識が必要であり業務負担は大きいですが、運用段階に入れば調整先が都道府県から事業者に変わるだけであり、統計作成プロセスは従来と同様にシステムに組み込まれるので、業務負担は大きく変わらないと考える。(物価統計室)
  
- 新技術の採用で統計精度の改善がメリットになると思うが、統計数値の変化にはどんなものが想定されるか。(構成員)
  - 航空運賃については、現在もデータはウェブサイトから取得しているため基本的に価格に変化は無いが、今後は75日前・45日前などと段階的なきめ細かい割引運賃の把握、及び路線数の10から15への増加というカバレッジの増加により、精度向上が見込めると考えている。一方で宿泊料については、データの取得頻度がこれまでの月2日間から毎日になり圧倒的にカバレッジが増すので、カレンダー要因による価格高騰などの影響を受けにくい精度の高い統計数値が作れると考えている。(物価統計室)
  
- Amazonや楽天等のECサイトのデータの使用についてはどう考えているか。(構成員)
  - CPIでは衣料品や食料品など品目毎の性質に着目して検討している。現在活用しているPOSデータにはAmazonのデータも入っており、ECサイトは複数あるサイトのデータの一つという位置づけになっている。(物価統計室)
  
- ウェブサイトはメニュー作成にかかるコストが低いので、メニュー価格の入れ替わりが激しく、特に低価格の情報が少なくなり、物価に影響を及ぼさないか。(構成員)
  - 外国パック旅行を例にすると、直前ではなくて余裕を持って半年以上前から価格情報を取得しているので、情報がなくなることも考慮して適切な価格を算出するよう工夫している。また売り切れた場合でも価格情報だけは載せているサイトも多い。(物価統計室)
  
- 日本のお正月や中国の春節の時期など、そのイベントに特化した商品しか存在せず、同じ商品の価格の動きが把握できなくなるという状況になると思うが、そん

などきのデータ収集や分析の苦勞・工夫があれば教えて欲しい。(構成員)

▶ 特定の季節に特殊なものしかないという商品の情報を前もって事業者から把握しているのので、例えば航空会社やホテルのグレードがどうなっているかといったそれと同じ品質の商品の価格を継続的に捉えることは可能だという感触を持っている。(物価統計室)

○ インターネットでは同じ在庫を複数の EC サイトでシェアできるので、シェアしているものとしていない商品を区別しないと適切な価格を出すのが難しいと思うが、どのように対応しているか。(構成員)

▶ CPI は取得した価格を単純平均しているのではなく、対象商品のシェアや販売数量を事業者や家計調査等で取得して重み付けをすることにより、適切な価格を算出している。また価格はまず日単位での平均価格を算出するなど、できるだけその影響を排除するように工夫している。(物価統計室)

○ ネット購入割合はここ数年で加速度的に増加しており、その基礎となるデータが 2014 年というのは古いのではないか。(構成員)

▶ EC 化率に関して、物販系については経済産業省が作っている 2018 年のデータがあるが、サービス系分野については同じデータがないため、航空運賃のネット購入率に関しては 5 年に 1 回の全国消費実態調査の情報を掲載している。(物価統計室)

○ 事業者との調整がかなり必要であるということだが、事業者の持つデータベースを直接参照させてもらうとか、実際の販売数量まで提供いただくとか、より実効的な方法はとれないのか。(構成員)

▶ 主流となる商品のトレンドやシェア等のタイムリーな情報に関しては、実際の生データではなく必要かつ提供可能なものを事業者に選別してもらっている。データベースそのものを直接参照させてもらう段階に至るには、決め毎・ルールなどの事業者間の更なる折衝や文書の取り交わしなどが必要なので、今後の検討課題だと考えている。(物価統計室)

○ 若い学生はインスタグラム等の写真を見ながらイメージで商品を選んでいるため、事業者側も画像情報を充実させているので、AI を使った画像検索などで情報の収集や整理をしてみたらどうか。(構成員)

▶ 例えばオランダで AI を使った衣料品指数の作成を始めており、我々も今年度からそうした研究に着手したところである。価格収集の段階で AI を使った画像検索を使えば効率的なので検討していく価値はあると思っているが、予算等の都合もあるので、段階的に進めていきたいと考えている。(物価統計室)

○ 今回のケースにおいて、実運用において民間リソースの活用を行うとしたとして

も、統計で求められる調査の継続性などの観点から、事業者の選定などで課題が生じるようなことはないと考えて良いか。(構成員)

▶ ウェブスクレイピングに限った話ではないが、事業者の選定等に関しては、普通に入札をかけてやることになるため、それを遂行できるという仕様をきちんと設けるということに尽きる。継続性については、国庫債務負担行為という仕組みで、単年度ではなくて長期的に予算を確保するなどの取組の必要があると考えている。(物価統計室)

○ 宿泊料については予約時期別の価格を調べているが、これは従来から行っていたのか。また重み付けのための販売額シェアなどはどうやって取得しているのか。(構成員)

▶ 今までは予約時期別の情報は把握しておらず、調査時点の価格を把握しているのみであった。また、旅行サービスに関するウェイト情報について、今までは主に5年に1回の消費者物価指数の基準改定時に大まかなデータを各事業者から提供を受けていたが、今回はチケットの種類ごとの購入割合などの詳細な情報の提供を受けている。(物価統計室)

○ 8ページの航空運賃のグラフでは試算値が公表値をずっと上回っており、「期間平均=100」になっておらず、過去を基準としたバイアスが現れているように見えるが、いかがか。(構成員)

▶ 表示している2018年1月～9月の期間平均を100とした数値を設定しているはずであるが、確認してみる。(物価統計室)  
→会議後、2018年1月を100とした数値であることがわかったため、期間平均を100としたものに修正してHP掲載している。

○ この数値は成約価格ではないので、同じ在庫の商品がそれぞれのECサイトで異なる価格で販売される場合には、平均化された価格になると考えて良いか。(傍聴者)

▶ ご理解の通りであり、価格が異なる場合には両方の価格を把握し、シェア等で重み付けをした上で最終的な価格として設定している。(物価統計室)

○ POSデータを入手するには、POSデータの販売事業者から購入する方法と、それぞれの小売事業者と交渉して自ら保有するデータを提供してもらうという2通りがあるが、広範囲に民間企業に働きかけて、後者の方法で集めていこうというビジョンを持っているか。(傍聴者)

▶ ヨーロッパの小さい国では、大企業の何社かのPOSデータを押さえれば十分なので、個別企業から入手しているところがあると推測する。日本でもその可能性を排除しているわけではないが、まずはPOSデータの販売事業者からの入手を進めていくのが現時点でのスタンスである。(物価統計室)

- 今回の外国旅行パックでは、全体で 217 万個の価格に対して、ノイズ処理等を行った結果として 22 万個の価格に絞り込んで最終的な価格を算出しているが、この後の実運用では改めて全体から抽出して算出するのか、それとも絞り込んだものを対象に価格を算出するのか。(構成員)
  - ▶ データを前もって取得しておくことにより、商品の変更や価格情報の消失に対応できるメリットがあるので、スクレイピングを許可されている時間に制約はあるが、全体から抽出していく予定である。(物価統計室)
  
- 諸外国と比較して、今回の取組は進んでいるのか。(構成員)
  - ▶ いろいろな評価があると思うが、個人的には、諸外国の中でも進んでいるヨーロッパの小さい国にも、それほど引けを取らないレベルではないかと推測している。(物価統計室)

## (2) その他

- 総務省統計委員会担当室より、参考資料 1 に基づき、これまで連携会議で採り上げてきた各府省におけるビッグデータの利活用に関する先行的事例と、それぞれの事例での活用のメリットや課題などについて説明した。

以上