

消費者物価指数（CPI）における ビッグデータの利活用案（イメージ）

令和3年12月13日
物価統計室

1. 2020年基準改定の概要

- 消費者物価指数は、消費構造の変化に合わせ、5年毎に基準年（指数を100とする年）を改定
- **本年8月、2015年基準（2015年=100）から2020年基準（2020年=100）に切替え**（前年同月比）

	2021年1月分	2月分	3月分	4月分	5月分	6月分	7月分	...
2015年基準 (2015年=100)	公表済	公表済	公表済	公表済	公表済	公表済		
2020年基準 (2020年=100)	8/6 公表（遡及改定） <small>（統計利用者の参考のためウエイト等の基礎情報を7/9に公表）</small>						8/20公表	以降 毎月公表

【基準改定の主な内容】

品目の改定

👉 **消費の重要度の変化に応じて品目を改定**（585品目→582品目^{注1)}）

追加品目 カット野菜、シリアル、宅配水、ドライブレコーダー、葬儀料 など（計30品目）

廃止品目 グレープフルーツ、固定電話機、辞書、ビデオカメラ、幼稚園保育料 など（計28品目）

ウエイトの改定

👉 **品目別ウエイト**（消費支出全体に占める各品目の消費支出の割合）を**2015年から2020年^{注2)}に更新**

ウエイト増加品目 調理食品、ルームエアコン、プリンタ、携帯電話通信料 など

ウエイト減少品目 外食、ガソリン、宿泊料、外国パック旅行費など

算式の改定

👉 **多様な料金体系からなる品目について、算入するモデルケース**（典型的な利用パターン）等を改定

例：携帯電話通信料の場合

新モデルケース 事業者別に8パターンの利用モデルケース：
通信量（**低・中1・中2・高**）×通話時間（低・高）

現モデルケース 事業者別に9パターンの利用モデルケース：
通信量（**低・中・高**）×通話時間（低・中・高）

⇒ より実態を反映できるようにパターン区分を改定

注1) 統合品目（10品目→5品目）がある。

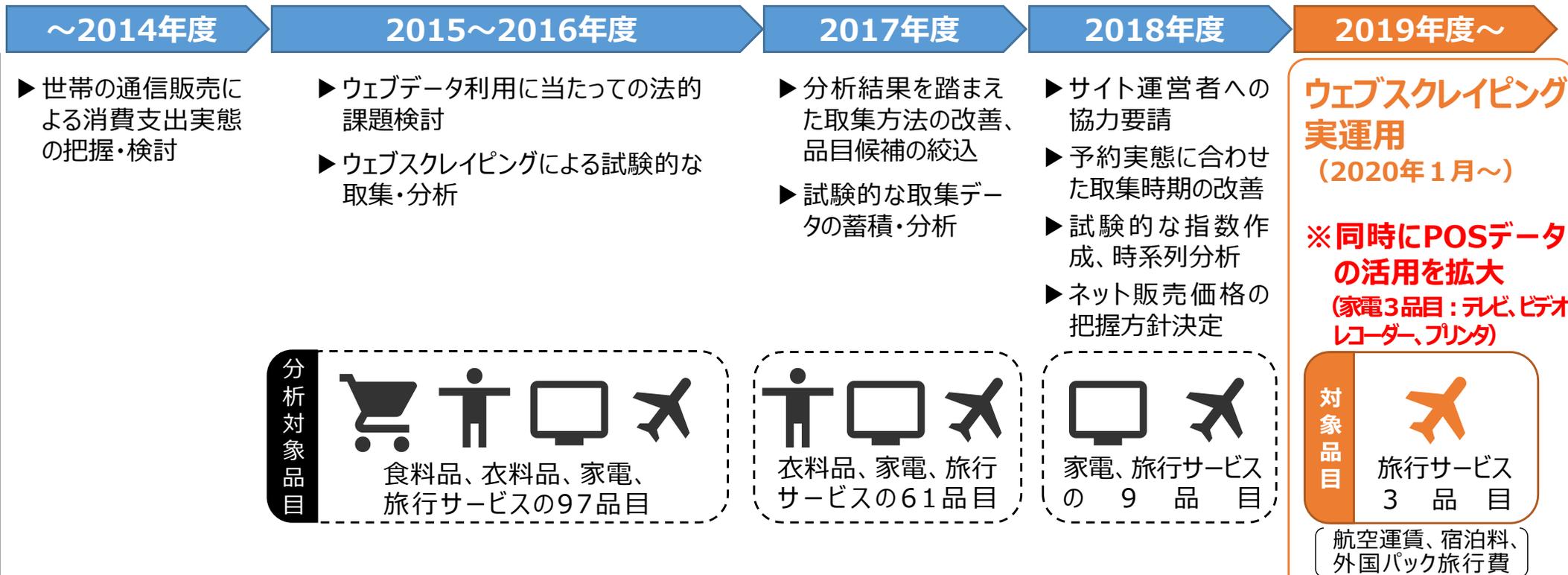
注2) 新型コロナウイルス感染症による急激な変化を緩和するため、2019年・2020年の平均消費支出を用いて作成

2. 取組の概要

ウェブスクレイピングを中心としたこれまでの経緯

主な取組経緯

背景となった政府決定等



**統計委
答 申**
(2012年1月)

☞ 全国物価統計調査の中止に伴い、同調査で把握していた**通信販売価格**(インターネット、カタログ販売等)について、その把握の必要性及び技術可能性について検討

**経済財政
諮問会議**
(2015年10月)

☞ ネット通販市場が拡大しているにもかかわらず、家電をはじめほとんどの商品について**ネット販売価格**が加味されていないのではないか

**統計改革
の基本方針**
(2016年12月)

☞ **インターネット販売価格の更なる捕捉及び(中略)採用の可否**を検討

**公的統計
基本計画**
(2018年3月)

☞ **インターネット販売価格の採用の可否**を検討

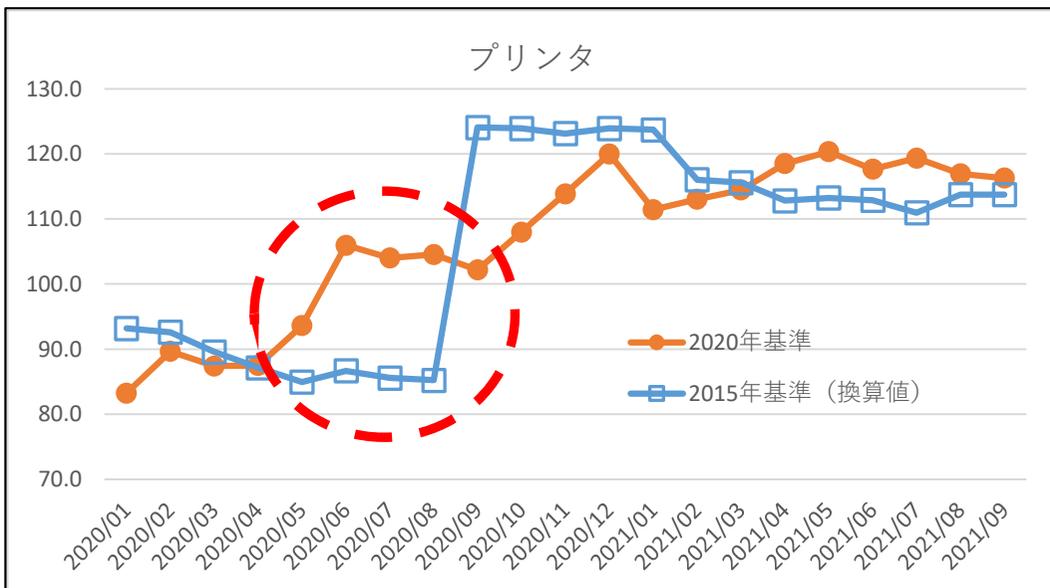
・カバレッジ(商標数、日数)拡大による、よりの確な価格動向の把握
 ・スペック情報を用いた品質調整等による、より適切な指数作成
 ・地方公共団体・調査員の事務負担の軽減

2. 取組の概要

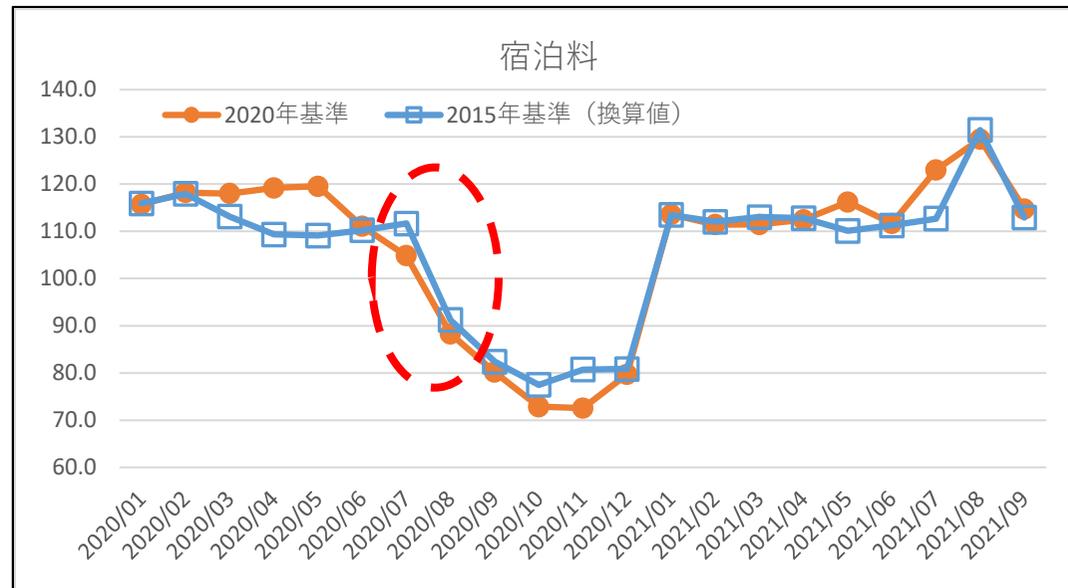
ビッグデータの活用による指数の作成状況

ビッグデータの活用により、2020年1月から2021年9月までの2020年基準における価格指数を2015年基準における価格指数（2020年 = 100に換算）と比較すると、次のような効果がみられた。

POSデータ



ウェブスクレイピング



カバレッジの拡大（**商標数の拡大**）やスペック情報を用いた品質調整により、より適切な指数作成が可能となった。

- 2015年基準指数は、特定1商標の価格を採用していたため、2020年9月の当該商標の価格上昇に伴い、指数が上昇
- 2020年基準指数は、選定スペックに該当する複数商標（ネット販売分も含む）を採用したため、同年5月頃から価格上昇した商標も把握

カバレッジの拡大（**把握日数の拡大**）により、よりの確に価格の動きを捕らえた指数作成が可能となった。

- 2015年基準指数は2020年8月に急落。これに対し、2020年基準指数は同年7月下旬開始のGo Toトラベル事業の影響を反映しており、変化がやや緩やか
- これまでは特定2日間の価格のみが対象であったため、調査日付近の日並びが指数に影響していたが、ウェブスクレイピングの導入により、全宿泊日を対象とし、よりの確に価格の動きを捕らえた指数作成が可能

3. ビッグデータの活用拡大に向けた検討の視点に対する主なご意見と対応状況

前回研究会後にいただいたご意見に対する対応状況、考え方は、以下のとおり。

ビッグデータの活用拡大に関する主なご意見	対応状況、考え方
データに問題があった際に、責任の所在問題が生じないよう、入手するデータの信頼度を確認することが重要	使用データの調達に当たっては、取集事業者数、地方区分別取集店舗数等、詳細な要件を定めて信頼度を担保しつつ、品目ごとの特性やデータの特徴を分析し、適切な指数作成を実施
大量に取集したデータからどのデータを使用するのか、外れ値、重複、特殊な商品の排除など、取捨選択が必要	<ul style="list-style-type: none">• POSデータ スペックを固定して対象データの絞り込み等を実施• ウェブスクレイピング 高額・低額プランの除去（航空運賃、外国パック旅行費）、平均値±3σ外の価格を除去（宿泊料）等、所要の対応を実施
リサーチ会社のみならず、小売店やスーパーのチェーンや各業界団体からデータを取集することも検討するべき。	リサーチ会社に対して、業界情報の収集、提供情報の拡充等についてヒアリングを優先的に検討
冷蔵庫や電子レンジ、エアコンなど、多くの耐久家電にもヘドニック等のPOSデータを活用すべきではないか。	品質調整（スペック固定、商標指定、容量調整等）に必要な情報が含まれていることなどを十分確認の上、POSデータの活用拡大を検討

4. ビッグデータの特徴

- ビッグデータの活用により、①膨大な量の価格データによる統計精度の向上、②報告者負担軽減・業務効率化、③ネット販売価格の把握拡大 が期待
- ビッグデータの活用拡大に当たっては、カバレッジ、品質調整（スペック固定、商標指定、容量調整等）を行うための情報の有無等、品目ごとのデータ特性などに照らし、費用対効果の慎重な見極めが必要

	実地調査の特徴	ビッグデータの特徴
把握できる販売形態	<ul style="list-style-type: none"> ● 店頭での実際の販売価格を把握可能（ネット販売価格は得られない） ⇒ 店頭販売が主流の商品については正確な実態把握が可能 	<ul style="list-style-type: none"> ● ネット販売価格を把握可能（店頭販売価格と異なる可能性） ⇒ ネット販売の割合が高い商品については正確な実態把握が可能
データ数	<ul style="list-style-type: none"> ● 得られるデータ数が限定的 ● 調査日が限定的（カレンダー要因が発生する場合あり） ⇒ 統計に振れが生じる場合がある 	<ul style="list-style-type: none"> ● 膨大な量の価格データ（POSデータは販売数量を含む。）を取集可能 ● 全日を対象とした価格や、日別の価格取集が可能 ⇒ より統計が安定することが期待できる
報告者負担、業務負担	<ul style="list-style-type: none"> ● 報告者負担が大 ● 職員や調査員の業務量大 ● 人件費などの費用が大 	<ul style="list-style-type: none"> ● 報告者負担はほぼゼロ ● 取集自動化により業務の効率化が可能 ● 比較的安価
その他の課題		<p>【POSデータ】</p> <ul style="list-style-type: none"> ● リサーチ会社（POSデータの販売元）の意向などによりデータ取集できなくなる場合の備えが必要 など ☞ 複数のリサーチ会社を取り扱っている品目を選択するなどの対応 <p>【ウェブスクレイピング】</p> <ul style="list-style-type: none"> ● バイアスが存在（店頭・ネット販売全体を代表していない、販売数量が不明） ● サイト休止やシステム障害などによりデータ取集できない場合の備えが必要 など ☞ 航空運賃（航空会社のサイトから取集）や宿泊料（情報が集約されたサイトから取集）など、一定程度のシステム安定性や障害発生時の迅速な復旧対応が見込まれる品目を選択 ● 法外な価格や不要な情報などのデータノイズが混在 ☞ 外れ値等の除去が必要（高額・低額プランの除去、平均値±3σ外の価格を除去等） <p>【POSデータ、ウェブスクレイピング共通】</p> <ul style="list-style-type: none"> ● 都道府県等地域情報の不足（消費者物価地域差指数への影響）

5. 今後の取組拡大の方向性

- ネット販売の拡大や情報収集技術の向上、新型コロナウイルス感染拡大による実査環境の更なる悪化も踏まえると、CPIにおいてこうした代替データの活用拡大を加速していく必要
- 新たに発生するコストや課題、調査員調査・都道府県調査との役割分担も踏まえ、予算上の制約等を考慮しつつ、費用対効果が見込める分野から優先順位をつけて取り組んでいく必要

検討対象

	品目例		候補となる価格収集方法
POSデータが 入手可能 と考えられる 品目	白物家電	電気炊飯器、電子レンジ など ☞ 既にテレビなどの家電をPOSデータに移行済。POSデータの活用範囲を白物家電に広げることで、より調査員の实査負担軽減に寄与する可能性	POSデータ ☞ 品目ごとに、品質調整の可能性・方法を検討・検証
	食料品、 医薬品、 日用品	うるち米、感冒薬、殺虫剤 など ☞ 調査員調査との役割分担の整理・検討の必要（例えば食料品の場合、惣菜などのPOSデータはないため、一部品目でPOSデータを用いても、調査員の実質的な実査負担軽減があまり見込めない可能性）	
ウェブサイトから の価格収集が 可能 と考えられる品目	衣料品	ワンピース、スラックス など ☞ サイトの代表性、製品区分などの格付け方法などに課題	ウェブスクレイピング ☞ POSデータが入手可能な品目は、POSデータの検討を優先 ☞ 衣料品について、製品区分などの格付けにAIの活用可能性を検証
	外食、 行政サービス	《調査員調査品目》 ハンバーガー（外食）、牛丼（外食）、ピザ（配達）、コーヒー（外食（セルフ）） など ☞ 大手チェーン店が高いシェア。同一企業の地域価格差は基本的になし。 《都道府県調査品目》 自動車免許手数料、パスポート取得料 など	ウェブクローラを活用した価格収集 ☞ 価格数、価格変動が少なく、ウェブクローラによる定期的なアクセス・差分検知により、効率的な価格変動の把握可能性

6. 課題と対応 | 製品区分などの格付けに係るAIの活用可能性の検証 (衣料品)

- 価格データが掲載されているサイトには、販売店舗を持つ個別企業のサイトや、多数の出店者の商品を扱うEC事業者のサイトが存在
- EC事業者のサイトでは、幅広い商品の情報が掲載されているものの、商品説明（商品名、素材など）や表記方法などの質・量が出店者によってバラバラな傾向

⇒ 同品質の商品の価格追跡が必要なCPIにおいては、ウェブスクレイピング活用は難易度が高い

商品名称にPR
情報等が混在

特定の商品名・
コードが存在しない

商品の品質に関する
情報の表記が不統一
で情報量が膨大

衣料品ECサイトの掲載情報の例 (背広服)

商品名	商品コード	出店者	商品説明の一部	商品説明の文字数
2つボタンスーツ ピンストライプ	—	A	… (略) …【生地】 Biella Finish イタリア技術者の指導を受け、原料選定から紡績、織布、仕上げまでの加工に、イタリアのノウハウを導入して作り上げた生地。高度な技術を実現化することで、イタリア製生地に遜色のない上質な生地に仕上がっています… (略) …【仕様：ジャケット】 2つボタン/本切羽/センターベント/胸ポケット/腰ポケット×2/総裏仕立て/内ポケット×2【仕様：パンツ】 ノータック/スリムフィットテーパード/エキストラローライズ/サイドポケット×2/ヒップポケット×2/ウォッチポケット/前ヒザのみ裏地あり【季節】秋冬モデル【洗濯表示】ドライオンリー ● 体型 YA (Drop8：スリム) A (Drop6：普通) AB (Drop4：ややがっしり) ● 身長【3】 160cm … (略) …	約1,500文字
1着は持っていたい!ウエストのアジャスター付きがうれしいシングルフォーマルスーツ メンズ	—	B	ビジネスばかりでなく、礼服・フォーマルにも使える黒無地スーツはマストアイテム。ウエストにはアジャスターを装備し、左右で±3cmずつ調整できるので、長時間の座り姿勢のときや、食後にお腹が張っているときなども快適。ベーシックなスーツなので、年齢も関係なく着用できるのがうれしい。■表地:毛50%ポリエステル50%(織物・背抜き)裏地:ポリエステル100% ▼スラックスの裾は半仕上げになっています。ご家庭でお好みの寸法にお直ください。ご家庭で簡単にできる裾上げテープを販売しておりますのでご利用ください。●ウエストアジャスター機能付き ●ドライのみ(ジャケット・パンツ) ●ミャンマー製… (略) …	約300文字
秋冬物ウオッシュャブル2ボタンスーツ【ブラック / ストライプ織柄】	—	C	… (略) …ご家庭でお洗濯可能! 2ボタンスーツ (8250-03) Item Information Dady Costa 2ボタンスーツのご紹介です。ウオッシュャブル生地を採用し、ご家庭でお洗濯可能なスーツです。スタイルは極端にラインを細く強調するのではなく、着心地を重視した2ボタンスーツで、細身のスーツが苦手なお客様にもご満足いただけるスーツになっています。ブラック地のストライプ織柄でオーソドックスながら品の良さがあり、どのような場面でも対応可能で幅広く着用可能なスーツです。ズボンの折り返し目が消えない耐久折り返し加工や、貴重品を守るフラップ付き内ポケットなど、充実した仕様となっております。お手頃な価格ですので、スーツを着る必要なお客様や家計のやりくりで苦労している奥様にも喜んでいただける商品です。Dady Costa Washable 2 Button Suit 8250-03 WEB Price (税抜き) 19,000 Information Size List A4 A5 A6 A7 AB4 AB5 AB6 AB7 BB4 BB5 BB6 BB7 Cloth Information ブラック / ストライプ織柄 ウール55% ポリエステル45% 日本製生地 ブラックのバイアス織柄地に4mm間隔でドライブが入っています… (略) …	約2,000文字

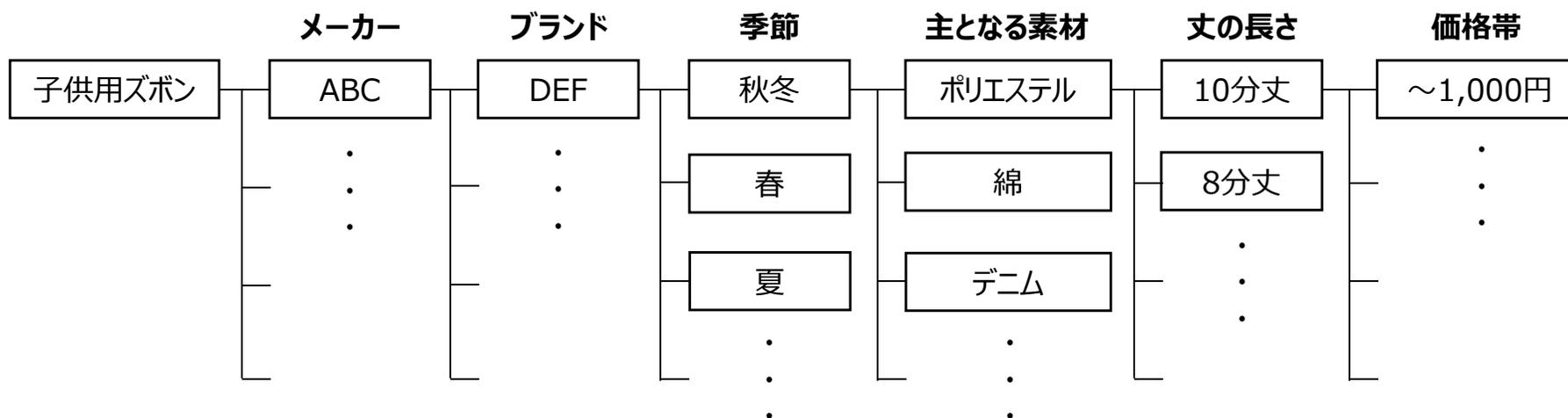
商品の同一性・同質性
の判別が困難

6. 課題と対応 | 製品区分などの格付けに係るAIの活用可能性の検証 (衣料品) (続き)

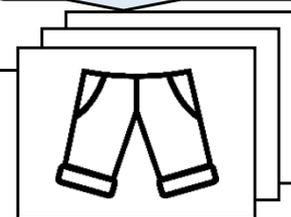
衣料品ネット販売データの製品区分などの格付けのため、AI（機械学習）の活用可能性について検証中

- ・ 2019年度～20年度： 商品情報（テキスト情報）を対象に、最適な機械学習アルゴリズムを検証
- ・ 2021年度～： 商品画像の併用可能性について検証、検証対象サイトの拡大

分類イメージ

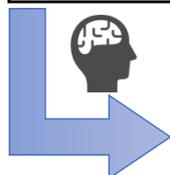


2021年度から、商品画像の併用可能性について検証中



AI機械学習 イメージ

商品名：**ABC 長ズボン** キッズ 子供服 男の子 **2～5歳 薄手のデニム** ジーンズ **青**
 説明：薄手なので**季節の変わり目**にぴったり！ キッズ 薄手のデニムジーンズ 長ズボン(2～5歳：**90・95・100・110**)
 薄手のデニムが入荷しました。薄手なので、季節の変わり目や蚊の多い公園など**アクティブ**に過ごすときに最適！
 半ズボンから長ズボンに変えるのを嫌がるお坊ちゃまにも。
【キッズ】【男の子】素材 ポリエステル100% サイズ 2T,3T,4T,5T
 注意点 ○洗濯機、乾燥機は弱で。○漂白剤不可。○**色落ち**する場合がございますので、最初の洗濯時は白いものとは分けてください。



7. 今後のスケジュール（案）

2021年度

2022年度

2023年度

2024年度

2025年度

物価指数研究会（年3回程度開催）

● 3月 基準改定計画案 ● 12月 基準改定計画

● 2025年1月

関係
POSデータ

〔白物家電〕

データ分析、指数の試算、指数作成手法の検討、実務運用面の確認
（2022年度は電気炊飯器、電子レンジなどを対象。2023年度はその状況を確認しつつ対象拡大を検討）

試算結果を物価指数研究会へ報告、採用の可否を検討

2025年1月分から、本番業務データ収集、運用

関係
ウェブスクレイピング

〔衣料品〕

2021年8月分から衣料品5品目※1)のデータ収集（大手3サイト）
収集サイトを拡大

衣料品先行3品目※1)を対象にAI格付け研究（画像活用の可能性）

衣料品5品目を対象に、AI格付け研究（サイト拡大対応、効率的な画像活手法※2)）

データ分析、指数の試算、指数作成手法の検討、実務運用面の確認

試算結果を物価指数研究会へ報告、採用の可否を検討

2025年1月分から、本番業務データ収集、運用

関係
ウェブクローラ

〔外食、行政サービス〕

品目の選定、現行指数作成方法の確認等、各種検討

データ分析、指数の試算、指数作成手法の検討、実務運用面の確認

試算結果を物価指数研究会へ報告、採用の可否を検討

2025年1月分から、本番業務データ収集、運用

※1) ワンピース（先行）、スラックス（先行）、子供用ズボン（先行）、スカート、ブラウス
※2) 画像データは、テキスト情報に比べてファイルサイズが非常に大きく、運用面に課題。このため、必要画像の点数削減（2021年度研究では最大10画像）、二値画像（白黒画像（2021年度研究ではカラー画像を使用））の活用可能性を研究