



都市再生におけるデータ活用推進WG④資料 オルタナティブデータによる効果検証について

株式会社サイバーエージェント
AI Lab リサーチサイエンティスト 森脇大輔

ウェブ企業におけるデータ活用

- 日常的なデータ分析

- ダッシュボードで日々の動きを表示(クリック率、リクエスト数 etc.)
- 想定通りサービスが動いているかの確認(=異常検知)
- 低コストのモニタリング=「軽い分析」

- 施策の効果検証

- ユーザーインターフェース(UI)変更の効果検証
- UI変更するか導入するかどうかの意思決定
 - 検索画面の表示を変えただけで1億ドルの収益改善した例(マイクロソフト)
- 厳密な効果検証=「重い分析」

- 低コストのモニタリングと手間のかかるが厳密な効果検証の
使い分けが重要

高頻度データによるモニタリングの例

(民間データ、いわゆるオルタナデータ)

- **消費**
 - POSデータ、決済データ、家計簿
- **モビリティ**
 - 位置情報(GPS、基地局、ビーコン、WiFi)
 - ICカードによるバスや電車の乗降客数
 - AIカメラによる人流把握
 - カーナビデータ
- **雇用**
 - ウェブ上の求人データ
- **経済活動**
 - 電気ガス水道
 - 通話
 - 衛星画像(駐車場など)

(行政記録情報)

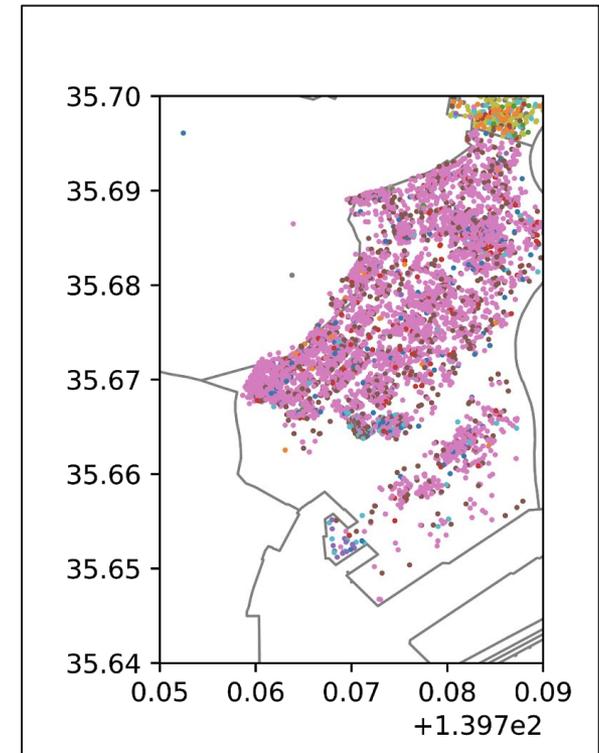
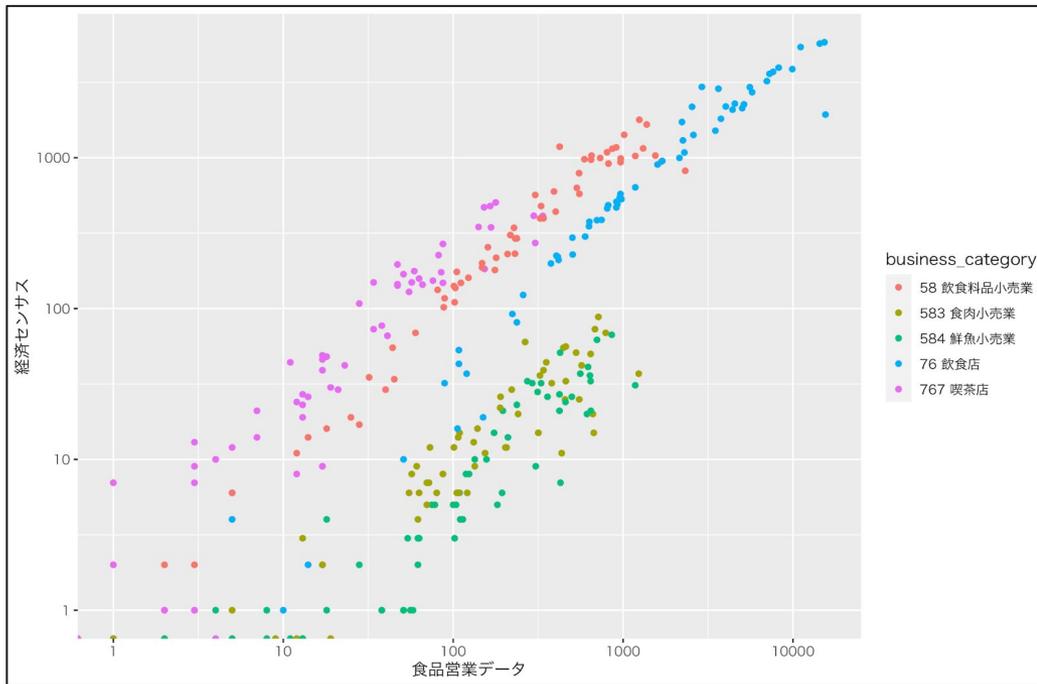
- 税込
- 転出入届
- 各種許可・届出(次ページ)
- ハローワーク
- 宿泊者名簿

(アンケート)

- **通行人アンケート**
 - 快適性総合スコア(国交省)
- **整備地域内のテナントの状況**
 - 賃料
 - 売り上げ など

行政記録情報の活用例

- 保健所の収集する飲食店の営業許可データはリアルタイムに飲食店の出退店を捕捉
- 経済センサス(市区町村別)とも一定の相関
- 行政の保有するデータの活用可能性は膨大



左図: 食品営業許可データにおける市区町村別・産業分類別事業所数と経済センサスの事業所数の比較。きれいな直線状にならんでいる

右図: 営業許可データから粒度の細かい事業所データが取得可能。色は事業所の業種を示す。

森脇(2021)「行政データによる統計作成について」日本経済学会春季大会

モニタリングにおける論点

● データの解釈

- 人流データや消費額を眺めていてもアクションにつながらない
- 解釈可能な指標に落とし込む
- 異常が見つければ仮説検証フェーズに(=重い分析)

● データのコスト

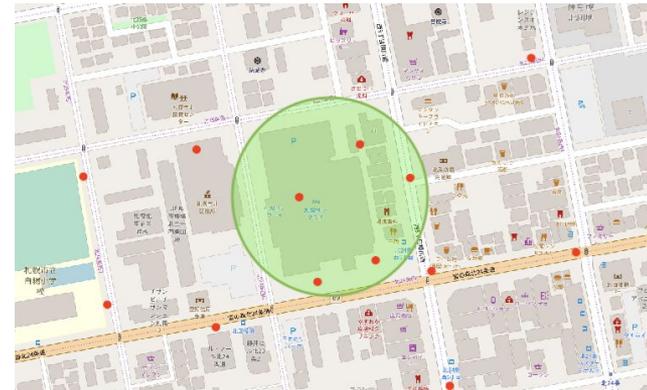
- データそのものを販売することに対する抵抗感
 - 保有データが無償公開されればビジネスが成り立たない
- データそのものではなくインサイトを販売する
 - Google Mobility Reportは位置情報そのものではなく解釈可能な人流データとして公開

● オルタナデータ由来の指標を作成することも一案

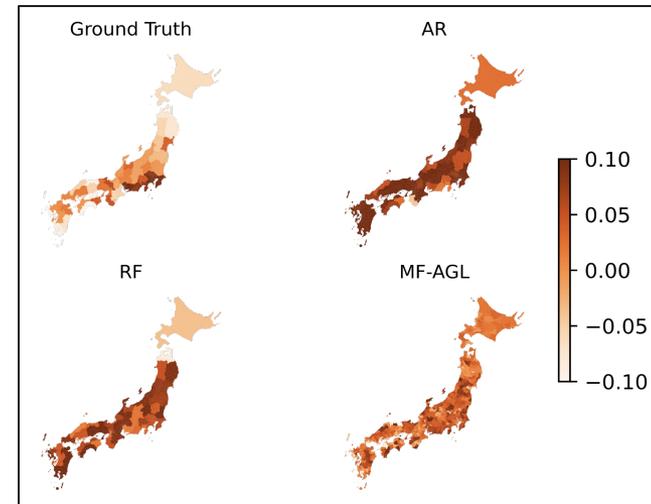
- データベンダー側はアドホックなレポート作成はコスト面で見合わない
- 長期的な収益が見込めれば指標作成への協力もありうるのでは

指標作成の例

- [Toda et al. \(2021\)](#) は本来観測できない小地域レベルの失業者数を粒度の細かい位置情報を用いて日次で予測
- 月次・都道府県別に公表される新規求職者数を位置情報で高頻度化・高解像度化
- ナウキャストイングも行うことでリアルタイムかつ詳細な動きを捕捉
- 同様の手法が業界別・地域別の売上動向や消費動向に適用できる可能性
- 様々なデータを活用することでリアルタイムな業界別売上や地域別消費指数など現実的には観測不可能な粒度の細かい指標を作成することが可能



人工的に作成した位置情報のイメージ。ハローワークを囲む緑円内に入ったログを来訪者としてカウント。背景はOSM。



月次失業者数の予測(昨年対比の増減)、左上から時計回りに正解データ、自己回帰モデルでの予測、位置情報を用いた詳細な予測、ランダムフォレストによる予測。MF-AGLモデルのみ高解像の小地域の予測ができていることがわかる

施策の効果検証(=重い分析)

- 継続的な施策の改善には効果検証が不可欠

- 検証結果を踏まえた改善、拡大、撤退
- 納税者を含めたステークホルダーに対する説明責任

(ウェブ企業でのワークフロー)



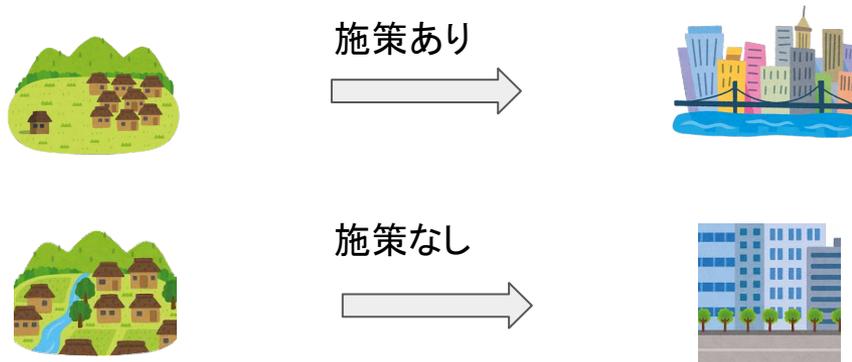
- 効果検証のフレームワーク

- ランダム化実験(RCT)
 - 倫理的、政治的、費用的に困難な場合が多い
 - イベント実施のようなコストの低い介入であれば可能
- 観察データによる因果推論
 - 差の差分析、マッチング回帰、回帰不連続デザイン、操作変数法など

因果推論による効果分析

- 自治体が地域内の緊急整備地域について検証

- 介入対象の地域とその他地域の施策導入後の変化の差分を計算し効果とする



- 中央政府(地方創生事務局)がこれまでの緊急整備地域について分析

- 介入対象と非介入対象が複数ずつ確保できるので上述に加え傾向スコアやマッチングによる分析が可能
- どのような地域において効果が高いか(個別因果効果)を推定する

差の差分分析・シンセティックコントロール

- 差の差分分析

- 人口や地価などのアウトカム指標について緊急整備地域とその他地域の差分を計算

- シンセティックコントロール

- 緊急整備地域が「対象にならなかった場合のアウトカム指標」をその他の地域のデータから予測してその差分を効果として計算する

- 必要なデータ

- 長期系列がとれているアウトカム指標
- 高頻度である必要はないが年次ではあったほうがいい
- 対象地域と比較地域のデータ
- 地理空間的に粒度が細かいデータ

- 分析例

- 地域別税収を所得指標として効果を検証
- 個別のハローワークにおける新規求人を雇用指標として効果を検証
- 住所つき食品営業許可等を事業所数の指標として効果を検証
- 住所つき電力消費や水道使用料を経済活動の指標として効果を検証 など

(参考) 差の差分析を利用した政策効果の分析

- 経済産業研究所における補助金の効果分析(関沢・牧岡・山口 2020、Hashimoto and Takahachi 2021, 牧岡 2021, 橋本・平沢 2021)はDIDやRDDを用いて中小企業政策の効果进行分析
 - 工業統計の個票、東京商工リサーチ、業務データなどを接続して利用
- 厚生労働省(2021)は[EBPMの分析レポート\(障害者雇用の促進\)](#)として、DID, RDDといった因果推論手法を用いた分析を公表
 - 法令に基づく報告データを利用

傾向スコア回帰・個別因果効果の推定

- 傾向スコア
 - 過去のデータからある地域が緊急整備地域に選定される確率(=傾向スコア)を推定
 - 選定確率がデータから予測できることが前提
 - 傾向スコアを用いて介入確率が近い地域をマッチングして比較あるいは傾向スコア回帰を用いて介入効果を推定する
- 個別因果効果の推定
 - アウトカム指標と介入データだけでなく、地域の属性(地価や人口密度、世帯属性、産業など)を用いてどのような地域において効果的だったかを分析
 - アップリフトモデリング

(参考) 広告配信データにおける効果検証

- 広告効果を分析するためには過去の配信データが必要
- 誰に広告を配信するかは ランダムではない
- 広告配信のされやすさとアウトカムが相関している場合、予測が過大もしくは過少になる

(例) 家電量販店の広告キャンペーン

定期的に来訪するAさん



リタゲロジックによって高値で入札

普段はライバル店を利用するBさん



来店経験がないので安値で入札

高値で入札されるAさんのほうが広告を配信されやすく来訪しやすい => 因果効果の上方バイアス

広告の配信されやすさを予測することでデータのバイアスを修正し効果を推定できる(=傾向スコア重みつけ法)

アウトカム指標の選定

- KPI

- 都市再生の目標となる指標(人口、地価、域内消費、所得、雇用)
- 「にぎわい」のような直接観測できない指標は代理指標や中間指標を設定する

- 中間指標

- 人口や地価など緩慢にしか変化しない指標に対してそれらの指標を予測するような指標を使う(検索数や問い合わせ数、ページ来訪数など)

- ガードレール指標

- 毀損したくない価値を設定し、その動向を確認する
 - 地価はあげたいが、学生が住めなくなるような賃料になると困る
 - 人口は増やしたいが、ワンルームマンションだらけになると困る
 - 開発は促進したいが緑地は維持したい

まとめと提言

- まとめ

- 日常的な地域動向はリアルタイムデータのダッシュボードによって把握する（＝軽い分析）
- 異常があったり、効果検証が必要な場合は厳密な分析（＝重い分析）
- データソースは公的統計や民間事業者だけでなく行政も重要なリソース

- 提言

- 民間データを何らかのインサイトを得られる指標として加工した上で事業者から購入・公開してはどうか
 - 位置情報そのものではなく失業者数指標とするなど
 - 事業者の不確実性を減らすために長期的な契約を結ぶ
- 行政データの活用に向け、自治体が保有するデータの活用を研究してはどうか
- 効果検証に必要な長期時系列を公的統計等から作成し、地方創生事務局で効果検証の例を提供・公開してはどうか

Thank You!

