

第4回 都市再生におけるデータ活用推進ワーキンググループ

「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

森脇リサーチサイエンティストによる説明内容

[スライド2]

- ・ウェブ企業におけるデータ活用ということで、簡単に説明させて頂く。広告であれば、我々は日々広告を出してクリックをしてもらったり、物を買ってもらったりして生計を立てているが、ダッシュボードというチャートを常に表示しているものをよく活用している。Tableauは有名であるが、ダッシュボードを置いてそれを見ながらサービスの状況を感じている。
- ・主に見ている指標としてはクリック率やリクエスト数である。リクエスト数というのはどれだけ広告を表示する回数があるのかというもの。それらを見て想定通りサービスが動いているのかをチェックしている。
- ・データ分析と言いながらもエンジニアやデータサイエンティストは何もしていなくて、ただデータが流れているだけである。それを見るのは、データサイエンティストよりはビジネス側の人間が多い。顧客に対してレポートするために常に数字を追いかけてたり、事業責任者やサービス全体の責任者が自分のサービスがうまく動いているのかを見たりしている状況であり、非常にコストの軽い分析である。
- ・もう一つの分析として施策の効果検証をやっており、例えばユーザーインターフェース、つまりウェブサービスやアプリの画面の見た目を変更するというのは結構大きな意思決定であり、これをした時にどうなるかという事を効果検証としてやっている。UIを変更したときに何が起こるのかというのを見て、変更時にうまくいったら変更を導入するという意思決定をする。
- ・マイクロソフトの公開情報によると、Bingという検索画面の表示を少し変えただけで1年間に1億ドルの収益が改善したという例がある。かなり大きな話である。よって1億ドルの収益が減少しないようちゃんと検証してやっている。
- ・これは重い分析であり、データサイエンティストも必要であるし、エンジニアにデータを用意してもらわないとならない。実験をするとなるとそのためのコーディングが必要となるのでかなり重い。これらを行うことにより収益を改善しているというのがウェブ企業である。
- ・日々のモニタリングと、手間はかかるが効果検証を使い分ける事が重要である。

[スライド3]

- ・都市再生の分野にどういうデータが使えるかどうか考えてみた。
- ・民間データ、オルタナティブデータは常に議論されていると思うが、POSデータ、決済アプリ、クレジットカード、家計簿アプリのデータなどは非常に良いと思われる。
- ・モビリティデータ、ICカードを使った分析や、バスや電車の乗降客数を集めている自治体もあると思うので、そういうものは使えると思われる。
- ・最近増えているものはAIカメラである。AIカメラを設置してどういう人が歩いているのか把握する。こういったデータも貴重である。
- ・内閣府が最近出しているワーキングペーパーがあるが、カーナビのデータを使って車がどこまで移動

しているのかというデータも使えるであろう。

- 雇用に関してはウェブ上のスクレイピングをして求人データを集めることや、経済活動については電話、水道、通話、つまりどこかに誰が電話したのか。衛星画像で写真を撮り駐車場に何台来たのか見ることもやっており、これらのデータも使えるであろう。
- オルタナティブデータは議論もあるが、注目しているのは行政記録情報である。行政が持っている情報であるが、例えば税収データを住民の所得の代理指標として使うという事も大事である。
- 色々な許可届け出が自治体には集まっているので、そういったデータを使って経済活動を捕捉する事もできる。転出入届を人口の移動として使う事もできる。
- 厚労省の紹介になるがハローワークの毎日の求人求職受付は高頻度かつ高粒度のデータとなる。
- ホテルや旅館がやらなければならない宿泊者名簿も旅行インバウンド、旅行のオルタナティブデータとして使える。
- オルタナティブデータや行政記録情報ではないが、アンケートも使えると思っている。国交省の快適性総合スコアはアンケートとして使える。緊急整備地域でアンケートをとり、テナントの状況を毎月報告する事もやってもらいたいと思われる。

[スライド4]

- 行政記録情報の活用事例であるが、保健所は全国にあり飲食店の営業許可データを持っている。飲食店に行くとき必ず許可証があるが、これが正に保健所が発行している許可証である。飲食店は必ず許可を取ったり届け出をしたりしないと出店できない。よって飲食店に関してはリアルタイムで事業所の出退店がわかるデータとなっている。
- 右図は中央区の例で、これだけの数の事業所のデータがある。このデータと経済センサスを散布図でプロットした。横軸が食品営業許可データから取った市区町村別の事業所の業種別の数である。経済センサスの方は、同じようなカテゴリーを見つけて取りだしたデータと関連付けてどういう関係するのか見たところ、ほぼ直線だったので正しいデータであると思っている。

[スライド5]

- モニタリングする際にいろいろ論点があると思うが、少しずつ考えてみる。
- 1つ目はデータの解釈で人流データや消費額というのは眺めていてもいいが、ではどうしたらよいか、というアクションはなかなかしづらい。
- 人流データというのは、例えばコロナで渋谷駅の人流が増えたので、おさえるために緊急事態宣言を出そうというのはわかるが、もう少し高度な意思決定をするために人流をどう考えるか。解釈可能な指標に落とし込むという事が重要であり、技術が必要なので、データを渡して各自治体にやってと言っても難しいので、政府などでやったほうがいい。
- もしその指標を受けて異常が見つかったら、検証フェーズに持ち込んで何が起きているのか、どうしたらいいのか考えるべきである。
- もう一つの問題点はデータのコストである。このワーキンググループでもデータが高いという発言があった。データベンダー側からも事情があり、データそのものを販売、つまり生データを渡してしまうと、そのデータを使って色々な事ができてしまい収益がなくなってしまうので、販売に対して抵抗が強い。

い。よってデータそのものを販売するのではなく、インサイトを販売するような形でデータ提供を受けた方がいいと思っている。

- ・ グーグルモビリティレポートというグーグルが出している業種別の人流データが公開されている。グーグルのデータを使った統計だが無料で公開されている。このデータを使って何か商売といってもなかなかできない。それは粒度が粗いということもあるし、人流の数値しかない。
- ・ これらのデータは使えないのかといえばそうでもなく、業種別、地域別にどういう人流が動いているのかというデータを使って、例えば内閣府の経済財政白書では緊急事態宣言の効果を分析している。よって用途を限ったデータの加工をすることでいいものができる。
- ・ オルタナティブデータで指標を作ることがいいと思っている。この時も注意点があり、データベンダー側はアドホックなレポートを作る事はすごく嫌がる。例えば 100 万円でレポートを作ってくれと言われても 1 カ月データサイエンティストをかかりきりにさせてレポートを作ったところで全然儲からないのでスケールするような販売の仕方がしたい。例えば長期的にこのデータを購入するという話をするとある程度収益を見込めるが、ワンショットでアドホックなレポートを作れとなると結構厳しい。

[スライド 6]

- ・ 手前味噌ではあるが、指標作成の例として私が行った分析を示している。
- ・ 求職者数のデータが厚労省から発表されているが、それを位置情報から高精度化したもの。位置情報で日ごとに非常に細かいデータがとれているので、求職者情報という統計を非常に細かい粒度で予測することができる。
- ・ 労働だけでなく色々な分野で実用可能であるのでこういうことも考えてはどうだろうか。

[スライド 7]

- ・ 施策の効果検証についてである。モニタリングをして問題があったり、あるいはある重い選択をしたりするときに、その効果を見たいとなり効果検証する必要がある。継続的な施策の改善には効果検証が必要で、検証結果をふまえて改善する、拡大、撤退するという判断が必要である。納税者を含めた施工面に対する説明責任である。
- ・ ウェブ企業の場合このように進んでいる。まずアイデアがあってそれを実験し、分析してリリースするかやめるか、改善するかという判断をする。
- ・ 効果検証は色々あるが一番ゴールドスタンダードといわれるものは、ランダム化比較実験でユーザーをグループに分けてやるかやらないかとやって結果をみてること。これは非常にコストが高く倫理的にも政治的にも難しい面があるので公共政策の分野でやるのは難しいと思われる。
- ・ イベントの実施であれば、ある地域はイベントするけど他の地域はしないということはできそうなので、そういうものはやってもいいだろう。メインになるのは観察データによる因果推論で、実験ができないのでとれているデータから因果推論手法を使って効果検証をしていく。
- ・ 例として挙げているのは、差の差分析とかマッチング回帰、回帰不連続デザイン、操作変数法など。

[スライド 8]

- ・ 因果推論による効果検証について簡単に説明している。地域内の緊急整備地域の効果が見たいという

事であれば、二つの地域があり一つは施策を施した緊急整備地域にして、もう一つはしなかった。これが何年か後にどうなったかを見る。元々どれくらい差があったのかという事と、何年か後にどのくらい差が広がったのかを見ると効果があったかどうか分かる。

- ・地方創生推進事務局がこれまでの緊急整備地域の分析をすることも可能である。この場合は介入地域と非介入地域が複数ずつ確保できるので傾向スコアの使用やマッチングによる分析が可能である。
- ・先に述べた方はある地域において効果があったかどうかはかわからないが、中央政府で全体のデータ分析をすることで、どういう地域だったら効果が高いか、個別の因果効果を追求することができる。

[スライド 9]

- ・分析手法について少し詳しく書いたが、差の差分分析、シンセティックコントロールというものについて説明する。差の差分分析は、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の地域で見えて差分を計算してその差分をとる。
- ・シンセティックコントロールというものは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予測する。カウンターファクチュアルというのが現実では起こらなかった架空の世界でのアウトカム指標の予測と現実の世界のアウトカム指標の差分を計算して評価の話をする。
- ・いずれにせよ効果の差分を見る。必要なデータは色々あるが長期比率がとれているアウトカム指標というのは非常に重要で、長期でとれていないとこういった分析はなかなか使えない。高頻度である必要はないが、年次である方が良く、5年ごとのデータだと5年間の間に施策が行われたりするといつから効果が発現するかわからないが効果が薄まってしまう。年次が最低限あったほうがいいと思われる。
- ・対象地域と比較地域と両方でとれている必要がある。2つのデータを組み合わせる必要がある。
- ・地理空間的に粒度が細かいデータである必要があり、対象地域は色々な大きさがあるという話であるが、すごく小さな地域でしか対象となっていない場合、その地域のデータが必要なため、ある程度地理空間的に細かいデータが必要である。
- ・例えば地域別の税収を所得の代理指標として効果を検証したり、個別のハローワークの求人求職を雇用指標として効果を検証、食品営業許可データを事業所ごとの指標として効果を検証したり、電気ガス水道の消費量を使うという事が考えられる。

[スライド 10]

- ・差の差分分析を利用した効果分析は RIETI や厚労省でも行っており、これは参考になると思われる。

[スライド 11]

- ・傾向スコア回帰と個別因果効果の推定について。過去のあらゆる地域のデータがとれていれば緊急整備地域に指定されるかどうかの確率を予測することができるかもしれない。これができれば傾向スコアという、どのくらい介入を受けるかという確率を予測するスコアを用いてマッチングしたり、傾向スコアを使った回帰分析をすることで介入効果を推定したりすることができる。
- ・個別の因果効果を推定する事で何ができるのかというと、アウトカム指標と介入データだけでなく地域の属性、例えば地価や人口密度、世帯属性、産業などを用いてどのような地域だったら効果的だったかということがわかる。これは機械学習の分野ではアップリフトモデリングという名前がよく使われ

ている手法である。

[スライド 12]

- 広告配信データの効果検証にこの手法を使っている。広告の効果は過去の配信データを使って分析をする必要があるが、広告はランダムに配信されておらず、配信のされやすさとアウトカム、つまり買ったか買わないかが相関している場合があり、この場合予測にバイアスがかかってしまう。
- 家電量販店の場合、定期的にご利用のある A さんはリタゲ（リターゲティング）と言って、何か買ったらまたその広告が打たれている。それによってかなり高頻度に広告を見る。これを勘違いして、広告を打ったから沢山買ったのだと思ってしまうと、効果を過大にしてしまう問題がある。
- 傾向スコアを使ってその人の広告を見る確率を予測できるとバイアスを除去する事ができる。

[スライド 13]

- 都市再生でも同じ事ができると思っている。アウトカム指標の選定について書いた。色々な指標があると思うが、人口、地価、域内消費、所得、雇用などである。にぎわいのような直接観測できないような指標は代理指標を使う事が考えられる。
- 中間指標も使えると思っており、人口や地価など緩慢にしか変化しないものについては、それらの指標を予測するような指標を使えばタイムリーに分析できる。例えば移住ならその地域に移住したいという人達の検索数、問い合わせ数、各自治体の移住サポートのページ来訪者数などが使えると思われる。
- 重要なのは KPI に対してガードレール指標を作った方がいいということ。毀損したくない価値を設定してその動向を確認する必要がある。例えば地価は上げたいが学生が住めなくなるレベルにはしたくない、人口は増やしたいがワンルームマンションばかりにしたくないなどである。毀損したくない価値をあらかじめ設定しておきその動向を確認することが大事である。もしおかしな傾向があればその選択を見直すことが大事である。

[スライド 14]

- 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を作って把握することが大事で、これは軽い分析である。異常があったり効果検証が必要であったりする場合は厳密な分析が必要で、これは重い分析である。
- データソースは公的統計、オルタナティブデータのような民間データも重要だが、行政記録情報のようなデータも重要なので活用したほうが良い。
- 提言であるが、民間データは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して公開したらどうだろうか。他の用途では使えないがある用途では使えるという指標を作ってしまうと事業者側もリスクを掴めるし良く使えるデータになる。
- 事業者側は不確実性があるので減らすためにも長期的な契約を結ぶ。
- 行政記録情報については自治体が保有するデータの活用を研究するのはどうか。ここが進んでない気がするのでやってみたらどうか。効果検証に必要な長期時系列を公的統計等から作成して地方創生事務局で効果検証の例を提供して公開すると、やり方がわかるので良いと思う。

以上