

運賃箱データ，携帯電話位置情報， GTFSデータを併用した バスのサービス改善に向けた分析手法検討

荒巻 凌¹・太田 恒平²・伊藤 昌毅³・星野 崇宏⁴

¹ 慶應義塾大学大学院 経済学研究科 E-mail: ryo_aramaki@keio.jp

² (株)トラフィックブレイン E-mail: kohei-ota@t-brain.jp

³ 東京大学生産技術研究所 E-mail: mito@iis.u-tokyo.ac.jp

⁴ 慶應義塾大学 経済学部 E-mail: hoshino@econ.keio.ac.jp

近年，交通や人流に関するビッグデータやオープンデータの流通により，地域公共交通網の検討に利用可能なデータは増加している．特にバスにおいてはGTFS形式に基づき国内で231事業者が詳細なバス情報を公開している．オープンデータのみならず，バス事業者は運賃箱に由来するODデータを保有しており，これを用いることでバスの乗降客数や乗降地点を精密に分析することが可能である．また，一部のソフトウェア企業はユーザの携帯電話に由来する位置情報ビッグデータを保持しており，都市レベルでの人流解析などに用いられている．本研究では，オープンデータに加えてバス事業者やソフトウェア企業の協力により得られたデータを用いて，バスサービスの改善に向けた分析手法の検討を行う．

Key Words : bus, GTFS, OD data, location data, bigdata

1. はじめに

バス等の地域公共交通の効率的な運営は，人口減少，運転手不足が進む現在の喫緊の課題である．その路線網の検討にあたっては，移動需要，交通網，利用実績のデータを揃えて効率的かつ精緻に行うことが肝要である¹⁾．

移動需要データについて従来主流であったパーソントリップ調査には，移動目的や個人の属性などを得られるメリットがある．一方で10年に1回といった長いスパンで実施されるアンケートベースであるため，日単位や時間単位などマイクロなデータを得ることができないという課題がある．近年はこれを補うため，高頻度でマイクロな情報を得ることの出来る交通ビッグデータが全国で利用可能であり，それを用いた手法が提案されている²⁾．

交通網データについては，バスにおいてはGTFSを拡張した「標準的なバス情報フォーマット」が国土交通省により制定された結果，国内231事業者³⁾が同形式に基づくオープンデータを整備，公開している．

利用実績データについては，ICカードや整理券により自動的に取得されたデータや，調査員によるOD調査・乗降調査データが主流である．しかし，そのデータ形式は交通事業者や調査ごとに異なっており，汎用的な分析手法の確立には至っていない．

そこで筆者らは，青森市営バスおよび八戸市営バスと

共同プロジェクトを発足し，複数のデータを統合的に利用したバスサービス改善に向けた分析手法の研究に着手した．本研究では，バスの運賃箱由来のODデータおよび通過実績データ（バス停単位），携帯電話由来の位置情報ビッグデータ，GTFS形式のオープンデータを揃えて入手可能な青森市営バスを対象として検討を行う．

2. 用いるデータとその特性

(1) 運賃箱のODデータ・通過実績データ

青森市営バスにおいて利用している運賃収受システムでは，整理券にバス停ごとに異なるバーコードが印刷されており，下車時に運賃を支払う際にこの整理券を投入するため，乗客ごとの乗車地点・下車地点のデータ（ODデータ）が記録される．この際，運賃種別（大人・子供）や乗車・下車の時刻なども記録される．なお，乗客が整理券を紛失した場合には乗車バス停が欠測となる．

また同システムでは，バス停ごとの通過実績が分単位で記録されている．

これらのデータは車両ごとに，運賃箱の金庫内に内蔵されている記録媒体に保存されている．バスが営業所（車庫）に帰着した際に運賃箱から金庫を取り外し，硬貨・紙幣を回収する装置に金庫をセットすることでデー

タが営業所内のDB サーバ (Microsoft SQL Server) に保存される仕組みとなっている。

本研究では、青森市営バスの2016年11月1日から2019年11月1日まで、3年間のデータを利用する。

(2) 携帯電話位置情報

共同研究先の企業の協力により入手した位置情報データである。スマートフォンアプリから5分おきのユーザの位置が匿名化された状態で記録されている。同種の分析においてはNTTドコモが基地局のデータを基に提供している「モバイル空間統計」もよく利用されるが、本位置情報データは各端末のGNSS測位データを用いているため、基地局由来のデータよりも高い解像度で位置が得られるため、停留所勢力圏の小さいバスの需要分析に適しているのが特徴である。一方、純粋な位置情報のみが記録されているため、モバイル空間統計には含まれている性別・年代といった属性を利用することはできない。

2018年4月から2019年10月にかけて全国レベルでの日次データが提供されている。データ量は日次によってばらつきがあるが、青森市周辺では約 6000件/日のユニークユーザ数があることが確認できている。

(3) GTFSデータ

General Transit Feed Specification (GTFS) は公共交通機関の時刻表や停留所データを格納するフォーマットで、複数のテキストファイルをZIP形式に圧縮したものがその実体である。当初はGoogle Transitのためのデータ形式として策定されたが、のちに交通オープンデータに関するデファクトスタンダードとなった。日本では国土交通省による拡張が行われ、「標準的なバス情報フォーマット」(GTFS-JP) としてバス事業者を中心に普及している。

青森市営バスではGTFS形式でバス情報データを2018年から提供開始しており、2020年3月現在で251路線1642便の時刻表、バス停818箇所の座標データなどが記録されている。

2. Google Maps 掲載の利用促進効果検証

(1) 背景

GTFS形式を中心とした公共交通機関のオープンデータ化は様々なイノベーションをもたらすが、目下のところバス事業者にとって最大の目的はネット上の乗換案内サービス、特に国際的にもシェアの大きいGoogle Mapsに反映されることである。バス路線は鉄道などに比べ、どこに路線が存在しているかが認知されづらいという性質がある。そのため、乗客にバス路線の存在を認知しても

らうには乗換案内サービスに表示されることが重要である。オープンデータ化を推進するバス業界の一部では「ネットで検索できなければ、バスは走っていないも当然」との表現もなされており、バス事業者においても乗換案内の重要性が認識されていることが分かる⁴⁾。

しかしながら、Google Mapsに掲載されることで利便性が向上すると言われている一方で、取り組みが始まって間もないこともあり、オープンデータについての便益分析に関する先行研究は少ない。

国内における先行事例として、岐阜県中津川市で行われた効果測定が挙げられる。この効果測定では、中津川市の中心部と観光地の「馬籠宿」を結ぶバス路線(馬籠線)の利用者を対象にしたアンケートが自治体によって行われた。アンケートにおいて「どのように馬籠線を知ったのか」という設問項目が設けられ、表-1のような結果が得られている⁵⁾。

表-1 馬籠行きのバスをどのように知ったのか、中津川市によるアンケート調査 (N=117)。中津川市(2018)による。

回答	日本人	外国人
Facebook	0	1
Instagram	0	1
Twitter	0	0
馬籠宿ホームページ	13	-
Trip Advisor	-	6
Google 経路検索	1	20
路線図ドットコム	0	-
Japan-guide.com	-	12
北恵那交通ホームページ	1	1
その他のホームページ	0	15
ガイドブック	0	20
その他・未記入	12	34

これより、中津川市は「外国人回答者では、20歳未満と70歳以上を除く全世代において2割以上が馬籠宿行きバスをGoogle 経路検索で知ったとの回答」があったと結論付けている。ただし、この調査はアンケート(N=117)のみを根拠としているほか、対象を観光地の路線に限定しているものである。それゆえ、統計的に信頼性の高い分析ならびに、観光地ではない一般的な都市や路線における便益分析が待たれている状況であるといえる。

(2) 分析手法

そこで本研究においては青森市の全路線を対象としてデータ分析を行った。分析手法はSynthetic Control Method (合成コントロール法)ならびにCausal Impactを用いた。いずれも目的変数として運賃箱ODデータ由来の乗客数を用い、説明変数としてAMeDAS由来の降水量データを用いた。説明変数として降水量を採用したのは夏季に自

転車を利用する学生などが冬季（降雪）や雨天にはバスを利用するようになるため、天候や季節による変動も生じるとの指摘があったためである。ただし、乗客数を決定する要因は多岐にわたるため、アウトカムのみでは分析の難易度は高いといえる。本来はGoogle Mapsにおける経路検索の件数があると望ましいが、データが入手できないため、乗客数を中心としたデータ分析を行うことになる。

Synthetic Control Method はマーケティングや公共施策評価に用いられる手法である。ある施策を実施した treatment groupがその施策を実施していなかった場合、その「実施していなかった」場合を観測することはできない。そのため、control groupからcounterfactualな「実施していなかった場合のtreatment group」を合成して生成し、実測値と比較する⁶⁾。当研究ではtreatment groupとして青森市営バスの各路線を、オープンデータ化を行わなかった八戸市のバス路線をcontrol groupとして分析対象とした。施策実施タイミングはGTFSデータの公開日とした。

Causal Impactは同様の目的であるが、control groupのデータを用いずtreatment groupのみからベイズ構造時系列モデルを用いてcounterfactualなデータを生成する⁷⁾。

(3) 分析結果

詳細な分析結果は追って当日に発表する。

3. その他分析手法の検討

(1) 利用状況の可視化と需給マッチング

本研究で用いる運賃箱のデータは、以下の課題によりこれまで活用が難しかった。

1. 時刻表データとの紐づけが無いため便別集計が不可
2. バス停座標が無いため地図上での可視化が困難
3. 3年間で2307万レコードのビッグデータである
4. 複数のマスタと紐づけが必要な複雑なデータ構造
5. 市営交通部内にデータ解析部門やIT部門が無くスキルやリソースが不足する
6. 実績データの分析だけでは潜在需要を見逃す

課題1,2についてはGTFSデータと対応付けることで解決する。この際、運賃箱とGTFSのマスタ同士の名寄せや、通過時分に基づく便の推定を行う。課題3,4,5については、共同プロジェクトを通じた技術支援により解決する。課題6については、携帯電話位置情報を利用することによる解決を試みる。

(2) 遅延分析

バスの遅延に関する情報提供やデータ分析のためには、バスロケーションシステムの導入が有効である。しかし同システムの導入には多額の費用を要するため、青森市営バスにおいては未導入である。青森市では降雪・除雪による遅延が大きいため、2019年度より冬ダイヤを設定したが、データ分析や合意形成に労を要し、遅延問題の解決には至っていない。

そこで本研究では、運賃箱の通過実績データを活用した遅延に関する分析を行う。同データはバスロケーションシステムの通過実績データに準じた活用が可能と考えられる。

具体的には、路線・時間帯・便別の遅延状況の可視化を行う。また今後のバスロケーションシステム導入に向けて、リアルタイム情報提供による待ち時間削減効果等の推計を行う。並行して、通過記録が分単位であること、時刻表データとの紐づけが無いといった、技術的な課題についても整理する。

4. 今後

本研究で示したように、公共交通に関連するデータはオープンデータ・クローズドなデータを問わず多種多様なデータが存在しており、分析を行えばサービスの向上や収益性の改善に繋がる可能性がある。既にマーケティングの分野においては産学問問わずにデータ分析が盛んに行われており、中にはデータサイエンスに関する専門部署を社内に設立している一般企業もある。しかしながら、公共交通の世界では十分なデータを蓄積していながら、計量的な分析を行っている取り組みはまだ多いとは言えない。本研究をさらに発展させ、既に存在しているデータを活用して公共交通の改善につなげたい。

参考文献

- 1) 国土交通省：「ビッグデータの活用等による地方路線バス事業の経営革新支援」について、
<https://www.mlit.go.jp/sogoseisaku/jouhouka/sosei_jouhouka_frl_000014.html>
- 2) 国土交通省都市局都市計画調査室(2018)：「パーソントリップ調査におけるビッグデータの活用について」
<https://www.soumu.go.jp/main_content/000588038.pdf>
- 3) 嶋田：GTFS・「標準的なバス情報フォーマット」オープンデータ一覧、2020.3.8現在
<<https://tshimada291.sakura.ne.jp/transport/gtfs-list.html>>
- 4) 青森市営バス：「標準的なバス情報フォーマット（GTFS準拠）」オープンデータのお知らせ（2018）
<<https://www.city.aomori.aomori.jp/kotsu-kanri/koutsu/oshirase/20180401gtfso-punnde-ta.html>>
- 5) 中津川市(2018)：「北恵那バス馬籠線利用者アンケート調査」
<http://www.city.nakatsugawa.gifu.jp/page/081195.html>
- 6) Alberto Abadie, Alexis Diamond, Jens Hainmueller.

- (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. 9, No. 1, 247-274.
- 7) Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL.(2015) Inferring causal impact using Bayesian structural time-series models. Annals of Applied Statistics, Vol. 8) 太田：バスロケータを基にした路線バスの遅延対策ダイヤ改正, 第 57 回土木計画学研究発表会, 2018 (2020.3.8 受付)

EXAMINATION OF ANALYSIS METHODS FOR BUS SERVICE IMPROVEMENT
WITH DATA FROM FARE BOX, LOCATION DATA FROM MOBILE PHONE
AND GTFS DATA

Ryo ARAMAKI, Kohei OTA, Masaki ITO and Takahiro HOSHINO