

4 都市圏パーソントリップデータの提供に向けたクレンジングと時空間内挿の実現*

Spatio-temporal Interpolation and cleansing for providing Person Trip Survey data in 4 major Metropolitan Areas.*

薄井智貴**・金杉洋***・関本義秀****・南佳孝****

By Tomotaka USUI**・Hiroshi KANASUGI***・Yoshihide SEKIMOTO****・Yoshitaka MINAMI****

1. はじめに

(1) 人の流れプロジェクト

「人の行動」を理解する上で、ダイナミックに時々刻々と変化する人々の詳細な動きを把握することは必要不可欠である。現在、都市圏内の人々の動きを把握するものの1つとして、パーソントリップ調査（以下、PT調査）がある。この調査は、住民のその日の行動を詳細に把握するためのアンケート調査であり、都市圏レベルでの都市交通計画や鉄道網計画、防災・防犯計画、新規路線等の需要予測などに用いられている¹⁾。しかし、PT調査は、主として行動の起終点に関する調査であり、詳細な行動経路やある時点における位置までを把握することはできない。さらに、データそのものの入手や加工に係るコスト、時間的・空間的な精度に関して、十分とは言えない。また、都市圏毎の調査集計であるため、都市圏間のデータ整合性が低く、全国規模での施策への利用は難易度が高い。

そこで、東京大学 空間情報科学研究センターでは、2008年7月より「人の流れプロジェクト」と題し、動線解析プラットフォームを立ち上げ、時空間クリーニングサービスと時空間データ提供サービスを開始している²⁾。このうち時空間データ提供サービスにおいては、平成10年10月に東京都市圏で実施された第4回東京都市圏PT調査データ72万人分のデータを処理し、「人の流れ」を把握するための社会基盤となる時空間データとして、1分刻みの都市圏内の人々の位置情報を提供している。しかし、現状の提供サービスは、東京都市圏の人の流れのみの対応となっており、

マーケティングや感染伝播、防災・防犯など全国規模のニーズには十分対応できていない。そのため、全国の「人の流れ」を把握するための時空間基盤データサービスの実現が急務となってきている。

(2) 本研究の目的

PT調査は、人の動きから都市を分析することを目的とし、各地方整備局や交通計画協議会を実施主体として10年毎に実施されている。調査は、調査区域内においてランダム抽出された5歳以上の住民を対象とし、調査日の詳細な行動についてアンケートを実施する。現在、全国45都市圏において実施されており、その地域特性や時代背景を活かした調査項目や属性分類によって行われている。しかし、人の動きを解析するという同一の調査目的にも関わらず、各都市圏において独自で実施されており、調査結果データ（以下、マスターデータ）に関してもフォーマットおよびデータの精度、品質に差がある。また、マスターデータ入手時に、利用を希望する都市圏の実施主体に対し、申請と許諾が必要である。さらに、マスターデータ入手後に各調査データ特有のデータクレンジング処理が必要となり、研究者にとって余分な労力を強いられている。

以上の背景から、本研究では、研究者が容易に利用可能な全国の人の流れに関する時空間データを整備し、サービスを実現することを目的とする。本稿では、全国対応に向け、まず東京都市圏以外の4つの都市圏PT調査データを詳細に比較分析し、マスターデータに関する問題点や処理方法を検討する。さらに、検討を行った処理方法にて4都市圏PTデータを動線解析プラットフォームに適用し、時空間データ提供サービスを実現する。対象となる4都市圏PT調査データには、三大都市圏である中京都市圏と京阪神都市圏、地方都市圏であり人口の多い道央都市圏と北部九州都市圏を用いた。

2. 4都市圏PT調査データとデータクレンジング

まず、本研究の全体の流れを図 1 に示す。本稿では、紙面の都合上、マスターデータからデータクレンジング処理方法、ジオコーデ

*キーワード: パーソントリップ, 時空間内挿, 人の流れ

**正会員, 博(工), 東京大学空間情報科学研究センター

(千葉県柏市柏の葉5-1-5, TEL:04-7136-4308,

E-mail:usui@csis.u-tokyo.ac.jp)

***非会員, 修(環境), 東京大学空間情報科学研究センター

****正会員, 博(工), 東京大学空間情報科学研究センター

ィングまでの検討結果を報告する。OD の経路探索以降の時空間内挿手法については、著者らの既往の研究を参照されたい³⁾。

(1) 4都市圏PT調査データの概要比較

4 都市圏における PT 調査データの概要を表 1 に示す。調査年や調査日、サンプル数は都市圏毎に様々である。また、調査範囲も、調査主体が独自に定めており、地方、県、市区町村の垣根を越え、様々であることが伺える。次に、各都市圏のマスターデータの詳細な比較を表 2 に示す。表より、同じ PT 調査においても都市圏毎に各項目の集計方法が異なっており、一括でデータ処理を行うことが難しいことがわかる。本稿では、この 4 都市圏において、詳細にマスターデータを比較し、各都市圏 PT 調査の特色を損なわないよう配慮しつつ、時空間内挿のためのクレンジング処理を行った。

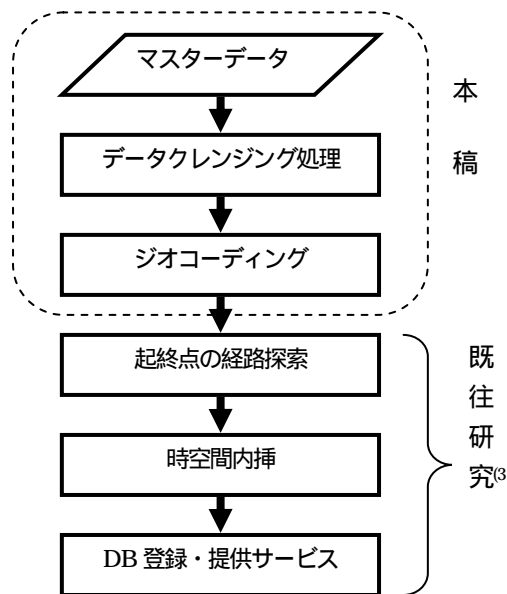


図 1 PT データの処理フロー

(2) マスターデータの特徴

PT 調査データは、地方整備局等へ利用申請することで直接入手することが可能である。提供されるマスターデータは、4 都市圏とも、数字のみの固定長アスキーデータである（一部、中京都市圏 PT 調査の中で可変長のものもある）。PT データを利用する際、まずこのアスキーデータの解読作業から始まる。今回のマスターデータの中には、データが右に 1 バイトシフトしたデータや個人情報保護の観点から意図的に削除されている（'9'で埋められている）データも存在していた。こういったデータを事前に処理しておくことは、誤データの利用を防ぐ上で、非常に重要なことである。

(3) トリップに関する記述と処理

個々の項目について見てみると、まず、個人 ID に関しては、調査票毎に固有の規則により割りふられていることがわかる。ただし、京阪神都市圏に関しては、個人情報保護の観点から ID が一意に割りふられていない。そのため、利用者側で一意に割りふる必要がある。

次に、トリップ日に関しては記載があるデータと記載がないデータがあるが、本稿では調査データを平日データとして一括に扱うため、同年の調査月の 1 日に修正し、統一処理を行っている。また、トリップ時刻に関しても、12 時間表記と 24 時間表記のデータがあるが、24 時間表記で統一して処理を行った。

(4) ゾーンコードについて

トリップの起終点および乗り換え地点を示しているゾーンコードの住所表記レベルは都市圏によって様々である。丁目レベルまで記載された詳細ゾーンレベルもあれば、町字を集約した C ゾーンレベルもある。また、起終点のみ精度が高く、乗り換え地点のみ住所精

表 1 本稿で使用した 4 都市圏 PT 調査データの概要

	道央都市圏	北部九州都市圏	中京都市圏	京阪神都市圏
調査主体	北海道開発局	九州地方整備局	中部地方整備局	近畿地方整備局
調査年	H18 (第4回)	H17 (第4回)	H13 (第4回)	H12 (第4回)
調査曜日	平日	平日	平日	平日・休日(全体の10%)
サンプル人数	10万人	21万人	26万人	47万人
総トリップ数	21万トリップ	48万トリップ	71万トリップ	120万トリップ
総ゾーン数	7,518	6,716	11,582	36,310
調査範囲	北海道札幌市	福岡県・佐賀県(一部)	愛知県・岐阜県・三重県	近畿圏

表2 4都市圏PTマスターデータの詳細比較

	道央都市圏	北部九州都市圏	中京都市圏	京阪神都市圏
個人ID	住所コード+バッチ番号+世帯番号+個人番号	住所コード+世帯番号+個人番号	中ゾーン+バッチ番号+世帯番号+個人番号	一意ではない
トリップ番号	個人毎に1から連番	個人毎に1から連番	個人毎に1から連番	個人毎に1から連番
サブトリップ	番号なし トリップ数も冗長	番号なし トリップ数も冗長	番号なし トリップ数も冗長	番号なし トリップ数も冗長
トリップ調査日	記載あり(平日のみ)	記載あり(平日のみ)	調査日不明	調査日不明
トリップ時刻	午前午後フラッグ 12時間表記	24時間表記	24時間表記	24時間表記
ゾーンコード	町字ゾーン	Cゾーン (町字集約レベル)	詳細ゾーン(自宅/勤務先) 町字ゾーン(それ以外)	町字ゾーン(自宅/勤務先) 市区町村ゾーン(それ以外)
乗換地点コード	鉄道駅・バス停・ 空港・港湾・IC	鉄道駅・バス停・ 空港・港湾・IC	鉄道駅・バス停・ 空港・IC	鉄道駅・高速バス・ 空港・港湾・IC
移動目的コード	17種類	22種類	16種類	12種類
交通手段コード	18種類	20種類	16種類	15種類
年齢コード	実年齢記載	実年齢記載	15段階(5歳刻み)	不明(入手不可)
職業コード	13+5種類	14種類	14種類	不明(入手不可)
住所コード	細目ゾーンコード	住所コード Cゾーンコード	詳細ゾーンコード	不明(入手不可)
拡大係数	拡大係数人口 拡大係数トリップ	拡大係数 補正拡大係数	拡大係数	拡大係数 補正後拡大係数

度の低い市区町村レベルのゾーンも存在する。ただ、提供されるマスターデータには、それ以外のデータは記載されていないため、基本的にはマスターデータに記載のあるゾーンレベルで処理を行っている。ただし、ゾーンレベルが直接時空間内挿後のデータの精度に影響を及ぼすことには利用の際、注意が必要である。

(5) ジオコーディング

PT調査データは、ゾーン化された起終点間の移動を記したデータである。ゾーンの精度は都市圏毎に異なっており、さらにそのゾーンコードは、その代表点における住所表記であるため、経緯度による時空間位置が把握できない。そのため、その代表点を経緯度に変換するための住所ジオコーディングを行う必要がある。ジオコーディングには、東京大学空間情報科学研究センターが提供する「CSVアドレスマッチングサービス」を利用したが、PT調査は過年度であるため、現在の住所表記におけるジオコーディングサービスでは、近年の市区町村合併や大字や字表記の有無などによる住所変更に対応できない。そこで、過年度の住所データや他社製の住所ジオコーディングシステム、デジタル地図などを用い、最終的には手作業を含め経緯度変換を行った。

(6) エラーデータ処理

マスターデータにおいては、入力データに不整合なエラーデータが少なからず含まれており、エラー処理は必要不可欠である。本研究では、主に表3の6項目については内挿処理の都合、削除対象データとしている。表より、北部九州については、全体の5%弱が削除対象データとなっており、データの品質向上が望まれる。

(7) 各調査データ固有データについて

調査項目のうち、移動目的や交通手段、職業については、地域特性や時代背景、経済状況によって、各都市圏間および調査年別においてもコードが異なっている。例えば、交通手段において、第4回道央都市圏PTでは、鉄道のコードに関して“路面電車”“地下鉄”“JR”と3つのコードに割りふられている。しかし、第4回北部九州都市圏PTでは、“地下鉄・モノレール”“西鉄電車・筑豊電鉄”“JR在来線”“JR新幹線”の4つのコードに分類される。また、北部九州において、S58年実施の第2回PT調査では、“地下鉄”“西鉄”“路面電車”“国鉄(在来)”“国鉄(新幹線)”の5つに分類されている。

以上のことから、これらの項目に関して、PTデータの都市圏と調査年毎に個別属性情報とし

表3 クレンジング処理により削除されたPTデータ数（紙面の都合上，2都市圏のみ記載）

削除対象となるデータ	道央都市圏	北部九州都市圏
1) 発着時刻の不明・欠落であるデータ	なし	4,719人（12,281トリップ）
2) ジオコーディングが不可なデータ	21人（64トリップ）	116人（331トリップ）
3) ゾーンコードが不明，または域外であるデータ	44人（48トリップ）	3,611人（10,255トリップ）
4) 交通手段が空路・海路，および調査域外のデータ	391人（678トリップ）	392人（837トリップ）
5) 鉄道の廃線・廃駅により補間が不可能なデータ	なし	1,341人（3,211トリップ）
6) サブトリップが欠落している（不連続）データ	なし	6人（29トリップ）
データ総数／総トリップ数	102,747人／212,399トリップ	211,019人／483,555トリップ

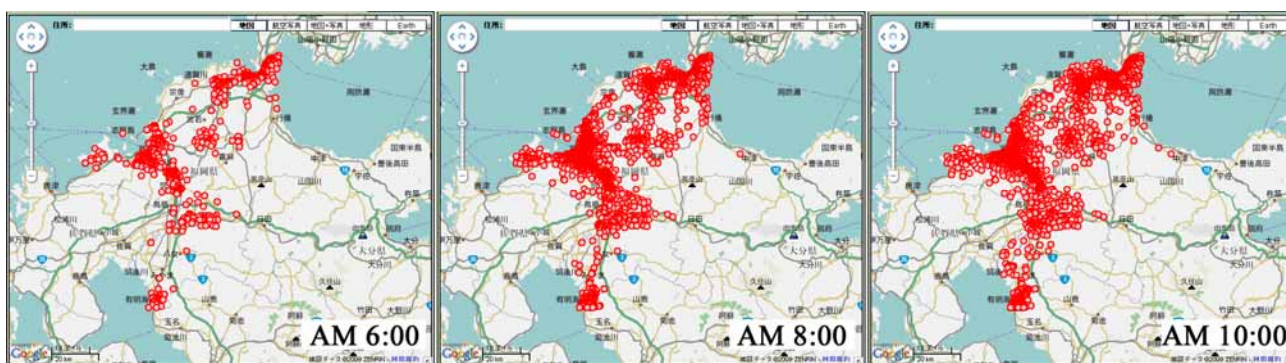


図2 AM6時からAM10時の人の位置の可視化（北部九州都市圏PTデータより）

て持つことが重要であり，多面活用のためにも統一処理すべきではないと考える。

3. 4都市圏PTデータの時空間内挿の実現と可視化

以上を踏まえ，4都市圏PTデータのクレンジング処理を行い，動線解析プラットフォームによって時空間内挿処理を行い，時空間データ提供サービスを実現させた。

図2は，時空間内挿処理結果を可視化したもので，北部九州地区の2万人のデータのAM6時～10時時点の人の位置を赤い点で表示したものである。図より，時間の経過に伴い，人の時空間分布の変化の様子を把握することが可能である。

4. まとめ

本稿においては，人の流れを俯瞰するためのデータの高度利活用を目的として，4つの都市圏を例にパーソントリップ調査データの違いを示し，調査結果の1分毎の人の位置を推定し，時空間内挿を行った。実装したデータは，当センターの共同利用データとして，東京PTデータ同様にサービスを開始する予定である。今後，本稿の知見を生かし，全国展開を行うと共に，人の行動を詳細に把握できるプローブパーソン技術とPT調査データの融合に関する研究⁴⁾についても検討を行い，今後，そういった様々な調査から得られるデータも動線解析プラットフォームにて

「人の流れ」データとして処理し，扱っていく予定である。

謝辞

本研究を行うにあたり，貴重なデータを提供して頂いた，札幌市交通計画課，九州地方整備局，中部地方整備局，近畿地方整備局殿に深く感謝致します。また，株式会社長大の増田様には大変お世話になりました。この場を借りてお礼を申し上げます。

参考文献

- 1) 中野敦：都市圏PT調査データの整備と活用，土木計画学研究・講演集，Vol.25，CD-ROM，2002年。
- 2) 人の流れプロジェクト：<http://pflow.csis.u-tokyo.ac.jp/>
- 3) 関本義秀，菊地英一，佐藤圭一，秋山祐樹：パーソントリップデータを活用した人の流れの時空間的な詳細化，交通工学研究発表会，第28回，pp197-200，2008年。
- 4) たとえば，牛尾龍太郎，倉内慎也，岡田拓治，森貴洋：PTデータとPPデータを同時に用いた交通手段選択モデルに関するパイロットスタディー，土木計画学研究発表会・講演集，Vol.38，CD-ROM，2008年。