

生活行動属性に基づく駅の分析手法の開発

相菌 敏子[†] 鈴木 敬[†]

人の移動が記録されたデータから人々の生活行動属性を抽出し、それらを用いて人や駅のような場所を分析する手法の開発を進めている。本稿では人々の一日をシーンの遷移と捉え、そのパターンを用いて駅を「利用のされ方」といった「質」的な観点から分析する手法について述べる。10万人の交通センサス・鉄道移動データから「勤務地に直行直帰」や「仕事帰りにプライベート」のような駅の利用パターンを抽出しそれらの出現頻度を用いて1185駅を特徴付けてクラスタリングした結果、駅を「住宅地」「勤務地」など5つに分類することができた。ユースケースでは本手法を用いて立川駅と吉祥寺駅を定量的に比較評価し、マーケティングや都市開発などで本手法が活用可能であることを示す。

Development of an Area Analysis Method based on Behavioral Characteristics

TOSHIKO AIZONO[†] KEI SUZUKI[†]

We develop an area and people analysis method based on behavioral characteristics, which are extracted from people's location data. In this paper, we represent one day as a sequence of scenes, then propose an area analysis method using patterns how people use stations. In an experiment using census of urban transportation data of over a hundred thousand people, patterns such as "straight home after work" and "private visit after work" are extracted, 1185 stations are categorized into 5 groups such as "residential area" and "office area". In use case, we compare Tachikawa sta. and Kichijoji sta., then our method can be applied to marketing, it's possible to analyze 'quality' of an area quantitatively.

1. はじめに

一般に、駅は一日の乗降客数や乗り入れている路線数などの数値を使って比較されることが多い。これら数値は駅の規模のような量的な特徴を表しており、その特徴に基づいた定量的な分析が行われている。例えば、国分寺駅は御茶ノ水駅と乗降客数がほぼ同じで、私鉄1社が乗り入れているJRの駅という点で類似している。

一方でマーケティングなどでは駅がどのような人々によってどのような目的で利用されているかを知りたいというニーズがある。先の例の国分寺駅は郊外に位置しており、周辺のマンションなどに住む人々が都心に通勤・通学するために利用されることが多いのに対して、御茶ノ水駅は都心の駅で、近隣にオフィスや学校が数多くあり、平日それらに通っている人々が利用している。このように「利用のされ方」といった駅の「質」的な特徴は乗降客数のような量的な特徴とは異なる駅の側面を捉えており、例えば駅ビルや近隣に出店を計画している企業や地域開発などでは施策上重要なデータである。

我々は移動データから人々の生活・行動面での特徴を表す指標値を生活行動属性として抽出し、それらを用いて人あるいは駅などの人々が生活する場所を分析する手法の開発を進めている。生活行動属性とは人の属性情報の一つで

あり、従来の年齢・性別等のような静的な属性情報（デモグラフィック属性）に対して、人の普段の生活や行動を表す動的な属性情報（ビヘイビアル属性）である。既報[1]では移動データとして個人の交通ICカードの利用履歴を用いて普段の生活圏やパターンを抽出し、空間（ナワバリ）、時間（リズム）、嗜好（テイスト）の3つの観点からカード利用者を定量的に分析できることを示し、生活行動属性のマーケティングなどへの適用可能性を述べた。

本稿ではこの生活行動属性を用いて駅を「利用のされ方」という「質」の観点から定量的に分析するモデルを提案する。実験では国土交通省が実施した平成17年の第10回大都市交通センサスデータ[2]を移動データとして用いて首都圏の鉄道利用者の一日の行動から生活行動属性を抽出し、駅を定量的に比較分析した結果を示す。

2. 関連研究と本研究の位置づけ

2.1 関連研究

人の移動データを用いてエリアを分析する研究として携帯電話の基地局接続情報を利用したものがある[3][4]。この研究では時間ごとに変化する人口の地理的分布を基地局接続情報から統計的に推計することで人口変動と都市空間の関係性を分析しており、まちづくり等での活用を狙っている。

また河本ら[5]は人がどのような生活をしているかといった「質」的側面を分析するため、人の生活を「典型日」の遷移に基づく特徴量で表現する手法を提案した。提案手

[†] 日立製作所 中央研究所
Central Research Lab. Hitachi Ltd.

法ではデータとして腕時計型センサによって取得された個人の加速度情報を用いており、それらから生活の質を表す特徴量を抽出し人の生活を定量化している。これにより分析対象者の長期間にわたる生活を「質」の観点から比較分析することができ、医療分野などでの活用を期待している。

2.2 本研究の位置づけ

本研究では、人の移動データから人の生活のパターンを行動属性として抽出し、人や駅などの場所をどのような人々がどのような目的で利用しているかといった「質」的観点から分析する手法の開発を行っている。既報[1]では移動データとして個人の交通 IC カードの利用履歴を用いたが、既存研究で用いられている携帯電話の基地局接続情報、あるいは自動車のプローブ情報なども移動データとして利用可能と考えている。よって本研究の分析対象は交通 IC カードの利用履歴に出現しうる鉄道利用者や駅に限定しておらず、広く人あるいは人々が生活する地域（エリア）を対象としている。

これらの点において本研究は、携帯電話の基地局接続情報を用いた人口変動分析と関連が認められるが、対象をどのような側面から分析するかといった点が大きく異なる。すなわち人口変動分析は人口といったエリアの「量」的側面をデータから推定しているのに対して、本研究ではどのような生活をしているかといった「質」的側面を扱う。また河本ら[5]の提案手法とは生活の質的側面を分析する点において類似しているが、利用するデータに差異がある。河本ら[5]の提案手法では比較的小規模な個人の生体情報を用いているのに対して、本研究では交通 IC カードの利用履歴や携帯電話の基地接続情報のような広域でかつ大量のデータを想定しているという点に特徴がある。

3. 生活行動属性に基づく駅の分析手法の提案

3.1 シーンに基づく生活行動属性と分析手法の考え方

本研究では、人々の一日の生活を「シーン」の遷移として捉える。シーンとは、人々が一日の生活においてどこでどのような目的で過ごしていたかを表す抽象的な概念である。例えば、「自宅で過ごす」「仕事をする」「余暇を過ごす」などは社会的な生活を共にする人々に共通して認められる活動であり、本研究ではこれらをシーンと呼ぶ。具体的には本研究では表 1 に示す 5 つのシーンを想定している。

既報[1]では交通 IC カードの利用履歴などからシーンを抽出し、一日におけるシーンの遷移を状態遷移図で表現してシーン間の遷移確率やシーンの遷移パターンの出現傾向などを指標値とすることで人の普段の生活を生活行動属性によって定量化できることを示した。既報[1]では分析対象は一人であるが、このようにシーンに基づいた生活行動属性を用いると、人々がどのような生活をしているか定量的に比較・分析することができる。図 1 にそのイメージ図を示す。図中の 2 人の人物は平日と休日でパターンの出現傾向が

表 1 本研究で想定するシーン

| # | シーン名 | 活動の内容 |
|---|---------|--------------------------|
| 1 | HOME | 自宅での活動(食事, 睡眠など) |
| 2 | WORK | 長時間のパブリックな活動(仕事, 学校など) |
| 3 | BRANCH | 短時間のパブリックな活動(出張, パートなど) |
| 4 | SPOT | 短時間のプライベートな活動(遊び, 買い物など) |
| 5 | LEISURE | 長時間のプライベートな活動(余暇など) |



図 1 シーン遷移の出現傾向による人の生活スタイル



図 2 シーンの遷移の出現傾向による駅の利用のされ方

異なりそれぞれ異なるスタイルで生活しているといえる。また分析対象は「人」だけではない。図 2 は人々のシーンの遷移においてどのようなシーンで人々に利用されたかによって駅を特徴付けた一例である。図中、上段と下段の駅は、それぞれ複数の人によって利用されているが、そのシーン(図中強調表示)が異なる。上段の駅は、WORK や BRANCH といったシーンで利用されることが多いのに対して、下段の駅は HOME において利用されることが多い。このことから上段の駅は勤務地の傾向が、下段の駅は住宅地の傾向が強いエリアに存在することがみてとれる。

3.2 提案手法の特徴

前述のようにシーンを用いると人や駅を定量的に分析することが可能である。これを実現するために我々は、まず人の移動データからルールを用いてシーンを抽出し、一日のシーンの遷移をデータに依存しない汎用的な形式で表現して一日のシーンの遷移のパターンを生成し、そのパターンを用いて分析対象の特徴量を求めて定量的に分析する手法を提案する。本手法の特徴は次の 3 つである。

3.2.1 推定ルールに基づくシーンの抽出

まず第一にデータからどのようにシーンを抽出するかが課題となる。これに対して本研究では、人の移動データからどこにどのくらいの時間人々が過ごしていたかといった滞在の情報を抽出し、滞在の時間帯や長さに基づきシー

表 2 シーンのタイプの推定ルール

| シーン | 出現順序 | 曜日 | 時間帯 | 時間長 | シーンの値の例 |
|---------|----------|----|--------|-------|---------|
| HOME | 一日の最初と最後 | - | - | - | 1, -1 |
| WORK | 上記以外 | 平日 | - | 7時間以上 | 5 |
| BRANCH | | | ~16:00 | 7時間未満 | 3 |
| SPOT | | | 16:00~ | 2 | |
| LEISURE | | 休日 | - | 7時間以上 | -5 |
| SPOT | | | - | 7時間未満 | -2 |

ンのタイプを推定する。シーンのタイプを推定するルールの一例を表 2 に示す。

なお前述の交通センサデータには人々がどの駅にどのような目的でいつ移動したかが記録されている。既報[1]では移動の目的（仕事や私事など）をすなわちシーンのタイプとし、人々が移動先に到着した時刻を滞在の開始時刻、次の移動の開始までを滞りの長さとし、それらを条件にシーンを分類できるか検証を行った。その結果、滞りの時刻や長さを用いると 95%以上の精度でシーンを分類できることを確認した。今後はこのような実データを用いて滞りの時刻・長さとの関係の分析し、シーンの推定ルールの改良を図っていく予定である。

3.2.2 24 次元ベクトルによるシーンの遷移の表現

シーンの遷移のパターンを抽出するには、人の一日をパターン抽出に適した形式で表現できることが望ましい。本研究では、シーンに対応する値をあらかじめ設定しておき、一日を 24 時間に分割して各時刻で過ごしていたシーンを値とする 24 次元ベクトルで表現する。なお時刻の単位はデータの特徴や分析の目的に応じて設定すればよく、人の行動が活発になる日中は 30 分単位、深夜は 90 分単位なども可能である。

シーンを表す値の設定はシーンのタイプと出現頻度を手掛かりに決めるなど、いくつか方法が考えられる。シーンの値の一列を表 2 の右欄に示す。表 2 に示す値は、筆者がシーンの意味を考慮し次のように設定した値である。すなわち、一般的に人に必須なプライベートなシーン HOME を基本の「1」とし、WORK はその対照、社会的なパブリックなシーンと位置付けられるため「5」とした。また出張などを想定したシーン BRANCH はパブリックであることから WORK に近い「3」、SPOT はそれより HOME に近く「2」とした。これら平日のシーンに対して休日は社会的にオフの状態であるとし、シーンの値をすべて平日とは対照となるようマイナスとした。このように一日のシーンの遷移を固定長のベクトルで表現することにより、膨大な人の移動データからシーンの遷移のパターンを効率的に抽出することが可能となる。

3.2.3 ベクトルのカスタマイズと多段階クラスタリング

人の行動属性に基づく人や駅の分析手法は、マーケティングのための顧客セグメンテーションなどを適用先として想

定している。そのような分析では同じデータに対して観点や条件を変えて試行錯誤的なプロセスを踏むことが多いため、人や駅も条件を柔軟にかえて分析できることが求められる。これに対して本研究では次の 2 つの手段を提供する。

(1) ベクトルのカスタマイズ

前項で示したように本研究では一日のシーンの遷移をベクトルで表現する。このベクトルを次の 2 つの方法でカスタマイズし、顧客ニーズにマッチした分析を実現する。

● ベクトルの重み付け

ある特定の駅に滞在したシーンあるいはある時刻に着目して人や駅の特徴を抽出したい場合、そのシーン/時刻に対応するベクトルの値に重みを付ける。図 2 の例を用いて説明する。ある特定の駅、例えば出店を計画している複数の候補駅を比較分析したい場合、それら候補駅を利用した人々の一日のシーンの遷移を前述のようにベクトルで表現し、さらに分析対象となっている駅を利用したシーンの値は 10 倍にする。例えば、自宅から勤務地に移動し勤務地からそのまま帰宅する「直出直帰」のシーン遷移において、分析対象の駅を WORK というシーンで利用した場合は WORK に対応するベクトルの値を 10 倍にする。これにより、一日のシーンの遷移が同じでもシーンを過ごした駅が分析対象であれば別のパターンとして抽出することができる。図 2 中、強調表示したシーンに 10 倍の重みが付けられる。

● ベクトルへの属性追加

前項で示したベクトルはシーンの遷移を表すが、他の属性にも着目してパターンを抽出したい場合ベクトルに属性を追加する。例えばシーンの遷移のパターンを性別や年代など人の属性も考慮して抽出したい、あるいは曜日やあるサービスの会員かどうかなどの観点でもよい。それらを表す数値をベクトルに追加することによりその属性にマッチする一日の出現傾向に特徴が見られれば各々パターンを抽出することができる。

(2) 多段階クラスタリング

シーンの遷移のパターンを一週間や特定の曜日など期間に関する分析目的に合わせて抽出できるよう、一日の遷移のパターンを基本とした多段階クラスタリングの手法を提供する。具体的には、まず基本となる一日のシーンの遷移のパターンを抽出しておき、次に分析期間の日を取得してマッチするパターンの番号（コード）に変換しパターン列を表すベクトルを生成して期間のパターンを抽出する。抽出条件は、期間のほか毎週水曜日、あるいは祝日を含む休日などでもよい。このように、まず基本となる一日のシーンの遷移のパターンを抽出しておけば期間に関するニーズに柔軟に対応することができる。と考える。

3.3 処理手順

図 3 は、移動データから駅の利用パターンを抽出し、その利用パターンを用いて駅を特徴付けて分析する処理の手順を示したイメージ図である。図 3 に示すように駅の分析は、

表 3 利用パターン一覧

| # | 利用パターン名 | ベクトル数 | 平均シーンベクトル | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|-------------------|---------|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | | | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 1 | 直行直帰(定時) | 61,694 | 1 | 1 | 1 | 4 | 23 | 44 | 48 | 50 | 50 | 50 | 50 | 50 | 47 | 38 | 15 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 2 | 直行直帰(残業) | 26,712 | 1 | 1 | 1 | 3 | 19 | 40 | 46 | 50 | 50 | 50 | 50 | 50 | 49 | 49 | 27 | 14 | 6 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 3 | 勤務地 直行直帰(午前のみ) | 8,083 | 1 | 1 | 2 | 4 | 15 | 26 | 28 | 30 | 27 | 23 | 17 | 12 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 4 | 勤務地 直行直帰(午後のみ) | 3,358 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 5 | 勤務地 勤務地の後私事 | 3,855 | 1 | 1 | 1 | 3 | 16 | 39 | 47 | 48 | 49 | 47 | 44 | 43 | 40 | 37 | 28 | 3 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 6 | 勤務地 勤務地の後業務先 | 5,059 | 5 | 5 | 5 | 7 | 16 | 26 | 27 | 23 | 20 | 17 | 13 | 10 | 8 | 6 | 5 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | |
| 7 | 業務先 午前中心 | 11,535 | 1 | 1 | 1 | 4 | 29 | 45 | 45 | 37 | 28 | 15 | 8 | 4 | 3 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 8 | 業務先 午後中心 | 8,878 | 1 | 1 | 1 | 1 | 3 | 5 | 9 | 19 | 25 | 29 | 30 | 30 | 30 | 24 | 17 | 10 | 6 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 9 | 業務先 夕方まで | 8,693 | 1 | 1 | 1 | 1 | 3 | 4 | 5 | 5 | 5 | 7 | 18 | 24 | 28 | 30 | 29 | 25 | 14 | 9 | 4 | 2 | 1 | 1 | 1 | 1 | 1 | |
| 10 | 業務先 深夜まで | 2,440 | 1 | 1 | 1 | 1 | 3 | 5 | 8 | 11 | 14 | 17 | 21 | 23 | 25 | 27 | 28 | 28 | 30 | 30 | 30 | 29 | 29 | 28 | 28 | 28 | 28 | |
| 11 | 私事 勤務地の後(午後) | 2,079 | 1 | 1 | 1 | 2 | 4 | 5 | 6 | 7 | 10 | 14 | 16 | 19 | 18 | 17 | 14 | 11 | 8 | 7 | 5 | 3 | 3 | 3 | 3 | 3 | 3 | |
| 12 | 私事 勤務地の後(夕方) | 6,116 | 1 | 1 | 1 | 2 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 9 | 15 | 17 | 15 | 12 | 7 | 4 | 3 | 3 | 3 | 3 | |
| 13 | 自宅 直行直帰(定時) | 57,041 | 10 | 10 | 10 | 10 | 8 | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | |
| 14 | 自宅 直行直帰(残業) | 22,622 | 10 | 10 | 10 | 10 | 8 | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | |
| 15 | 自宅 直行直帰(午前のみ) | 6,082 | 10 | 10 | 10 | 10 | 3 | 8 | 8 | 8 | 8 | 8 | 9 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | |
| 16 | 自宅 直行直帰(午後のみ) | 3,843 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 9 | 7 | 6 | 6 | 5 | 5 | 5 | 6 | 7 | 8 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| 17 | 自宅 勤務地の後私事 | 3,022 | 10 | 10 | 10 | 10 | 8 | 6 | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 2 | 2 | 2 | 4 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | |
| 18 | 自宅 勤務地の後業務先 | 4,866 | 10 | 10 | 10 | 10 | 8 | 6 | 5 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 6 | 7 | 8 | 9 | 10 | 10 | 10 | 10 | 10 | |
| 19 | 出発 勤務地の後業務先 | 1,919 | 10 | 10 | 10 | 10 | 8 | 7 | 6 | 6 | 5 | 4 | 4 | 3 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 20 | 出発 直行直帰(定時) | 4,199 | 10 | 10 | 10 | 10 | 8 | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 21 | 到着 勤務地(定時) | 896 | 1 | 1 | 1 | 3 | 4 | 5 | 6 | 6 | 7 | 7 | 8 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | |
| 22 | 到着 勤務地(残業) | 3,945 | 1 | 1 | 1 | 3 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 7 | 8 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | |
| 23 | 到着 業務先 | 1,438 | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 6 | 7 | 8 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | |
| 24 | 不明 | 117 | 1 | 1 | 1 | 3 | 4 | 4 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| | | 262,092 | | | | | | | | | | | | | | | | | | | | | | | | | | |

表 4 利用スタイル一覧

| 利用スタイル | 駅数 | 特徴 | 例 |
|-------------|-------|-----------------------------|--------------------|
| 住宅地 | 396 | 自宅パターンで利用されることが多い | 大泉学園, 上大岡, 二俣川 |
| 勤務地 | 229 | 勤務地パターンで利用されることが多い | |
| 勤務地/業務先(日中) | 199 | 直行直帰中心とした勤務地パターンと業務先パターンが多い | 新宿, 東京, 渋谷, 新橋, 田町 |
| 勤務地(午前中心) | 30 | 「勤務地の後業務先パターン」が多い | ひばり丘, 新百合丘, 鶴ヶ峰 |
| 混在 | 560 | 自宅と勤務地のパターン両方で利用 | |
| 住宅地より | 364 | 自宅パターンのほうが多い | 青葉台, 町田, 綱島, 日吉 |
| 勤務地より | 196 | 勤務地パターンのほうが多い | 大宮, 大森, 本厚木, 浦和 |
| 総計 | 1,185 | | |

どで自宅以外に当日または前日宿泊した場合がこのようなパターンにマッチするが、自宅から出発し帰宅した場合でもこれらパターンにマッチすることがある。すなわち交通センサスでは1人あたり1日3回までの移動しか記録しない。そのためそれ以上移動した人は帰宅のデータが記録されず、結果として一日の最初と最後の駅名が一致しない「出発パターン」「到着パターン」が生成された。最後の「不明パターン」は移動目的データが欠損しているパターンである。そのため次のステップである特徴ベクトルの生成では利用しないこととする。表3中ベクトル数を見ると、利用パターンのうち90%は「勤務地パターン」(42%)「業務先パターン」(12%)および「自宅パターン」(39%)であり、「私事パターン」が少ない(3%)。この結果は、既報[1]の「交通センサス鉄道移動データの大半が直行直帰型の通勤である」という分析結果と一致する。

4.4 特徴ベクトルの生成

前記抽出した利用パターンを用いて滞在データに出現した駅1,335駅のうち利用した人の数が10回以上の駅1,185駅(89%)を対象として駅の特徴ベクトルを生成する。具体的には、各駅の利用を表すシーンベクトルが表3に示すどのパターンに当てはまるかチェックして各パターンの出現頻度をカウントし、利用パターン数23を次元数、パターンの出現頻度を値とするベクトル(1,185件)を生成した。

なお3.3で述べたように分析対象の駅や利用パターンによって出現頻度に偏りがある場合がある。そこでカウントした出現頻度を、全駅における利用パターンの出現頻度およびその駅の利用者数に基づき重み付けを行った。重み付けにより、各駅における利用者総数、すなわち駅の規模の影響をできるだけ少なくしている。

4.5 利用スタイルの抽出

前記生成した1,185駅の特徴ベクトルをクラスタリングする。クラスタリングは、利用パターンの抽出同様、Rを用いてk-means法で行う。クラスタ数は試行でのクラスタの評価値に基づき5とした。その結果を表4に示す。なお表4中「利用スタイル」は、結果を参照して筆者が付与した名称である。表4に示すように利用スタイルは大きく3つに分かれる。すなわち「住宅地」は表3の利用パターンのうち「自宅パターン」の出現頻度が高いクラスタ、「勤務地」は「勤務地パターン」と「業務先パターン」の出現頻度が高いク

4.3 利用パターンの抽出

前記生成した265k件のシーンベクトルをクラスタリングする。クラスタリングには、統計用フリーソフトR[7]を用いて非階層クラスタリングの代表的なアルゴリズムであるk-meansを適用した。クラスタ数は、試行によるクラスタの評価値を参考に24とした。クラスタリングの結果を表3に示す。表中「利用パターン名」とは筆者がクラスタの特徴を参照して付けた名称であり、「ベクトル数」とはクラスタに属するシーンベクトルの数、すなわちそのパターンで駅を利用した人の延べ数に相当する。また「平均シーンベクトル」とはクラスタに属するシーンベクトルの平均ベクトルであり、ベクトルの値が大きければ濃く、値が小さければ薄くグラデーション表示している。

表3に示すように生成された24パターンは、大きく7つに分類できる。表中「勤務地パターン」と「業務先パターン」は主な利用の目的がそれぞれ「勤務先」、「業務先」であるパターンであり、利用した時間帯によって各々細分化されている。また「私事パターン」は私事を主な目的として利用したパターンであり、勤務地に行った後、午後早めの時間帯に私事で駅を利用したパターンと、夕方に利用したパターンの2つに分けられる。「自宅パターン」は一日の最初と最後に駅を利用したパターンであり、一日において上記「勤務地パターン」「業務先パターン」「私事パターン」の組み合わせで出現する。この「自宅パターン」に対して、「出発パターン」と「到着パターン」は一日の最初に利用した駅と一日の最後に利用した駅が一致しないパターンである。「出発パターン」は自宅から出発して同じ駅に帰宅しなかったパターン、「到着パターン」は他の駅から出発して帰宅したパターンに相当する。旅行や出張な

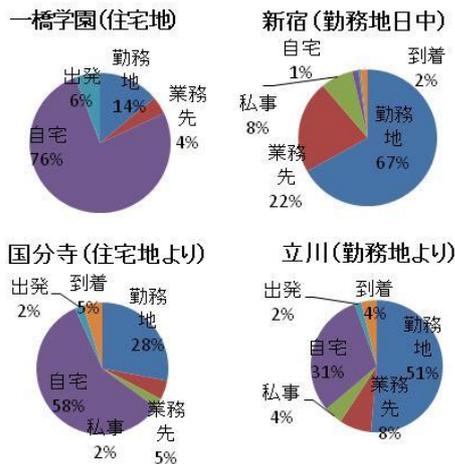


図 4 利用スタイル別利用パターンの出現割合の一例

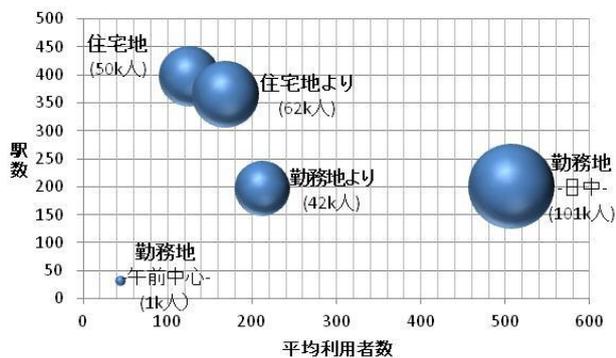


図 5 利用スタイルにおける駅数と利用者数の関係

ラストである。「勤務地」にはさらに2つのクラスタがあるが、このうち「勤務地(午前中心)」は表3の#6「勤務地の後業務先」(その駅を勤務地として利用し、午後は業務先に移動したパターン)の頻度が高いことを特徴とするクラスタである。このクラスタに属する駅の数30と少なく、このクラスタは1,185駅中、特異な駅が集まったクラスタであると言える。「混在」は「自宅パターン」と「勤務地パターン」の頻度が共に高いクラスタであり、比較的「自宅パターン」の出現頻度が高い「住宅地より」と勤務地パターンの出現頻度が高い「勤務地より」に分かれる結果となった。

図4に「勤務地(午前中心)」を除く4クラスタに属する駅の一例を挙げ、各駅の利用パターンの出現頻度の割合を示す。ただし利用パターンは表3に示すように23あるが、図4では見やすさのため「通勤先」、「業務先」などの6つにまとめて表示している。

5. 適用結果の考察

5.1 利用スタイルと駅の利用者数

表4に示す利用スタイルに関し、各スタイルにマッチする駅の数とその平均利用者数をプロットしたものが図5である。図5中、バルーンの大きさは駅の利用者数(=各スタイルにマッチする駅の利用者数合計)を表す。

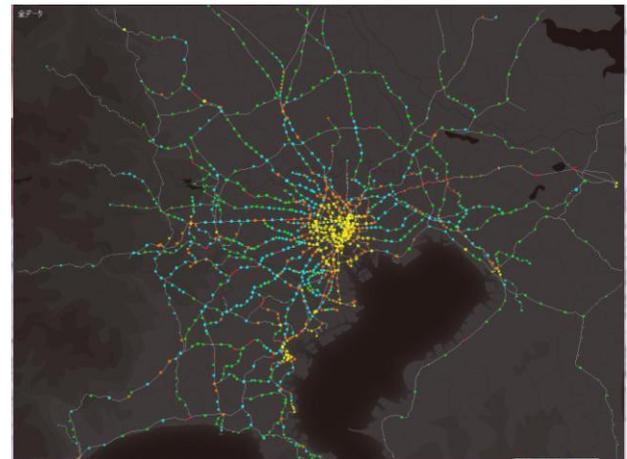


図 6 利用スタイル別駅の所在地

図5より、「勤務地(勤務地/業務先)」は駅数199駅と少ないが、駅の平均利用者数が多い(508人)。このことから通勤・通学、および業務で移動する利用者はいくつかの駅に集中していることが分かる。一方で「住宅地」は駅の数396駅と「勤務地」よりも多いが、平均利用者数は少ない(126人)。これは、自宅は通勤・通学、業務先より駅が分散していることを表している。「住宅地より」と「勤務地より」は、平均利用者数は「住宅地」と「勤務地」の中間の「住宅地」近くに位置しているが(それぞれ169人と212人)、マッチする駅の数「住宅地より」は「住宅地」に、「勤務地より」は「勤務地」に近い。このことから勤務地は利用者が少数の駅に集中し、住宅地は分散する傾向があることが分かる。

5.2 駅の所在地と利用スタイル

クラスタリングした駅を利用スタイル別の色で地図上にプロットしたのが図6である。図6と表4より「新宿」「渋谷」など「勤務地(勤務地/業務先)」(図中黄色の点)の駅は都心および横浜の中心地に集中しているのに対して、「大宮」「浦和」「立川」のような「勤務地より」(オレンジ)は都心から郊外へ伸びる路線上に間隔をおいて存在している。前者が首都圏全体の中心地であるのに対して、後者は都心から伸びる路線の急行の停車駅あるいはさらに郊外に伸びる路線のターミナル駅であり、郊外における中心地という位置づけであることが見て取れる。このような駅に通勤/通学している人が多いと思われるのが「大泉学園」「上大岡」のような「住宅地」(緑色)の駅および「青葉台」、「町田」のような「住宅地より」(水色)の駅である。前者が都心からかなり離れた郊外に分散しているのに対して、後者は都心へのアクセスがよい路線上に位置することから、「住宅地より」の駅は同じ路線にあるいは都心等を経由した通勤/通学先でもあると予想される。

5.3 ユースケース

本節では、利用パターンに基づく駅の特徴量と利用スタイルによる駅の種類を活用した分析のユースケースを示す。

具体的には次のような分析ニーズを想定する。

| | |
|-------|---|
| 業種 | サービス系企業 |
| 背景 | 仕事をしている人をターゲットとした店舗を郊外に出店する計画がある。その駅を勤務先最寄駅とする人が同僚等と一緒に平日の仕事の帰りがけに立ち寄るイメージ。まず中央線沿線の比較的大きな駅に出店を予定しており、立川駅と吉祥寺駅が候補に挙がっている。また中央線の他に同様の駅があれば出店を検討したい（東急線など） |
| 分析ニーズ | ①立川駅と吉祥寺駅のうち、出店に適しているのはどちらか？ ②中央線以外にも出店に適した駅はないか？ |

表4に示す利用スタイルのうち、仕事を持った人が勤務先の近くで仕事帰りに立ち寄ることが多そうな駅は、#2「勤務地（日中）」か、#5「混在」の「勤務地より」であるが、このうち後者は郊外に多いことから「勤務地より」に出店に適した駅が含まれると考える。現在候補となっている立川駅と吉祥寺駅はいずれも「勤務地より」の駅であり、ともに適している可能性が高いが、さらに詳しく利用パターンによる駅の特徴量を見ると2つの駅で違いが出てくる。

図7に立川駅と吉祥寺駅の利用パタンの出現頻度を示す。図7より、立川駅と吉祥寺駅はともに「勤務地（直行直帰定時）」のような勤務地のパターンと、「自宅（直行直帰定時）」のような自宅パタンの両方の出現頻度が高いが、立川駅は前者の出現頻度のほうが高い（①）のに対して、吉祥寺駅はその2つの出現頻度がほぼ同程度という点で異なる。

表3の利用パターンを見ると仕事帰りの時間帯（18時～20時）に対象駅に滞在するパターンとして、「勤務地（勤務地の後私事）」、「業務先（夕方まで）」、「私事（勤務地の後夕方）」などがある。これらの出現頻度を比較すると立川駅と吉祥寺駅である程度の差がみられるのは「私事（勤務地の後夕方）」である。吉祥寺駅のほうが若干他の駅での仕事の帰りにプライベートで立ち寄る人が多い（②）。①、②のことから顧客企業がターゲットとしている「仕事の帰りがけに勤務先最寄駅で立ち寄りする人」は立川駅のほうに多いと言える。

また同じ「勤務地より」タイプの駅のうち、利用パタンの出現傾向が立川駅と類似している駅の一つとして武蔵小杉駅がある（図7参照）。一方で三軒茶屋駅のように吉祥寺駅と類似している駅は出店の候補としての優先度は低くすべきである。

6. まとめ

人の移動データからシーンを抽出して一日のシーンの遷移のパターンを抽出し、その出現傾向で駅を特徴付けて分析する手法を開発した。国土交通省の交通センサス鉄道移動データを使って10万人のシーンの遷移から24個の駅の

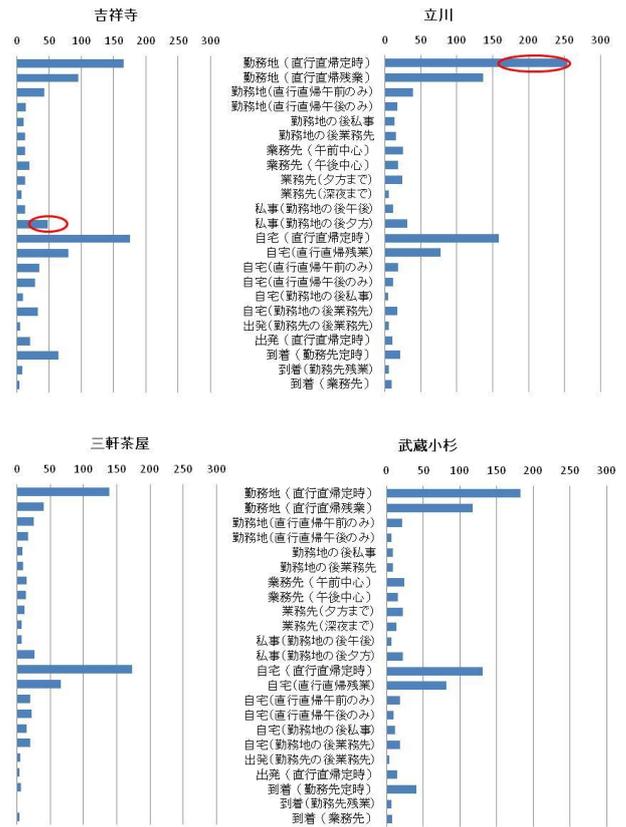


図7 利用パタンの出現頻度による4駅の特徴付けの結果

利用パターンを抽出し、その出現頻度によって1,185駅を特徴付けてクラスタリングしたところ「住宅地」、「勤務地」（2タイプ）、「住宅地と勤務地の混在」（2タイプ）の計5つの駅クラスターが生成できた。駅の特徴を比較分析したいというニーズに対して、本手法を適用することにより、次のような効果が期待できる。

(1) 質的な側面からの分析が可能

従来は利用者数のような「量」的な分析が主であるのに対して、本手法は利用の目的やシーンといった人の行動属性に着目しており、対象を「質」的な側面から分析できる。利用者が同じでも利用の目的やシーンによってニーズが変わる。例えば、同じ駅でも仕事帰りにプライベートで立ち寄った場合は娯楽性重視、勤務地から出張（業務先）できた場合は利便性重視となる。企業の出店戦略やマーケティングなどを行う企業にとって、エリアの質的な分析は有用な情報であると考えられる。

(2) 定量的な比較分析が可能

本手法は、分析対象の駅がどのような使われ方をしているかを利用パタンの出現頻度で表現する。利用のパターン（一日のシーン遷移）のバリエーションは無限ではなく、高々数十個であると考えられる。そのような利用パターンを使って駅を特徴付けることにより、駅の使われ方の季節ごとの変化や使われ方が類似している駅の抽出（利用スタイル分析）などを定量的かつスケーラブルに行うことができる。

参考文献

- 1) 鈴木敬,相菌敏子:交通 IC カード利用履歴を用いた生活行動属性指標の提案,信学技報,LOIS2011-84,pp67-72(2012).
- 2) 国土交通省:平成 17 年大都市交通センサス第 1 0 回,
http://www.mlit.go.jp/sogoseisaku/transport/sosei_transport_tk_000007.html
- 3) NTT ドコモ:モバイル空間統計に関する情報,
http://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/
- 4) 日経 B P : N T T ドコモが巨大マイニング設備構築,日経コミュニケーションズ,2009 年 10 月 1 日号.
- 5) 河本健,栗山裕之,田中毅,新谷隆彦,鈴木敬:典型日遷移列に基づく生活モデルの提案,信学技報,LOIS110-207, pp.19-24,(2010).
- 6) 北 研二,津田 和彦,獅々堀 正幹:情報検索アルゴリズム,共立出版(2002)
- 7) 統計解析ソフトウェア R プロジェクト :
<http://www.r-project.org/>