

# 人の移動データに基づく地域のクラスタリングとその応用

落合 桂一<sup>1,a)</sup> 寺田 雅之<sup>1</sup>

概要：人々の社会生活において移動は基本的な行動であり、人々の移動は経済活動や交通、公衆衛生など様々な分野と関わりが深い。人々の移動データと応用先の分野のデータを組み合わせて解析することで、経済活動の予測や交通の最適化など各分野での分析精度の向上や、より詳細な分析が期待される。そこで本研究では、人の移動データに基づいて地域をクラスタリングすることの有用性を実データを使って評価する。まず、約 11 万ユーザの半年間の移動データから地域をクラスタリングする。次に、クラスタリングの有用性を、(1) 市町村別観光消費額の年間推移の類似性、(2) 入込観光客数の予測問題という 2 つの評価実験により検証する。予測問題では、行政が決めた地域区分と比較して予測誤差を削減できることを確認した。

## 1. はじめに

人々の社会生活において移動は基本的な行動であり、人々の移動は経済活動や交通、公衆衛生など様々な分野と関わりが深い。人々の移動データと応用先の分野のデータを組み合わせて解析することで、経済活動の予測や交通の最適化など各分野での分析精度の向上や、より詳細な分析が実施可能になると期待される。例えば、小売業の販売額を都道府県ごとに予測することを考える。このとき、予測対象の都道府県の過去の販売額推移だけでなく、近隣の都道府県の推移を活用することで予測精度を向上させることが考えられる。具体的には、隣接している都道府県や同一の地域の情報を利用することができる。しかし、隣接している都道府県でも地理的な制約で 2 つの都道府県を結ぶ交通機関が少ないことに起因して人の移動が少ない場合、隣接している都道府県の情報を加味しても予測精度の向上には寄与しないと考えられる。そのため、都道府県の地理的な配置のみではなく、人の移動に基づいて都道府県の結びつきを考慮することが有用であると考えられる。

そこで本研究では、人の移動データに基づいて地域をクラスタリングすることの有用性を実データを使って評価する。本研究は、(1) 移動データの収集、(2) 移動データに基づくクラスタリング、(3) 有用性の評価の 3 つのパートから構成される。人の移動データには、Foursquare 社が提供する Swarm[1] というチェックイン共有 SNS でのチェックインデータを利用する。期間は 2018/10/1-2019/3/31 の半年間で、約 11 万ユーザの約 730 万件のチェックインデータ

を収集した。市町村をノードとし、市町村間の移動があった箇所にエッジを設定することで市町村グラフを構築した。エッジの重みには市町村間を移動したユニークユーザ数を利用する。そして、市町村グラフに対してノードのクラスタリングを行う。グラフのクラスタリング手法の違いによる結果への影響を考察するため、Newman Algorithm[2] および Clauset-Newman-Moore (CNM) Algorithm[3] を利用してクラスタリングを行い結果の比較を行う。最後に、クラスタリングの有用性を観光消費額の推移の類似性、および入込観光客数予測問題で検証する。観光消費額の推移の類似性の評価では、神奈川県内の各市町村の観光消費額の推移を、行政が定めた地域区分と移動データに基づく地域区分で比較することで評価する。入込観光客数予測では、神奈川県内の各市町村の入込観光の予測に市町村のクラスタリングの結果を特徴量として利用することで、行政が定めた地域区分との予測精度への影響を比較し評価する。

本研究の貢献は以下の通りである。

- 大規模な移動データの統計データに基づき地域のクラスタリングを行い、行政により決められた地域区分と移動に基づく地域区分では異なることを明らかにした。
- 移動データに基づくクラスタリングにより、観光消費額の推移がより類似している市町村をクラスタリングできることを定量的に示した。
- 移動データに基づくクラスタリングを地域に関する統計指標の予測に活用することを提案し、神奈川県の入込観光客数の予測問題において有効性を検証した。

本稿の構成は以下の通りである。次章にて人の移動データの分析の既存研究について概観する。次に、3 章で提案手法の詳細を説明し、4 章では実データを用いた評価を行

<sup>1</sup> 株式会社 NTT ドコモ  
NTT DOCOMO, INC.  
<sup>a)</sup> ochiaike@nttdocomo.com

い、既存手法と提案手法の精度について検証する。最後に5章にて本研究のまとめおよび今後の課題について述べる。

## 2. 関連研究

人の移動データの分析に関わる研究は(1) 個々人の移動を対象にした研究と、(2) 個人のデータを集約し集団としての移動を分析した研究の2つに大きく分けられる。本章ではそれぞれの関連研究についてまとめる。

### 2.1 個人の移動データの分析

個人の位置情報を解析する研究では、過去の位置情報を元に将来の目的地予測 [4]、移動手段の推定 [5]、訪問した施設 (POI: Point of Interest) の推定 [6] など、様々な研究が行われている。また、過去に訪問した POI を元にユーザの嗜好や場所の特性から POI 推薦を行う研究も盛んに研究されている [7], [8], [9], [10]。これらの研究では、推定モデルの入力が個人の位置情報履歴であり、出力 (予測や推定の結果) も個人の位置や移動に関する情報となる点が共通している。

### 2.2 集団の移動データの分析

本研究とより関連するのは本節で述べる集団の移動に関する分析である。Thiemann ら [11] は、ある場所  $i$  から  $j$  への紙幣の移動量が人の移動量と相関するという仮定のもと、wheresgeorge.com というアメリカでの紙幣の場所の追跡ゲームサイトのデータを利用し、人の地域間の移動データを収集した。郵便番号単位の地域分割を行い、各地域をノード、地域間にエッジを設定し、エッジの重みには移動量を利用し人の移動をグラフとして表現した。そして、Modularity に基づくコミュニティ検出手法を適用することで地域のクラスタリングを行なった。その結果、行政による地域区分と異なるクラスタが見つかったと報告されている。Rinzivillo ら [12] は、車載 GPS のデータセットを利用して、各位置をノード、移動があった位置同士にエッジを設定することで移動を表すグラフ構築した。構築したグラフに対して、infomap というネットワークのコミュニティ検出手法を適用することで地域のクラスタリングを行なった。移動に基づくクラスタリング結果と行政区分がマッチしていたと報告している。移動データではないが、地域の結びつきを分析した研究に Ratti らの研究がある [13]。Ratti らはイギリス国内における固定電話の発信/着信のデータから、各場所をノード、場所間の結びつきを通話時間としてグラフを構築し、Modularity に基づくコミュニティ検出手法を適用することで地域のクラスタリングを行なった。クラスタリングの結果を活用した分析の事例として、イギリスの各地域が独立した場合の影響を分析し、スコットランドが独立した場合の影響が、他の地域が独立した場合と比べて相対的に少ないことを試算した。こ

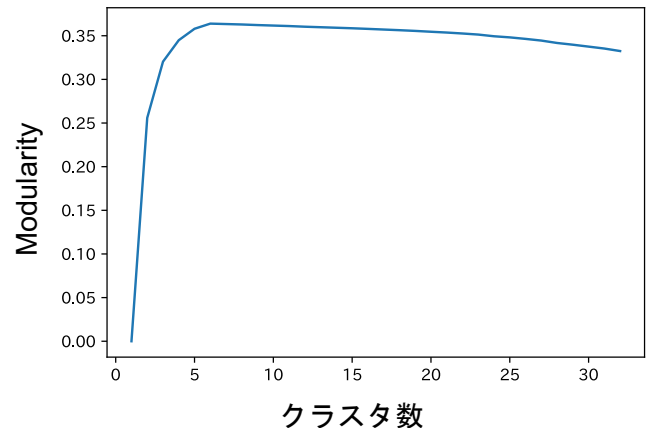


図 1 CNM Algorithm においてクラスタ数を変えたときの Modularity の推移

れらの研究では、移動などの人の生活のデータに基づいて地域をクラスタリングし、行政区分との比較や各地域の国全体への影響を評価している。本研究でも、既存研究と同様に移動に基づくクラスタと行政区分との比較は行いが、日本を対象に移動に基づいて地域をクラスタリングした研究は著者らの知る限り行われていない。また、地域のクラスタリング結果を予測問題に応用し評価した研究はない。

## 3. 移動データに基づく地域のクラスタリング

### 3.1 データ

人の移動データとして、Foursquare 社が提供するチェックイン共有 SNS である Swarm [1] のチェックインデータを利用する。Swarm の API では自分にチェックイン情報は取得可能であるが、プライバシー保護のため他のユーザのチェックイン情報は取得できない。Swarm では、ある場所でチェックインする際、Twitter にチェックイン情報を連携することでツイートとして投稿することができる。本研究では、Twitter API [14] を利用して Swarm のチェックインが Twitter に投稿されたデータを収集した。期間は 2018/10/1-2019/3/31 の半年間で、約 11 万ユーザの約 730 万件のチェックインデータを収集した。

### 3.2 移動データに基づくグラフ構築

収集したチェックインデータから各ユーザの連続するチェックインをユーザの移動とみなす。実際には Twitter 連携されなかったチェックインや、すべての場所でチェックインを行っていない可能性があり、その点は制限事項である。4 章の評価において、神奈川県の入込観光客数の予測問題で評価を行うため、チェックインデータを神奈川県内のデータにフィルタリングした。ユーザ数は約 3 万 8 千で、チェックイン数は約 53 万件である。位置情報の粒度は、各チェックインが行われた市町村レベルに統一し、

表 1 神奈川県が定めた地域区分

地区名	市町村名
横浜地域	横浜市
川崎地域	川崎市
横須賀三浦地域	横須賀市、鎌倉市、逗子市、三浦市、葉山町
県央地域	相模原市、厚木市、大和市、海老名市、座間市、綾瀬市、愛川町、清川村
湘南地域	平塚市、藤沢市、茅ヶ崎市、秦野市、伊勢原市、寒川町、大磯町、二宮町
県西地域	小田原市、南足柄市、中井町、大井町、松田町、山北町、開成町、箱根町、真鶴町、湯河原町

表 2 Newman Algorithm による地域区分

地区名	市町村名
横浜地域	横浜市
川崎地域	川崎市
横須賀三浦地域	横須賀市、逗子市、三浦市、葉山町
県西地域	二宮町、南足柄市、大磯町、小田原市、平塚市、真鶴町、箱根町、開成町
湘南地域	茅ヶ崎市、藤沢市、鎌倉市
県央・県西地域	中井町、伊勢原市、厚木市、大井町、大和市、寒川町、山北町、座間市、愛川町、松田町、海老名市、清川村、湯河原町、相模原市、秦野市、綾瀬市

表 3 CNM Algorithm による地域区分

地区名	市町村名
横浜地域	横浜市
川崎地域	川崎市
横須賀三浦地域	横須賀市、逗子市、三浦市、葉山町
県央地域	中井町、伊勢原市、厚木市、大和市、山北町、座間市、愛川町、海老名市、清川村、相模原市、秦野市、綾瀬市
湘南地域	大磯町、寒川町、平塚市、茅ヶ崎市、藤沢市、鎌倉市
県西地域	二宮町、南足柄市、大井町、小田原市、松田町、湯河原町、真鶴町、箱根町、開成町

市町村間の移動を抽出した。政令指定都市では区レベルの情報まで取得可能であるが、3.4 節で評価に利用する入込観光客数のデータが市レベルであったため、粒度を合わせ市レベルに統一した。その後、各市町村をノードとし、移動があった市町村間にエッジを張ることで市町村のグラフを構築した。エッジには移動したユニークユーザ数による重みを設定する。

### 3.3 地域のクラスタリング

グラフのクラスタリング手法は多数提案されている。本研究では、クラスタリング手法の違いによる結果への影響を考察するため Newman Algorithm[2] および Clauset-Newman-Moore (CNM) Algorithm[3] を利用してクラスタリングを行なった。どちらのアルゴリズムもクラスタリングの質を Modularity[15] により評価する手法である。

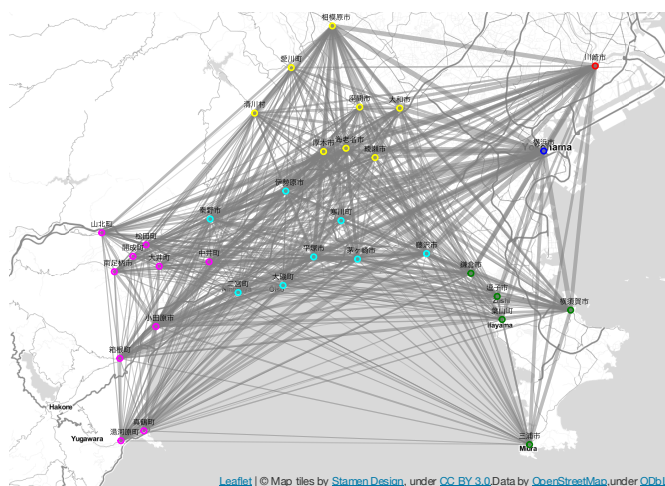


図 2 神奈川県が定めた地域区分

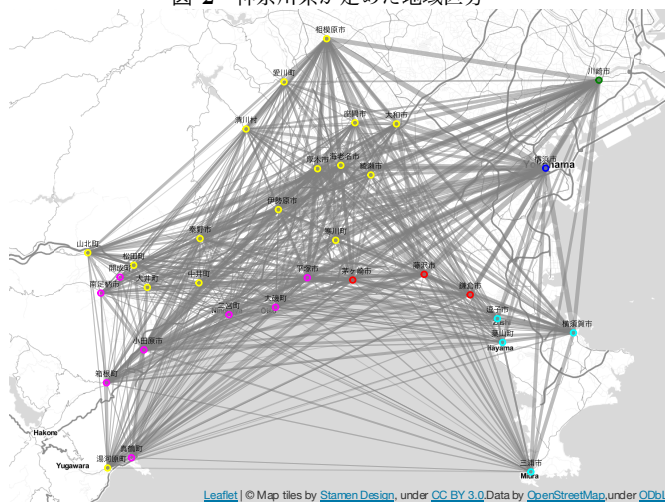


図 3 移動データに基づく市町村のクラスタ結果 (Newman Algorithm)

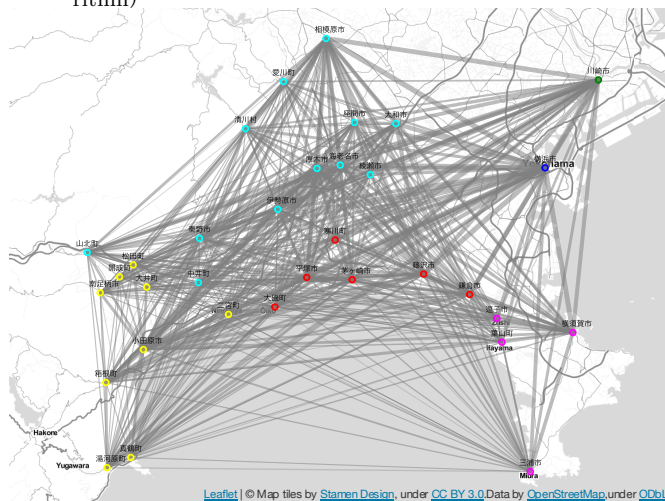


図 4 移動データに基づく市町村のクラスタ結果 (CNM Algorithm)

Modularity はクラスタ内とクラスタ間のエッジ数の割合に基づきクラスタリングの質を評価する指標である。図 1 に CNM Algorithm においてクラスタ数を変えたときの Modularity の推移のグラフを示す。この推移から Modularity が最大になるクラスタ数を 6 と決定した。

各アルゴリズムでの市町村のクラスタリング結果を図3, 図4および表2, 表3に示す。ノードの色がクラスタを示し、エッジの太さがエッジの重み（市町村間を移動したユニークユーザ数）を示す。表1および図2に神奈川県が定めている地域区分を示す\*1。神奈川県が定めた地域区分と移動データに基づく地域区分を比較する。Newman Algorithm の場合は、平塚市、大磯町、二宮町が小田原市など同一の地域と区分されている点異なる。CNM Algorithm の場合は、山北町と二宮町が神奈川県が定めた地域区分と異なっている。また両方のアルゴリズムに共有して、鎌倉市が横須賀三浦地域ではなく、藤沢市や茅ヶ崎市と同様の湘南地域に分類されていることが特徴的である。

### 3.4 クラスタリングの評価

移動データに基づくクラスタリング結果をクラスタリングの活用先の指標で定量的に評価する。本研究では、移動データに基づく地域のクラスタリング結果を経済指標や公衆衛生のデータ分析などに活用することを想定している。そこで、一例として観光消費額の推移を利用して評価を行う。海や山などの観光資源が地域ごとに類似しているため、観光消費額は地域ごとに推移が似ていると想定できる。そのため、各地域区分に含まれる市町村の観光消費額の推移が類似しているほど、市町村のクラスタリングが適切に行えていると考えられる。そこで、観光消費額の推移を行政で定めた地域区分と移動データに基づく地域区分で比較する。観光消費額は神奈川県の全市町村でデータが提供されているわけではないため、全市町村のデータが揃っている入込観光客数を利用する。ここで、入込観光客数と観光消費額は非常に強い相関があると考えられる。実際に神奈川県の2018年の入込観光客数データ\*2で相関係数を計算したところ、相関係数0.911,  $p$  値 $<0.001$ であった。そのため、入込観光客数の推移で評価した結果から観光消費額の推移についても同様の傾向があると考えられる。

具体的な計算方法を説明する。

- (1) 2つ以上の市町村を含むクラスタから任意の2市町村を抽出
- (2) 抽出した2市町村間で入込観光客数の年間推移を各市町村の最大値で割って正規化した上でユークリッド距離を計算（各市町村が12次元のベクトルとして表現される）
- (3) 全組み合わせで計算したユークリッド距離の平均値を計算

入込観光客数のデータは神奈川県が公表している2018年の入込観光客数データを利用した。以上の手順で距離を

表4 入込観光客数推移の平均距離の比較

クラスタリング方法	平均距離
県で定めた地方の分け方	1.323
Newman Algorithm	1.275
CNM Algorithm	1.238

計算した結果を表4に示す。なお、上記手順で距離を計算したため、クラスタに1つの市町村のみであった横浜市と川崎市は距離計算の対象から除外されている。人の移動に基づいて市町村をクラスタリングした方が分布間の距離が短いという結果であった。移動データでクラスタリングすることで入込観光客数の推移が似た市町村をまとめることができ、予測精度の向上が期待できる。

## 4. 入込観光客数予測への応用

本章では、移動データに基づく地域のクラスタリング結果の応用事例として、入込観光客数予測問題で有効性の評価を行う。

### 4.1 入込観光客数予測問題

本節では、移動データに基づく地域のクラスタリング結果が予測精度の向上に寄与するかを検証する。予測モデルを構築する際、クラスタリングの結果を考慮する簡易な方法の1つとして、教師あり機械学習モデルに特徴量としてクラスタリングの結果を利用することが考えられる。そのため、ここでは教師あり機械学習モデルによる回帰モデルを検討する。一方、入込観光客数は時系列データのため、時系列モデルの自己回帰モデルをベースとしたARIMA、季節性を考慮したSARIMAなどのモデルがよく用いられている[16]。そこで、これらの時系列モデルで考慮されている特徴を特徴量として予測モデルに利用する。時系列モデルの1つである季節調整モデルでは、観測値=トレンド成分+季節成分+変動として要素を分解しモデル化を行なっている[17]。本研究では、季節調整モデルを参考にそれらの要素の類似した特徴量を機械学習モデルの基本特徴量とし、移動データに基づく地域分けを追加の特徴量としてクラスタリングの結果を活用することの効果を検証する。

基本特徴量として以下の特徴量を利用する。

- 予測対象の市町村を示す one-hot ベクトル ( $feat_{city}$ )
- 予測対象の月を示す one-hot ベクトル ( $feat_{month}$ )
- 予測対象の市町村の過去直近  $N$  ヶ月の入込観光客数 ( $feat_{trend}$ )
- 予測対象の市町村の前年同月の入込観光客数 ( $feat_{seasonal}$ )

上記の基本特徴量に、地域（クラスタ）を示す one-hot ベクトルを追加の特徴量 ( $feat_{cls}$ ) として加える。この追加特徴量には、行政で決めた地域区分または移動データに基づくクラスタを利用し、それぞれでの予測精度を

\*1 <https://www.pref.kanagawa.jp/docs/ie2/cnt/f530001/p780102.html>

\*2 <https://www.pref.kanagawa.jp/docs/ya3/cnt/f80022/p1202218.html> (2020/3/25 参照)



比較することで移動に基づくクラスタの有用性を検証する。機械学習モデルには  $feat_{city}$ ,  $feat_{month}$ ,  $feat_{trend}$ ,  $feat_{seasonal}$ ,  $feat_{cls}$  を連結したベクトルを特徴量として入力する。教師データとして各市町村の入込観光客数を利用する。機械学習モデルには様々なモデルが利用可能であるが、本研究では XGBoost 回帰モデル [18] を利用する。

#### 4.2 評価条件

入込観光客数の予測には、前章で利用した移動データに基づく神奈川県内の各市町村のクラスタリング結果と、2015年から2018年の神奈川県入込観光客調査結果のデータを利用する<sup>\*3</sup>。例として横浜市の入込観光客数の推移を図5に示す。毎年大まかな傾向は同じであるが年によって大きく増減する月も存在する。機械学習モデルの学習データとして2016年と2017年の2年分を利用し、2018年の1年分をテストデータとして利用する。

神奈川県は33市町村あるため、 $33 \times 2 \times 12 = 792$  件を学習データとして利用でき、 $33 \times 12 = 396$  件がテストデータとなる。テストデータにおける入込観光客数の分布を図6に示す。図6より、実際の入込観光客数の分布には偏りがあることがわかる。テストデータの中央値を調べたところ、142,500であった。そこで、全テストデータでの予測誤差と、正解データの値が1ヶ月あたり15万人未満のみ、および1ヶ月あたり15万人以上のデータのみでの予測誤差を評価指標とした。

追加特徴量には、神奈川県が定めた地域区分、Newman Algorithm によるクラスタリング、CNM Algorithm によるクラスタリングを利用し、それぞれの予測性能を比較することで、クラスタリング結果を活用することの有用性を評価する。予測性能の評価指標は、回帰問題の研究でよく用いられる二乗平均平方根誤差 (RMSE: Root Mean Square Error) と平均絶対誤差 (MAE: Mean Absolute Error) を用いる。XGBoost モデルのパラメータは、木の深さ (max\_depth) および木の数 (n\_estimators) をグリッドサーチにより探索し、利用する各特徴量ごとに最適なパラメータを設定した。

#### 4.3 評価結果と考察

全テストデータでの予測誤差の評価結果を表5に示す。表中の太字は最も誤差が小さいことを示す。全テストデータでは、RMSE では Newman Algorithm を利用してクラスタリングした特徴量を利用した場合が最も誤差が小さく、行政が決めた地域区分と比較して誤差を 4.75%削減した。MAE では CNM Algorithm を利用してクラスタリングした特徴量を利用した場合が最も誤差が小さく、行政が決めた地域区分と比較して誤差を 0.84%削減した。

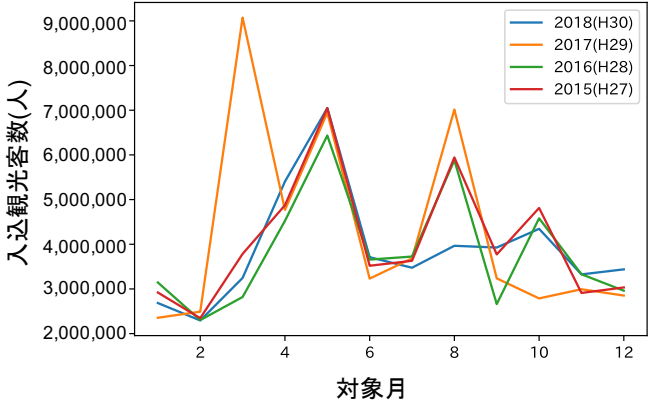


図 5 横浜市における 2015 年から 2018 年の入込観光客数の推移

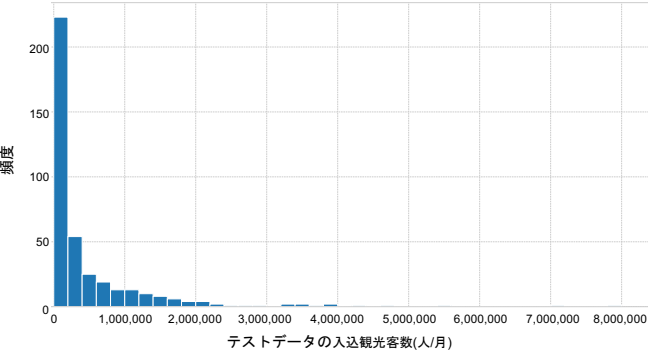


図 6 テストデータの入込観光客数の分布

特徴量	RMSE	MAE
基本特徴量のみ	272,962	91,608
基本特徴量+行政が決めた地域区分	266,454	90,408
基本特徴量+ Newman Algorithm	<b>253,802</b>	89,857
基本特徴量+ CNM Algorithm	258,048	<b>89,647</b>

表 6 正解データが 15 万人/月未満での入込観光客数予測の結果

特徴量	RMSE	MAE
基本特徴量のみ	25,077	15,078
基本特徴量+行政が決めた地域区分	25,469	15,197
基本特徴量+ Newman Algorithm	<b>24,457</b>	<b>14,887</b>
基本特徴量+ CNM Algorithm	24,572	15,121

表 7 正解データが 15 万人/月以上での入込観光客数予測の結果

特徴量	RMSE	MAE
基本特徴量のみ	389,146	171,294
基本特徴量+行政が決めた地域区分	379,800	168,722
基本特徴量+ Newman Algorithm	<b>361,753</b>	167,919
基本特徴量+ CNM Algorithm	367,826	<b>167,246</b>

15 万人/月未満のデータのみの結果を表6に示す。この条件では、Newman Algorithm を利用してクラスタリングした特徴量を利用した場合が最も誤差が小さかった。次に、15 万人/月以上のデータのみの結果を表7に示す。この条件では、RMSE では Newman Algorithm を使って移動データにより市町村をクラスタリングした結果を活用し

<sup>\*3</sup> <https://www.pref.kanagawa.jp/docs/ya3/cnt/f80022/p27746.html>

表 8 鎌倉市の入込観光客数予測の結果

特徴量	RMSE	MAE
基本特徴量のみ	338,653	259,254
基本特徴量+行政が決めた地域区分	326,575	250,791
基本特徴量+ Newman Algorithm	<b>315,282</b>	<b>240,671</b>
基本特徴量+ CNM Algorithm	325,031	257,990

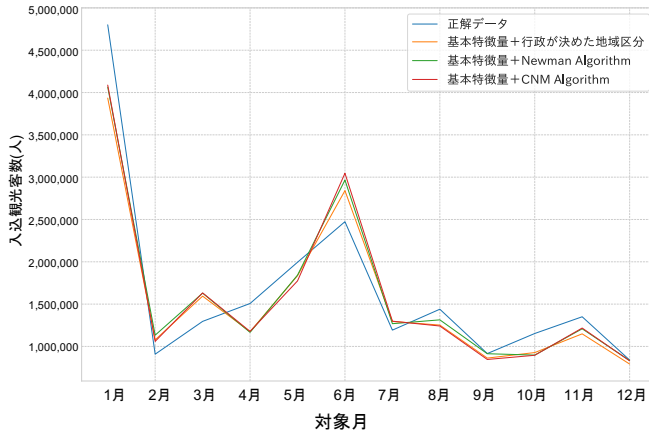


図 7 鎌倉市の入込観光客数の推移

た場合が最も誤差が小さく、MAE では CNM Algorithm のクラスタリング結果を特徴量とした場合が最も誤差が小さかった。これらの結果から、移動データによる市町村のクラスタリングの結果は、入込観光客数予測問題において、特に観光客数が多い場合において有用であると考えられる。

入込観光客数が多い市町村の例として、鎌倉市の正解データと予測結果をプロットした図を示す (図 7)。鎌倉市では 1 月は初詣、6 月は紫陽花のスポットとして有名で観光客が多くなっている。6 月は行政が決めた地域区分を利用した場合が正解データに近いが、それ以外の月については移動データによるクラスタリングを利用した場合が正解データの推移に近いことがわかる。表 8 に鎌倉市のみでの予測誤差の結果を示す。この結果では、RMSE, MAE のいずれにおいても Newman Algorithm アルゴリズムの結果を特徴量として利用した方が誤差が小さい結果となった。

## 5. おわりに

本研究では、人の移動データに基づいて地域をクラスタリングしたデータを地域に関する統計指標の予測問題に活用することを提案した。Swarm の約 3 万 8 千ユーザ、約 53 万件のチェックインデータを利用し、市町村をノードとし、市町村間の移動があった箇所にエッジを設定することで市町村グラフを構築し、市町村をクラスタリングした。1 つめの評価として、各クラスターに含まれる市町村の観光消費額の推移が類似しているほど市町村のクラスタリングが適切に行えていると考え、観光消費額の推移の距離を行政が定めた地域区分と移動データに基づく地域区分で比較

した。その結果、移動データに基づく地域区分の方が観光消費額が類似した市町村をまとめることができることを確認した。2 つめの評価として、移動に基づく地域のクラスタリングの結果を神奈川県内の各市町村の入込観光客数の予測問題において有用性を確認した。行政が決めた地域区分を特徴量とした場合と比較して予測誤差を 4.75% 削減できることを確認した。

今後の課題は、感染症など観光客数予測以外の地域に関する統計指標の予測での有用性を確認し、手法の一般性を評価することが挙げられる。また、移動データに本研究では SNS のチェックインデータを利用したが、ユーザに偏りがあると考えられるため、より偏りのない位置情報データ (例えば、携帯電話の基地局在圏情報に基づく人口データ) での評価を行いたい。図 3 および図 4 を比較すると、Newman Algorithm では同一クラスターの中の市町村が飛び地のようにならば 1 つだけ離れた場所にあり、CNM Algorithm の方が市町村が塊として抽出できている。また、3.4 節の評価でも、平均距離が CNM Algorithm の方が短かったため、これらの結果からは CNM Algorithm の方が移動データのクラスタリングに適していると考えられる。しかしながら、4 章の予測問題の評価では、概ね Newman Algorithm の方が誤差が小さかった。そのため、クラスタリング手法によって適した応用先が異なる可能性がある。今後、クラスタリングのアルゴリズムの特性と分析内容の関係についても明らかにしていきたい。

## 参考文献

- [1] Foursquare: Swarm, Foursquare (online), available from <https://www.swarmapp.com/> (accessed 2020-04-28).
- [2] Newman, M. E.: Finding community structure in networks using the eigenvectors of matrices, *Physical review E*, Vol. 74, No. 3, p. 036104 (2006).
- [3] Clauset, A., Newman, M. E. and Moore, C.: Finding community structure in very large networks, *Physical review E*, Vol. 70, No. 6, p. 066111 (2004).
- [4] Wang, Y., Yuan, N. J., Lian, D., Xu, L., Xie, X., Chen, E. and Rui, Y.: Regularity and conformity: Location prediction using heterogeneous mobility data, *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1275–1284 (2015).
- [5] Zheng, Y., Liu, L., Wang, L. and Xie, X.: Learning transportation mode from raw gps data for geographic applications on the web, *Proceedings of the 17th international conference on World Wide Web*, pp. 247–256 (2008).
- [6] Nishida, K., Toda, H., Kurashima, T. and Suhara, Y.: Probabilistic identification of visited point-of-interest for personalized automatic check-in, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 631–642 (2014).
- [7] Ye, M., Yin, P., Lee, W.-C. and Lee, D.-L.: Exploiting Geographical Influence for Collaborative Point-of-interest Recommendation, *Proceedings of SIGIR '11*, pp. 325–334 (2011).
- [8] Yuan, Q., Cong, G., Ma, Z., Sun, A. and Thalmann,

- N. M.: Time-aware Point-of-interest Recommendation, *Proceedings of SIGIR '13*, pp. 363–372 (2013).
- [9] Kurashima, T., Iwata, T., Hoshide, T., Takaya, N. and Fujimura, K.: Geo Topic Model: Joint Modeling of User's Activity Area and Interests for Location Recommendation, *Proceedings of WSDM '13*, pp. 375–384 (2013).
  - [10] Ochiai, K., Fukazawa, Y., Yamada, W., Manabe, H. and Matsuo, Y.: Gravity of Location-based Service: Analyzing the Effects for Mobility Pattern and Location Prediction, *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)* (2020).
  - [11] Thiemann, C., Theis, F., Grady, D., Brune, R. and Brockmann, D.: The structure of borders in a small world, *PloS one*, Vol. 5, No. 11 (2010).
  - [12] Rinzivillo, S., Mainardi, S., Pezzoni, F., Coscia, M., Pedreschi, D. and Giannotti, F.: Discovering the geographical borders of human mobility, *KI-Künstliche Intelligenz*, Vol. 26, No. 3, pp. 253–260 (2012).
  - [13] Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R. and Strogatz, S. H.: Redrawing the map of Great Britain from a network of human interactions, *PloS one*, Vol. 5, No. 12 (2010).
  - [14] Twitter: API リファレンス, Twitter (オンライン), 入手先 (<https://developer.twitter.com/>) (参照 2020-04-28).
  - [15] Newman, M. E.: Modularity and community structure in networks, *Proceedings of the national academy of sciences*, Vol. 103, No. 23, pp. 8577–8582 (2006).
  - [16] 大井達雄：自己回帰実数値と分移動平均モデルを使用した観光需要予測に関する考察, *観光学*, No. 6, pp. 1–7 (2012).
  - [17] 北川源四郎：時系列解析入門 (2005).
  - [18] Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794 (2016).