

タクシー軌跡クラスタリングに基づく都市の人気区域及び経路分析

Analysis of Urban Hotspots and Routes Based on Taxi Routes Clustering

談 栩辰

TAN,XUCHEN

概要: 衛星利用測位やカーナビゲーションなどの技術が急速に発展し、膨大なタクシーデータが生まれた。タクシー軌跡にクラスタリング分析を行い、都市の人気区域と人気経路を抽出できれば、人々の生活習慣などの研究ができ、スマートまちづくりや交通計画の策定に役に立てる。本研究では、2012. 11. 1-11. 7 北京の約 12,000 台のタクシー軌跡データおよびクラスタリングアルゴリズムを用いてタクシーの乗り降りの位置情報データをクラスタリングし、都市の人気乗下車地点と人気経路を分析する。

Summary: With the high-speed development and widespread use of technologies such as satellite positioning and car navigation, a large number of taxi route data have been recorded. Clustering analysis of taxi routes to discover hot spots and route in cities can provide reference for city intelligent traffic planning, intelligent travel guidance for passengers, and analysis of people's daily travel patterns. The research data in this paper is the GPS route data of about 12,000 taxis in Beijing in 2012.11.1-11.7. Firstly, the original data needs to be preprocessed. Then, based on the general rules of the residents' travel, combined with specific application scenarios, utilizes a suitable clustering algorithm from point data clustering and route clustering to get the hot spots of getting on or off and the hot routes of taxis.

キーワード: タクシー軌跡; 類似性距離; 軌跡クラスタリング; 人気区域; 人気経路

Keywords: taxi trajectory; similarity measure; trajectory clustering; hot spot area; hot spot passenger route

1. 研究背景および目的

都市交通システムの一部として、タクシーは都市の経済発展を促進し、住民の生活に利便性をもたらす効果を果たしている。最近、ほとんどのタクシーに GPS ナビゲーション装置が取り付けられており、位置、速度、方位、顧客の有無など軌跡にかかわるビッグデータデータが容易に取得できるようになってきている。

軌跡ビッグデータには人々出入りの多い人気地区の情報や利用頻度など数多くの有用な情報が含まれており、それを有効に利用すれば、スマートなまちづくりや人々により良いサービスの提供などに役に立てると思われる。

本研究では、タクシー運行軌跡オリジナルデータの予備処理を行ったうえで、タクシー軌跡を分割し、ノイズに強い SSPD 距離アルゴリズムと密度ベースの快速クラスター分析の組み合わせ処理により、都市人気区

域および人気経路を抽出する。

2. 国内外研究現状

現在、移動対象軌跡データに対する研究は主に軌跡データの時空分布分析、軌跡類似性計量分析と軌跡クラスタリング方法に集中している。そのうち、タクシー軌跡データの分析にもっとも適切な方法はノイズに強い密度ベースの快速クラスター方法である。軌跡類似性研究について、重要なのは軌跡類似距離の計量であり、主にユークリッド距離^[1]、DTW 距離^[2]、LCSS 距離^[3]、編集距離と他の類似性計量距離に分けることができる。

密度ベースの快速クラスター方法とは、ポイントが集中しているエリアとポイントがないまたは少ないエリアを快速分離する方法である。クラスターに含まれないポイントは、ノイズとしてラベルが付けられる。この方法は、空間位置および指定された近傍数までの

距離のみに基づいてパターンを自動的に検出する。

軌跡類似性計量は軌跡クラスタリングの基礎であり、軌跡クラスタリングユニットによって軌跡クラスタリングを全体軌跡クラスタリングと子軌跡クラスタリング^[4]に分けることができる。

3. 人気区域の抽出

3.1 研究対象地域とデータについて

本研究の対象地域は図1に示す北京市中心領域である。利用するデータは2012年11月1日一日間のタクシーデータである。タクシーデータでは、OperatingStatus というフィールドがあり、1はタクシーに乗る地点、0はタクシーから降りる地点を示す。

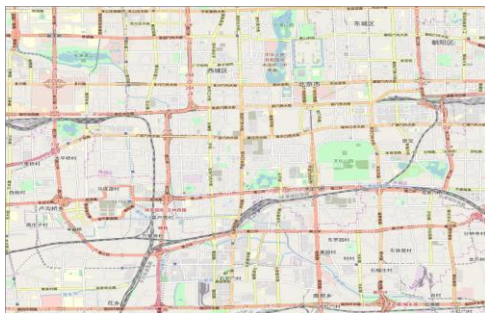


図1 研究対象地域（北京中心地）

3.2 快速クラスタリング手法

Rodriguez^[5-6]らは2014年に密度ピークに基づくクラスタリング方法を開発した。このアルゴリズムは任意のクラスター形状を検出しやすく、クラスターの数量も自動識別できる。ただし、この手法では、クラスタリングの閾値を手動でしか設定できず、効率と精度があまり高い欠点がある。

本研究では、局部密度 ρ 、と局部密度が高い対象間の距離 δ を自動計算できるとともに、任意の領域に対して、自動的に最適閾値を設定することができる方法を開発した。

3.3 結果

北京中心地区のタクシー乗車地点及び下車地点データを使い、快速クラスタリングを行った結果を図2から図7に示す。

図3にタクシーに乗車地点の ρ 、 δ 関係図を示す。図4は乗車地点のクラスタリングの結果を示したものである。図5は地図に反映したクラスタリングの結果である。

図6から8はそれぞれ、タクシーから下車地点の ρ 、 δ 関係図、クラスタリングの結果、地図に載せた結果を示したものである。

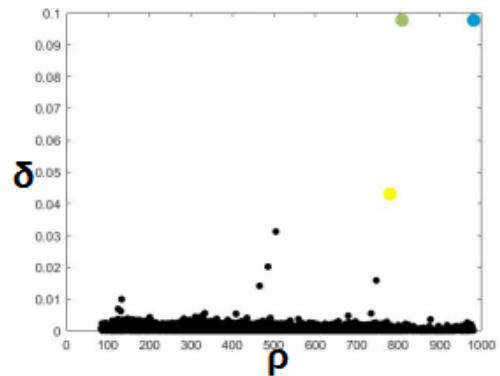


図2 乗車地点の ρ 、 δ 関係図

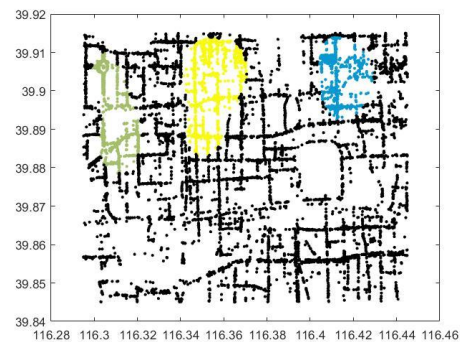


図3 乗車地点クラスタリング図

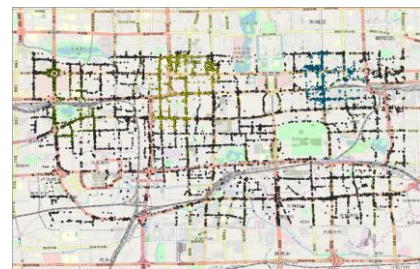


図4 地図に反映した乗車地点クラスタリング結果

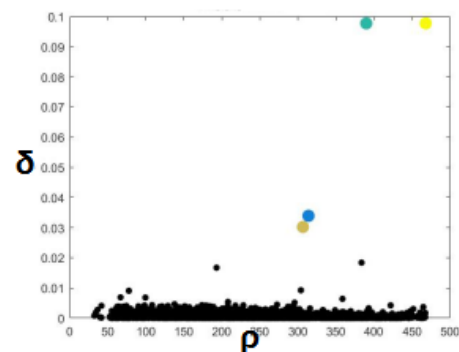


図5 下車地点の ρ 、 δ 関係図

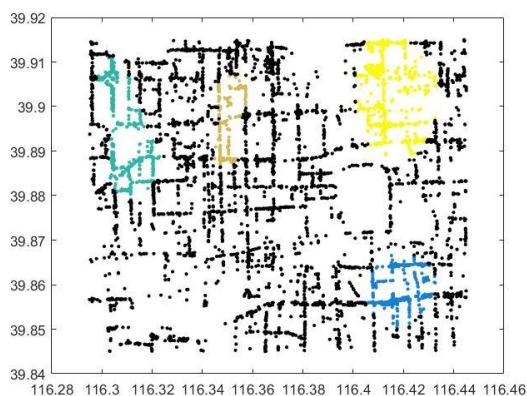


図6 下車地点クラスタリング図

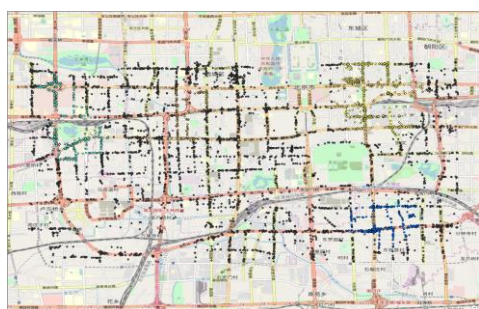


図7 地図に反映した下車地点クラスタリング結果

表1はクラスターの中心座標を示したものである。

表1 クラスター中心の座標

	ラベル	数量	クラスター中心点
乗車地点	3	1725	116.3532, 39.9065
	1	1442	116.4130, 39.9051
	2	1367	116.3532, 39.9065
下車地点	4	916	116.4129, 39.9051
	2	688	116.3107, 39.8865
	1	410	116.4159, 39.8579
	3	205	116.3508, 39.9020

4. 人気経路の抽出

Philippe Besse ら^[7]は2016年に軌跡形状ベースの軌跡類似性計量アルゴリズム SSPD (Symmetrized Segment-Path Distance) を提案した。SSPD 距離は計算効率が高く、ノイズの干渉を効果的に回避できるため、タクシー軌跡距離の計算に利用されている。本研究で

は、SSPD 距離をベースに、密度子軌跡快速クラスタリング方法を開発し、人気経路の抽出に適用する。

迟相松等^[8]は形状ベースの Hausdorff 距離を用いて子軌跡間の類似性距離を計算し、DBSCAN に基づく HDBSCAN クラスタリングアルゴリズムと高速探索発見密度ピーククラスタリングアルゴリズムに基づく HFDST アルゴリズムを提出した。HFDST は多くの類似子軌跡を抽出でき、計算率も高い。

4.1 アルゴリズムのプロセスについて

本研究では、SSPD 距離に基づく高密度子軌跡の快速クラスタリング SHDST を提出する。図9に密度子軌跡快速クラスタリングアルゴリズム SHDST のプロセスを示す。

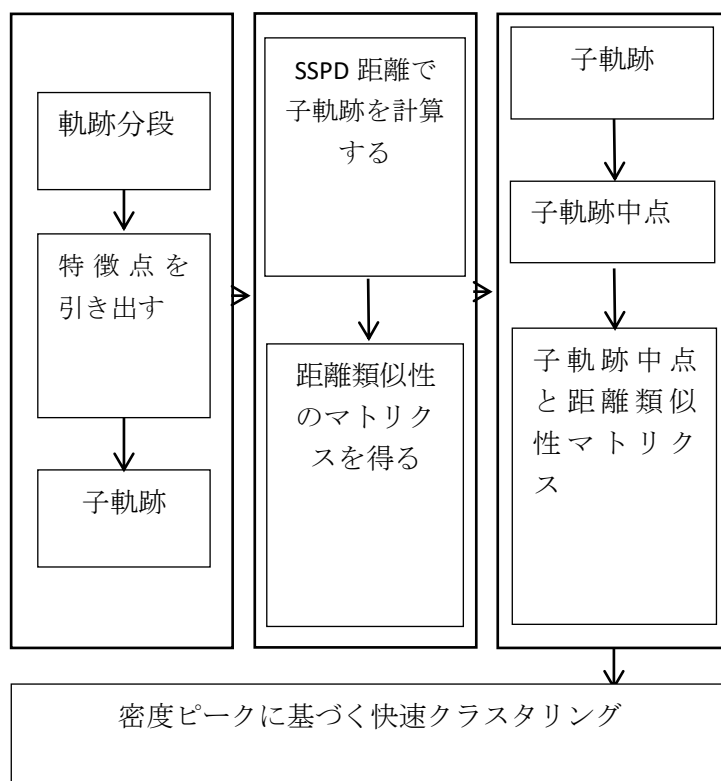


図8 アルゴリズムプロセス

軌跡分段と特徴点を引き出す目的はタクシー軌跡中の冗長な点を取り除くことと、クラスタリングの計算効率をアップさせることである。

タクシーの運営状態によりタクシー子軌跡は貸走り軌跡と空車子軌跡に分け、貸走り軌跡データだけを用いる。特徴点の引き出しは曲がり角の閾値（例えば20度）を設定し、閾値より高い特徴点を保留し、抽出する（図10を参照）。

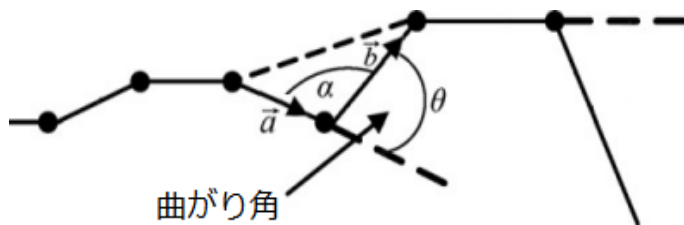


図9 軌跡間の曲がり角

4.2 実験の結果

実験用データは2012年11月1日から7日までのタクシー運行データからランダムにより100台を抽出したものを利用する。実験では、賃走り軌跡データだけを用い、曲がり角の閾値を10度に設定し、それぞれをSHDSTとHFDSTでクラスタリングする。結果を表2と図10、図11に示す。

表2 子軌跡クラスタリング結果

クラスターラベル	SHDST	HFDST
1	375	228
2	170	314
3	1914	392
4	129	471
5	317	686
6	203	126
7	244	N/A
8	803	N/A
9	173	N/A
Noises	359	2470



図10 SHDSTで抽出した人気経路



図11 HFDSTで抽出した人気経路

実験の結果では、SHDSTの計算時間がHFDSTより約4秒早くなり、より多くのクラスターを抽出でき、また、ノイズの数量も少なくなっていることが分かった。

5. まとめ

本研究では、密度ピークに基づく快速クラスタリング方法およびSSPD距離に基づく密度子軌跡の快速クラスタリング方法SHDSTを提案し、HFDSTと比較して実験を行った。実験の結果では、改良型SHDSTが従来のHFDSTより優れていることが分かった。しかし、本研究での特徴点の選別条件は軌跡間の角度だけにしてしているため、タクシーが真っ直ぐ走る場合、特徴点が二つしかなくなり、精度があまり高くないのが課題として残っている。

参考文献

- [1] Priestley M B. STATE-DEPENDENT MODELS: A GENERAL APPROACH TO NON-LINEAR TIME SERIES ANALYSIS[Z].
- [2] Bendt D J. Using dynamic time warping to find patterns in time series[C]Kdd Workshop. 1994:359-370.
- [3] Agrawal R, Lin K I, Sawhney H S, et al. Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases[J]. Proc. VLDB, Sept. 1995, 1995:490-501.
- [4] Lee J, Han J, Whang K. Trajectory Clustering: A Partition-and-group Framework[C]. New York, NY, USA: 2007:593-604.
- [5] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. Science. 2014, 344(6191): 1492.
- [6] Ester M. A density-based algorithm for discovering clusters in large spatial Databases with Noise[J]. Proc. 1996 Int. Conf. Knowledge Discovery and Data Mining (KDD '96). 1996: 226-231.
- [7] Besse P C, Guillouet B, Loubes J M, et al. Review and Perspective for Distance-Based Clustering of Vehicle Trajectories[J]. IEEE Transactions on Intelligent Transportation Systems. 2016, 17(11): 3306-3317.
- [8] 迟相松, 皮德常, 关鹏. 移動対象高密度子軌跡クラスタリングアルゴリズム [J]. 2016(09): 2014-2018.