HOKKAIDO UNIVERSITY

Title	Studies on Geospatial Mobility Analysis of Regions and Trajectories from Location Data [an abstract of dissertation and a summary of dissertation review]
Author(s)	稲越, 宏弥
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第14123号
Issue Date	2020-03-25
Doc URL	https://hdl.handle.net/2115/78405
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Hiroya_Inakoshi_abstract.pdf, 論文内容の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学　位　論　文　内　容　の　要　旨

博士の専攻分野の名称　　博士（情報科学）　　氏名　稲越 宏弥

学　位　論　文　題　名

Studies on Geospatial Mobility Analysis of Regions and Trajectories from Location Data
（位置データからの領域と軌跡に関わる地理空間移動分析の研究）

産業の発展により多く人が豊かな生活をおくることができるようになった一方で, 人口増加, 環境汚染, インフラ老朽化など新たな問題に我々は直面している. 今後も安全で快適なくらしを継続していくために, 社会の効率化や利便性の向上が大きな課題になっているといえるだろう. 特に, 人やモノの動きや, これに伴う資源や資本の消費を把握し, 分析し, 理解するために, 高度な計算機やデバイス技術を活用した, これまでになった実世界を対象にした空間情報処理に期待が集っている.

これには, 次の技術が含まれる. 第一の技術は米国が開発した人工衛星を用いた測位システム GPS である. GPS はスマートフォン, カーナビ, デジタルカメラ, アクティビトラッカーなどに搭載され, 時刻と座標を刻々と計測することができる.GPS は 1993 年に民間での利用に開放されたうえ, 先進国らは独自の GNSS (Global Navigation Satellite System) を開発・運用している. 第二の技術はデジタル道路地図 DRM である. DRM は国家の主導のもとで専門の官庁が, あるいは民間企業が商用目的で整備されただけでなく, オープンソースによっても日々メンテナンスされている.

これらの技術が可能にした実世界のリアルタイムかつ広域な測位データのひとつひとつはほぼ無意味であるが, これらを分析すれば, 従来とは異なる人口統計, 交通センサスなど実現が期待され, 社会の効率化や利便性の向上に役立てられるだろう. 我々は, 人やモノを対象とした空間・時間に関わるデータ分析のなかで 2 つの問題に注目した. ひとつは, 観測座標点に対するイベント発生の比が高いエリアの発見である. 例えば, 人やモノの滞留, 流入, 流出などをイベントと考えることで, オープンな空間上での人の動線分析が可能になる. もうひとつは, ネットワーク上での交通流分析のためのマップマッチングである.

地理空間情報に係わる研究は Cressie らによって体系化されており, 主に空間データを空間上の連続量として扱う地球統計学と, 離散的な空間におけるデータ間の関係を扱う計量地理学に分類される. 発展の経緯の違いから相互に参照されることは少なかったが, これらは総称して空間統計学と呼ばれ, 公衆衛生, 都市・交通計画, 出店計画, 犯罪発生から画像解析に至るまで幅広く応用されてきた. ところが, 我々の関心である GPS データは, ひとつひとつは無意味なので, 前述の学問体系の成果を適用するためには, 集計して量的データに変換する必要がある. しかし, 人やモノと, それらに対するイベント発生は動的で複雑であるため, そもそも集計単位を決めることが難しい. また, ネットワーク上での交通分析は比較的新しい問題だが, アルゴリズムの提案に主眼がおかれ, 不正確で不完全な GPS で観測された軌跡データに対して確率統計的な視点から真のルートを推定する提案は乏しい.

そこで本学位論文では, 前述の課題の解決に向けて次の内容を研究する. エリア発見問題に関しては, メッシュにより分割されたセル囲う閉じた領域の中から, 高いイベント発生比を持つもの発見する従来の枠組みである CDT (Cluster Detection Test), 最適相関ルール発見,bichromatic descripancy,box rule induction の関係を示したのち, 第 3 章では最適相関ルール発見の枠組みのも

とで確信度最適な領域の探索アルゴリズムを提案し, 第 4 章では前述の解を初期解として領域の記述長に基づく正則化つきの解へ拡張を行う. 一方, マップマッチングについては, 同時マップマッチング問題を事後確率最大化の枠組みで定式化することで, これに動機づけられる真ルートを推定するための探索的アルゴリズムを提案する.

第 3 章では, 与えられた位置情報データ集合に対して, 相関ルールの確信度を極大化するエリアを発見する新たなアルゴリズムを与える. GPS を搭載したデバイスからのデータは, 有望なアプリケーションを創出する可能性を秘めているが, これらの GPS 観測点は 2 次元の空間上を一様には分布せず, 地域的な偏りが見られるのが普通である. そのため, 一様な散布を仮定した, 単調に接続された領域や, 矩形領域を探索するマイニングアルゴリズムでは, 重要な知見を見落とすリスクがある. 提案アルゴリズムは, 相関ルールの確信度を極大化する, 不規則な格子上で推移的に接続されたセルを囲う領域を探索するので, 一様な散布を仮定しておらず, 特定の形状に縛られずに領域を探索することができる. セルに関する極大化条件に従って, 反復的にセルを削減していくことで, 提案アルゴリズムは先行手法に対して 50 倍高速であった.

第 4 章では, 空間バイアスが強いデータに対して, 期待応答値を最適化する領域を発見する新しい提案を行う. 従来手法は散布および応答が, 空間上で一様および等方的であることを仮定しているため, 前記の性質を持つデータに対して有効に機能しない問題があった. このような仮定を緩和するために, 提案手法では, 不規則な格子中の, 推移的に隣接したセルを囲む領域を仮説空間とする. ところが, このような仮説空間のサイズは指数的であるため, これ探索することは, 膨大な計算を要し, しかも過適合に陥りがちである. そこで二つの問題を解くことでこの不具合を克服する. まず, 仮説空間を列挙するのではなく, 極大領域に含まれることのないセルを判定する条件を証明し, これに従ってセルを除外することで仮説空間の実質的なサイズを削減する. さらに, 推移的に隣接したセルを囲む領域に対して記述長を定義し, これによって期待応答値を正則化する. 実験によれば, 提案手法は GPS データが非一様, 非等方的な確率分布に従う場合にも有効に機能し, 正則化を導入したことで期待応答値を 20% 向上させることができた.

第 5 章では, 本来観測することができない物体が通過したパスを推定する同時マップマッチングの枠組みを提案する. 提案手法は, 事後確率最大化する軌跡データの生成モデルを考えるという点で初めての研究であり, ここで事後確率とは, 例えば自動車が道路に沿って走行するときに, 観測された軌跡が出力される確率のことである. 我々は, EM アルゴリズムを用いて, この生成モデルのパラメタと隠れ変数を最適化するが, この隠れ変数こそが不可観測なパスに相当する. この EM アルゴリズムは, あらゆる観測された軌跡データを生成する確率が大な幾何グラフの探索アルゴリズムに還元される. このような確率的な枠組みを採用したことにより, GPS データのサンプリングノイズが大きいときに本手法は力を発揮し, 実験によれば, サンプリングにより GPS 観測を 40% まで削減しても, 真のパスとのマッチングされたルートの残差は 7% にとどめることができた.

第 6 章では, これらの研究を総括し, 今後の研究の方向性を示す.