

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 10 日現在

機関番号：24402

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25420633

研究課題名(和文) 半教師ありクラスタリングに基づく空間の曖昧性を考慮した狭域空間分析手法の開発

研究課題名(英文) Development of the Small Space Analysis Method Considering Spatial Ambiguity Based on Semi-Supervised Clustering

研究代表者

瀧澤 重志 (Takizawa, Atsushi)

大阪市立大学・大学院工学研究科・准教授

研究者番号：40304133

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究では建築空間や街路レベルの比較的小さな空間スケールを対象として、汎用性が高い新しい空間分析手法を開発した。本手法は本来曖昧な性質を有する空間上で意図せず発生する行為やイベントが、どのような空間的状况のもとで発生するかを、統計的に高精度に説明・予測できる。さらに、それら行為やイベントが発生する可能性が高い領域を、半教師クラスタリングの考え方をを用いて自動的に決定できる。この手法を人工データと梅田地下街のデータに適用して検証を行い、良好な結果を得た。

研究成果の概要(英文)：In this study, I have developed a novel and general spatial analysis method for relatively small space such as architectures and streets. This analysis method can statistically explain and predict the place where human's unconscious activities might occur in the space which essentially has vague boundary. In addition, this method can automatically decide the shape of the area in which such activities might occur based on the concept of semi-supervised clustering. I applied this method to the artificial data and the actual data of pedestrians rendezvousing in Umeda Underground Mall and got good results.

研究分野：工学

キーワード：空間分析 イベント 領域分割 半教師ありクラスタリング 分類問題 混合整数2次計画問題

1. 研究開始当初の背景

建築空間や街路のオープンスペースといった、都市スケールから見れば比較的小規模な空間を設計する際には、その中で発生する様々な人々の行為やイベントの発生を期待して、設計者は設計を行っていると考えられる。建築計画や都市計画の研究でも、例えば、滞留行動やオフィス内での行動分析など、行為やイベントに関する研究が盛んに行われている。こうした人々の行為やイベントには、設計者が意図するものだけでなく意図しないものも含まれる。空間が思わぬ使われ方をされることは、どのような空間でも多かれ少なかれ生じることである。こうした認識は、研究者や設計者の間で広く共有されている問題意識だと思われる。

しかし意図しない行為・イベントは必ずしも人々にとって歓迎されるものだけではない。その典型的な例が街頭犯罪である。犯罪発生と空間構成に関連があることは1970年頃に指摘され、今日では防犯環境設計(CPTED)として世界中で実践されている。その動きに追随するように、地理情報システム(GIS)の発達により、犯罪と空間の関係を実証的に分析する地理的犯罪分析などの研究が広く行われている。申請者は研究助成金の補助により、これまで主に京都市伏見区などを対象地域として、街頭犯罪と空間分析の関係を分析してきた。これら一連の研究では、ビッグデータの時代を見越して最新のデータマイニング技術を積極的に空間分析に活用し、その主要な成果が、世界的に著名なデータマイニングの研究者である合衆国Wright大学のG. Dong教授らが2012年9月に出版した著書に掲載されるなど、研究成果が国際的に認知されるに至っている。

加えて近年の空間デザインは、コンピュータショナル・デザインやアルゴリズムック・デザインなどの考え方も反映して、機能分割と空間構成が一見対応しない曖昧さを積極的に許容する空間となってきた。例えば、伊東豊雄らによるせんだいメディアテーク、石上純也による神奈川工科大学 KAIT 工房、藤本壮介の House O、平田晃久の杵屋本店など、明示的に分節化されていない空間が現代建築の世界で広まってきており、こうした空間を分析するための新たな方法が必要とされてきている。

空間分析を行う場合、実証性を高めるために統計解析手法を援用することは、今日では当然のように行われている。例えば申請者も著者の一人となっている文献(研究業績6)では、建築や都市計画のための調査・分析方法として、多変量解析などの様々な統計的手法と応用例が解説されている。建築・都市計画の空間分析での統計解析の使われ方を申請者なりに概観すると、注目する行為やイベントが発生する場所の特徴を、クラスター分

析によりタイプ分けする研究が非常に多いと感じる。しかしクラスター分析は、判別分析などと異なり教師データ(目的変数)を使用しないので、行為・イベントが、場所の違いによりなぜ発生したりしないのかを説明することはできない。この原因として、先に述べた空間の曖昧性が大きく関わっていると考えられる。判別分析では、まず空間的範囲を定め、各範囲にクラスラベルを付与して教師データを作成した上で分析を行うが、空間の曖昧性や領域形状の多様性の問題から、事前にどのように空間的範囲を定めたらよいのかについては、未解決の問題として残っている。

またこの問題は、主に計量地理学の分野で考慮されてきた空間的自己相関の問題と強く関連する。空間的自己相関とは、ある注目している空間の周辺も注目している空間と類似の空間的特徴を持っている性質である。空間データに対して、誤差項の相関を考慮しない通常の回帰/判別分析手法を適用すると、モデルの精度が低下してしまうことが知られている。そのため、誤差項の共分散成分に空間的な制約を入れた地理的回帰分析手法などが発展してきた。しかしこの方法は主にマクロな地理学的スケールの事象を扱うために開発されてきたため、基本的に空間の広がりを目点からの距離だけで表現してモデルを生成する。これは領域を単純な円として想定することを意味しており、本研究が対象とする狭域の分析にそのまま応用するには問題がある。

このような問題意識を、申請者は助成金の補助の下で行っている研究の実施過程で強く抱き、本研究へとつながった。具体的にいえば、既往研究においてひたたくり犯罪と街路の空間構成の関係を分析した際に、当該点だけでなく、ひたたくりが潜在的に発生すると考えられる近隣部も考慮して分類を行うという、本研究につながる新しい分析方法を構想した。しかしこの研究では、領域を線形なネットワーク空間上に限定しており、さらに理論展開が十分でないことやイベント発生点の判別精度が低いなどの問題があり、本研究課題を申請するに至った。

2. 研究の目的

以上の背景から本研究では、既往研究を発展・一般化させ、建築空間や街路レベルの比較的小さな空間スケールを対象として、本来曖昧な性質を有する空間上で意図せず発生する行為やイベントが、どのような空間的状況のもとで発生するかを統計的に高精度に分類・予測し、同時にそれら行為やイベントが発生する可能性が高い領域を、半教師クラスタリングの考え方をういて自動的に決定する、汎用性・柔軟性が高い新しい空間分析手法を開発することを目的とする。半教師ク

ラスタリングとは、通常のクラスタリングに加えて、部分的に教師データ(クラスラベル)が付与されたデータを用いてクラスタリングを行う手法である。この手法は近年、主に画像認識の分野で領域分割を行う場合などに利用されて有効性が確認されているが、本研究で扱う問題は、特徴ベクトルの次元数が場合によっては100以上になること、それぞれの特徴ベクトルの重みが一律でないこと、領域形状をある形状クラスに制限する必要があるなどの点で、既存の方法をそのまま利用することができないため、目的に合った方法を研究する。そして検証のための対象地域として、これまで街頭犯罪の研究を行ってきた京都市伏見区の街路上の線的空間に加えて、今回新たに、面的空間として駅等の公共空間を調査し、そこで発生する様々な行為を抽出・分析する。そして、研究成果を反映した分析用ソフトウェアを作り、一般公開を目指す。

3. 研究の方法

平成25年度は、既報で提案した基本的な分析手法をベースにしつつ、関連研究を通じて、より一般性が高く性能が良く、さらに豊かな情報を提供できる空間分析手法の開発を行う。まず、半教師学習手法、画像の領域分割、地理的重帰帰モデル、空間データマイニングなど、本研究に関する最新の関連文献や書籍を参照し、提案手法を理論的、アルゴリズム的な面から検討していく。特に関連する手法は画像の領域分割である。これは、グラフカットなどのテクニックを用いて、ピクセルを類似性の高い領域にボトムアップに分割するクラスタリング手法の一種であり、画像認識の分野で近年広く用いられている技術である。しかしピクセルの特徴量は、高々3次元の色情報ベクトルの情報として表現可能であり、さらに領域分割される形状にも制約が無い場合、本研究には導入できない。本研究では領域の形が重要になる。前述したように、地理的重帰帰では暗に領域を円として考えているが、これは狭域分析では自由度が小さすぎる。本研究ではビットマップとして平面を離散化して分析するが、それでも領域の形状は無数に考えられる。そこで、数学的に良い性質を持っている図2のような領域クラスを考える。イベント発生の潜在的な領域はその点からみて凸であることが望ましいと考えられるので、直方凸領域で領域を近似することを考える。直方凸領域は、 x 単調かつ y 単調な領域である。この領域を ZDD などの新しい圧縮アルゴリズムを使って列挙・索引化する方法を開発する。次に各イベント発生点において、列挙した領域形状を当てはめ、領域とそれ以外の部分を分割してクラス分類を行い、分類精度が最高になるような領域形状の最適化を行う。一般に注目するイベント発生領域は、考慮する空間全体の中心

で相対的に狭い範囲に集中する傾向がある。従って、本研究のクラス分類問題は、機械学習の分野の言葉で **Imbalanced Classification Problem** と呼ばれる、データ分布が偏った難しい問題となる。このような問題に対して頑健な分類モデルとして、**Contrast Pattern Based Classification** モデルが提唱されており(参考文献1)、筆者もその初期型の **CAEP** と呼ばれる方法をこれまで使用してきた。現在は新しい方法が複数提案されているので、それらを実装し応用可能性を検討する。また、比較対象として一般化線形モデルも導入し、性能の違いを把握する。対象空間として、これまで研究を行ってきた京都市伏見区の街頭犯罪の発生データを分析する。このデータは図3(a)に示すように街路上の線的空間であり、開発した領域形状のうち使用するものが限定されるが、研究の初年度としてこれまでの懸案を解決することを目指す。

H26年度は、前年度に構築した分析方法を広がりのある空間に適用するために、対象地域として京都駅北側の駅前広場を選定し分析を行うとともに、計算量等の面から方法の改良を行う。この場所は図3(b)に示すように、通常の駅前広場と異なり、比較的複雑な空間構成となっている。また、利用客も通勤通学以外に、買い物客や観光客など多様な目的を有する人々が集まっており、多様な行為が観察される。まず、京都駅前広場の詳細な 3D モデルを作る。作成を効率化するために、高解像度型の一眼レフカメラを購入して駅前を撮影し、その画像データを最新の写真測量ソフトウェアに入力し、3次元の建物モデルを自動的に作成する。モデルの作成のために高速なワークステーションを購入して利用する。なお、このワークステーションは当年度の空間分析にも利用するが、線的空間よりも複雑になる面的な凸領域の列挙のために大容量メモリが必要となる。その作業と平行して、駅前の人々の行動を小型のビデオカメラを利用して撮影する。天候の影響が少ない春の曇天の日を選んで、データを作成する。この作業には学生アルバイトを利用して行う。これらのモデルとデータから、ビットマップのピクセル毎の空間特徴量を算出し、提案手法による分析を行い、妥当な結果が得られるまで、分析方法等を改良する。

最終年度は、前年度までの研究をさらに改良し纏めるとともに、これまで作成した分析方法をソフトウェアとして実装・公開することを目指す。領域形状の列挙はメモリを多く使うため、一般の PC でも使用できるように、効率的に情報圧縮して格納する方法を開発する。また、**Contrast Pattern Based Classification** は性能が高いが、これまでほとんど実装が公開されていないため、本研究で実装を行い広く世界に発信する。実装する環境としては C++ によるコンソールアプリケーションを予定している。マルチコアといった最新の計算機技術に対応し、できるだけ省メモリで高速

に解を求めることができる、実用的なソフトウェアの開発を目指す。

4. 研究成果

平成 25 年度は、既報で提案した基本的な分析手法をベースにしつつ、関連研究を通じて、より一般性が高く、高精度で、豊かな情報を提供できる空間分析手法の基礎的な開発を行った。

まず、半教師学習手法、画像の領域分割など、本研究に関する関連文献や書籍を参照し、提案手法を理論的、アルゴリズム的な面から検討した。特に画像の領域分割手法が深く関連するが、しかしピクセルの特徴量は高々 3 次元の色情報であり、さらに領域分割される形状にも制約が無いため、本研究には導入できない。本研究では領域の形が重要になる。本研究ではビットマップとして平面を離散化して分析するが、それでも領域の形状は無数に考えられる。イベント発生の潜在的な領域はその点からみて凸であることが望ましいと考えられるので、直方凸領域の一種である矩形和楕円形領域でイベント発生領域をカバーする手法を開発した。

次に各イベント発生点において、列挙した領域形状を当てはめ、領域とそれ以外の部分を分割してクラス分類を行い、分類精度が最高になるような領域形状の最適化を、PBIL を用いて行った。一般に注目するイベント発生領域は、考慮する空間全体の中で相対的に狭い範囲に集中する傾向がある。従って、本研究のクラス分類問題は、機械学習の分野の言葉で Imbalanced Classification Problem と呼ばれる、データ分布が偏った難しい問題となる。このような問題に対する頑健な分類モデルとして、Contrast Pattern Based Classification モデルを実装し応用可能性を検討した。21×21 グリッドの中央に一つだけイベント点がある仮想平面を考え、3つの空間属性を与えて判別実験を行ったところ、判別精度の高い領域分割が達成できた。

平成 26 年度は、前年度に提案した半教師学習型の狭域空間分析手法を改良するとともに、対象地域の空間のデータ化を行った。まず、狭域空間分析手法の改良についてであるが、前年度に開発した分析手法にはふたつの改善点があった。一つは、最適領域の発見にメタヒューリスティクスの一つである PBIL (Population Based Incremental Learning) を用いており、(準)最適領域の発見に時間がかかることと不確実性が存在することであった。もう一つは、最適領域のサイズを陽にコントロールできないことであった。一般に、イベント発生箇所と非発生箇所の中間にある非ラベル領域の面積は少ない方が望ましいと思われるが、昨年度に提案した方法ではこのサイズをコントロールできないことが懸案であった。そこで両者の問題を同時に解決する手法として、整数計画問題として最適

領域を求めるための定式化を提案し、実装と数値実験を行った。求解の手順としては、非ラベル領域を除いて一度分類モデル (CAEP) を用いて学習を行い、学習済みのモデルを用いて非ラベル領域のラベルを予測する。その結果を用いて、全体の誤分類精度 (BER) が最小になるように、イベント点に対する凸領域制約を満たす非ラベル領域形状を整数計画問題で発見し、見つかった解を元にして、学習モデルの学習・予測、最適領域形状の改善を繰り返すというものである。整数計画問題として本問題を定式化するためには、最適領域サイズをパラメータとして事前に与える必要があり、これが結果として領域サイズのコントロールにつながっている。C++ と CPLEX12.6 を用いて実装を行い、昨年度と同様に 21×21 の領域サイズにて数値実験を行った結果、従来よりもサイズが小さく精度の高い領域形状を発見することができた。また、対象地域を梅田地下街に決定し、分析に必要な基本的な空間構成要素のデータ化も行った。

平成 27 年度は、前年度までに開発した手法を改良し、梅田地下街の歩行者滞留データに適用した。分析手法の主な改良点は、誤差関数の変更と、整数計画問題の凸領域制約の精密化である。これまでは誤差関数として Balanced Classification Rate, $BCR = (\text{Sensitivity} + \text{Specificity}) / 2$ を用いて誤差関数を構成したが、評価がアンバランスになることを防げなかった。そこで今年度は、両者のバランスを考慮した Root Squared Balanced Error Rate, $RSBER = ((1 - \text{Sensitivity})^2 + (1 - \text{Specificity})^2) / 2$ を提案した。また、潜在的なイベント発生領域が複数のイベント発生点にまたがるセルの場合、混合整数 2 次計画問題における凸領域形状の制約が、関連イベントセルの or 条件となるので、その点について改良を行った。対象地域は、地下鉄御堂筋線南改札口を出た、方向感覚が狂いやすいことで有名な広場である。対象地域を 50cm のセルに分割し、H27 年 5 月 12 日の 11:00~11:30 に現地調査を行い、約 30 秒以上立ち止まっている人を滞留者として、その位置を最寄りのセルにプロットした。各要素について 2 次元の可視領域解析を行い、得られた Isovist Field に関する特徴量を各セルの中心で求めた。検証を行った結果、バッファを設けない場合の分類精度が全体精度で約 0.55 だったのに対して、提案手法によりバッファを最適化した場合、それが 0.8 程度まで上昇し、実データで良好な結果が得られた。

最終的な分析手法をソフトウェアとして実装している。実装は C++ で行い、インターフェースの整備やデバッグが完了次第、申請者の HP にて公開する予定である。

5. 主な発表論文等

[雑誌論文] (計 1 件)

瀧澤重志, 狭域空間におけるイベント発生領域の識別的クラスタリング-整数計画法を用いた最適領域形状の決定手法と梅田地下街での検証, 日本建築学会・情報システム技術委員会 第38回情報システム利用技術シンポジウム論文集, 論文, pp.25-30, 2015 (査読有)

[学会発表] (計 5件)

A. Takizawa, Estimating Potential Event Occurrence Areas in Small Space based on Semi-Supervised Learning, eCAADe 34th Annual Conference, Oulu, Finland, 2016.8.24-26

瀧澤重志, 狭域空間におけるイベント発生領域の識別的クラスタリング 整数計画法を用いた最適領域形状の決定手法と梅田地下街での検証, 地理情報システム学会第23回研究発表大会, 慶應義塾大学, 東京都港区, 2015.10.10

A. Takizawa, A Basic Study on the Region Partitioning based on Semi-Supervised Classification, Proceedings of the GIScience 2014, pp.420-423, Sept. 24-26, 2014, Vienna University of Technology, Vienna, Austria, 2014.9.26

瀧澤重志, イベント発生に基づく空間の領域分割に関する基礎的研究, 日本建築学会大会学術講演梗概集 (近畿), 神戸大学, 兵庫県神戸市, 11067, pp.143-144, 2014.9.14

瀧澤重志, 半教師あり学習の考え方に基づく空間の曖昧さを考慮した領域分割に関する研究, 日本建築学会第36回情報・システム・利用・技術シンポジウム論文集, 報告, 建築会館, 東京都港区, pp.141-144, 2013.12.5

[その他]

ホームページ等

<https://sites.google.com/site/drtakizawa/>

6. 研究組織

(1)研究代表者

瀧澤 重志 (TAKIZAWA, Atsushi)

大阪市立大学・大学院工学研究科・准教授

研究者番号: 40304133