



オープンデータを組み込んだ因果分析と離散分析による 土地価格の積み上げ予測

2120031 中島健希 情報基盤工学講座

指導教員 奥原浩之

要約

本論では、富山県における住宅地地価の空間的分布を観察することを通じて、地価に有意な影響を与える不動産特性がヘドニック・アプローチによってどの程度捕捉され得るのか、またその場合に具体的にどの要因が価格差を決定づけるうえで重要となるのかを定量的に整理する。

キーワード：ヘドニック・アプローチ、地価予測、オープンデータ、離散選択モデル

1はじめに

土地価格形成要因は一般的要因、地域要因、個別的要因に分類され、特に住宅地域に関連する要因が重要です。これらの要因は相互に関連し、住宅地は自然的特性や再生産不可能性、用途の多様性からユニークな商品とされます。こうした要因の重要度を定量化するために、本論ではヘドニック・アプローチを採用します。このアプローチでは、住宅地の価値はアメニティや環境質、利便性、規模などの特性の合成物と見なされ、これに基づいて住宅地の価格が決定されると考えます。そこで本論では、ヘドニック・アプローチを用いて、住宅地における土地価格に影響を与える要因を特定し、それらの影響度を定量化することで、土地価格の予測精度を向上させることにあります。

2多様な要因を考慮したデータセットの作成

2.1 サイバー空間からのデータ取得

土地価格の形成には、多くの要因が影響を与えると考えられる。そのため、土地価格を予測するモデルを作成する際には、これらの要因を表す説明変数を多く考慮する必要がある。しかし、一般的に取得できるデータ、特にオープンデータには限界が存在する。例えば、地価の形成に重要な地理的データや施設の位置情報は、統計データと比較して容易に取得できるものではない。そこで本研究では、施設データを補完するために、ナビゲーションサービス「NAVITIME」からスクレイピングを用いてデータを取得した。スクレイピングは、Webサイト上から必要なデータを収集し、それを分析しやすい形で加工する手法である。NAVITIMEでは、施設情報がジャンルごとに分類され、都道府県ごとのデータを取得できるため、土地価格の要因として特定施設の影響を分析するための有効なデータ源となる。このデータを用いることで、土地価格形成要因の精度向上を目指す。

2.2 データクリーニングによる前処理

データ分析において、欠損値や異常値が含まれるデータセットは、モデルの予測精度を著しく低下させる可能性があるため、データクリーニングによる前処理は不可欠である。数値で表せないカテゴリ変数を0と1で表現することで、数値データとしてモデルに入力できるようにする手法である。これを行うことで、より洗練されたデータセットを構築し、予測精度向上につなげることが可能である。

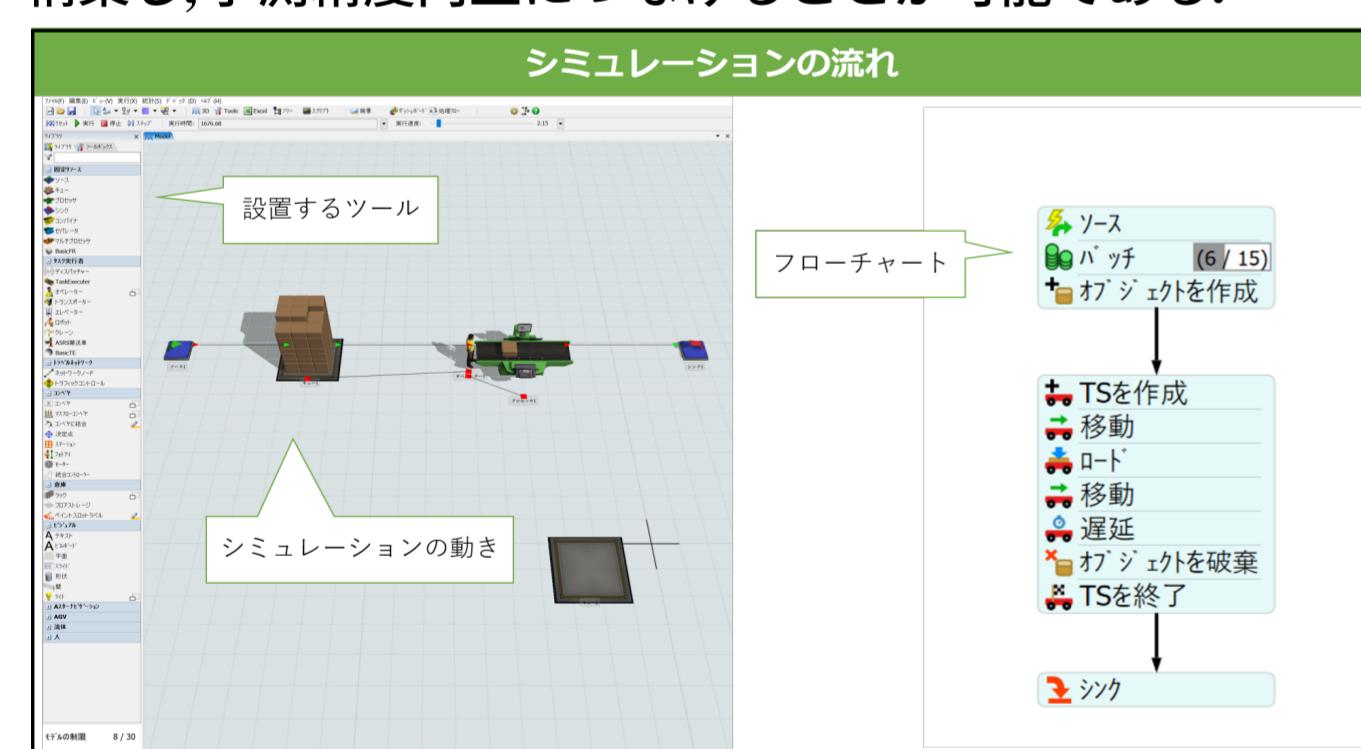


図1 FlexSimのシミュレーションの流れ

2.3 説明変数の選定

土地価格の決定にはさまざまな要因が考えられる。そのため、できるだけ多くの説明変数を考慮することが望ましい。しかしながら、むやみやたらに目的変数と相関がない説明変数を追加しても、予測精度が上がるどころか、計算コストが増大するだけであろう。また、本システムで統計データを取得するために利用しているe-Statは、グリッドセル・小地域ごとのデータに絞っても、200以上公開されている。さらに、それらを別のデータと計算し、新たな説明変数を作成することを許せば、その組み合わせは考慮しきれない。そこで、機械学習による犯罪予測モデルを作成する際に、説明変数を容易に選定することができるようになっている。選定できるデータは、富山市内の犯罪発生データ、および、NAVITIMEで公開されている施設データである。

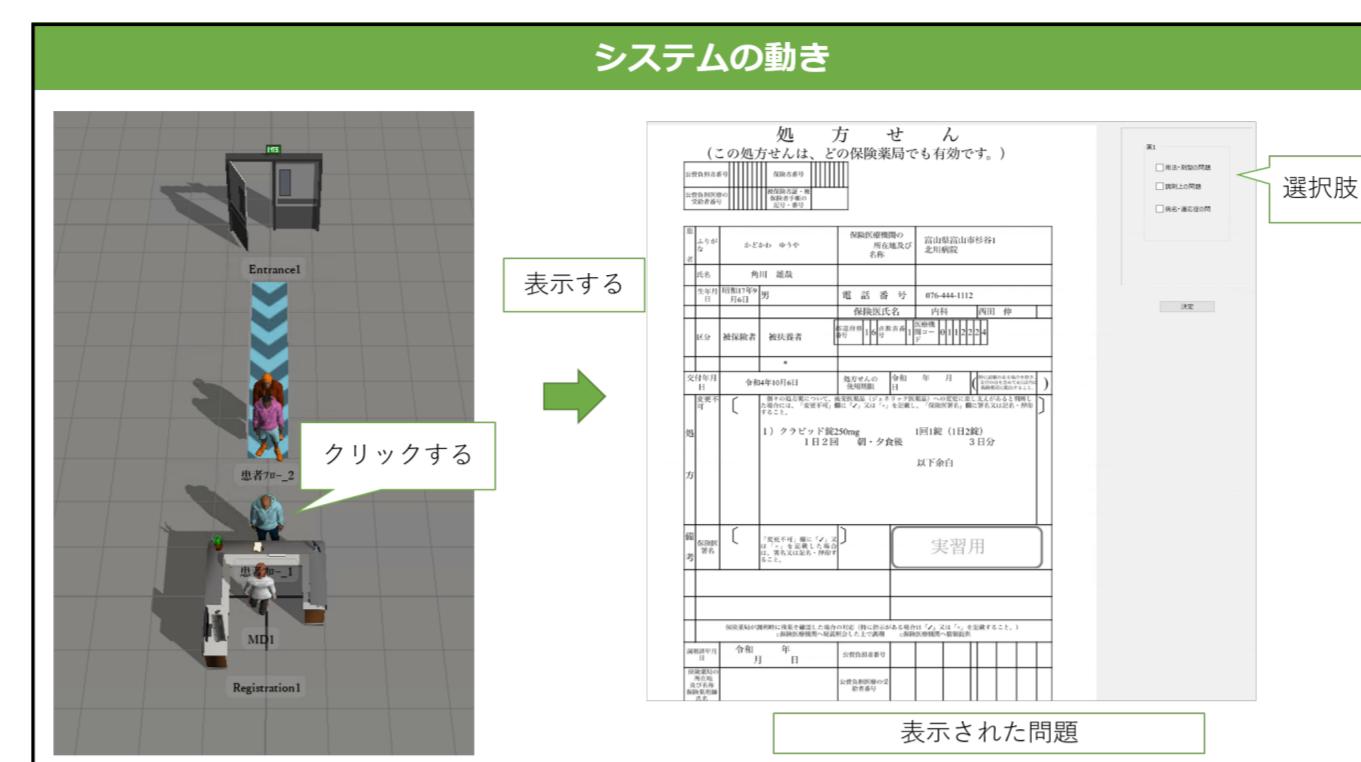


図2 FlexSimのシステムの動き

3

3.1 富山県住宅地地価ヘドニック関数の推計

富山県における住宅地の地価形成要因を明らかにするため、ヘドニック関数の推計を行う。ヘドニック・アプローチは、財の価格がその属性の総和として決定されるという仮定に基づき、土地価格を説明するために広く用いられている。本研究では、富山県内の住宅地における土地価格データを用い、地価に影響を与える各種の要因を定量的に評価する。具体的には、土地の物理的特性（面積、形状）、立地特性（最寄駅までの距離、交通アクセス）、環境要因（公園や教育施設の近接性）といった説明変数を用い、これらが地価にどのように影響を与えるかを分析する。また、土地の再生産不可能性や自然的特性が地価に与える影響も考慮し、地域特有の要因についても評価を行う。推計結果を基に、富山県における住宅地地価の形成要因を明らかにし、地域特性を反映した地価の予測精度向上を目指す。

3.2 未観測の交絡因子への対処

データ分析において、未観測の交絡因子は、モデルの予測精度や因果関係の推定において重大な問題となる。これらの交絡因子は、説明変数と目的変数の両方に影響を与えるにもかかわらず、データセットに直接観測されていないため、モデルが本来の関係性を歪めるリスクがある。未観測の交絡因子を無視すると、バイアスのある推定結果が得られ、結果の信頼性が低下する可能性が高い。これに対処するために、操作変数法を用いて、観測されていない要因の影響を制御し、推定のバイアスを最小限に抑えることができる。

3.3 構造推定

構造推定を行うことで、消費者が特定の属性に対してどの程度の選好を有しているかを明らかにすることが可能である。構造推定は、消費者の選択行動を理論モデルに基づいて推定し、選好の強さやその背後にある要因を定量的に把握する手法である。本研究においては、土地や住宅に関連する属性（例：立地条件、建物の構造、環境アメニティなど）に対する消費者の選好を効用関数の形でモデル化し、そのパラメータを推定することで、各属性が土地価格に与える影響を評価する。さらに、異なる年齢層や所得階層の消費者の間で、属性に対する選好がどのように異なるかについても分析することで、消費者の嗜好の多様性を評価し、需要構造の理解を深めることができる。

4 提案手法

本研究では、富山県警が公開している「犯罪発生マップ」から犯罪発生データを、総務省が公開しているe-StatsやRESAS APIなどから関連データを取得する。また、Mapbox APIを利用して地図画像を生成し、地理的な特徴量を抽出する。適切な機械学習アルゴリズムの決定、ハイパラメータのチューニング、特徴量の選択を行い、どのエリアにいつ犯罪が発生するかを精度高く予測するモデルを構築する。そのモデルに対してSHAPを適用し、交錯因子・相関関係のある特徴量の排除など、適切な処理を行つたうえで、犯罪発生に寄与する詳細な要因を地域ごとに可視化する。

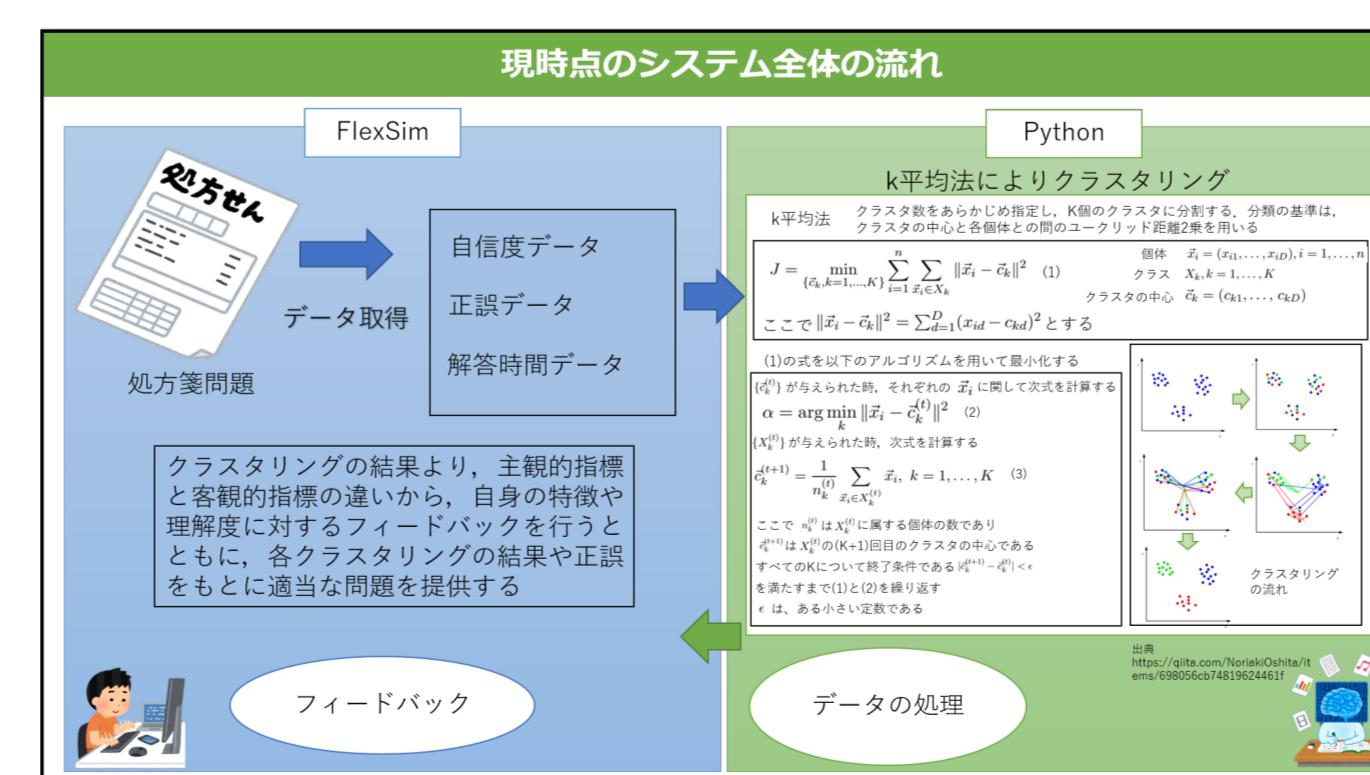


図3 提案手法の概要

正誤データの蓄積方法に関しては、問題を解いてもらい、正解の場合は1、不正解の場合は0を入力し、カンマ区切りのTXTファイルとして解答者ごとに保存する。保存されたデータは、Pythonを用いて拡張子をCSVに変換し蓄積する。その様子を図4に示す。



図4 データの蓄積の流れ

データの傾向と理解度を可視化する方法を説明する。

データの傾向を可視化する方法としては、ソートによるプロック表示法というものがある。まず、横軸を受験者縦軸を問題とし、右から点数の高い順として並べる。正解を白色とし不正解の場合は、それぞれの選択肢ごとに色を決めその色とする。そしてソートを行い、全体のデータの傾向を見やすくするというものである[7]。

理解度を可視化する方法としては、正答率、解答時間、選択肢ごとの0から10段階の自信度を合計が10になるように選択し、この3つの要素からクラスタリングを行い、解答者ごとにグループに分ける。その結果から、それぞれのグループに対して理解度診断を行う[8]。

5 数値実験並びに考察

今回の数値実験では、実際に1日ごとの犯罪が発生するリスクを予測し、その精度と犯罪発生に寄与している変数について考察を行う。予測する地域は、富山駅を中心とする東西に10km、南北に12kmの正方形の内側とし、予測する空間的な単位は、それを分割する一辺が500mのセルとした。対象地域における2009年1月1日から2018年8月31日までを学習用データとし、同年9月1日から9月30日までを予測した。

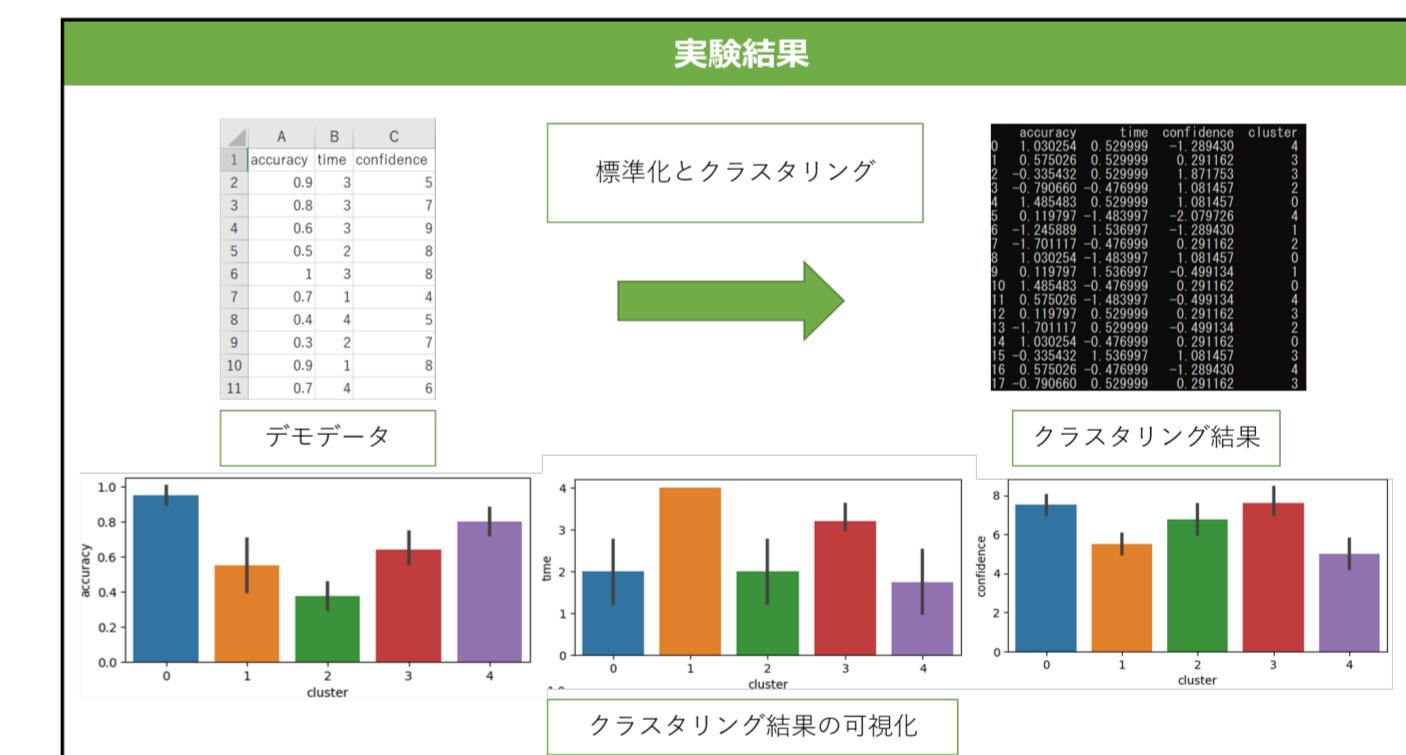


図5 実験結果

6 おわりに

本研究では、異なるアプローチから犯罪を抑止することを支援するため、過去の犯罪発生データや犯罪発生の要因となり得るデータを用いて、犯罪が発生するかどうかを時空間的に予測するとともに、犯罪発生の寄与する要因を地域ごとに抽出する。今後の課題として、特徴量や前処理などの追加などで予測精度を向上することや、交錯因子・相関関係のある特徴量の排除などで適切な因果分析を行うこと、他の地域にも適用することなどが考えられる。

参考文献

- [1] 得田 雅章「ヘドニック・アプローチによる滋賀県住宅地の地価形成要因分析」、山崎一真教授退職記念論文集（第381号）, pp. 183-205
- [2] 沢辯 隆司、山鹿 久木、水谷 徳子、大竹 文雄、「犯罪発生の地域的要因と地価への影響に関する分析」、日本経済研究 No56, pp. 70-91, 2007
- [3] 谷下 雅義、長谷川 貴陽史、清水 千弘「景観規制が戸建住宅価格に及ぼす影響—東京都世田谷区を対象としたヘドニック法による検証—」、計画行政 32(2), pp. 71-79, 2009
- [4] 金本 良嗣、中村 良平、矢澤 則彦、「ヘドニック・アプローチによる環境の価値の測定」、環境科学会誌 Vol.2, No.4, pp. 251-266, 1989