



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

価格形成要因は一般的な要因、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。需要者が効用最大化を図り、供給者が利益最大化を目指すことで均衡価格が成立するため、この手法は不動産市場分析や社会資本評価で広く利用されています。

2 サイバー空間からのデータ取得

2.1 多様な要因を考慮したデータセットの作成

地理的犯罪予測は、車上あらし、自動車や自転車などの盗難、空き巣など、犯罪を行いやすい時間や場所、状況を選んで行われるような罪種に対して行われる。その予測結果は、警察などに提示し、そのエリアに対して集中的にパトロールを行うなどといった判断を助けることを想定している。そのため、予測の範囲を、時間的には月や週、日や時単位、空間的には数十から数百メートルのグリッドセル単位としている研究が多い。このような地理的犯罪予測に関する研究は、その手法によって、大きく4つに分類することができる[3]。

2.2 データクリーニングによる前処理

機械学習による地理的犯罪予測では、その説明変数として、過去に発生した犯罪のデータのほかに、犯罪発生の要因となり得るデータを用いる。それらは、「数年、あるいは数十年間固着しうる長期のリスク」と「むこう数日から数週間にかけて増大する短期のリスク」の2つに分類することができる[4]。以降、前者を「長期的リスク」、後者を「短期的リスク」と呼ぶ。長期的リスクとして考えられるものに、「社会経済的要因」、「犯罪誘引要因」、「犯罪可能要因」が挙げられる。短期的リスクとして考えられるものに、「直近の犯罪」、「環境要因」が挙げられる。

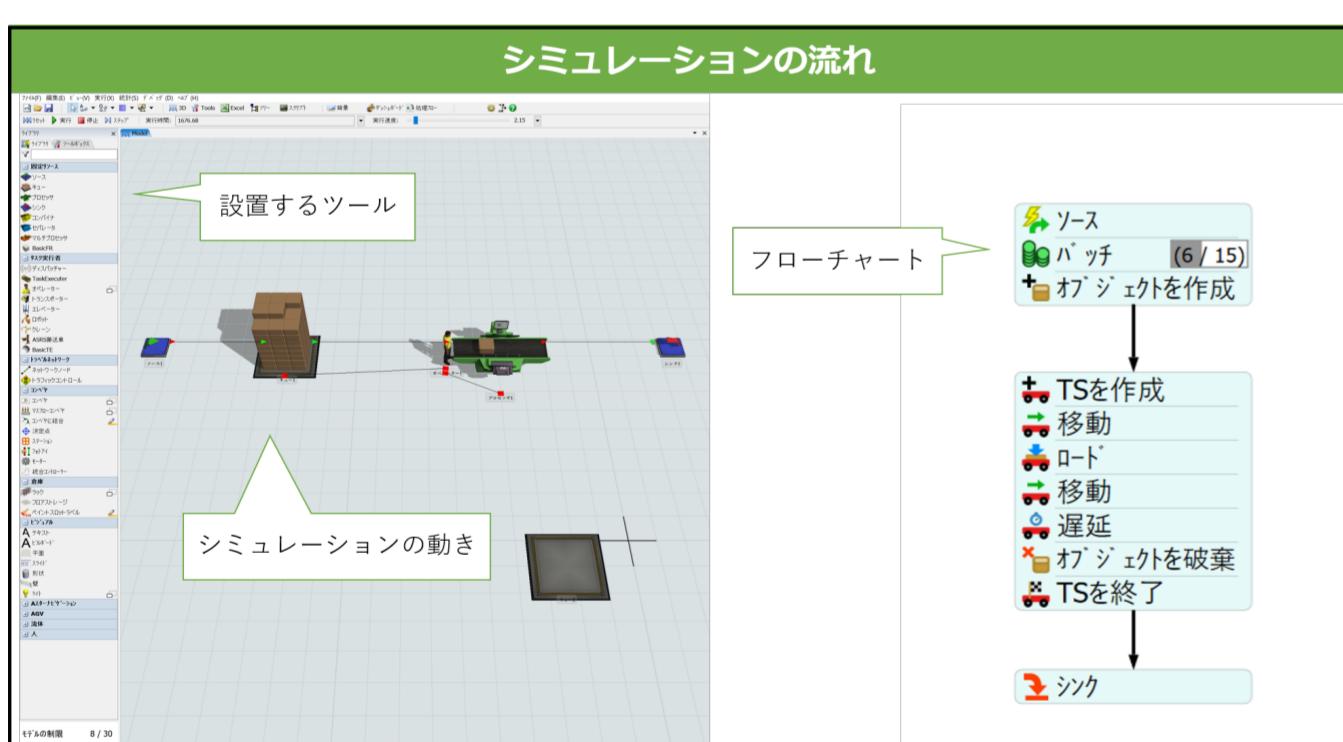


図1 FlexSimのシミュレーションの流れ

2.3 ...

わが国においては、欧米の諸外国のように早急な対策が求められているとはいえない。しかし、単身世帯の増加や、海外からの労働者の受け入れによる人種・民族の多様化など、将来的に犯罪発生リスクを増大させる要因が潜在している。また、少子高齢化・人口減少の進行により、警察官の「大量退職期」を迎えており、採用試験の受験者数も年々減少している。そのため、警備力そのものが弱まっていくことが危惧されている。こうした背景により、わが国においても、地理的犯罪予測の導入に向けた取り組みが見られはじめている。2016年には京都府警が、米国で導入されている手法を参考に犯罪予測システムの運用を開始している。また、2018年には神奈川県警、長野県警が、同様のシステムの導入を検討している[3]。

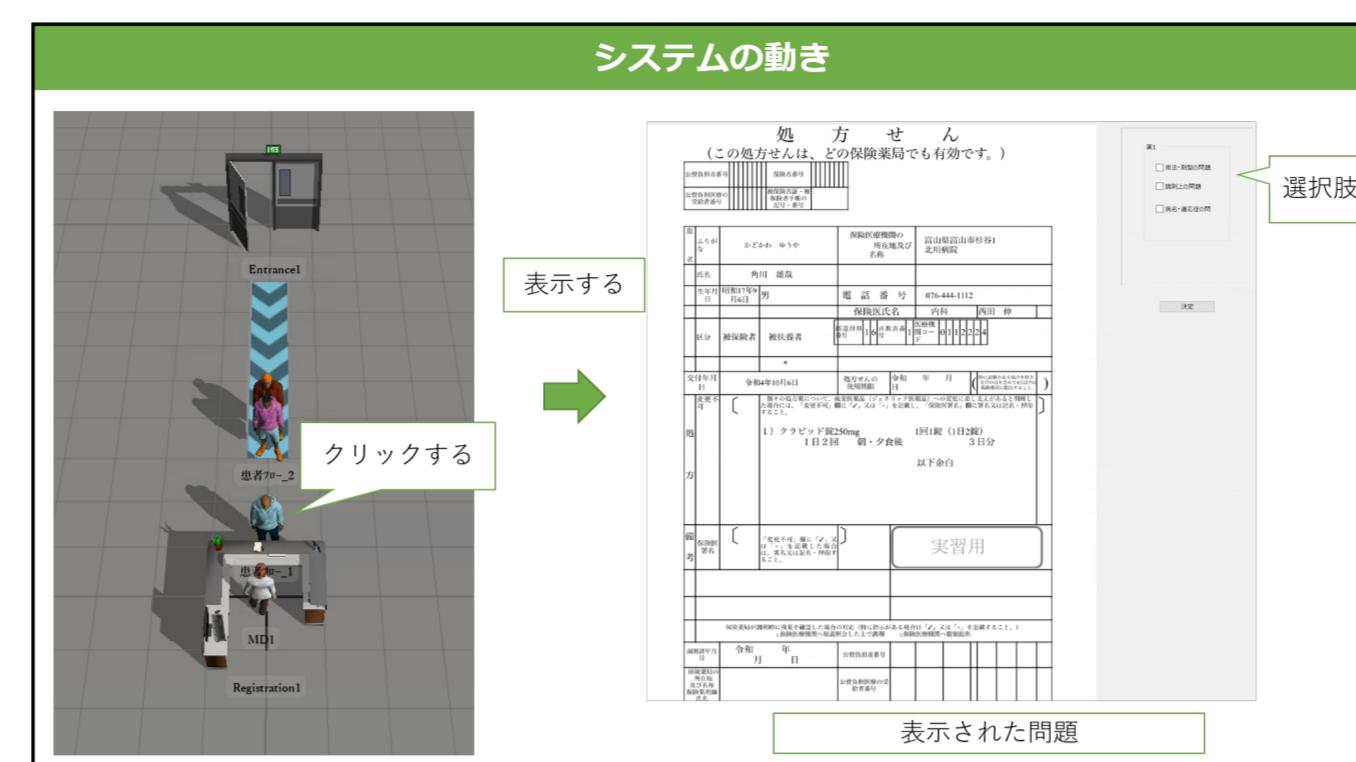


図2 FlexSimのシステムの動き

— 3 —

3.1 富山県住宅地地価ヘドニック関数の推計

機械学習は、ある変数とそれと関係がある変数との法則性を、人間の介入少なく推定する手法である。よって本来ならば必要となる、犯罪学、社会環境学などの専門的な知識や、数理モデルを考案・評価できる高度な数学的教養がなくとも、必要なデータさえあれば予測できる点で優れている。犯罪が発生するかしないかの2値分類であること、正確性をより重視することから、非線形サポートベクターマシン、ランダムフォレスト、ニューラルネットワーク、勾配ブースティング決定木の4つを選定した。それについて同様のデータで予測した結果、最も精度が大きかったニューラルネットワークを本研究では用いる（図3参照）。

3.2 离散選択モデル

2.2節で述べた犯罪が発生するリスクのうち、犯罪誘引要因や犯罪可能要因は、地理的な特徴が大きく影響している。しかし、一般的な統計データは、そのエリアの代表値を掲載しているものが多く、詳細な情報を取得するには限界がある。そこで本研究では、より詳細な地理的特徴を取得するために、地図画像を利用する。地図画像の生成には、Mapbox社が提供しているStaticTile APIを利用する。あらかじめMapbox Studio上で、道路や駐車場、建造物などといった地理的情報を追加・編集することで、分析に利用しやすい形式で生成することができる（図4参照）。生成した地図画像は、ピクセル数による面積比の算出や、道路ネットワークの抽出などを行い、犯罪発生の要因となり得る特徴量を抽出する。

3.3 予測モデルの構築

機械学習によって算出された予測モデルは、その精度が大きくなるほど、その解説可能性が小さくなる関係にある。これを「オッカムのジレンマ」という。これを解消に近づけるひとつの手法として、SHAP（SHapley Additive exPlanations）がある[5]。SHAPは、協力ゲーム理論において、プレイヤをモデルの説明変数、貢献度の期待値（Shapley値）を目的変数への影響度に見立てたものである。そのモデルの予測精度が十分に大きく、各説明変数の相関関係が十分に小さい場合は、ブラックボックス化したモデルであっても、入出力関係を解説することができる。

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。

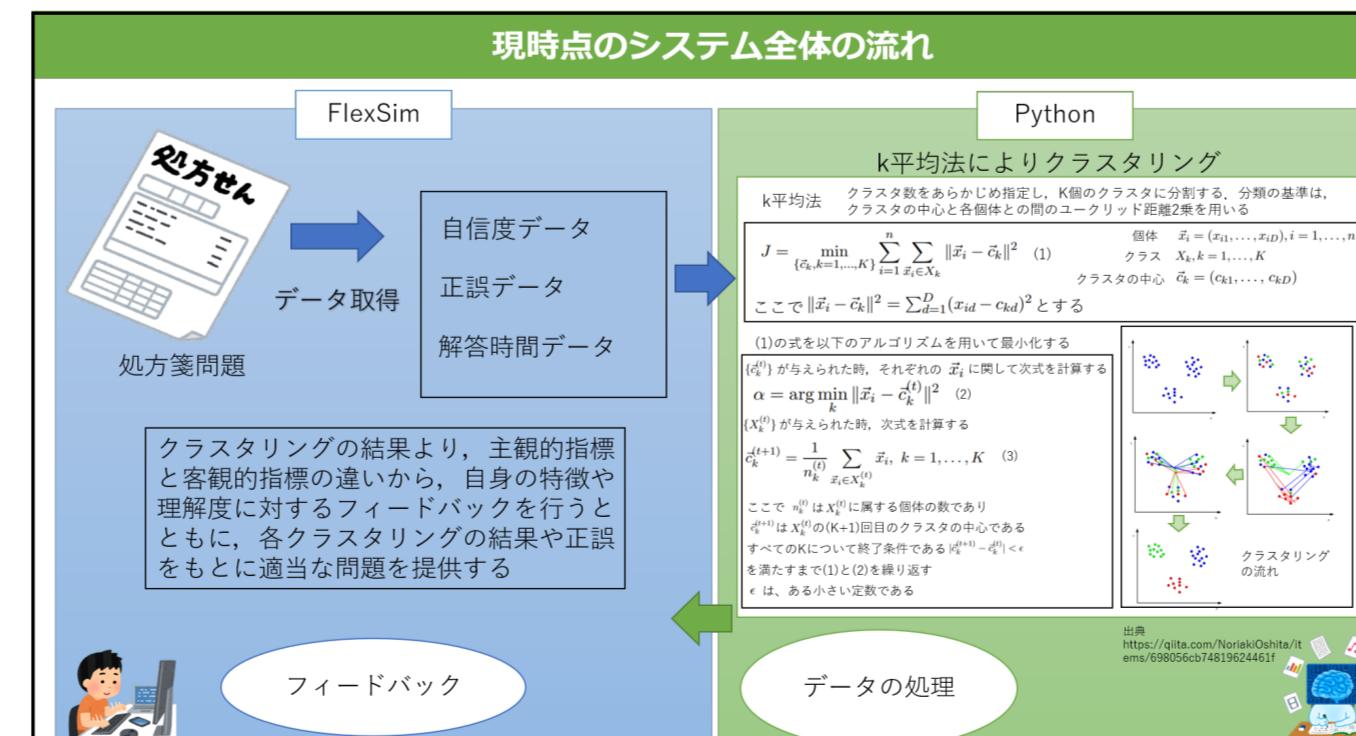


図3 提案手法の概要

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。

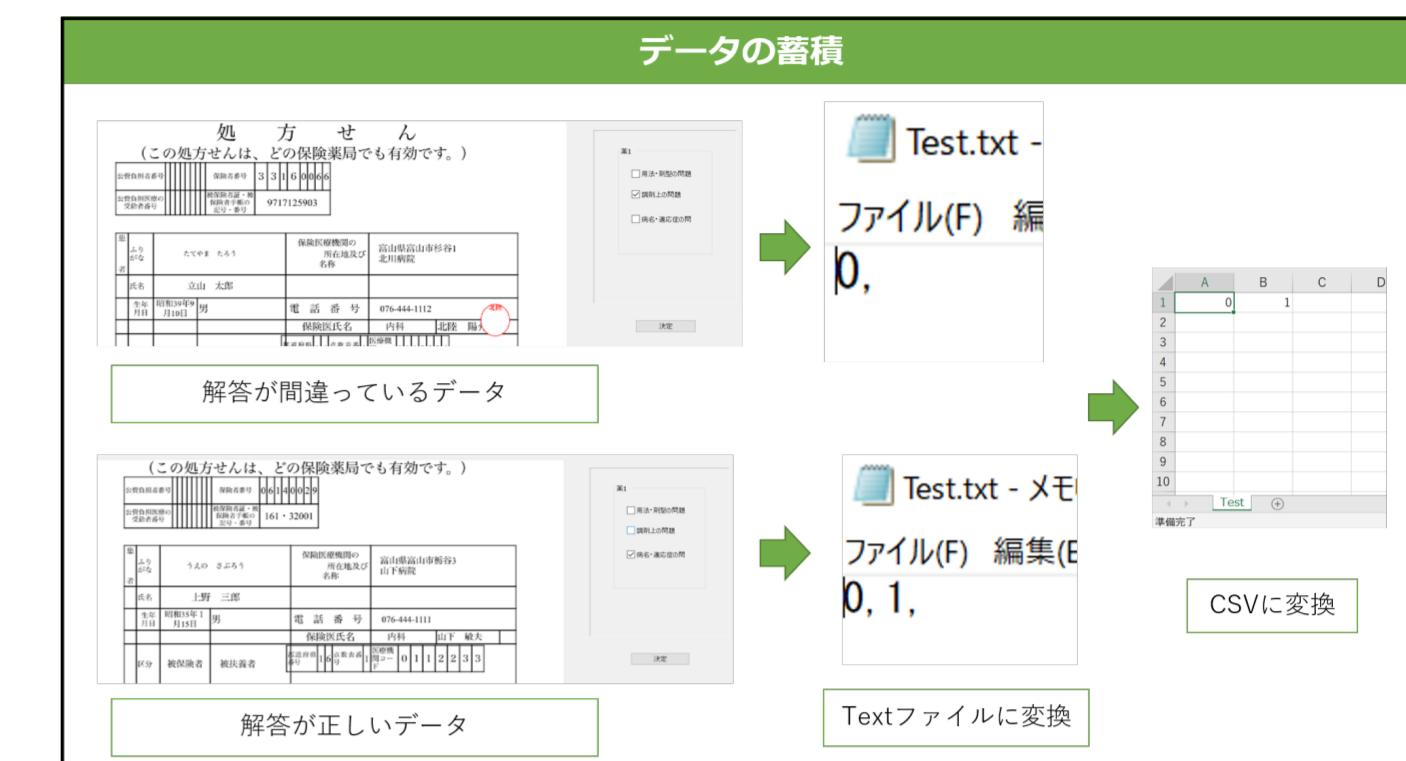


図4 データの蓄積の流れ

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、ソートによるブ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。

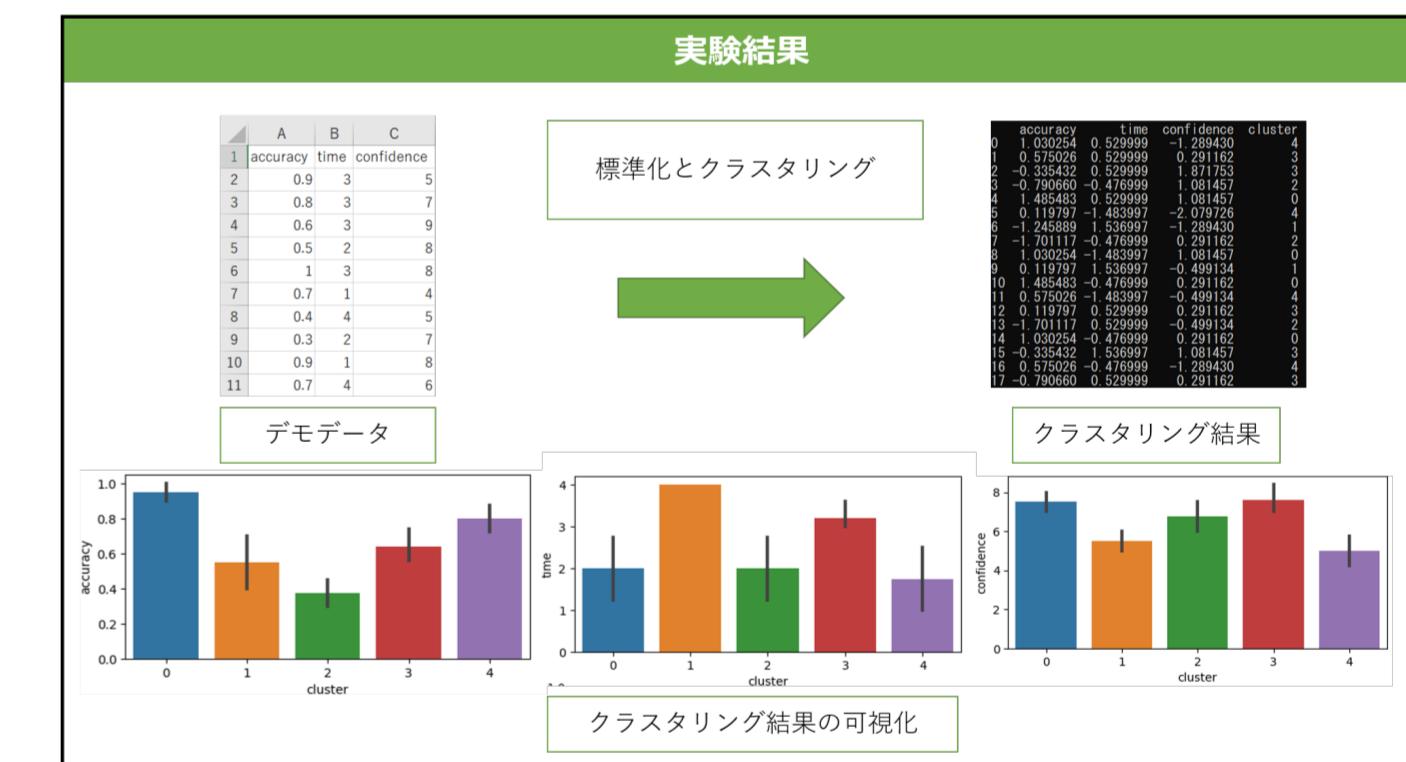


図5 実験結果

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] 岱津 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