

東京都住宅地価格モデルにおける地理空間情報と画像利用の有効性の検証と
住宅地価格の形成メカニズムの解明

Validation of Effectiveness of Geospatial Information and Aerial
Imagery in Construction of Residential Land Price Model in Tokyo /
Analysis of Price Structures in Tokyo Residential areas

不動産鑑定士 河畑靖宏

Yasuhiro Kawabata, Real Estate Appraiser

所属：公益社団法人石川県不動産鑑定士協会

勤務先：河畑不動産鑑定士事務所

連絡先：〒922-0242

石川県加賀市山代温泉 1 7 - 3 - 1

電話番号：0761-76-0569

FAX：0761-76-0969

Mail：kwbt-rea@kagacable.ne.jp

要旨

わが国の不動産の実務上、土地は建物から独立して重要な意味を持っている。令和6年地価公示の東京都住宅地価格を研究対象とし、地理空間情報と上空画像を利用して、住宅地価格の予測モデルを構築した。その過程において地理空間情報により、各公示地から一定バッファ内の便利施設等をカウントして利便性等に基づく特徴量（以下、地理空間特徴量）を作った。緑化や街並みの状態のような抽象的な地域要因は、上空からの画像からセマンティック・セグメンテーションⁱにより抽出した地物の占有比率を特徴量（以下、画像特徴量）とした。これを lightGBM 回帰ⁱⁱに与えて学習させることで、高精度の価格予測モデルを構築することができた。特に地理空間特徴量による精度向上が著しかった。一方で画像特徴量の精度向上に対する寄与度は小さかった。構築したモデルを説明変数の解析ツール SHAP を使用して説明変数の重要度を確認したところ、都心との距離の影響が特に大きく、東京都住宅地価格を一義的に決定づけていることが分かった。また都心との距離により連鎖的に価格が変化することは、空間自己相関により統計的にも確認できた。都心要因を反映した価格を前提とし、地域的な要因が影響してそれぞれの地域の価格が形成されるが、その際に接近条件に基づく利便性が強く影響している。一方で画像特徴量による影響は小さかった。そして最後にそれぞれの土地の個別属性が作用して、住宅地価格が形成されることが分かった。高精度住宅地価格モデルは価格の予測だけでなく、不動産鑑定評価における公示価格との規準や都市開発の分野での利用が期待できる。

キーワード： 地価公示、地理空間情報、セマンティック・セグメンテーション、lightGBM、SHAP、モラン I

Abstract

In the practice of real estate in Japan, the price of vacant land is important. We used geospatial information and aerial imagery to build a prediction model for residential land prices, based on the Tokyo residential land prices notified in 2024. In the process, we used geospatial information to count convenience facilities within a certain buffer from each notified land, and created features based on convenience. For abstract regional factors such as greenery and street landscaping, we extracted the occupancy ratio of ground objects from aerial images using semantic segmentation and used them as features. By feeding these features to lightGBM regression and training it, we were able to build a highly

ⁱ 画像分類の一種。第4章で詳説する。

ⁱⁱ 木形式の機械学習アルゴリズム。第5章で詳説する。

accurate price prediction model. In particular, the accuracy improvement due to geospatial features was remarkable. On the other hand, the contribution of image features to the improvement in accuracy was small. Using the explanatory variable analysis tool SHAP to check the importance of explanatory variables in the constructed model, we found that the influence of distance factors from the city center was particularly large, and that it uniquely determined the residential land prices in Tokyo. Furthermore, spatial autocorrelation was used to statistically confirm that prices change in a chain reaction depending on the distance from the city center. Prices in each region are determined based on prices reflecting city center factors, with convenience, particularly accessibility, having a strong influence. Image features, on the other hand, had little impact. Finally, it was found that residential land prices are determined by the individual attributes of each piece of land. This high-precision residential land price model is not only useful for predicting prices, but can also be used as a benchmark for publicly listed prices in real estate appraisals and in urban development.

Key words : land price public notice, geospatial information, semantic segmentation, lightGBM, SHAP, Moran's I

1. はじめに

ヘドニック価格理論においては、財は属性の集合体であり、財の価格はそれぞれの属性価値の総合体である（Lancaster. K 1966¹⁾、Rosen. S 1976²⁾）。同一時刻に同一の場所において同種の二つの財が存し、かつ両者の価格が異なる場合、その価格差はそれぞれの財の属性差を表す。

鑑定評価の手法のうち、取引事例比較法は価格差のある二つの不動産（対象不動産と取引事例）について、それぞれの不動産の価格を価格形成要因レベルに分解し、標準補正、地域要因の比較および個別的要因の比較を行うものであり、価格形成要因の差異から二つの不動産の価格差を説明する手法といえる。この意味で取引事例比較法は、ヘドニック価格理論を理論的根拠としている。

ヘドニック価格理論を現実の財の価格予測や価格メカニズムの解析に応用する場合、回帰の手法をとる場合が多い。例えばある財について、その属性が既知で価格が未知の場合、ヘドニック価格理論の考え方からは、価格は属性に対する一次式で表すことができる。

$$y = w_0 + \sum_{i=1}^n w_i x_i + e$$

y は予測価格、 w_0 は定数項、 x は財の有する属性、 w_i は属性 i に対する重みパラメータを表す。 e は価格のうち説明変数により説明することができない部分を表す。このような直感的明解さから、ヘドニックアプローチにより不動産価格を説明する研究も多く行われている（T. L. Chin et al³⁾、前田ほか⁴⁾、山田ほか⁵⁾、興ほか⁶⁾、C. Wei et al⁷⁾）。

ヘドニックアプローチにより不動産価格を説明する際、不動産の地域性を説明変数としてどのように織り込むかが重大な問題となる。地域性には、道路幅員、駅までの距離、供給処理施設など、数値で表現されているものや数値化が容易なものがある。一方で、例えば地域の名声、街路の配置、街並みの状態・美しさ、騒音・振動など抽象的・主観的で数値化が困難な地域要因がある。このような抽象的地域要因は、何らかの方法で数値化しない限りヘドニックモデルに組み込むことはできない。そのため、ヘドニックアプローチによる不動産価格の形成メカニズムに関する過去の研究は、抽象的地域要因を用いないものがほとんどであった。

現在ではコンピュータの処理能力の向上、データの普及、処理のためのモジュール開発等が進み、AI技術の応用により地理空間情報や画像データを用いて価格モデルに抽象的な地域要因を反映させる研究が活発に行われている。

先行研究により地理空間情報の利用が住宅価格モデルの精度向上に有効であることが示されているが、更地価格モデルに対する有効性の検証はほとんど行われていない。画像利用はストリートビューを用いたものが多い。ストリートビューは対象住宅の前面道路の両側の画像を使用しているが、対象物件の個別的な特徴量抽出には向いているが、面的に成立する地域の特徴量を抽出するには画像の範囲が狭い。画像利用のうち上空画像を利用

したものは、深層畳み込みニューラルネットワーク（DCNN）ⁱⁱⁱを用いて、価格や賃料に応じて画像をカテゴリー化した結果を特徴量としている。この方法は抽象的な地域要因をまとめてカテゴリーとして表現するので、そのカテゴリーがどのような特徴量を表現しているか（景観、緑化、騒音・振動など）はブラックボックス化する。

我が国では慣行上及び法律上、土地と建物は別個の不動産であると認識されている。欧米では土地と建物が一体として一つの不動産であるとの認識から、住宅売買などの一般的な不動産取引や住宅ローン申請目的の担保価値の査定などの不動産鑑定評価実務でも、土地と建物のそれぞれに別個に価格が成立しているという意識が低い。一方日本ではバブル崩壊以前は、土地は長期的には必ず値上がりする安全な資産であり、土地神話などという言葉も生まれた。また木造住宅の価値は有限であるとの意識が強く、普及レベルの住宅は長くても 50 年で価値はゼロとなり、更地価格マイナス取り壊し費などで取引されている。米国では 100 年前の木造住宅でも価値を保ち、住宅市場で取引されていることと対照的である⁸⁾。日本でも建売住宅などは土地建物一体としての価格で売買されているが、消費税は建物価格だけに課されることから、取引価格を土地と建物に分解する必要性がある。固定資産税は土地と建物のそれぞれの価値に応じて課税され、土地と建物を別々に評価している。このように、我が国では様々な不動産実務において、不動産のうち更地や更地価格が重要な意味を持っている。

本研究では地価公示の東京都住宅地価格を用いて住宅地価格モデルを構築するが、地理空間情報と上空画像を利用して住宅地域の特徴量を作る。その後モデルの精度確認において、地理空間情報を利用することの更地価格モデルに対する有効性とその程度を測定する。また街並みの状態や騒音といった抽象的な地域特徴量を、上空画像にセマンティック・セグメンテーション等を用いて計測した地物比率（道路、建物等が画像ピクセルに占める比率）により表現することを試みる。そして構築した住宅地価格モデルの説明変数の影響度を分析して、画像特徴量の有効性とその可能性について検証する。最後に価格モデルの説明変数の重要度から、東京都の住宅地価格の形成メカニズムを明らかにする。

第 2 章では、データの前処理の後、使用するデータを概説する。第 3 章では、地理空間情報を利用して利便施設等との接近条件に基づく利便性等を特徴量化する。第 4 章では、公示地の上空からの画像を処理して特徴量化する。第 5 章では、lightGBM 回帰を用いて各データの組み合わせによる精度比較を行い、地理空間情報の利用が更地価格モデルの精度向上にも有効であることを検証する。第 6 章では、第 5 章で構築したモデルによる説明変数の影響度から画像特徴量の価格に対する影響を計測し、上空画像を価格モデルに利用することの可能性について検討する。最後に東京都住宅地価格の形成メカニズムについて考察する。第 7 章は、まとめと価格モデルの実務上の応用、そして将来展望を示す。

ⁱⁱⁱ 畳み込みニューラルネットワーク（Convolutional Neural Network : CNN）の一種で階層が深いものを指す。

以下の作業は、Python を使用した。作業環境は以下のとおりである。

CPU : Intel 13th Gen Corei5-13400 ×16

メインメモリ : 32GB

GPU : NVIDIA GeForce RTX3060 Ti ×1

Python : 3.12.5

Numpy : 1.26.4

Pandas : 2.2.2

Geopandas : 1.0.1

PyTorch⁹⁾ : 2.4.0

Matplotlib : 3.9.2

Folium : 0.17.0

SHAP : 0.46.0

2. データの取得と対象とするデータの理解

令和6年地価公示東京都住宅地価格データを取得し、機械学習モデルの訓練に使用することができるデータ形式にする（前処理）。その後、東京都住宅地価格について、探索的データ分析（Exploratory Data Analysis : EDA）と探索的空間データ分析（Exploratory Spatial Data Analysis : ESDA）を行う。

2-1. データの取得

国土数値情報ダウンロードサイト [10\)](#)から東京都の令和6年地価公示データを取得した。データには島しょ部の公示地も含まれる。島しょ部の公示地は、例えば都心との距離等が超巨大化すること、駅関係のデータを持たないこと等の理由から、島しょ部の公示地は研究対象から除外した。その結果、採用するデータは1,685地点になった。

ダウンロードしたデータは「Table 1」の形式をとる。

	経度	緯度	行政区域コード	標準地名	番号用途区分	番号連番	地積	給排水等状況	土地形状	前面道路幅員	側道状況	駅距離	都市計画用途地域	防火地域	建蔽率	容積率	価格R06
0	503081.334	128484.504	13101	千代田	0	1	969	111	—	11.0	—	500	2住居	防火	60	400	3600000
1	503055.072	128452.320	13101	千代田	0	2	535	111	—	5.4	—	300	2住居	防火	60	400	2260000
2	503038.343	128477.308	13101	千代田	0	3	2266	111	—	15.0	—	330	2住居	防火	60	400	4390000
3	503087.317	128515.087	13101	千代田	0	4	142	111	—	8.0	側道	360	1住居	防火	60	400	1740000
4	503086.147	128505.888	13101	千代田	0	5	1511	111	台形	12.5	側道	450	1住居	防火	60	400	3340000

Table 1 ダウンロードした地価公示東京都住宅地

2-2. データの前処理

取得したデータは、次のとおり前処理した。

- ① 緯度経度は秒単位であるので、度単位に変換した（÷3600）。

- ② 給排水等状況は1とゼロで表現されている。このままでは数字として認識されてしまうので、111を「水道・ガス・下水」、11を「水道・下水」、101を「水道・ガス」にそれぞれ置き換えた。
- ③ 土地形状の「一」は、長方形または正方形である。ここでは「整形」と置き換えた。
- ④ 商業地には前面道路の幅員がゼロで駅前広場となっているものがあるが、住宅地には幅員ゼロ地点がないことを確認した。
- ⑤ 側道区分の「一」は「中間画地」に置き換えた。
- ⑥ 用途区分の「一」は「無指定」に置き換えた。
- ⑦ 防火地域の「一」は「無指定」に置き換えた。
- ⑧ 「R6価格」は対数化した。
- ⑨ 使用する列は「緯度、経度、地積、給排水等状況、土地形状、前面道路幅員、側道状況、駅距離、都市計画用途地域、防火地域、建蔽率、容積率、価格R6、log価格」を選択した。

前処理後のデータは「Table 2」のとおり。

	行政区画コード	標準地名	地積	給排水等状況	土地形状	前面道路幅員	側道状況	駅距離	都市計画用途地域	防火地域	建蔽率	容積率	価格R06	log価格
0	13101	千代田	969	水道・ガス・下水	整形	11.0	中間画地	500	2住居	防火	60	400	3600000	15.096444
1	13101	千代田	535	水道・ガス・下水	整形	5.4	中間画地	300	2住居	防火	60	400	2260000	14.630875
2	13101	千代田	2266	水道・ガス・下水	整形	15.0	中間画地	330	2住居	防火	60	400	4390000	15.294840
3	13101	千代田	142	水道・ガス・下水	整形	8.0	側道	360	1住居	防火	60	400	1740000	14.369396
4	13101	千代田	1511	水道・ガス・下水	台形	12.5	側道	450	1住居	防火	60	400	3340000	15.021481

Table 2 前処理後のデータ

このデータうち「給排水等状況、土地形状、側道状況、都市計画用途地域、防火地域」をダミー変数としたものを、ベースデータとする。

2-3. 探索的データ分析（EDA）

探索的データ分析は、データが持つ特徴や傾向を明らかにすることである。数値データの基本統計量は次のとおりである [11\)](#)。

	価格R06	log価格	地積	前面道路幅員	駅距離	建蔽率	容積率
count	1685.00	1685.00	1685.00	1685.00	1685.00	1685.00	1685.00
mean	479365.22	12.77	243.88	6.09	1052.99	52.75	163.51
std	461084.35	0.78	499.62	3.68	943.68	9.68	89.58
min	28900.00	10.27	47.00	2.00	0.00	30.00	50.00
25%	207000.00	12.24	117.00	4.00	500.00	40.00	80.00
50%	367000.00	12.81	150.00	5.40	780.00	60.00	150.00
75%	587000.00	13.28	198.00	6.00	1200.00	60.00	200.00
max	5350000.00	15.49	11464.00	40.00	8500.00	80.00	600.00

Table 3 地価公示東京都住宅地価格の基本統計量

対象とする公示地は1,685地点、平均価格は約479,000円、最低価格は28,900円、最

高価格は 5,350,000 円で、最低価格と最高価格との間には約 185 倍の差がある。

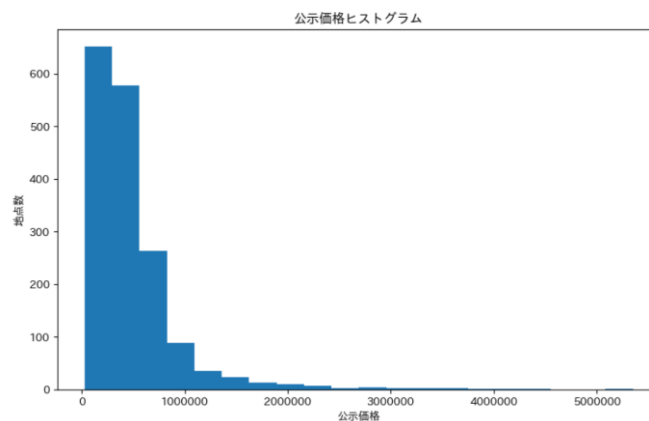


Figure 1 R 5 価格の分布

対数前の価格（真数価格）でヒストグラムを作ると、グラフは左に偏り右のすそ野が広がる（Figure 1）。R 6 価格の平均が 40 万円台であることから分かるが、ほとんどが 100 万円未満に集中している。R 6 価格を対数化すると、正規分布に近づく（Figure 2）。

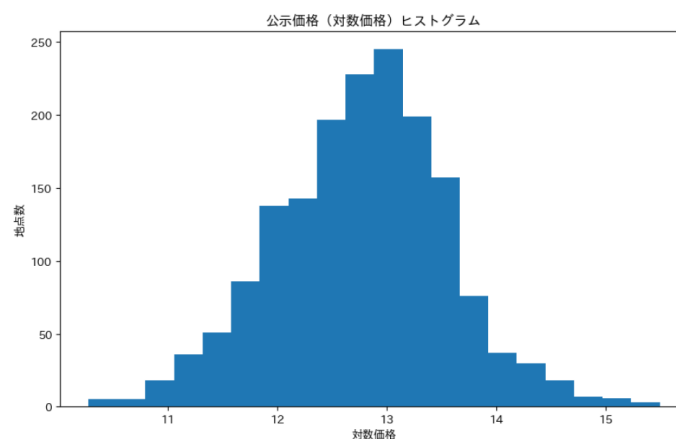


Figure 2 対数価格の分布

「Figure 3」は、採用した説明変数間の相関を表す。log 価格と前面道路の幅員、建蔽率、容積率は正の相関があるようだ。log 価格と駅距離は負の相関が認められる。log 価格と地積との間には負の相関があるようにも見えるが、あまり明確ではない。

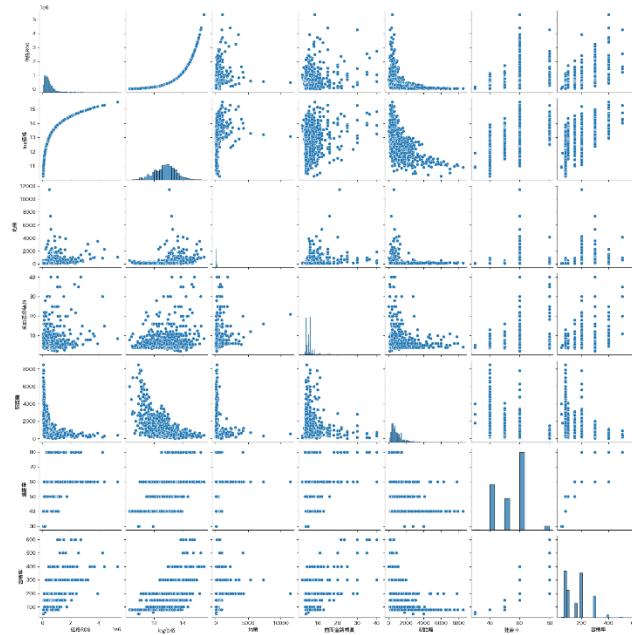


Figure 3 相関行列

2-4. 探索的空間データ分析 (ESDA)

探索的空間データ分析は、データが持つ空間構造を探り出すことである。具体的には地理空間上に分散するデータを分析し、地理空間における分散の特徴や傾向を明らかにすることである。

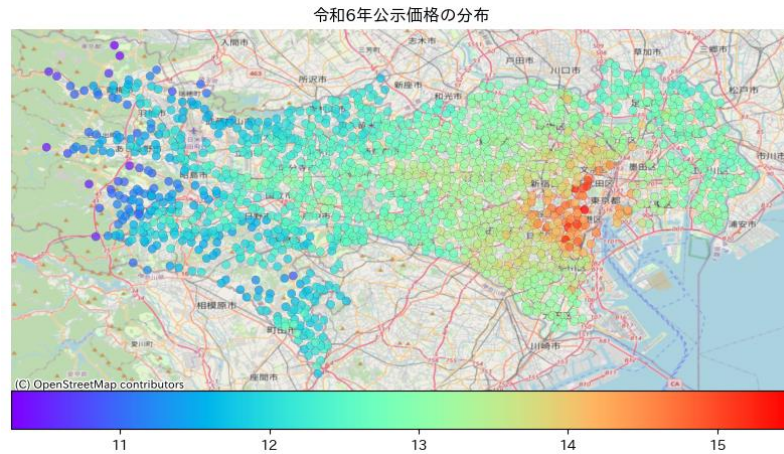


Figure 4 公示地の分布

公示地をプロットしたものが「Figure 4」である。都心に最も高い地点が集中し、離れるにつれて価格は低下している。千代田区、中央区、港区に高い公示価格のクラスターがあり、西端部に安い価格クラスターが発生している。中間価格帯でも、中の上、中、中の下など、地理的にまとまりがあるように見える。

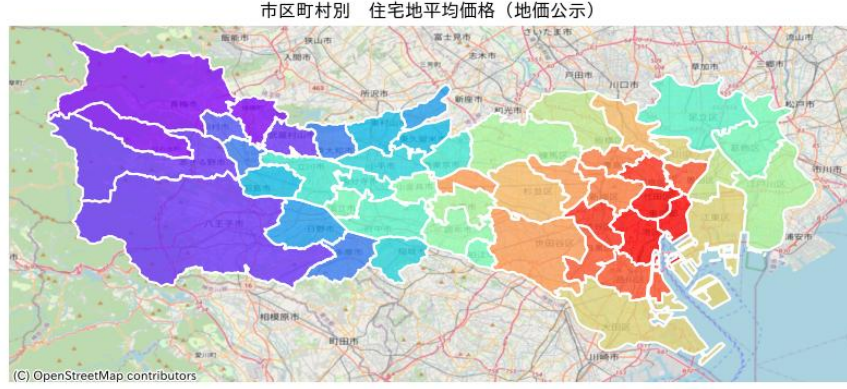


Figure 5 市区町村別平均価格

市区町村別の平均価格をプロットするとより分かりやすい（「Figure 5」）。千代田区と港区が最も高く、この2区を取り囲むように次の価格帯があり、さらにその外周には次の価格帯がある。価格が安い市町村は都の西部の県境付近に集中している。住宅地価格は、都心部が最も高く郊外に向かうにつれて連鎖的に下落していく様子が分かる。

二つの図から、地域的な価格クラスターの発生、価格の連鎖的变化など、地理空間上でまとまりや傾向を示している可能性がある。

2-5. 空間自己相関の検証

ある物の属性（または属性集合）が他の物の属性と系列的、あるいは連鎖的に相関することを自己相関という。自己相関が地理空間上で認められることを、空間自己相関という。空間自己相関は、公示価格の空間分布やその背後のプロセスを理解するために重要な役割を担っている [12\)](#)。

可視化資料から東京都の住宅地価格に空間自己相関が認められる可能性があるので、統計的に検証する。空間自己相関の程度を表す統計値の一つに **Moran's I** [13\)](#)（グローバル Moran）がある。Moran's I 統計量は次の式により表すことができる。

$$Moran's I = \frac{n}{W} \frac{\sum_{i=1}^n \sum_{j=1}^n w(i,j) (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

$(y_i - \bar{y})$: 地点*i*の公示価格の偏差

$(y_j - \bar{y})$: 地点*j*の公示価格の偏差

\bar{y} : 分析対象地域全体の公示価格の平均

$w(i,j)$: *ij*間の重み

n : 観測値の数

W : 重み行列の全要素の合計

$\frac{W}{n}$: *ij*間の重みの平均

$i = j$ のとき、 $\sum_{i=1}^n \sum_{j=1}^n w(i,j) (y_i - \bar{y})(y_j - \bar{y}) = \frac{W}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ となるので、**Moran's I** = 1になる。つまり周辺地点（*j*）と類似性が高い対象地点（*i*）の数が多いほど、**Moran's I** は1に近づき、正の空間自己相関が強くなる。その逆のとき、つまり近くにあるほど相反する値をとる地点が多いときは、負の空間自己相関が強くなりマイナス1に近づく。ゼロ

付近は空間的にランダムであることを意味する。「空間的ランダム」とは、空間上のどの位置においても周辺からの影響を受けないし、自分自身も周辺に対して影響を与えないことを意味する。

令和6年地価公示の住宅地データから、R6対数価格と緯度・経度情報を抽出し、pysalのMoran関数に渡してMoran's I統計量を求めた結果は、「Figure 14」のとおりである。

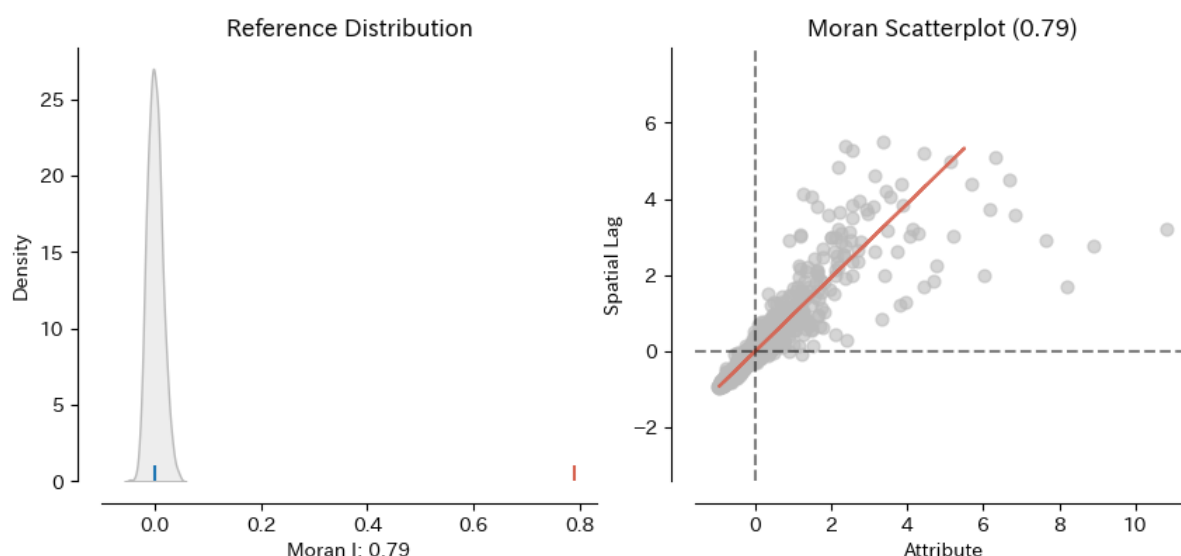


Figure 14 グローバル Moran の結果とローカル Moran の散布図

左の図は、空間自己相関がないと仮定してシミュレーションした相関係数の分布を表す。例えば渋谷区内に存在するある公示地に対して、「同一需給圏内の類似地域」を完全に無視してランダムに数十個の公示地を選択し、渋谷区内の対象公示地との間の相関を測定することである。その結果はほぼゼロ、つまり無相関となった。Moran's I 統計量は左図中の赤線の位置 0.79 となった。東京都住宅地価格はランダムに成立しているのではなく、近傍に位置する別の公示価格と影響を与え合いながら成立しており、住宅地価格が都心等から連鎖的に変化しているのは偶然ではないことを統計的に確認した。

右の図は相関の方向を表す。右上の第一象限と左下の第三象限にデータが集中すると、右肩上がりの正の空間自己相関である。

3. 地理空間情報を利用した特徴量の作成

公示地データは地理空間情報（緯度経度などの座標や住所）を持っている。もし駅や食品スーパーなどの便利施設等も地理空間情報を持っている、あるいは位置情報を付加することができるならば、両者の座標を利用して地理空間上の位置関係を説明変数として抽出することができる。

3-1. 地理空間情報の活用に関する先行研究

小川ほか [14\)](#) は建物の位置情報を含むデータを地図と結合して作成した地理空間特徴量

と建物の画像データのそれぞれを用いて、建物の築年と構造を分類するモデルを構築しその精度検証を行った。

Wei et al (2022)¹⁵⁾ は、不動産鑑定評価に関係するデータとデータソース並びにモデリングのアルゴリズムをまとめた。データに関連するものには、利便施設の地図データ、画像データ、リモートセンシングデータ、GPS データなど、座標情報を含むものも多い。

Hill et al (2013)¹⁶⁾ はオーストラリアのシドニーの住宅価格指数を用いて、価格モデルに地理空間特徴量による内挿を行うことで、フィッシャー価格指数が 15～30% 高まることを明らかにし、地理空間特徴量を利用することの有効性を示した。

Das et al (2020)¹⁷⁾ は、利便施設が対象不動産（住宅）に与える影響を、geo-spatial network embedding と呼ばれるフレームワークで特徴量化することで、予測精度が平均絶対誤差で 8.1%、平均平方二乗誤差が 7.2% それぞれ改善することを示し、価格予測に地理空間特徴量が有効であることを明らかにした。

鶴田ら(2021)¹⁸⁾ は、緯度経度を利用して取得することが可能な情報（災害マップや行政情報など）を説明変数に組み込むことにより、賃料の予測精度が向上することを示した。

3-2. 地理空間情報の特徴量化

公示地と利便施設等との空間上の位置関係を数値化し特徴量を作る。

3-2-1. 利便施設等情報、座標情報の取得、距離マトリックスの作成

住宅地価格に影響を及ぼす可能性がある利便施設等として、都心、副都心、コンビニ、ドラッグストア、ホームセンター、食品スーパー、公園、小学校、駅の 9 つの利便施設と、嫌悪施設として污水处理施設を加えた 10 施設を選択した。駅に関してはベースデータに最寄り駅までの距離があるが、複数駅利用の利便性を表現するために「駅」も利便施設に加えた。

都心（東京駅丸の内口）と副都心（新宿駅）は Google Map から座標情報を取得して公示地までの直線距離を測定した。

コンビニは（株）アインツから「コンビニエンスストアリスト」を購入した。

このリストはコンビニ大手 3 社（セブンイレブン、ファミリーマート、ローソン）だけの全国リストであるが、大手 3 社の店舗数シェアは 2024 年 1 月時点で 90.4%¹⁹⁾ であり、コンビニに関する利便性は大手 3 社だけで表現することが可能と判断した。購入したコンビニリストは座標情報を持っていないので、Google Map Platform²⁰⁾ の Geocoding API を利用して、住所を座標に変換した。

その他の 7 施設については、同じく Google Map Platform の Google Place API を使用して名称と座標を取得した。Geocoding API や Google Place API は一定の使用量を超えると有料であり、東京都と隣接県の県境付近に存在する 7 施設のデータを取得するには 5 万円以上課金された。

これで 10 施設と公示地の両方に座標情報を持たせることができた。次に双方の座標情報に基づいて直線距離を測定し、すべての公示地を行、すべての利便施設等を列とした

1685 行×15981 列（10 施設の総数）の距離マトリックスを作成した 21)。

3-2-2. 利便施設等との距離情報の特徴量化

利便施設等の特徴量化する方法としては、従来は最も距離が近いものまでの距離を採用する方法が一般的であった。しかし従来の方法では、例えば施設が複数存在することによる利便性の差が全く考慮されないなどの弊害がある。そこで現在では、対象からの一定距離（バッファ）内に存在する利便施設の数の特徴量とする方法が活用されている（Figure 6）。



Figure 6 施設の数え方

作成済みの距離マトリックスを使用して、各公示地について、250m、500m、750m、1000m、1250m、1500mのそれぞれのバッファ内に存在する施設等（列）の数をカウントし、公示地 1685 行×（8 施設×6 距離）の利便施設等の地点数マトリックスを作成した 21)。これが地理空間特徴量になる。

「Table 4」は、インデックス 50～54 の公示地について、コンビニに関する利便性の特徴量化したものの例示である。数字はそれぞれのバッファ内のコンビニの数を表す。

	conveni_250m	conveni_500m	conveni_750m	conveni_1000m	conveni_1250m	conveni_1500m
50	0	2	10	33	53	97
51	1	2	10	23	43	70
52	0	6	16	23	37	53
53	0	8	13	43	77	108
54	2	14	31	45	83	117

Table 4 地理空間情報の例

4. 画像データの利用

ここまでの作業により取得した特徴量は、道路幅員、都市計画情報、利便施設等の数など、客観的で数値化が容易なデータである。住宅地価格の形成は、このような客観的要因だけでなく、不動産鑑定評価基準が住宅地域の地域要因として例示している『住宅、生垣、街路修景等の街並みの状態』や『眺望、景観等の自然的環境の良否』のような、具体化が困難な要因も反映している。

このような抽象的価値、あるいは需要者の主観的価値を測定する手段としては、アンケートなどがあるが時間と費用の点で現実的ではない。そこで本研究では公示地の上空からの画像を利用して、地域の抽象的な情報を抽出することを試みる。具体的には、公示地を中心とする上空画像に含まれる「建物、道路、駐車場、水域、軌道、緑地」の6種類の地物のピクセル数が画像全体に占める比率を求め、これにより抽象的な地域要因を表現する。

4-1. 画像データの利用に関する先行研究

Archith J. Bency et al (2017) 22) は、英国のロンドン、リバプール、バーミンガム

の 3 都市の住宅価格と賃料を研究対象として、6 段階のズームレベルの上空画像を CNN で分類したカテゴリ特徴量、地理空間特徴量、建物の特徴量の 3 種類の特徴量を用いて、線形回帰、ランダムフォレスト回帰^{iv}、深層学習モデルにより価格（賃料）を予測した。これにより上空からの画像データが、近隣環境の特徴量化に有効であることを示した。

Yuhao Kang et al (2020) ²³⁾ は、ストリートビュー画像から深層学習により抽出した特徴量とその他の特徴量（建物の特徴、施設の整備状況、交通機関との接近性、地域の社会経済的特徴など）を融合して作り出した特徴量に、機械学習モデルと地理的加重回帰を適用して住宅価格の予測を行った。

Wei Li et al (2020) ²⁴⁾ は、街路の美しさや景観といった数値化が困難な特徴量が価格形成に重要な影響を与えているとの仮説をたて、ストリート画像にセマンティック・セグメンテーションを適用して求めた 18 カテゴリの占有比率を特徴量として、陝西省西安市の住宅価格の予測を行った。

Xiang Xu et al (2022) ²⁵⁾ は、住宅の価格形成要因を①建物に関する属性、②地域要因、③地域より狭い近隣要因、ストリート画像の④主観的印象と⑤客観的印象に分類し、線形回帰により上海市の住宅地価格を説明した。これによって画像データから作成した説明変数の有効性を示した。

環境に関する特徴量である街区の美しさや名声のような数値化が困難あるいは主観が入り込みやすい特徴量を表現するために、画像データの活用は有効である。Stephen Law et al (2019) ²⁶⁾ は、通常の特徴量（建物の特徴量、接近条件など）、ストリート画像、衛星画像の 3 種類の特徴量を組み合わせたモデルが、最も精度が高くなることを示した。

Zona Kostic et al (2020) ²⁷⁾ は、通常の特徴量、ストリート画像と上空画像に CNN を適用して抽出した画像特徴量、建物内部の画像の 3 つの特徴量を使用し、価格モデルの精度比較を行った。

Cordts, M et al (2016) ²⁸⁾ は、ストリート画像をセマンティック・セグメンテーションにより解析するときに使用するデータセットは、Cityscapes Dataset²⁹⁾ が他の学習済みモデル (PASCAL VOC³⁰⁾、KITTI³¹⁾ など) と比較して、より有効であると述べている。

Marco, M et al (2017) ³²⁾ は、スペインのバレンシア州において、Google のストリート画像に含まれる路上のごみや壁面の落書き、廃屋などを計測し、ストリート画像が地域の荒廃度を測定する手段として有効であることを示した。

Kucklick et al (2021) ³³⁾ は、GIS を使用して収集した住宅の特徴量、住宅売買データに含まれる情報、建物の外観画像の 3 種類の特徴量を組み合わせてモデルを構築した結果、売買データ情報と画像データを組み合わせた場合と、3 つの特徴量をすべて採用した場合に精度が高まることを示した。

^{iv} 木形式の機械学習アルゴリズム。回帰と分類がある。

鈴木（2023）³⁴⁾は、東京郊外の低層住宅地を対象として、ストリート画像から抽出した緑・植栽、空、建物の構成比率が、戸建住宅の成約価格に与える影響を明らかにし、画像データの利用の有効性を示している。

浅井ら（2015）³⁵⁾は、北海道の富良野周辺地域において、ストリート画像を用いて、道路景観を定量的に把握する研究を行った。

4-2. 画像データの取得

公示地の上空画像は、「国土地理院タイル・ベースマップ・写真」からを取得した（URL：<https://cyberjapandata.gsi.go.jp/xyz/seamlessphoto/{z}/{x}/{y}.jpg>）。

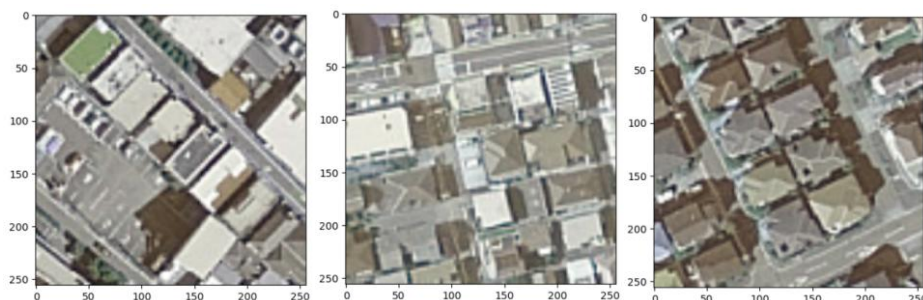


Figure 7 画像データの例（左から品川-30、葛飾-7、町田-8）

4-3. 地物抽出モデルの自作と画像データの特徴量化

画像データは色の強弱値の集合であり、そのままの状態では意味を持たない。そこで各画像中の全ピクセルを、「建物、道路、駐車場、水域、軌道、緑地、いずれにも属さないもの」のいずれかに分類し、全ピクセルに対するそれぞれの地物の比率をもって画像の特徴量とする。

緑地に関しては、画像中の緑系ピクセルを1、その他をゼロとすることで容易に緑地比率を測定することができた。その他の「建物、道路、駐車場、水域、軌道」は、画像から建物等の地物を認識するための機械学習モデルを自作する必要がある。

画像ピクセルが何かに属することを判定することをセマンティック・セグメンテーションという³⁶⁾。セマンティック・セグメンテーション技術は、急速に発展している物体検知に関するAI技術で、自動運転をはじめとする画像認識タスクや生成AIに広く応用が進んでいる³⁷⁾。

4-3-1. セマンティック・セグメンテーションと Unet

セマンティック・セグメンテーションとは、画像を構成するピクセルをカテゴリーに分割することである。

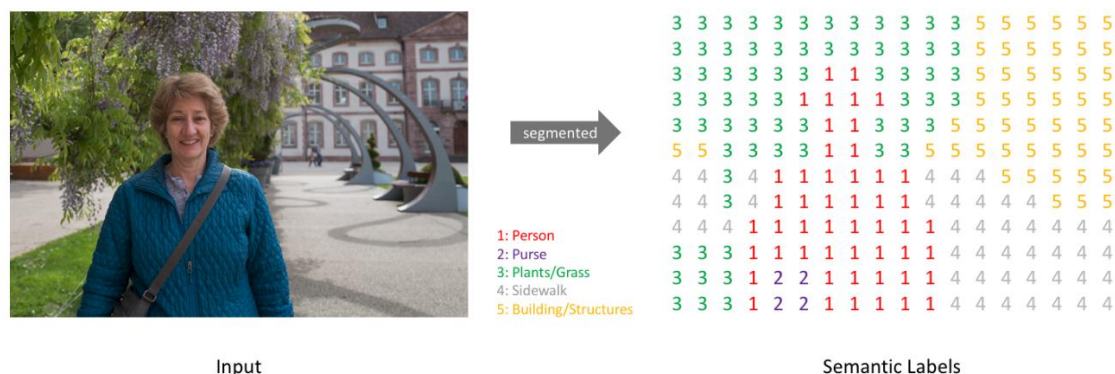


Figure 8 セマンティック・セグメンテーションの例
(Jeremy Jordan, An overview of semantic image segmentation,
<https://www.jeremyjordan.me/semantic-segmentation/>)

「Figure 8」は 12×18 ピクセルの画像をセマンティック・セグメンテーションにより 5 つのカテゴリに分割した例である。セマンティック・セグメンテーションモデルは、グレースケールの 2D 画像や 3D カラー画像を入力すると、縦横は元画像と同サイズの 2 次元配列を出力する。

本研究の自作モデルの構築では、畳み込みニューラルネットワーク (CNN³⁸⁾) をベースとして開発されたセマンティック・セグメンテーション用のモデルである Unet³⁹⁾を使用した。Unet の構造を図示すると「Figure 9」のようになる。

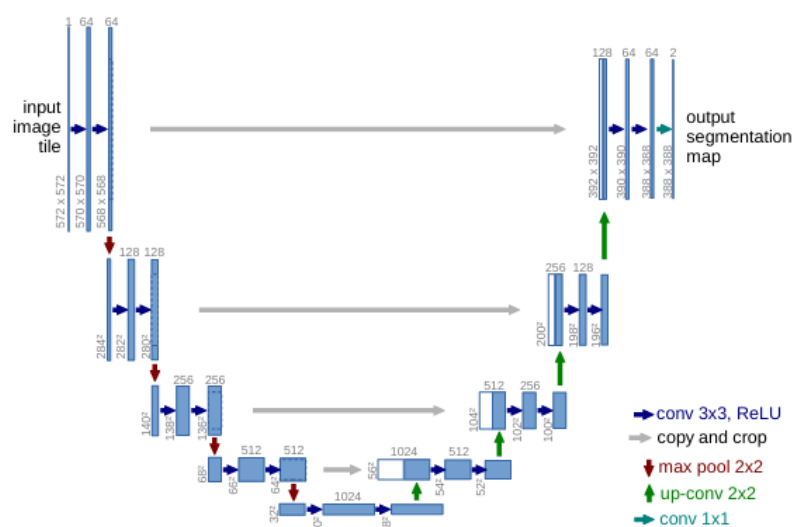


Figure 9 UNet の構造³⁹⁾

Unet は、画像を入力すると、ある地物の形状や色等の特徴とそれが存在する画像の位置情報の両方を入力する事前学習済みモデルである。中心より左側が画像の入力から各物体の特徴を抽出する部分であり、エンコーダーと呼ばれる。右側はエンコーダーが抽出した特徴を入力サイズまで拡大しながら、その物体が存在した入力データの位置に特徴を配置する部分でありデコーダーと呼ばれる。UNet ではエンコーダー部分を他の事前学習モデルと付け替えることが可能で、本研究ではエンコーダーに ResNet34⁴⁰⁾を使用した。

ResNet34 の 34 は層の深さを表す。従来 CNN モデルは層を深くすると精度は向上する

が、勾配消失問題によりモデルの学習が進まなくなるという欠点があった。ResNet はスキップコネクションという仕組みを用いて、勾配喪失問題を解消することに成功し、CNN モデルの多層化を可能にした⁴¹⁾。

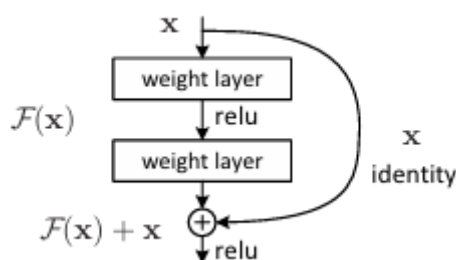


Figure 10 スキップコネクションの構造⁴²⁾

4-3-2. 国土地理院 GIS データセット

エンコーダー部分を resnet34 に組み替えた Unet（以下、自作モデル）を国土地理院 GIS データセット⁴³⁾ から取得した画像データで訓練し、「建物、道路、駐車場、水域、軌道」のそれぞれを抽出する 5 つのモデルを構築した。

国土地理院 GIS データセットは、国土地理院の特別研究「AI を活用した地物自動抽出に関する研究（平成 30 年度～令和 4 年度）」の研究成果として公表されている。このデータは、道路、水部、防波堤、太陽光発電設備、タンク、透過水制、普通建物、茶畑、墓地・墓碑、水田、駐車場、軌道、分離帯、高塔、竹林の 18 のカテゴリについて、訓練用画像と教師用画像がセットになっており、ホームページから無償でダウンロードすることができる。



Figure 11 道路の訓練用画像（左）と教師用画像（右）の例

自作モデルの学習に要する時間は 1 モデルあたり約 8 時間～12 時間であり、「建物、道路、軌道、水域、駐車場」の 5 つのモデルを学習させるにはトータルで約 50 時間を要した。モデルの精度確認には Dice Loss⁴⁴⁾ を評価指標として用いた。Dice Loss はセマンティック・セグメンテーションモデルを学習させる際の損失関数で、次の式により表すことができる。

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p} + 1}{y + \hat{p} + 1}$$

y は正解ラベル、 \hat{p} は予測値、右辺第 2 項は Dice 係数と呼ばれ、分母がゼロにならないように分母と分子に 1 を加えている。例えば本件のような 2 値分類（例えば道路か、それ以

外化) の場合、 y と \hat{p} は 0 または 1 の値をとる。 $y = \hat{p}$ のとき $DL = 0$ になる (誤差がゼロ)。 $y \neq \hat{p}$ のならば $DL = 0.5$ になる。自作モデルの損失は最終的に 0.2 台後半から 0.3 台になった。

4-3-4. 画像データの特徴量化

セマンティック・セグメンテーション・モデルは、入力した画像に対して各ピクセルが道路等のカテゴリーに該当する確率を返す。そこで確率 50% を超えるピクセルをそのカテゴリーを表すピクセルとして 1、50% 以下をそのカテゴリーではないピクセルとしてゼロで表現することで、画像内の各ピクセルを 1 とゼロで二値化した。「Figure12」は自作モデルで抽出した道路と建物の画像である。

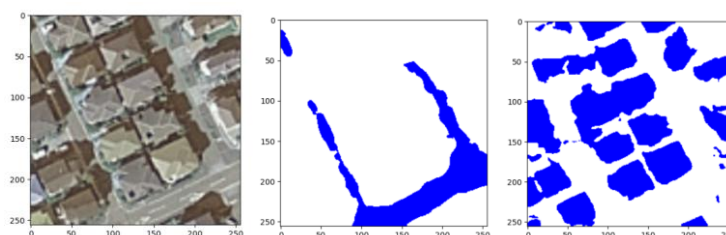


Figure 12 自作モデルによるセマンティック・セグメンテーションの例
(左から原図、道路、建物)

対象画像内のピクセルを 1 (対象物) とゼロ (対象物以外) に分類し、「1」ピクセルが画像全体に占める比率を求めて、全公示地 1685 行×施設等の数 6 列の地物比率マトリックスを作成して特徴量とした。この作業も演算の負荷が大きく、終了までに約 17 時間連続でコンピュータを稼働させる必要があった。

「Table 3」は画像特徴量の例である。数字はその公示地の上空画像に占める地物の占有比率を示す。

	RoadRatio	BuildingRatio	ParkingRatio	WaterRatio	Railroad	GreenRatio
30	0.094070	0.479935	0.000000	0.000000	0.000000	0.27184
31	0.000000	0.159348	0.000000	0.000000	0.000046	0.36697
32	0.199295	0.125320	0.000000	0.00322	0.000000	0.47774
33	0.074066	0.243729	0.000000	0.000000	0.000015	0.36859
34	0.004684	0.330673	0.066635	0.000000	0.000000	0.36201

Table 3 特徴量化した画像データ

5. 各データを使用した試行

①ベースデータ、②ベースデータ+地理空間情報、③ベースデータ+画像特徴量、④ベースデータ+地理空間情報+画像特徴量のそれぞれのデータの組み合わせに回帰アルゴリズムを適用して精度を比較する。

5-1. 評価指標

① 決定係数 R^2 (R Squared)

決定係数 R^2 は、観測値に対して予測値がどれほどよく当てはまっているかを表す。 $0 \sim 1$ の値を取り、 R^2 が1に近いほど予測値が実際のデータに近い値であることを表す。

$$R^2 = 1 - \frac{S_e}{S_{yy}} = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

② 平均平方二乗誤差 (Root Mean Squared Error : RMSE)

平均二乗誤差の平方根をとったものである。小さいほど良いモデルである。モデル全体としての誤差の平均を表すものであり、複数のモデル間の精度比較に有効である。反面、分析対象の個別地点（それぞれの公示地）についてどの程度の誤差が発生しているかを示すことはできない。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

③ 平均絶対誤差 (Mean Absolute Error : MAE)

予測値と観測値の差（誤差）の絶対値の平均であり、何を表す数値であるかを直感的に理解しやすい。小さいほど良いモデルである。平均平方二乗誤差と同様の長所短所がある。

$$MAE = \frac{1}{n} \sum_{i=1}^n (|\hat{y}_i - y_i|)$$

④ 平均絶対パーセント誤差 (Mean Absolute Percentage Error : MAPE)

MAPE は予測値と観測値の差（誤差）を観測値に対するパーセントとして求めたものであり、どの観測値に対しても、誤差の大きさが理解しやすいという長所がある。小さいほど良いモデルである。

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left(\frac{|\hat{y}_i - y_i|}{y_i} \right)$$

5-2. 回帰アルゴリズム～lightGBM～

回帰アルゴリズムとして lightGBM⁴⁵⁾を使用する。lightGBM は木形式の勾配ブースティングフレームワークの一種で、勾配ベース片側サンプリング (Gradient-based One-Side Sampling : GOSS) と排反的特徴量選択 (Exclusive Feature Bundling : EFB) という特徴的なサンプリング方法を採用している⁴⁶⁾。

GOSS と EFB により、LightGBM は

- ① 学習の高速化と高効率化
- ② 少ないメモリ使用量
- ③ 高精度
- ④ GPU を使用した並列分散処理が可能

⑤ 大規模データに対応
を実現している。

5-3. 試行結果

lightGBM を用いて 5 分割の交差検証^vを行い、平均スコアにより精度確認を行った。
試行結果は次のとおりである。

採用したデータ	R2	RMSE	MAE	MAPE
ベースデータ	0.6898	0.4359	0.3407	0.0268
ベースデータ+地理空間情報	0.9429	0.1866	0.1417	0.0110
ベースデータ+画像特徴量	0.7480	0.3932	0.3029	0.0238
ベースデータ+地理空間情報+画像特徴量	0.9403	0.1910	0.1453	0.0114

Table 4 試行結果

ベースデータに地理空間特徴量を加えた場合、ベースラインスコアに対して、 R^2 が 36.7%、RMSE が 57.2%、MAE が 58.4%、MAPE が 59.0%それぞれ向上した。

画像特徴量を加えた場合の精度は、 R^2 が 8.4%、RMSE が 9.8%、MAE が 11.1%、MAPE はが 11.2%それぞれ向上した。

ベースデータに地理空間情報と画像特徴量の両方を加えた場合、 R^2 が 36.3%、RMSE が 56.2%、MAE が 57.4%、MAPE が 57.5%それぞれ向上した。

住宅（土地+建物）を対象とした先行研究では、ベースデータ+地理空間特徴量で lightGBM 回帰を訓練することで、ベースデータのみときと比較して MAE が 0.135 から 0.127（5.3%）に、RMSE が 0.190 から 0.180（5.3%）にそれぞれ改善している⁴⁷⁾。また家賃を対象として回帰アルゴリズムにランダムフォレストを使用した研究では、 R^2 が 0.7190 から 0.9133 に改善（21.3%）している⁴⁸⁾。

採用する施設の種類と数、バッファの区切り方等による精度に対する影響は考えられるが、地理空間特徴量を採用して更地のモデルを作った場合でも、土地建物モデルと同等、あるいはそれを上回る精度の向上が期待できることが確認できた。

ベースデータに地理空間特徴量を加えた場合と画像特徴量を加えた場合、いずれも精度の改善が見られた。特に地理空間情報を加えることで、精度は大きく向上することが分かった。一方で、ベースデータに地理空間情報と画像特徴量の両方を加えた場合、地理空間情報だけを加えた場合と比較して、精度にほとんど変化は見られなかった。

高精度（ $R^2 > 0.94$ ）の住宅地価格予測モデルを構築するためには、地理空間情報を利用して利便施設等の特徴量化することが有効であることが分かった。また上空画像を本研究の方法で特徴量化した場合も精度は向上するが、地理空間特徴量を使用した場合と比較して精度向上に対する寄与度は小さかった。

^v 訓練データを 5 分割し、4 つを訓練用、1 つを精度検証用に用いる。訓練用データと精度検証用のデータを組み替えることで、試行回数は全部で 5 回になる。

6. 特徴量の影響度の確認と東京都の住宅地価格の形成メカニズムに関する一考察

6-1. 特徴量の価格に対する影響度

「ベースデータ+地理空間情報」が最も高い精度を示した。このモデルは、街路・街並みの状態、緑化の程度など抽象的な地域要因の影響は全く反映しておらず、不動産鑑定評価実務の立場からは受け入れ難い。そこで本研究では「ベースデータ+地理空間情報」モデルと同程度の精度を示した「ベースデータ+地理空間情報+画像特徴量」で学習したモデル（以下、最終モデル）を使用して、各説明変数の影響度を確認する。

「Figure 13」は lightGBM による価格の予測過程を表している。四角はノードで中の数字は yes/no に分割する閾値と分類されたデータ数である。楕円はリーフ（葉）であり予測値を表す。

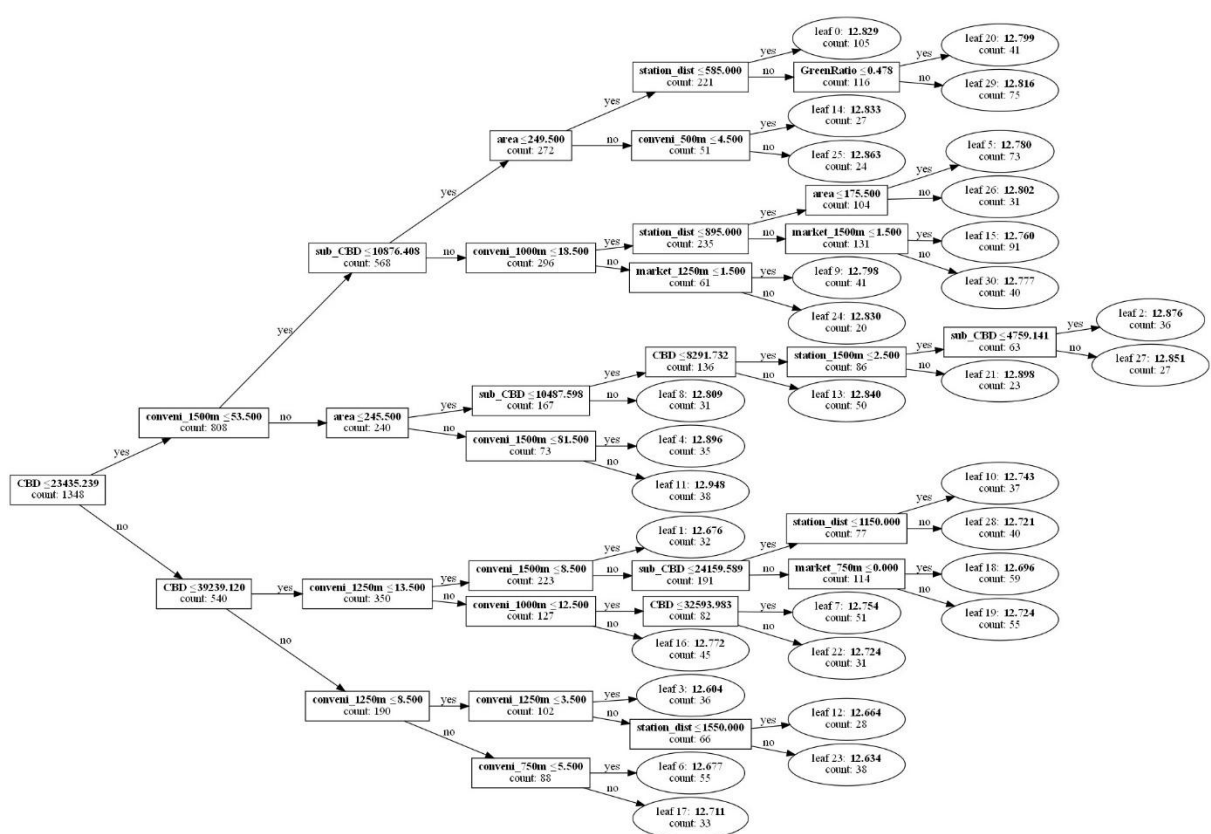


Figure 13 lightGBM による価格予測過程

説明変数の影響の解析には Python ライブラリの SHAP から出力した可視化資料を使用する。SHAP (SHapley Additive exPlanations⁴⁹⁾) は、機械学習モデルの予測値を協力ゲーム理論により説明するオープンソースライブラリである。「Figure 14」は、SHAP から出力した説明変数の影響度の絶対値平均（単位：対数円）と相関の方向を表している。

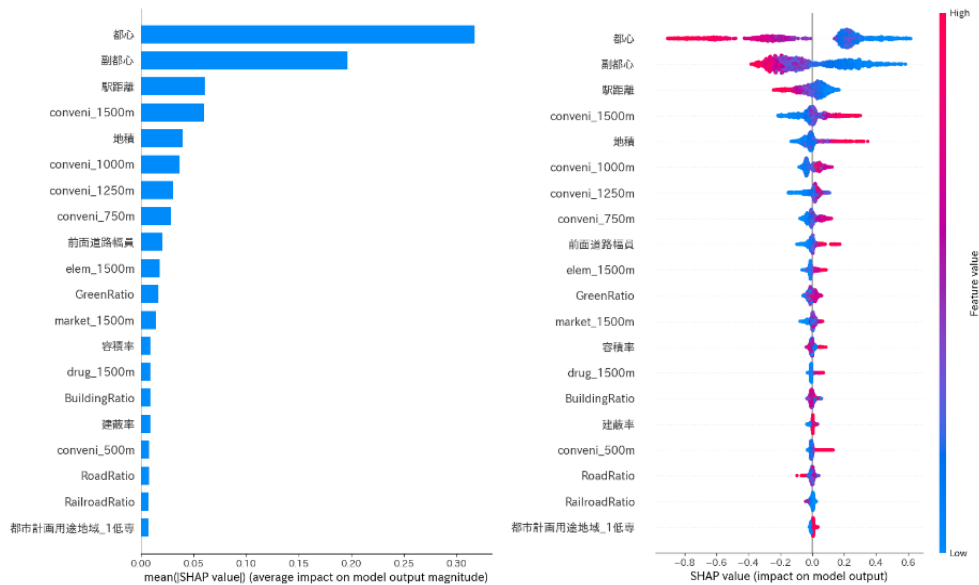


Figure 14 説明変数の影響度と相関

左のグラフは各説明変数の予測価格に対する寄与度（影響の強さ）を表す。東京都の住宅地価格に最も強い影響を与えているのは、都心・副都心までの距離（時間）であることが見て取れる。また利便施設との距離を表す説明変数名が目立ち、接近条件に基づく利便性が住宅地価格に影響を与えていることが分かる。特にコンビニの影響が大きく、コンビニが社会インフラであるとの認識⁵⁰⁾を裏付けることとなった。

一方で図中に抽象的な環境要因として、GreenRatio（緑地率）、BuildingRatio（建物率）、RoadRatio（道路率）、RailroadRatio（軌道率）の4特徴量が現れているが、価格に対する影響力は比較的小さい。

右の図は、その説明変数が価格に影響を与えた公示地の数、説明変数の影響の方向と強さ、最終結論の価格（赤：高い、青：安い）を表す。中心のゼロは、その説明変数の平均値を表す。例えば都心は「Figure 13」のとおり起点ノードであり、訓練用データ（1685の5分の4で1348）の全てが影響を受けるので点の数が多い。左の方向へ向かう、つまり都心までの距離が平均より近くなるほど価格は高くなっている。その逆に都心までの距離が平均より大きくなると価格は安くなる。つまり価格と都心までの距離に負の相関を認めることができる。conveni_1500mは1500m圏内のコンビニ数が多いと価格が高くなり、正の相関があることを示している。

4つの抽象的な地域要因は、影響を与えた公示地の数はコンビニ等の利便性の影響を受ける公示地と比較して少ない。GreenRatioは正の相関が認められる。BuildingRatioには正の相関、RoadRatioとRailroadRatioには負の相関があるように見えるが、あまり明瞭ではない。

価格モデルは、GreenRatioとBuildingRatioが大きくなると価格は高くなる傾向があり、RoadRatioとRailroadRatioが大きくなると価格は低下するという傾向を表しているが、これら4つの説明変数が価格に影響する理由は教えてくれない。これはモデルの

開発者やユーザーが解釈しなければならない問題である。

GreenRatio は緑化の程度であり、一般的に環境条件にプラスに作用すると考えられている。都心のマンション街の価格は郊外の戸建住宅地域の価格よりも高い。マンション街が都市化を表すと仮定すると、**BuildingRatio** は都市化の程度を反映して正の相関が生じたと解釈することができる。**RoadRatio** は、道路が多くなると自動車の通行量が多くなり、騒音や振動の影響が背景にある可能性がある。**RailroadRatio** の負の相関は、軌道に近いことによる騒音等の影響が考えられる。このように解釈するならば、これらの4特微量の影響の方向は、不動産鑑定評価実務の認識と一致する。

説明変数の影響度を確認した結果、緑化等の価格に対する影響度は利便性要因と比較して小さいという結論となったが、従来は数値化が困難であった抽象的あるいは主観的な地域要因を、上空画像から抽出して特徴量化する可能性を示すことはできたと考える。

6-2. 東京都の住宅地価格の形成メカニズムに関する考察

「Figure 14」に表れる説明変数は次のように分類することができる。

- ① 都心・副都心までの距離（時間）：都心、副都心
- ② 地域の価格水準に影響する変数：駅距離、conveni、elem、GreenRatio、market、容積率、drug、BuildingRatio、建蔽率、RoadRatio、RailroadRatio、都市計画用途地域
- ③ それぞれの公示地の個別的属性：地積、前面道路幅員

説明変数を上記のように分類し、東京都の住宅地価格は以下の過程を経て成立していると考察した。

- ① 東京都住宅地価格の持つ空間自己相関により、都心・副都心からの距離により、例えば市区町村のような不動産鑑定評価の近隣地域よりも広い範囲での地域の価格水準が定まる。そしてこれが個々の住宅地価格の形成に決定的な影響を与えている。
- ② ①の価格を前提として、主として利便施設等との距離に基づく利便性を反映して、例えば近隣地域のような細分化された地域の価格水準が定まる。この時、緑地、建物の密集度、幹線道路や軌道に接近することによる振動等の具体化するのが困難な要因の影響度は、接近条件により計測した利便性要因と比較して小さい。
- ③ 最後に面積や前面道路の幅員などの個別的な属性により、それぞれの住宅地の価格が定まる。

不動産鑑定評価において鑑定評価額を求める過程は、一般的要因の分析により市区町村レベルの大きな地域の価格水準を把握したうえで、地域分析により近隣地域の価格水準を求め、最後に個別的要因の比較により対象不動産の価格を求めるという手順をとる。この過程は本研究で考察した住宅地価格の形成メカニズムと一致しており、不動産鑑定評価における価格形成過程を説明変数により実証的に確認することができた。

最後に公示地の価格形成過程を具体的に例示する。「Figure 15」は SHAP から出力し

た「杉並－25」のウォーターフォール図である。この図は価格の形成過程を、住宅地平均価格を起点として、「どの要因が、どの方向へ、どの程度」の影響を与えているかを可視化したものである。

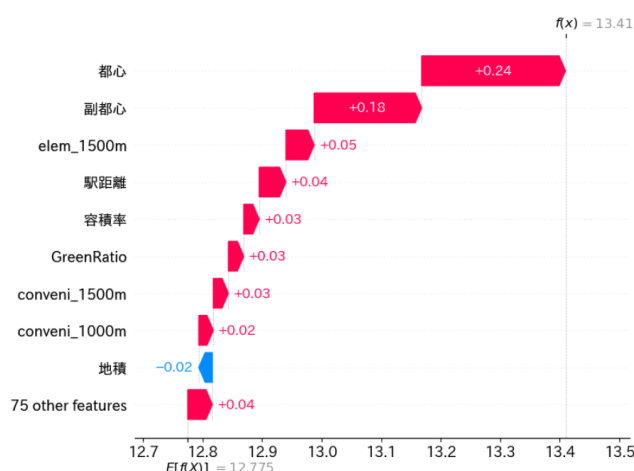


Figure 15 杉並－25の価格形成過程

横軸の下での $E[f(X)] = 12.775$ は、対数化した公示価格の平均を表す。右上の $f(x) = 13.41$ は、杉並－25の公示価格である。杉並－25の価格は、12.775円から出発し、75 other features（その他75の属性）は平均よりも優れていて+0.04（単位は対数円）、地積は平均より劣るので-0.02、conveni_1000mから都心までの要因は、すべて平均よりのプラスに作用しており、+0.02・・・+0.24となり、最終的に杉並－25は13.41円になる。ここからも、都心・副都心の影響が特に大きいことが分かる。

7. まとめ、実務上の応用、将来展望

7-1. まとめ

令和6年地価公示住宅地価格を使用して、高精度の価格モデルを作ることができた。特に地理空間特徴量をモデルに追加することによる精度の向上が顕著であり、更地価格モデルに地理空間情報を活用することの有用性を確認できた。説明変数の価格に対する影響度を測定すると、都心・副都心までの距離が最も影響度が大きく、次いで接近条件に基づく利便性であった。その中でもコンビニの利便性の影響度が高いことが分かった。上空画像にセマンティック・セグメンテーションを行い、地物比率を抽象的な地域要因の代替変数にすることを試みた。緑地の価格に対する影響は、不動産鑑定評価実務上の影響の方向と一致した。建物、道路、軌道の相関の方向は、解釈によって実務上の価格に対する影響の方向と一致するが、緑地と比較するとあまり明瞭ではなかった。これら4特徴量の影響度は比較的小さかったが、上空画像を利用して緑化、都市化の程度、騒音や振動の影響を数値化することの可能性を示すことができた。

7-2. 実務上の応用

7-2-1. 不動産鑑定評価における応用

地価公示法第8条は

『不動産鑑定士は、公示区域内の土地について鑑定評価を行う場合において、当該土地の正常な価格（略）を求めるときは、第六条の規定により公示された標準地の価格（略）を規準としなければならない。』

と定めている。「規準とする」とは、鑑定評価手法の適用により求めた土地の正常価格と、同一需給圏内の類似地域内に存在する公示地の価格と均衡を保たせることを意味する。

「規準とする」方法は、一般に比準の方法をとる。具体的には、不動産鑑定士が適切と判断した公示地を選択し、その公示価格に時点修正、標準化補正、地域要因の比較、個別的要因の比較を行って求めた価格を鑑定評価による土地の正常価格と比較する。

時点修正率と標準化補正率については、不動産情報ライブラリ [51](#)で毎年1月1日時点の公示価格、対前年変動率及び鑑定評価書が公表されており、鑑定評価書のユーザーは不動産鑑定士が採用した時点修正率や標準化補正率の根拠を確認することができる。

地域要因の比較は、一般に街路条件、交通接近条件、環境条件、行政的条件、その他の条件に細分し、格差率を具体的に説明する形をとる場合が多い。地域要因の格差率は不動産鑑定士の判断を根拠とする場合が多く、時点修正率や標準化補正率のようにユーザーが検証するために利用できる資料は存在しない。不動産鑑定士は不動産鑑定書に記載された内容について説明する義務を負うが、比準の地域格差の理由やその根拠は、不動産鑑定士にとっても説明が困難な部分の一つである。また公示地の選択理由も十分な説明が難しい場合がある。

もし、公示地で訓練した価格モデルによる予測値と不動産鑑定士が公示価格を規準とした価格が近似した場合、東京都の全公示地からAIが求めた価格と不動産鑑定士が比準した価格が近似することを意味する。両方の価格が近似するということは、不動産鑑定士の判断に基づく地域格差率とAIが大量データから自律的に学習した格差率が近似するということと解釈できる。したがって両者の価格が近似することにより、例えば街路条件等の内訳までは踏み込めないとしても、不動産鑑定士が査定したトータルとしての地域格差が適切であることを、間接的ではあるが検証することができる。また、規準のための公示地の選択が恣意的ではなく、適切であることも示すこともできる可能性がある。

7-2-2. 都市開発等における応用

東京都住宅地価格の形成メカニズムを考察した結果から、住宅地価格に決定的に影響を与えているのは都心との距離であった。都心との地理空間上の距離を縮めることはできないが、距離＝時間と考えると、高速鉄道網の整備や急行や特急ダイヤの見直しにより、将来的に時間空間上の距離を短縮することができる可能性がある。

施設の利便性は、食品スーパーのような中大規模の商業施設よりも、コンビニのような小規模だが数が多い商業施設の影響が大きいという結果となった。また上空画像からの特微量の影響度は小さかったが、その中において緑地率の影響が比較的大きく正の相関が認められた。商業施設の種類と配置、緑化などのプランニングにより、住宅地の価値を高め

ることができる可能性があることを価格モデルにより示すことができた。

住宅地価格モデルは、価格の予測や影響度が大きい要因の抽出に留まらず、都市計画、市街地（再）開発などの分野における住宅地価格（≡地域の価値）のシミュレーションツールとして利用することができる。

7-3. 将来の展望

本研究においては、上空画像から抽象的な地域要因を抽出することが目的の一つであった。結果としては、緑地率に正の相関を認めることができ、画像から抽出した緑地率を地域特徴量として採用できる可能性があることが分かった。しかし緑地率以外の地物の占有比率の影響は小さかった。抽象的な地域要因の影響が本当に小さいのか、それとも地物占有率によって抽象的な地域要因を十分表現することができないのかは、本研究の中では明らかにすることはできなかった。この点については、将来の研究の課題である。

今回構築したモデルは $R^2 \cong 0.94$ と、鑑定評価実務の「公示価格との規準」の場面で活用できる程度の精度を持っている。このモデルで予測した価格は、あくまでも公示地から求めた価格であり、取引事例から求めた比準価格の代替や検証手段にすることはできない。取引事例を使って価格予測モデルを作ることができれば、不動産鑑定士の判断が主要な部分を占める取引事例比較法の検証手段として利用することができる。しかし地理空間情報を持った取引情報を、高精度モデルを作ることができるほど大量に取得するのは困難である。またモデルの構築に必要なプログラミング等の学習コストも高い。このような障害から、個人が現実の取引事例を対象とした研究を行うことは難しい。費用面や学習面の障害をクリアすることが可能な不動産鑑定に関わる研究機関において、実際の取引事例を使用した研究が進むことにも期待したい。

8. 謝辞

旧不動産鑑定士試験制度の当時、平成7年春から平成14年の夏にかけて福井県の（株）林不動産鑑定所に勤務していた。在職中、不動産鑑定士林逸男氏の薫陶を得て、不動産鑑定士としてあるべき姿勢を学んだ。不動産鑑定士は常に最新技術を取り入れて、不動産鑑定評価の進歩のために前へ進まなければならない。これは氏の教えであり本稿執筆のための原動力となった。林逸男氏に心より感謝申し上げる。

【参考文献】

- 1). Kelvin J. Lancaster, A New Approach to Consumer Theory, The Journal of Political Economy, Vol. 74, No. 2 (1966), p. 132-157
- 2). Rosen, Sherwin, Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition (1974), The journal of political economy, Vol. 82, Issue 1, p. 34 - 55
- 3). K.W. Chau, T. L. Chin, A Critical Review of Literature on the Hedonic Price Model (2003), International Journal for Housing Science and Its Applications 27 (2), p. 145-165

- 4). 前田 翠, 関本 義秀, 瀬戸 寿一, 異なるタイプの不動産データを用いた不動産価値推定モデルの研究 ヘドニックと Deep Neural Network の比較 (2019), ジャーナル オープンアクセス 54 巻 3 号 p.818-825
- 5). 山田 慧, 岡本 一志, 柴田 淳司, カルマンフィルタによる住宅価格の説明変数の逐次推定, 人工知能学会全国大会論文集, 2022, JSAI2022 巻, 第 36 回 (2022)
- 6). 興星陽, 木村泰知, 最寄り駅との標高差が不動産価格に与える影響について、情報処理学会全国大会講演論文集 (2022) 、84 号、p4.447-4.448
- 7). Wei, Cankun; Fu, Meichen; Wang, Li; Yang, Hanbing; Tang, Feng; Xiong, Yuqing, The Research Development of Hedonic Price Model-Based Real Estate Appraisal in the Era of Big Data, Land (2012), 2022, Vol 11, Issue 3, p334
- 8). Jan-Peter Kucklick, Jennifer Müller, Daniel Beverungen, Oliver Müller, QUANTIFYING THE IMPACT OF LOCATION DATA FOR REAL ESTATE APPRAISAL-A GIS-BASED DEEP LEARNING APPROACH, European Conference on Information Systems (ECIS) 2021
- 9). A Paszke, S Gross, F Massa, A Lerer, J Bradbury, G Chanan, T Killeen, Z Lin, N Gimelshein..., Pytorch: An imperative style, high-performance deep learning library, Advances in neural information processing systems, 2019·proceedings.neurips.cc
- 10). <https://nlfpt.mlit.go.jp/ksj/> (2024.6.3 ダウンロード)
- 11). Bonny P, McClain 著, 廣川類訳, Python による地理空間データ分析初版第 1, オライリー・ジャパン, 2023, ISBN978-4-8144-0032-4
- 12). 山田 育穂, 岡部 篤行, 都市事象の空間解析のための Moran 空間自己相関検定における棄却限界値推定式の構築, 都市計画論文集, 2016, 51 巻, 3 号, p. 923-928, 公開日 2016/10/25, Online ISSN 2185-0593, Print ISSN 0916-0647
- 13). Sergio J. Rey, Dani Arribas-Bel, Levi J. Wolf, Geographic Data Science with Python 1st edition (2023), CRC Press, ISBN 978-0-367-26311-9
- 14). 小川 芳樹, 沖 拓弥, 関本 義秀, 柴崎 亮介, 不動産に関する地理空間情報と建物画像を用いた建物の構造・築年代推定手法の検討, 人工知能学会全国大会論集, 2020, JSAI2020 巻, 第 34 回
- 15). 前掲 7)
- 16). Robert J. Hill, Michael Scholz, Incorporation Geospatial Data into House Price Indexes: A Hedonic Imputation Approach with Spline, 2013
- 17). Das, S.S.S., Ali, M.E., Li, YF. et al. Boosting house price predictions using geo-spatial network embedding. Data Min Knowl Disc 35, 2221-2250 (2021)
- 18). 鶴田 大, 豊島 裕樹, 機械学習と地理空間情報を活用した周辺の過去の賃料情報を用いたアパート賃料推定 (2021), 人工知能学会全国大会論文集, JSAI2021 巻, 第 35 回, セッション ID 2H3-GS-3b-03, p. 2H3GS3b03
- 19). 日本ソフト販売株式会社, 【2024 年度版】コンビニエンスストアの店舗数ランキング、<https://www.nipponsoft.co.jp/blog/analysis/chain-conveniencestore2024/>
- 20). <https://developers.google.com/maps?hl=ja>
- 21). David S. Jordan, Applied Geospatial Data Science with Python (2023), ISBN:978-1-80323-812-8
- 22). A. J. Bency, S. Rallapalli, R. K. Ganti, M. Srivatsa and B. S. Manjunath, Beyond Spatial Auto-Regressive Models: Predicting Housing Prices with Satellite Imagery (2017), IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 2017, p.320-329.

- 23). Yuhao Kang, Fan Zhang, Wenzhe Peng, Song Gao, Jinmeng Rao, Fabio Duarte, Carlo Ratti. Understanding house price appreciation using multi-source big geo-data and machine learning (2021), Land Use Policy Volume 111
- 24). Li, Wei (2020) The relationship between street visual features and property value using deep learning.
- 25). Xu, X.; Qiu, W.; Li, W.; Liu, X.; Zhang, Z.; Li, X.; Luo, D. Associations between Street-View Perceptions and Housing Prices: Subjective vs. Objective Measures Using Computer Vision and Machine Learning Techniques (2022), Remote Sens.14, p.891
- 26). Law, Stephen, Brooks Paige and Chris Russell, Take a Look Around (2018), ACM Transactions on Intelligent Systems and Technology (TIST) 10 p.1 - 19.
- 27). Kostic, Zona and Aleksandar Jevremović, What Image Features Boost Housing Market Predictions? (2020), IEEE Transactions on Multimedia 22 p.1904-1916.
- 28). Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, Bernt Schiele, The Cityscapes Dataset for Semantic Urban Scene Understanding, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- 29). CITYSCAPES DATASET: <https://www.cityscapes-dataset.com/>
- 30). PASCAL Visual Object Classes Homepage: <http://host.robots.ox.ac.uk/pascal/VOC/>
- 31). KITTI Vision Benchmark Suite: <https://www.cvlibs.net/datasets/kitti/>
- 32). Marco M, Gracia E, Martín-Fernández M, López-Quílez A. Validation of a Google Street View-Based Neighborhood Disorder Observational Scale. J Urban Health. 2017, p.190-198
- 33). 前掲 8)
- 34). 鈴木雅智, ストリートビュー画像を用いた景観の指標化 -不動産価値評価への活用可能性- (2023) 不動産研究 65 (1), p.14-22, 一般社団法人日本不動産研究所, ISSN 0532-7776
- 35). 浅田拓海・亀山修一, Google ストリートビュー画像を用いた道路景観の地域内分布特性の分析, 土木学会 (土木計画学研究発表会) 2015, 土木計画学研究・講演集, Vol.52, p.1991-1998
- 36). 小川雄太郎, つくりながら学ぶ! PyTorch による発展ディープラーニング 初版第 8 刷, マイナビ, ISBN 978-4-8339-7025-3
- 37). Garcia-Garcia, Alberto, Sergio Orts, Sergiu Oprea, Victor Villena-Martinez and José García Rodríguez. "A Review on Deep Learning Techniques Applied to Semantic Segmentation." ArXiv abs/1704.06857 (2017): n. pag.
- 38). O' Shea, Keiron and Ryan Nash. "An Introduction to Convolutional Neural Networks." ArXiv abs/1511.08458 (2015)
- 39). Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015
- 40). 前掲 36)
- 41). 田村正人, 中村克行著, Python で学ぶ画像認識 初版(2023), (株) インプレス, ISBN978-4-295-01599-4

- 42). K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Figure 2, Las Vegas, NV, USA, 2016, pp. 770-778
- 43). 国土地理院GISデータセット: https://gisstar.gsi.go.jp/gsi-dataset/index_ja.html (2024.9.2ダウンロード)
- 44). S. Jadon, A survey of loss functions for semantic segmentation, 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, 2020, pp. 1-7
- 45). 毛利拓也, モデル評価・制度改善・可視化 実務活用のためのLightGBM予測モデル 実装ハンドブック 第1版第1刷、秀和システム、ISBN 978-4-7980-6461-2-C3055
- 46). Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye and Tie-Yan Liu. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." Neural Information Processing Systems (2017).
- 47). 前掲17)
- 48). 前掲22)
- 49). Scott, M., Lundberg., Su-In, Lee. (2017). A unified approach to interpreting model predictions. 30, 4768-4777.
- 50). 経済産業省社会インフラとしてのコンビニエンスストアのあり方研究会, 競争と協働の中で社会と共に進化するコンビニ 研究会報告書, 2009
- 51). 不動産情報ライブラリ, <https://www.reinfolib.mlit.go.jp/>