重回帰分析による土地価格推定の一手法

加藤　暢之† 　新妻　弘崇†† 　太田　学††

† 岡山大学工学部情報系学科　〒 700–8530 岡山県岡山市北区津島中三丁目 1 番 1 号
†† 岡山大学大学院自然科学研究科　〒 700–8530 岡山県岡山市北区津島中三丁目 1 番 1 号
E-mail: †phxx4bmw@s.okayama-u.ac.jp, ††niitsuma@de.cs.okayama-u.ac.jp, ††ohta@de.cs.okayama-u.ac.jp

あらまし　本研究は Web 上に存在する不動産物件情報を分析することを目標としている．不動産物件情報の中でも特に土地価格に注目した分析を行なう．土地価格の分析のために不動産物件情報を蓄積している不動産ポータルサイトのデータを取得して整形し，土地価格の予測タスクを行なう．土地価格を予測する手法として重回帰分析を利用する．重回帰分析によって推定されたモデルが実際の土地価格をどの程度正確に推定できるか評価し，重回帰分析のモデルの偏回帰係数などについて考察を行なう．

キーワード　土地価格，重回帰分析，価格予測

1. はじめに

近年 Web 上にはあらゆる情報が蓄積され，その量は加速度的に膨れ上がっている．不動産情報もその一つであり，不動産仲介業者のもつ情報がいくつかの不動産ポータルサイトに集約されており，不動産情報の分析が進んでいる．米国では，現在進行形で不動産データのデジタル化が進んでおり，教育や小売業のデジタル化と並んで注目されている．図 1 に Mckinsey Global Institute による米国におけるデータのデジタル化状況調査の抜粋を示す．図 1 中のカラーチャートに対応した部分は，デジタル化の進行度を示しており，赤はデジタル化が進んでおらず緑に近づくにつれデジタル化が進行していることを表している．GDP 割合は米国内での GDP の内訳，雇用率は全業種内での雇用人数の割合，成長率は 2005 年から 2015 年にかけての GDP の増減率を示している．また，デジタル化促進分野とは米国でデータのデジタル化が重要視されている分野を指しており，図 1 において不動産，教育，小売業が該当する．メディアは通信媒体を用いた配信を行うためデータのデジタル化が特に進んでいる．対して，エンターテインメントはデータを定量化することの難しさから全体としてデータのデジタル化が進んでいない．デジタル化促進分野の特徴は，データをデジタル化することに対する需要が大きく，かつこれまでデジタル化されていなかったことである．不動産，教育，小売業の分野はこれまでデジタルデータとして利用されることは少なかったが，多くの情報がデジタル化されるにあたって注目されるようになった．

一方，日本でも今後の不動産業界に影響を与えるであろう動きが起きている．例を挙げると，国土交通省が不動産会社の業務効率化及び情報透明化を目的として試験運用した不動産総合データベースがある．この施策は 2016 年 10 月に静岡，大阪，福岡の不動産流通機構会員を対象に行われた．

しかし，ここ数年で進んでいる不動産に関する研究には問題点がいくつか存在する．例えば，適切な物件価格を予測することの難しさがある．不動産価格には不動産の価値や周辺の状況だけに限らず，売り手買い手の個人的な事情なども影響するため，類似した条件でも価格のずれが生じている．そのため不動産の価格推定においてはこのような個人的な事情等の余分な要素を排除することが重要になる．

問題の二つ目は，物件データベースが網羅的に整備されていない点である．不動産データは不動産ポータルサイトを通じて売り手から買い手に伝達されるが，不動産情報が不動産ポータルサイトに掲載されている期間は短い．そのため過去全てのデータを入手することは難しい．

その他にも，売り手と買い手のニーズを分析することの難しさが挙げられる．不動産ポータルサイトでは，売り手と買い手の情報は提供されていない．そのためどのような客層がどのような物件を購入したかを把握することは困難である．

本研究ではこれらの問題点のうち物件価格の推定に焦点をあて，不動産ポータルサイトの情報をもとに土地価格の推定を行う．

本稿は次のような構成になっている．2. 節で関連研究について述べ，3. 節で提案する土地価格推定方法について述べる．4. 節では評価実験の内容と結果を示すと共に，その結果について考察する．5. 節では本稿のまとめと今後の課題について述べる．

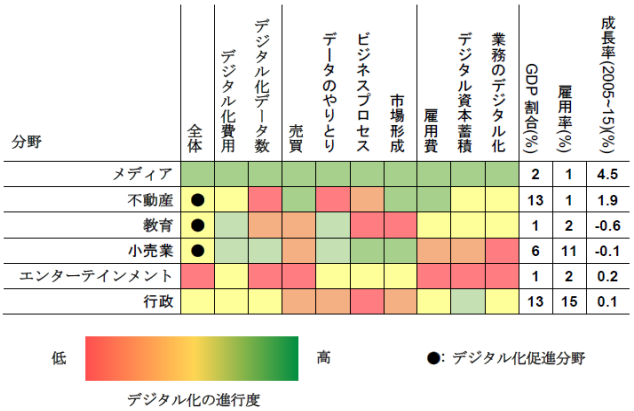

図 1　MGI 調べ [11]

2. 関 連 研 究

2.1 重回帰分析

重回帰分析 [1] では，推定したい変数を目的変数，推定の根拠となる変数を説明変数と呼び，説明変数と目的変数の関係を回帰的に分析することで説明変数ごとの重みを算出し，目的変数を推定する．

一般的な重回帰分析は，目的変数の実測値 y に対して，これの変動を説明すると考えられる説明変数 $x_1, x_2, \cdots, x_n$ により目的変数の予測値 y' を式 (1) の形で推定する．

$$y' = a_0 + a_1 x_1 + \ldots + a_p x_p \tag{1}$$

ここで $a_1, a_2, \ldots, a_p$ は偏回帰係数と呼ばれ，予測値である y' と実測値である y の各標本における差の二条和を最小にするように定める．

$$(a_0, a_1, \ldots, a_p) = \mathrm{argmin}_{a_0, a_1, \ldots, a_p}(< (y' - y)^2 >) \tag{2}$$

式 (1) において a_0 は定数項である．重回帰分析の目的の一つは，重相関係数から目的変数と説明変数の関係について洞察を深めることである．また，このように構成された予測式を用いて，測定したデータを説明変数として目的変数を予測することも行われる．偏回帰係数は式 (2) のように予測値の残差平方和を最小にするものとして定められるが，これは同時に目的変数の実測値 y と予測値 y' との相関を最大にするときの係数であり，このときの相関係数の値が重相関係数になる．

重回帰分析を用いた推定問題として商品の値段や評判を推定する研究が行われている．松尾ら [2] は商品レビュー文の分析に重回帰分析の結果の一部を利用した．松尾らは極性辞書を用いて商品レビュー文の単語ごとに感情極性値を割り振り，その値を利用した重回帰分析を行った．松尾らの研究では，レビュア評価を目的変数に，商品の属性ごとの感情極性値を説明変数とした重回帰分析で求まった t 値を，レビュア評価を決定づける商品属性を特定するための根拠としている．

2.2 不動産情報のマイニングに関する研究

本節では，不動産情報のマイニングにより，一般の不動産情報利用者や研究者が情報を利用しやすくするための研究について述べる．

不動産情報のマイニングに関する研究として，国土交通省が公開している公示地価から，特定の地点の地価を予測する研究がおこなわれてきた．一般的に地価予測の研究には，地球統計学の分野で考案されたクリギングとよばれる空間回帰的な手法を用いることが多い．クリギングとは，多くの場合土壌学や地質学に使用される手法である．クリギングでは測定値のない位置の測定値を周囲との空間的位置を考慮し重み付き平均として推定する．

2005 年には井上ら [6] が，2007 年に増成ら [7] がクリギングの土地価格推定への応用を提案した．増成らは地価が立地条件と密接に関係しているため，空間的相関関係を考慮し空間統計学的方法が有効であると考えた．また，東京都の地価分布予測と説明変数の係数，土地資産総額の 3 点からクリギングが経済データの分析に有効であることを示した．

2008 年に李ら [8] が，予測する変数と関連が強い，または観測数が多いとより有効に働く共クリギングによる地価モデルを提案した．2009 年に井上ら [9] は，東京 23 区全用途地域を対象とした公的地価の分布と変遷の視覚化をクリギングで行った．

これらの研究を踏まえて，2016 年には井上ら [10] が公示地価指標と取引価格の比較をもとに地価情報提供の提案をした．公示地価は，近隣地域を代表する属性を持った標準地を定め，不動産鑑定評価から周辺の標準的な価格を標準地の価格として定期的に公表しているものである．公示地価を分析することで地価の地域差や時間的変化に関する情報を得られ，地価の地域的分布や長期的変遷の把握に利用できる．しかし，情報の時間解像度の低さ，時間遅れが一つの原因となり，市場取引価格からの乖離が指摘されており，短期的な不動産市場の動向把握には適していない．一方，取引価格情報は短期的な市場動向を含んだ情報であるが，取引当事者の個別事情や取引物件属性の個別性が反映されており，加えて定点観測情報ではない．そのため，個々の取引価格から価格の地域差や時間的変化などの市場動向をとらえることは困難である．しかし，取引物件属性の個別性や取引当事者の個別事情の影響を除去，緩和できれば市場の活況，不況を把握できる可能性がある．井上ら [10] は，取引物件の個別性の影響除去を目指し，取引物件の場所や取引日，その他の属性を考慮した標準的価格を公的地価指標内挿値として算出した．次に各取引にかかわった投資者の個別事情が取引価格に与えた影響の緩和を目指し，取引価格水準を空間，時間で集計した分布情報を作成した．この取引価格水準の分布情報が有する，局所的，短期的な取引価格の上下などの市場動向把握に対する利用可能性を確認した．

3. 提 案 手 法

Web 上に存在している不動産情報は土地価格などの一部を除いて数値データの形式になっておらず，そのまま重回帰分析に使用することはできない．そのため本研究では不動産情報を土地価格推定に利用できるよう整理したデータを用いる．3.1 節ではデータの収集方法について述べる．また 3.2 節では 3.1 節のデータを土地価格推定のために整理する手法と土地価格推定モデルへの適用の手法について述べる．

3.1 データの収集

本研究では不動産ジャパン [13] に掲載されている不動産情報を利用する．これは HTML 形式のため Web スクレイピングにより文字列として収集する．不動産ジャパンは他の不動産ポータルサイトと比べて，ページのフォーマットが固定されており，データ収集がしやすいため利用した．図 2 に示したのはこの HTML の一例であり，この中から必要な情報を Web スクレイピングにより収集する．HTML の解析には BeautifulSoup [14] を用いた．

本研究では以下に示す物件に関する情報を土地価格推定で使用する．

（1） **所在地**　物件の住所．本研究では建物名などを除く番地までの住所を使用する．

（2） **土地価格**　売り手から提示されている土地の価格であり，本研究ではこの価格を推定する．

（3） **土地面積**　敷地の面積．土地の大きさは土地の所在地に関わらず土地価格に直結するため土地価格の推定に使用できると考え，特徴の一つとして利用する．

（4） **接道状況**　対象の土地の接している道路の状況を示す特徴量として不動産ジャパンで利用されている特徴である．例えば，接道状況:二方, 南東:27m(公道) 接面:24.5m, 北:6m(公道) といった記述が不動産ジャパンでは使われている．一般的には接道が多いほど価格は上昇し，1 方向しか接道がない場合や袋小路などは価格が下降する傾向にある．

（5） **用途地域**　都市計画法で定められた地域地区の一つで，用途の混在を防ぐ目的で住居，商業，工業など 13 種類に土地を分類したものである．この 13 種類とは，第一種低層住居専用地域，第二種低層住居専用地域，第一種中高層住居専用地域，第二種中高層住居専用地域，第一種住居地域，第二種住居地域，準住居地域，田園住居地域，近隣商業地域，商業地域，準工業地域，工業地域，工業専用地域である．用途地域の分類によって建てられる家屋の大きさ等に制限が存在するため土地価格に影響を与えると考えられる．

図 2　物件情報 (不動産ジャパン [13]) の HTML

（6） **容積率**　土地面積に対する延床面積の割合のことで 50%から 1500%の範囲で制限が定められている．また，これは土地に対する建築物の大きさの制限を表しているため，家屋の有無を問わず価格に影響を与えると思われる．

（7） **地目**　土地の用途による区分のことで 23 種類に分類される．この 23 種類とは具体的には，原野，山林，畑，田，宅地，学校用地，鉄道用地，塩田，鉱泉地，池沼，牧場，墓地，境内地，運河用地，水道用地，用悪水路，ため池，堤，井溝，保安林，公衆用道路，公園，雑種地である．地目は土地に対しての制約等は存在しないが，これは周辺の環境を表しているため土地価格に影響を与えると考えられる．

（8） **地勢**　地理学で一般的に用いられる用語であり，標高や勾配などその土地の特徴や在り様を意味している．周辺を含めた地形が把握できるため推定に使用する．

（9） **都市計画区域区分**　都市計画法における都市計画区域の区域区分であり，優先的に市街化を行う区域を市街化区域，市街化すべきでないとされている区域を市街化調整区域，その他の合計 3 種類に分類されている．市街化調整区域は建物の建築が制限されており，市街化区域は推進されている．そのため土地に建築物を建てようとする場合には少なからず影響を与えると考えられる．

（10） **古家の有無**　土地が取引されている段階で古家が存在しているかどうかを示している．古家の有無で価格が変動するのは明らかであり，この項目は有るか無いかの 2 値で表現できる．

（11） **公示地価**　国土交通省が選出した標準地の価格を客観的に鑑定した価格であり一般に公開されている．標準地とは，国土交通省が一定範囲内で平均的な特徴をもつと判断した土地のことであり，全国で約 26,000ヶ所が選出されており毎年更新される．公示地価はその公平性から土地物件取引における価格判断の材料になることが多い．しかし，土地価格の推定に必要と考えられる公示地価は不動産ポータルサイトには記載されていないため，国土交通省の土地総合情報ライブラリ [5] から物件情報と同様に Web スクレイピングにより収集する．

3.2　土地価格推定モデル

本研究では 3.1 節の方法で得た土地データを整形し，推定モデルの説明変数または目的変数とする．以下にその整形手法と土地価格推定モデルへの適用手法について述べる．

3.2.1　外れ値の除去と対数化

本節では土地価格の推定を容易にするために土地価格に対して施した処理について述べる．土地価格データの分布は，2017 年 7 月 6 日の岡山県のデータ 3071 件をヒストグラムで表すと，図 3 示したようになる．重回帰分析は正規分布のデータの分析に適した手法であるため，土地価格の対数を求め，データを正

規分布に近付ける．土地価格を対数化したデータの分布は図 4 のようになる．この土地価格の対数を重回帰分析の目的変数にする．

また，収集したデータには全体のなかで大きく外れた値が存在し，外れ値は重回帰分析において誤差を生じる大きな原因となる．そのため本研究では四分位数を用いて土地価格が四分位範囲の 1.5 倍を超えるデータを外れ値としてデータから除外した．

3.2.2 one-hot 表現

重回帰分析の説明変数として使用するためには，文字列の特徴は何らかの数値に変換する必要がある．そのため文字列の項目を one-hot 表現を用いて数値ベクトルへ変換した．one-hot 表現は，カテゴリ情報を数値ベクトルで表現する方法の一つであり，該当するカテゴリのベクトルの要素のみを 1 に，その他を 0 にする表現方法である．この one-hot 表現を利用した項目以下にまとめる．

接道状況 表記方法が均一でないため，頻出の角地，二方，三方の 3 次元の one-hot 表現で表した．

用途地域 13 種類の用途地域を 13 次元の one-hot 表現で表した．

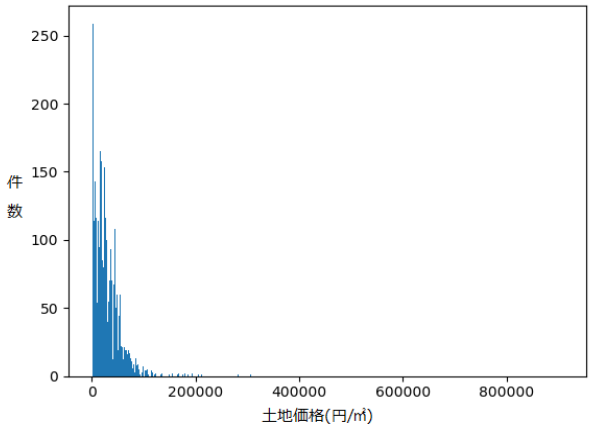

図 3　岡山県内土地価格の分布

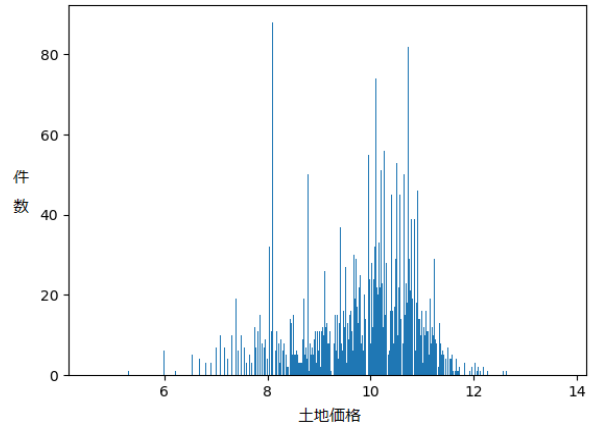

図 4　岡山県内土地価格の対数の分布

地目 頻出の原野，山林，畑，田，宅地，雑種地の 6 次元の one-hot 表現で表した．

地勢 不動産ジャパンに存在する，傾斜地，平坦，高台，低地，ひな段の 5 種類の表現を用い 5 次元の one-hot 表現で表した．

都市計画区域区分 2 種類の都市計画区域区分を 2 次元の one-hot 表現で表した．区分指定がない場合は 2 要素を 0 とした．

古家の有無 古家があるかないかの 1 次元の one-hot 表現で表した．

3.2.3 補間公示地価

公示地価は全国に約 26,000ヶ所の標準地を設定し鑑定を行うため，地価を知りたい土地と標準地が完全に合致することはほとんどない．そのため公示地価がない土地はその所在地周辺の公示地価から補間して代用する．本研究では以下の手順でこの補間公示地価を求めた．

（1） 国土交通省の土地総合情報システム [5] から得た標準地の所在地と公示地価の組の辞書を作成する．

（2） 公示地価を推定したい土地の所在地を正規表現により都道府県と市，それ以降の二つに分割する．東京都など市がない県では都道府県と区，それ以降の二つに分割する．

（3） （2）で分割した文字列全体を含む標準地の所在地が（1）の辞書にあればその公示地価を利用する．そのような標準地が複数あればそれらの平均を補間公示地価とする．公示地価が求まれば終了する．

（4） （2）で分割した市区以降が空であれば終了，そうでなければ市区以降の文字列の最後の 1 文字を削除し（3）に戻る．

3.2.4 土地価格推定モデルへの適用

3.2.2 項と 3.2.3 項で述べた方法で整形した土地データを推定モデルの説明変数として扱う．実験では，不動産ジャパンに数値が記載されている土地面積，容積率のみを説明変数とするモデルをベースラインとする．本稿で提案するモデルとしてベースラインに one-hot 表現の項目を追加したモデルを one-hot モデルと呼ぶこととする．また，one-hot モデルに公示地価を追加したモデルを公示地価モデルと呼ぶこととする．本実験で扱う説明変数は 3.1 節と 3.2.2 項で述べたもので合計 33 あるため，式 (1) は式 (3) となる．

$$y' = a_0 + a_1 x_1 + \ldots + a_{33} x_{33} \tag{3}$$

本研究で重回帰分析に使用する変数と対応する項目を表 1 に示す．

4. 評価実験

本節では評価実験とその結果について述べる．本実験では 10 分割交差検定を用いてモデルを評価する．k 分割交差検定では，データを k 個に分割し，そのうち一つを除いたデータを学習に使用し，残りのデータをテストに用いて評価する．これにより分割された各データについてランダム性の少ない評価が得られる．使用した不動産情報は不動産ジャパンから抽出した物件情

報のうち岡山県のデータ 3,071 件，広島県のデータ 3,602 件，東京都のデータ 4,031 件である．

4.1 推定価格と誤差

ここでは 3.2.4 項で述べたように土地価格推定モデルから算出した価格と実際の値の誤差を評価する．実験の結果を表 2 にまとめる．

表 2 より，この三つの評価データの分析結果全てにおいてベースラインモデル，one-hot モデル，公示地価モデルの順に誤差が小さくなっている．しかし，県内の平均土地価格から見た誤差の割合は岡山が 53%，広島が 52%，東京が 39%と誤差が大きい．

4.2 自由度調整済み決定係数

重回帰分析の結果を評価する指標として決定係数 R^2 がある．これは，実際の値と推定した値の相関の大きさを表す指標である．しかし決定係数は説明変数の数が増加すると値が大きくなる．本研究では土地価格推定モデル間で説明変数の数が異なるため，説明変数の数を考慮した自由度調整済み決定係数を重回帰分析の結果を評価する指標として用いる．また，一般的に決定係数が 0.7 を超えると説明変数と目的変数の間に高い相関があり，0.2 を下回るとほとんど相関がないとされている．自由度調整済み決定係数 R_f^2 は式 (4) で表される．

$$R_f^2 = 1 - \frac{\frac{\sum_{i=1}^{n}(y_i - y_i')^2}{n-1-k}}{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}} \tag{4}$$

式 (4) において，y_i は実際の土地価格，y_i' は推定した土地価格，$\overline{y}$ は y_i の平均である．k は説明変数の数である．各モデルにおける自由度調整済み決定係数を表 3 に示す．

表 3 から，説明変数が二つしかないベースラインは，誤差も大きいがこの決定係数の値も大きいことがわかる．一方，one-hot モデルと公示地価モデルは，説明変数が 33 個で 0.7 を超えているため，モデルとしての精度もある程度保っているといえる．

表 1　使用する説明変数の一覧とモデル

		ベースライン	one-hot	公示地価
x_1	公示地価	-	-	○
x_2	土地面積	○	○	○
x_3〜x_5	接道状況	-	○	○
x_6〜x_{18}	用途地域	-	○	○
x_{19}	容積率	○	○	○
x_{20}〜x_{25}	地目	-	○	○
x_{26}〜x_{30}	地勢	-	○	○
x_{31}〜x_{32}	都市計画区域区分	-	○	○
x_{33}	古家の有無	-	○	○

表 2　誤差 (円) と平均土地価格に対する割合 (%))

	岡山	広島	東京
県内の平均土地価格	26,460	28,453	436,285
ベースライン (割合)	16,139(61.0)	18,249(64.1)	246,221(56.4)
onehot モデル (割合)	14,293(54.0)	14,965(52.6)	230,015(52.7)
公示地価モデル (割合)	14,028(53.0)	14,874(52.3)	169,322(38.8)

唯一東京のみ，one-hot モデルに比べて公示地価モデルの自由度調整済み決定係数決定係数が小さくなっている．

4.3 標準偏回帰係数

2.1 節で述べたように，重回帰分析の利用目的の一つに，どの説明変数が目的変数をよく説明しているかを知ることができるという点がある．しかし，より目的変数を説明している変数は偏回帰係数の大きさでは判断できない．説明変数どうしを比較するときは標準偏回帰係数を用いる必要がある．標準偏回帰係数は，目的変数を平均値 0，分散 1 に標準化したときの偏回帰係数である．

標準偏回帰係数を算出するために各項目の各要素を標準化する．その際に東京都の工業専用地域など，全要素にデータがない項目が一定割合で出現するため，そのような項目は予め除外して分析する．

標準化済みのデータに対して重回帰分析により標準偏回帰係数を算出したものが表 4 に示した結果である．表 4 をみると，

表 3　自由度調整済み決定係数

	岡山	広島	東京
ベースライン	0.969341	0.984503	0.904133
onehot モデル	0.768843	0.731727	0.806315
公示地価モデル	0.739689	0.721930	0.570400

表 4　標準偏回帰係数

		岡山	広島	東京
公示地価		0.168684	0.129834	0.531365
土地面積		0.058592	0.037809	0.018591
接道	角地	0.017400	0.056735	0.075365
	二方	0.029424	0.010778	0.017626
	三方	0.012009	0.009722	0.011731
用途地域	第一種住居地域	0.062931	0.011827	0.028780
	第二種住居地域	0.077377	0.054087	0.003182
	準住居地域	0.010320	0.019482	0.039665
	近隣商業地域	0.014090	0.071339	0.058589
	商業地域	0.003424	0.008425	0.121164
	準工業地域	0.092876	0.045806	0.046449
	工業地域	0.010709	0.011054	0.040818
	工業専用地域	0.016219	0.008185	-
容積率		0.056543	0.000652	0.235365
地目	原野	0.046624	0.017590	0.007546
	山林	0.041475	0.054998	0.010153
	畑	0.026612	0.011731	0.023373
	田	0.035121	0.016503	0.006149
	宅地	0.298713	0.174234	0.137965
	雑種地	0.111019	0.031082	0.032147
地勢	傾斜地	0.020689	0.004461	0.007025
	平坦	0.045367	0.089306	0.056951
	高台	0.027199	0.061146	0.030040
	低地	0.074243	0.069745	0.032536
	ひな段	0.027564	0.008925	0.007278
都市計画区域区分	市街化区域	0.259742	0.369174	0.015046
	市街化調整区域	0.041221	0.100860	0.024328
古家の有無		0.031728	0.065999	0.088597

大きな値を示しているのは東京都の公示地価，岡山県と広島県の市街化区域である．これはそれぞれの評価データにおける土地の特徴を反映している．東京では公示地価に重きをおいて土地価格が決定され，岡山と広島では都市計画区域区分が土地価格の決定に影響を与えている．この岡山，広島，東京の評価データにおいて one-hot 表現の標準偏回帰係数の値が全て小さい接道状況や地勢の項目については，データの整形方法などに改善の余地がある．

4.4 考　　察

本実験は Web 上に存在する土地物件情報から土地価格推定モデルを作成することを目的とするが，結果として土地価格推定の誤差は平均土地価格の 5 割程度であった．テストデータのうちこの推定誤差が 1 割以内であったのは全体の約 1 割，2 割以内であったのは約 2 割しかなく改善の余地をまだ残している．特に本稿で one-hot 表現として扱ったデータのいくつかは，標準偏回帰係数としてはかなり低い値ばかりとなった．one-hot 表現として扱ったデータのうちいくつかは土地価格に影響を与えることが本実験で明らかになったため，データの扱いの改善が土地価格推定モデルの改善につながると考えられる．

5. ま と め

本研究では，Web から得られる土地物件情報をもとに土地価格推定を行なう線形なモデルを重回帰分析により作成した．土地物件情報は不動産ポータルサイトから Web スクレイピングにより取得した．本稿では，この情報で数値以外のデータは one-hot 表現を利用することで数値データに変換した．また，不動産ポータルサイトになかった土地の公示地価を求めるため，国土交通省の土地総合情報システムの公示地価も使用した．これらの情報を使って重回帰分析による土地価格の分析を行った．重回帰分析による土地価格の推定値の誤差の割合は 平均土地価格の 50%程度あり，線形モデルのみでは土地価格を推定するのが困難なことを確認した．また，実験から公示地価と都市計画区域区分が土地価格の推定において有用であることを確認した．今後の課題は物件情報を数値データに整形する手法の改善と非線形モデルを用いた土地価格の推定である．

文　　献

[1] F. Pedregosa, G. Varoquaux, A. Framfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine Learning in Python, pp. 2825-2830, 2011.

[2] 松尾哉太，新妻弘崇，太田学，“レビュー文書における重要文選択の一手法”，情報処理学会研究報告，Vol.2015-DBS-162，No.12，pp.1-7，2015.

[3] 新井晃平，新妻弘崇，太田学，“Twitter を利用した観光ルート推薦の一手法”，第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM2015)，pp.1-8，2015.

[4] 中川智也，新妻弘崇，太田学，“マイクロブログを利用した観光ルート推薦における移動効率の改善”，第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016)，pp.1-8，2016.

[5] 国土交通省:土地総合情報システム，
http://www.lad.mlit.go.jp/webland/

[6] 井上亮，木越尚之，清水英範，“時空間クリギングの地価推定への適用可能性の検討”，地理情報システム学会第 14 回学術研究発表大会講演論文集，Vol.14，pp.39-42，2005.

[7] 増成敬三，“kriging による公示地価の分析”，計算機統計学，Vol.18，No.2，pp.107-122，2007.

[8] 李勇鶴，井上亮，清水英範，“土地取引価格の空間内挿への共クリギングの適用可能性の検討”，地理情報システム学会第 17 回学術研究発表大会講演論文集，Vol.17，pp.245-248．2008.

[9] 井上亮，清水英範，吉田雄太郎，李勇鶴，“時空間クリギングによる東京 23 区・全用途地域を対象とした公示地価の分布と変遷の視覚化”，Theory and Applications of GIS，Vol.17，No.1，pp.13-24，2009.

[10] 井上亮，杉浦綾子，米山重昭，中西航，“公示地価指標と取引価格の比較に基づく地価情報提供の提案-不動産市場の透明性向上に向けて-”，土木学会論文集 D3(土木計画学)，Vol.72，No.1，pp.1-13，2016.

[11] McKinsey Global Institute，“THE US ECONOMY: AN AGENDA FOR INCLUSIVE GROWTH”，https://www.mckinsey.com/ /media/McKinsey/Global%20Themes/Employment%20and%20Growth/Can%20the%20US%20economy%20return%20to%20dynamic%20and%20inclusive%20growth/MGI-US-Economic-Agenda-Briefing-paper-November-2016.ashx

[12] 不動産総合データベース資料，http://www.mlit.go.jp/report/press/totikensangyo16_hh_000139.html

[13] 不動産ジャパン，http://www.fudousan.or.jp/

[14] Beautiful Soup, L.Richardson, Beautiful Soup Documentation, 2016.