

地域における不動産価格情報および公表のための Web システム の提案

Local Real Estate Price Information and the Web System for Publication

植杉 大^{1*} 生島高裕^{2†}
Dai Uesugi¹ Takahiro Ikushima²

¹ 摂南大学経済学部

¹ Setsunan University, Faculty of Economics

² 株式会社 数理先端技術研究所

² Mathematical Science Advanced Technology Laboratory CO., LTD.

Abstract: It is often said that real estate price information in Japan is less sufficient than that in the US. Especially, we need price estimates of a specific real estate that reflects local characteristics. In this paper, we suggest a local price estimate method using GWR and a web system for publication.

1 はじめに

一般的に、地価形成の要因分析や地価水準の変動、特定地点の地価の予測等を行う場合、ヘドニック回帰モデルが用いられる。ヘドニック・アプローチは、財・サービスの有する各品質の合成がその価格を与えると考えるものである。したがって、土地や不動産のように、市場に同質的なものが存在せず差別化や多様化の著しい財の場合、ヘドニック回帰モデルは非常に強力な分析ツールである。さらにその実証分析方法は、対象となる土地や不動産の価格を、CBD への時間距離、最寄駅からの時間距離や面積、属する地域の様々な特性などに回帰する極めて単純なものであるため、ヘドニック回帰モデルを通常の最小二乗法 (Ordinary Least Square: OLS) で推定するといった研究が従来なされてきた。

しかし不動産は同一用途に基づいて社会的・地域的にまとまりを有している。したがって個別土地の価格形成も地域的なまとまりを示し、近隣地域に属する土地や不動産の価格は強い相関を持つことが予想される。そのためサンプルの独立性という、OLS 推定の前提が満たされないという批判から、空間的相関を考慮した計量経済モデルの研究がおこなわれてきた。この空間計量

経済学における代表的なモデルとして、空間自己回帰モデル (Spatial Autocorrelation Regression model: SAR model) や空間誤差モデル (Spatial Error Model: SEM) などが挙げられる。上記のいずれも、研究の対象となる地域に属する標本データを用いて、その地域全体に係るパラメータ推定を行う。そこでは、対象地域の空間的な同質性を強調し標本データに基づく情報を集約するとともに、当該地域における一般的な法則性を見出すことに重点が置かれる。したがって、土地の価格を各説明変数に回帰した場合、例えば CBD への時間距離に係る係数パラメータの推定量は地域全体で一つしか求められない。このようなモデルは対象地域全体の推定を行うことから、グローバル回帰モデルと呼ばれる。

これに対して、従来のグローバル回帰モデルにかわり空間的非定常性を仮定したローカル回帰モデルによる研究が行われている。これらの研究は、研究の対象となる地域に属する標本データを用いて、局所的なパラメータ推定を行うことを目的としており、対象地域の一般的な法則性の発見や情報の集約という観点よりも、研究地域内での差異の発見や例外的な地点の発見などを重視する。したがって、このようなモデルは局所的に係数パラメータなどを推定することとなるため、ローカル回帰モデルと呼ばれる。本研究で用いる地理的加重回帰モデル (Geographically Weighted Regression model: GWR model) は、局所多項式回帰モデル (Local Polynomial Regression model: LPR model) などと並び、ローカル回帰モデルの代表的なモデルである。

*連絡先：摂南大学
大阪府寝屋川市池田中町 17-8
E-mail: uesugi@econ.setsunan.ac.jp

†連絡先：株式会社 数理先端技術研究所
東京都千代田区神田錦町 1-13 宝永錦町ビル
E-mail: ikushima@symbiotic.jp

ところで近年、大都市圏と地方との地価動向が二極化しており、大きな経済的問題とされている。このような二極化現象は一国全体だけではなく、非常に小さなエリア内においても見られる現象である。この地点あるいは地域に関する特性に関してグローバル回帰モデルでは必要な情報が得られない。さらに空間的非正常性が存在するならば、当該地点あるいは地域について、ヘドニック回帰モデルの属性に係るパラメータがグローバルな推定パラメータと大きく異なる可能性がある。すると、属性値を代入して求めた土地価格の予測値も大きく異なってしまう可能性がある。以上の考えのもと、本発表ではローカル回帰モデルを用いて小地域レベルでの不動産価格推定方法を提案する。

本論文の構成は以下のとおりである。第2節では、我が国の不動産に関する情報整備の現状を概説する。第3節では、GWRによる推定モデルを概説する。第4節では、GWR推定を行った事例について例示を行う。第5節では、GWR推定に基づく結果を公表するためのWebシステム構築の試作について提案を行う。

2 不動産に関する情報整備の現状

最近になってようやくではあるが、不動産市場における価格情報等を公開する動きが出始めている。価格情報の量や正確性は、適切な市場価値の形成には欠かせない。しかしその制度整備は未だ不完全であるといわざるを得ない。

不完備情報市場における逆選択の問題として、レモン市場の原理が有名であるが、不動産市場(特に中古不動産市場)については、価格情報や物件の質の情報など、さまざまな面で買い手が必要とする情報を提供していない。したがって買い手は売り手の提示する募集価格や近隣地域に存する対象不動産と類似の不動産の価格情報、さらにマクロ経済動向などを加味しながら大雑把に評価額を推定しているにすぎない。

マクロの不動産価格動向を示すものとして、2011年4月から東証住宅価格指数の試験公開が始まった¹。この指数は中古マンションの成約価格に基づいた指数で、データは東日本不動産流通機構(東日本レインズ)が登録不動産業者から収集したデータを用いている。公開頻度は月1回で、2か月前の指数が公開される。作成方法は、アメリカのS&Pケース・シラー住宅価格指数と同様のリポートセールス法である²。指数の基準日もケース・シラー住宅価格指数と同様の2000年1月と

しており、国際比較が可能となっている。このことから、日本国内の不動産市場の停滞を受けて、指数作成方法をワールドスタンダードに合わせることによって海外からの投資マネーを期待しているものと思われる。いずれにせよ、マクロな不動産価格動向を(東日本のみであるが)知ることができるようになることで、中古不動産流通の情報の不透明さが改善されることになろう。また何よりも、指数推定のためのデータが成約価格であることは、情報の信頼性を裏付けることになろう。しかし不動産の買い手が必要とする個別不動産の価格情報は明確にはならない。

また、国土交通省土地総合ライブラリーにおいて、2012年8月より、不動産価格指数(住宅)の試験公開が始まった³。対象は住宅一般であり、更地や建物つき土地、マンションである。対象地域は全国に及び、ブロック別や都市圏別での指数算出を行っている。産出も基準時点は2008年4月から2009年3月までの算術平均を100としている。情報ソースは取引当事者へのアンケートによる。推定方法はヘドニック法の時間ダミーを使った指数の推定を行っている。これも先の東証住宅価格指数と同様、個別不動産の価格情報は明確にならず、大きなトレンドを把握するにとどまる。

それ以外の不動産価格情報としては、国土交通省の土地総合情報システムにおける不動産取引価格情報検索がある⁴。これは、平成18年4月に始まったサービスであり、不動産取引当事者へのアンケートに基づく不動産の実際の取引価格に関する情報を四半期毎に提供しているものである。したがってデータの数は必ずしも十分とは言えないが、成約価格ベースでのデータである。しかし、取引当事者の個人情報保護の観点から取引不動産の所在地情報が枝番まで明らかにされていない。また、公表される取引価格が有効数字2ケタまでの概数であるため、あくまでも目安にしかならない。

また、レインズが公表している成約価格情報サイト「REINS Market Information:RMI」もある⁵。これはマンション、戸建住宅の成約価格情報でレインズに登録されたものをエリアごとに検索できるシステムである。やはり個人情報保護の観点から取引不動産の所在地は分からず、価格も有効数字2ケタまでの概数表示となっているのは、国土交通省の不動産取引価格情報検索と同様である。

我が国の成約価格情報をもとにしたデータは以上の4つが主なものである。その他はさまざまな研究所等で公表する有料の価格データやインターネットの物件検索サイト等での募集価格情報となる。

¹<http://www.tse.or.jp/market/data/homeprice.indices/index-dex.html>

²不動産価格指数の作成に当たっては、主にヘドニック法とリポートセールス法がある。リポートセールス法については「住宅市場動向に関する指標のあり方の検討業務報告書」(http://www.mlit.go.jp/report/press/sogo16_hh_000030.html)を参照のこと。

³<http://tochi.mlit.go.jp/secondpage/6993>

⁴<http://www.land.mlit.go.jp/webland/servlet/MainServlet>

⁵<http://www.contract.reins.or.jp/search/displayArea-ConditionBLogic.do>

翻ってアメリカでは、不動産価格情報はどのようになっているのかというと、日本では考えられないほどの情報量と精度で公開されている。例えば、Zillow⁶は、アメリカの主要都市に所在する各個別の不動産の価格がすべてわかるようになっているウェブサイトである⁷。公開情報は成約価格履歴のほか現在にかけての推定価格の推移、推定利回りの推移、写真なども含めた物件の概要などがある。その他、google map の street view 機能を使つての近隣エリアの調査もできるし、ローンの試算結果の表示、果てには取引に際してのエージェント紹介とその人物評価など、不動産購入に際してのありとあらゆる情報が一度にそのウェブサイトで得られるようになっている。

このようにアメリカでは当たり前のように公開される成約価格情報であるが、日本では個人情報保護の観点から秘匿事項となっている。業者も成約価格情報は一種の既得権益のように業者以外には一切公表していない。しかし現実には、不動産市場における情報の未整備は、前節で説明したような混乱を招く。大雑把な価格動向を把握するにとどまらず、小地域や地点を限定した不動産価格の推定とその公表システムの構築の必要性はいうまでもないことであろう。

3 GWR について

本発表では、ローカル回帰モデルとして地理的加重回帰モデル (Geographically Weighted Regression model :GWR) を用いる。このモデルは、Brunsdon, et al.(1996) や Fotheringham, et al.(2002) によって考案されたものであり、対象地域内における全体的傾向を研究する一般のグローバル回帰モデルと異なって、むしろ対象地域内における例外的挙動を示す小地域の探索や地域差を強調する目的に用いられる。しかしグローバル回帰モデルに属する空間計量経済学のアプローチと比較して、GWR の当該分野への適用例はそれほど多くはない。

グローバル回帰モデルとローカル回帰モデルの相違を端的に言えば、回帰式の説明変数と被説明変数との関数関係が対象地域内で同一であるか、あるいは対象地域内で局所的に変動するかである。より具体的に言えば、回帰分析により対象地域全体で各パラメータが唯一推定されるか、あるいは地点ごとに異なるかということである。ここで特に、空間的に変動する場合を空間的非定常性 (spatial nonstationarity) という。

土地価格や不動産価格に関するヘドニック回帰モデルでは、都心への時間距離や最寄駅からの時間距離、地理的特性を表すダミー変数が説明変数として用いられる。もし空間的非定常性が存在する場合、これらに係る局所的なパラメータ値の間に無視できない差異が生じる可能性が高い。さらに対象地域内に位置するある地点の予測値を求める場合、グローバル回帰モデルに基づいて予測を行うのが妥当でない可能性もある。

空間的非定常性が存在する場合、対象地域で局所的に変化する土地価格や不動産価格を推定するために、パラメトリック手法とセミパラメトリック手法が用いられる。パラメトリック手法の代表的なものとしては、多項式展開モデル (parametric polynomial expansion model) が挙げられる。このモデルは局所的な当てはまりを重視し高次元多項式を用いてパラメータ推定を行う。しかしどの程度の次元の多項式を用いるかの基準は曖昧であり、また次元が増えれば推定パラメータ数も増加するので推定の自由度が低下する。一方セミパラメトリック手法としては、局所回帰モデル、スプライン関数を用いたモデル、LOWESS 回帰モデル、カーネル回帰モデル、そして本論文で用いる GWR などが挙げられる。これらセミパラメトリックあるいはノンパラメトリック手法は、回帰モデルに先験的な関数形を与えず、空間的非定常性を各標本点の位置情報 (座標値) のみで除去することを特徴としている。中でも、GWR を用いる最大の利点は、なじみの深い通常の回帰分析の枠組みを用いているため直観的で解釈が明快であることだといえる。

いま、対象地域内の n 個の地点で収集された標本データセットがあるとする。説明変数の数は定数項を含め $K+1$ 個あるとする。被説明変数を y_i 、説明変数を X_{ki} とする。ここで、 $1 \leq i \leq n, 0 \leq k \leq K$ である。 ε_i を誤差項とし、正規分布に従うと仮定する。

すると、GWR は以下のようにあらわされる。

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^K \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \quad (1)$$

ここで、 u_i が緯度、 v_i が経度である。この式からわかるように、GWR と通常のグローバル回帰モデルとの相違点は、推定パラメータが標本点の位置情報である空間座標の関数となっているため、標本点ごとのローカルパラメータが求められる点である。さらに、標本地点 i からすべての標本地点までの距離を反映した重み行列を w_i とすれば、上記の式は以下のように行列表記される。

$$w_i y = w_i X \beta_i + \varepsilon_i \quad (2)$$

ここで、各行列およびベクトルを以下のように定義する。

⁶<http://www.zillow.com/>

⁷google map によるエリアの上空写真をみると、不動産 1 戸 1 戸に価格が表示されている。同様の大きなサイトとしては、Trulia(<http://www.trulia.com/>), Roost(<http://www.roost.com/>) がある。

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1K} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nK} \end{pmatrix},$$

$$\beta_i = \begin{pmatrix} \beta_{0i} \\ \vdots \\ \beta_{Ki} \end{pmatrix} = \begin{pmatrix} \beta_0(u_i, v_i) \\ \vdots \\ \beta_K(u_i, v_i) \end{pmatrix}, \epsilon_i = \begin{pmatrix} \epsilon_i \\ \vdots \\ \epsilon_n \end{pmatrix},$$

$$w_i = \text{diag}\{w_{i1}, \dots, w_{in}\}$$

したがって、パラメータ推定量は以下の式で求められる。

$$\hat{\beta}_i = (X'W_iX)^{-1}X'W_iy \quad (3)$$

ここで、 $W_i = w_i'w_i$ である。以降では、この W_i を空間加重行列と呼ぶ。空間加重行列は、対角成分が標本地点 i からすべての標本地点までの距離を反映した重みであり、非対角成分が 0 の行列である。したがって、以下のようにあらわされる。

$$W_i = \text{diag}\{w_{i1}^2, \dots, w_{in}^2\} = \begin{pmatrix} W_{i1} & 0 & \cdots & 0 \\ 0 & W_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{in} \end{pmatrix} \quad (4)$$

この空間加重行列の対角要素 $W_{ij}(1 \leq j \leq n)$ を決定するのがカーネル関数である。カーネル関数は、回帰を行う各地点から標本地点の距離が遠くなるに従い加重を低減させるという発想に基づく。カーネル関数としていくつか提案がなされているが、特に以下のようなガウス型関数がよく用いられている。Brunsdon, et al.(1996) で推奨されているガウス型のカーネル関数は以下のようにあらわされる。

$$W_{ij} = \sqrt{\exp\left(-\left(\frac{d_{ij}}{\theta}\right)^2\right)} \quad (5)$$

ここで d_{ij} は標本地点 i から他の標本地点 j までのユークリッド距離を表わす。また、 θ はカーネルのバンド幅を表わす。最適なバンド幅は、一般的にクロスバリデーション (Cross Validation:CV) に基づき、ヘドニック回帰モデルの理論値と実績値の平方和 (CV 値) を最小化することで求められる。

また、カーネルのバンド幅の設定方法は、固定型カーネルと適応型カーネルとがある。固定型カーネルは、一般化交差確認法 (Generalized Cross Validation:GCV) において GCV 値を最小にするバンド幅を求めた後にこれを固定して用いる方法である。一方適応型カーネ

ルは、初めに回帰を行う地点の近傍にある標本地点の個数を決定し、回帰地点から最も遠い標本地点までの距離をバンド幅とする方法である。したがってバンド幅について、前者は固定的だが、後者は回帰地点ごとに可変的である。

4 GWR を用いた分析例

本節では、OLS と比較して GWR がいかに小地域特性を把握できるかを、我々が行った他の調査研究を用いて確認する。

本発表では、大阪府における都市緑化の経済価値を調査するため、公示地価における価格および物的情報、町丁目別の緑地率を用いて、GWR に基づく緑地率上昇の限界的な経済効果を計測した調査研究を用いる。

推定式は以下のとおりである。

$$\begin{aligned} \log(\text{price}) = & \beta_0 + \beta_1 * \log(\text{dist}) \\ & + \beta_2 * \log(\text{area}) + \log(\text{walk}) \\ & + \beta_4 * \text{com} + \beta_5 * \text{ind} + \beta_6 * \text{other} \\ & + \beta_7 * \text{fire} + \beta_8 * \text{gas} + \beta_9 * \text{sew} \\ & + \beta_{10} * \text{kenpei} + \beta_{11} * \text{youseki} \\ & + \beta_{12} * \text{lg} + \beta_{13} * \text{hg} + \beta_{14} * \text{rg} \\ & + \beta_{15} * \text{cg} + \beta_{16} * \text{ig} + \beta_{17} * \text{og} \\ & + \beta_{18} * \text{fg} \end{aligned}$$

ここで、 price は公示地点の地価、 dist は大阪駅までの移動距離 (km)、 area は面積、 walk は最寄り駅までの時間距離 (分)、 com は商業地域ダミー、 ind は工業地域ダミー、 other は用途地域外ダミー、 fire は防火準防火地域ダミー、 gas はガスダミー、 sew は下水ダミー、 kenpei は建ぺい率、 youseki は容積率、 lg 、 hg 、 rg 、 cg 、 ig 、 og 、 fg は用途地域等 (低層住居専用地域、中高層住居専用地域、その他住居地域、商業地域、工業地域、その他用途地域外、防火地域) と緑地率の交叉項である。

この基本モデルに基づき、OLS と GWR で推定を行った結果、それぞれの修正済み決定係数は OLS が 0.7613 である一方、GWR は 0.9159 と大幅に改善された。図 1 が OLS、図 2 が GWR によって推定された予測地価と実際の地価の推定誤差の分布である。推定誤差が小ければ全体として色が薄くなる。GWR では推定誤差の大幅な改善が見られる。

この大阪府を対象とした分析では、緑地の経済価値の地域的分布の様子は、図 3 のようになることがわかった。ここで注目してほしいのは、エリア特性と分布の様子が多様であることが可視化されている点である。

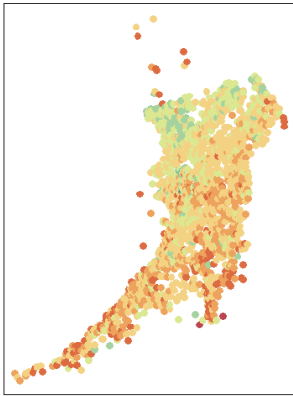


図 1: OLS による推定誤差

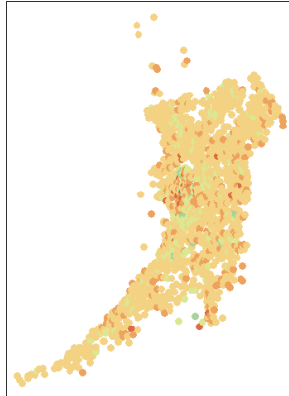


図 2: GWR による推定誤差

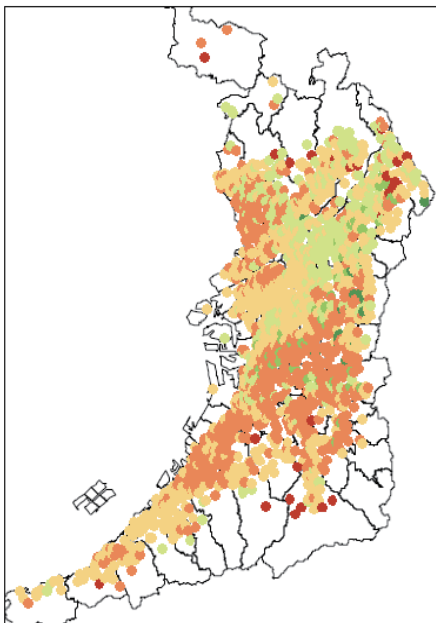


図 3: GWR による緑地の経済価値

以上の例でもわかるように、GWR を用いることでエリア限定、あるいは地点限定の価格やパラメータが求められることができる。また推定されたパラメータに基づき、サンプル地点以外の地点の価格も求めることができる。

5 GWR 推定にもとづく Web 公表システム

本システムは、国土交通省の『不動産取引価格情報』（オープンデータ）を使用し、全国の不動産取引情報を可視化し、GWR モデルにより推定されたパラメータを用いて、希望場所、地積を指定した時の、推定価格

を表示するシステムである。以下、システム仕様と実装について記述する。

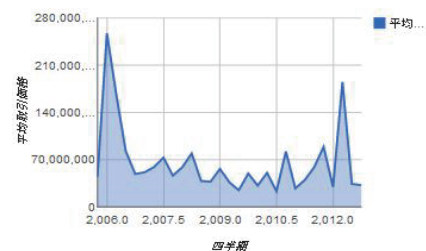
5.1 Web システム仕様

全国の不動産取引データを地域（都道府県、市区町村、町大字、最寄り駅）、四半期年、取引物件種類（更地、新築一戸建て、中古一戸建て、新築マンション、中古マンション）単位に、7つの不動産情報項目（平均取引価格、平均一平米単価取引価格、物件数、平均築年数、平均面積、最寄り駅からの距離平均、間取り）を計算しデータベース化した。そして、ユーザインターフェイスでの可視化項目は、地域、取引物件種類を固定にして7つの不動産情報項目の時間変化を捉えたグラフ表示、四半期年、取引物件種類を固定して、空間的变化を表示した地図表示、不動産情報項目の2項目関係を見える化するために、不動産情報項目の組み合わせを取り、クロス集計表として可視化した。

ユーザはこれらの情報から、時間的、空間的に状況を把握する。

最後に地域不動産価格インデックス表示機能であるが、これは上記のデータを基に、GWR モデル計算を行い、各地域でのパラメータを算出し、最寄り駅から都心までの時間距離、最寄り駅からの時間距離、地積を入力し、価格推定を行う。これによってユーザは自分の希望する場所に、地積を入力するだけで推定価格を知ることができるようになる。このことは、不動産についての基準を与えられることになり、不動産取得に関する重要なサービスとなる。

時系列グラフ

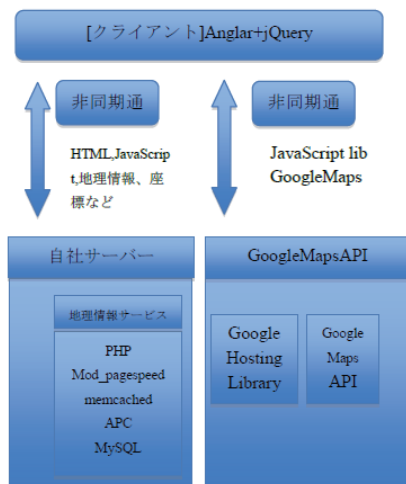
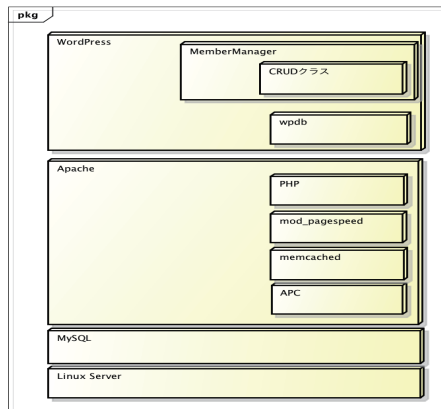


5.2 Web システム実装仕様

本システムでは CMS 管理ソフトのシェアと開発効率の面から Wordpress を採用し、これのプラグインおよびテーマを開発することとする。Web サーバソフトウェアは Apache、開発言語は PHP、JavaScript、Wordpress を基盤として開発する。



DB は MySQL を使用し、DAO は Wordpress モジュールを使用する。Web サービスの IO を最小化するため、memcached を使用、PHP の PHP の中間コードのキャッシュや最適化を行うため、APC を使用、開発効率のため、Apache モジュールである mod_pagespeed を使用する。



5.3 オープンデータについて

オープンデータを利用したシステムは今後増えてくると思われる。動向としては「世界最先端 IT 国家創造」宣言に始まる、「オープンデータ・ビッグデータの活用の推進」があげられる。今回のシステムで使用したデータは、土地総合情報システムの不動産の取引価格情報を使用した、「地価公示」の詳細を基にデータ構造設計を行った。

ビッグデータは現在、クローズデータが主であるが、今後オープンデータの活用が進むと思われる。

オープンデータの推進についての情報は以下の情報を参考にされたい。

- LOD チャレンジ 2014
<http://lod.sfc.keio.ac.jp/challenge2014/>
- オープンデータ流通推進コンソーシアム
<http://www.opendata.gr.jp/>
- アーバンデータチャレンジ 2014
http://aigid.jp/?page_id=421
- Open Knowledge Foundation Japan
<http://okfn.jp/>
- リンクト・オープン・データ・イニシアティブ
<http://linkedopendata.jp/>
- 「世界最先端 IT 国家創造」宣言
<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>
- 土地総合情報システムの不動産の取引価格情報
<http://www.land.mlit.go.jp/webland/servlet/MainServlet>
- 「地価公示」の詳細
http://nlftp.mlit.go.jp/ksj/jpgis/datalist/KsjTmplt-L01-v1_0.html

謝辞

本研究は JSPS 科研費 (課題番号: 24618016, 研究代表者: 植杉大) の助成を受けたものです。

参考文献

- [1] Anselin, L.(1988) "Spatial Econometrics: Methods and Models," Dordrecht: Kluwer.
- [2] Brunsdon, C., Fotheringham, A.S. and Charlton, M.E.(1996), "Geographically weighted regression : A method for exploring spatial nonstationarity", *Geographical Analysis* 28, pp.281-298.
- [3] Brunsdon, C., Fotheringham, A.S. and Charlton, M.(1999), "Some notes on parametric significance tests for geographically weighted regression", *Journal of Regional Science* 39, pp.497-524.
- [4] Clapp, J.M., Kim H.J. and Gelfand A.E.(2002), "Spatial Prediction of House Price Using LPR and Bayesian Smoothing," *Real Estate Economics*, 30:4, pp.505-532.
- [5] Fotheringham, A.S., Brunsdon, C. and Charlton, M.(2002), "Geographically Weighted Regression: The Analysis of Spatially Varying Relationships", Chichester, UK: Wiley.
- [6] Leung, Y., Mei, C.-L. and Zhang, W.-X.(2000), "Statistical tests for spatial nonstationarity based on the geographically weighted regression model", *Environment and Planning A*32, pp.9-32.