

## Doc2Vec と GWR を用いた不動産価格分析の新たなアプローチ

### A New Approach for Property Price Analysis with Doc2Vec and GWR

植杉 大\*  
Dai Uesugi\*

摂南大学経済学部\*

本研究の特徴は、ヘドニックモデルに基づく不動産価格分析において、テキストなどの非構造化データから抽出された特徴量を説明変数として用いている点である。地価公示データに付属する鑑定評価書のテキストについて、Doc2Vecを用いて各鑑定地点の特徴量を抽出し、さらに緯度経度を考慮した地理的加重回帰モデル(GWR)を併せて推定することにより、不動産価格が予測可能かを実験した。その結果、従来の構造化データによる推定結果と同等かそれ以上の説明力が観察され、高い有用性が示された。

**Keywords:** Doc2Vec, GWR (Geographically Weighted Regression), 非構造化データ (Unstructured Data), ヘドニックモデル (Hedonic Model)

#### 1. はじめに

本研究の目的は、従来の構造化データによる不動産価格の推定方法と比較して同等あるいはそれ以上の精度で、非構造化データによる不動産価格推定が可能か検証し、最適な方法を提案することである。

本研究の問題意識及び発想と同様に、ヘドニックモデルの説明変数として、非構造化データによる外生変数の利用を行っている研究はいくつか存在している。

Miura and Asami(2011)[1]では、東京 23 区内の約 800 地点においてインターネットで地名検索し、Yahoo! Blog に書き込みされた地域の評判に関するテキスト情報を極性分析し、ヘドニックモデルの説明変数として用いている。結果的に、一般的な構造化データのみで推定された場合には決定係数が 0.528, 評判に関するテキスト等の非構造化データの情報を組み合わせた場合は決定係数が 0.560 と報告している。

鶴山・諏訪・小川・荒川・安本(2019)[2]では、Doc2Vec を利用して、物件に付与されているキャッチコピーを収集して特徴量ベクトルに変換し、それらを説明変数とした重回帰分析を行った結果、決定係数が

0.611 となったと報告している。

植杉(2020)[3]では、確率モデルに基づくトピックモデルを利用して、地価公示に付属する鑑定評価書を用いることでテキスト情報を特徴量に変換している。それらのみをヘドニックモデルの説明変数として利用し重回帰分析を行うことで、0.6 程度の決定係数が得られることが報告されている。

以上の既存研究をみると、ヘドニックモデルの説明変数に各地点のテキストから抽出される特徴量を含めることで、モデルの説明力および予測精度が向上することが示されている。

そこで本研究は地価公示における各鑑定地点の鑑定評価書の文書データを利用し、ニューラルネットワークによる自然言語処理手法の Doc2Vec を用い、さらに地理的加重回帰(GWR)を用いて、ヘドニックモデルにおける各鑑定地点の価格に対する説明力および予測精度が向上するかを検証する。

#### 2. データ

本研究のデータソースは、主に国土交通省土地総合情報システムの地価公示情報<sup>1</sup>である。当該サイトから各鑑定地点の評価

<sup>1</sup> <https://www.land.mlit.go.jp/webland/>