

シャムネットワークを用いた賃貸物件のクラスタリング

発表者: 先端工学基礎課程 IMC コース 学籍番号 1525059 服部 凌典
指導教員: 岡本 一志 助教

1 はじめに

既存の賃貸物件の価格は、築年数やロケーション、などの変えられない属性の影響を強く受けるために、部屋ごとの価格差を出しにくい課題がある。本研究では、不動産業者が物件の部屋ごとの価格差をつけやすくしたり、より大きな付加価値を提供できるための可視化モデルを提案する。提案手法として、本研究ではシャムネットワークを用いて、家賃や築年数、専有面積などを入力し、間取り画像の特徴ベクトルの類似に応じたクラスタリングを行う。シャムネットワーク [1] とは、出力が結合された 2 つの同一のサブネットワークで構成されたもので、統一されたフレームワーク内でテキスト特徴及びメトリックを共同で学習することができる技術である。本研究では、学習したモデルに新規のデータを入力し、そのデータの分布場所及び、周辺の物件を可視化する。また、分布の位置が、変えられる変数（家賃や、共益費など）を変えた時の分布遷移を可視化する。

用いるデータセットについて本研究では、国立情報学研究所が株式会社 LIFUL から提供を受けて研究者に提供しているデータセットを使用する。データセットは 2015 年 9 月時点から 2017 年 4 月 1 日までの賃料、面積、立地（市区町村、郵便番号、最寄り駅、徒歩分）、築年数、間取り、建物構造、諸設備などの全国約 533 万件の賃貸物件データ及び、それら物件に紐付けされた、大横 120 ピクセル×縦 120 ピクセルの JPEG 形式の間取り図や室内写真などの画像データ約 8,300 万ファイルを用いる。本研究では、路線、徒歩距離、建物面積/専有面積、建物階数（地上）、築年月、部屋階数、間取部屋数、間取部屋種類及び、賃料/価格、賃料+管理費が欠損している物件を取り除いたデータ及び、間取り画像のみを使用した。また、河合伸治 [2] の論文を参考にデータを加工した。

2 word2vec による次元圧縮と重回帰分析

3 次元圧縮が与える影響の評価実験

3.1 実験環境

3.2 予測精度の比較と考察

3.3 推定したパラメータの比較と考察

4 おわりに

参考文献

- [1] JJ.Bromley, I. Guyon, Y. LeCun, and R. Shah: Signature verification using a "Siamese" time delay neural network, in NIPS, 1993, pp.737-744.
- [2] 河合 伸 治: ヘドニック・アプローチによる賃貸住宅価格の価格決定要因の推定: ソシオサイエンス vol.14 2008 年 3 月.