

賃貸物件の価格と属性を利用したレジームスイッチングモデル

齊藤 光† 神戸 聖寧† 佐藤 亘† 能代 哲太† 山岸 祐己†‡ 橋本 正洋§

†静岡理科大学 情報学部 ‡理化学研究所 革新知能統合研究センター §法政大学大学院 政策創造研究科

1 はじめに

一般に、賃貸情報サイトの検索結果や地価公示価格を参照することによって、ある程度各地域の家賃相場が推定でき、さらに精緻な家賃相場や価格に影響する物件の特徴などを推定したい場合には、価格に基づいた教師あり学習アプローチも存在している [1]。しかし、ライフイベントの変化にともなって家賃がどのように増減するか、もしくは、家賃を増減させることによって間取りをどのように変化させることができるか、といった、地理的制限と家賃と間取りの比較考量は、依然として手間がかかることが推察される。本研究では、家賃を系列、間取りをカテゴリとしたときの、多項分布型レジームスイッチングモデルを構築し、教師なし学習アプローチによって、このような比較考量における意思決定を助けるための情報可視化を目指す。また、提案モデルの系列やカテゴリは、主要駅からの距離や専有面積、築年数等にも拡張が可能であることも示す。

2 分析手法

賃貸物件データを $\mathcal{D} = \{(s_1, t_1, a_1), \dots, (s_N, t_N, a_N)\}$ とする。ここで、 s_n と t_n と a_n は、 J カテゴリの間取り $j \in \{1, \dots, J\} = \mathcal{J}$ と n 番目の物件の家賃と専有面積をそれぞれ表す。 $|\mathcal{D}| = N$ を物件数とすると、 $t_1 \leq \dots \leq t_N \leq \dots \leq t_N$ となる。なお、家賃は最安値のものが t_1 、最高値のものが t_N となるように昇順ソートを行い、同一の家賃についてはさらにデータ取得順序によって昇順ソートを行っているものとする。ここから、家賃に基づいた系列を考えるため、 n はタイムステップとし、 $N = \{1, 2, \dots, N\}$ をタイムステップ集合とする。

2.1 単純な可視化方法

間取り j の専有面積 a_n あたり価格を

$$\mu_j = \frac{\sum_{n:s_n=j} t_n}{\sum_{n:s_n=j} a_n}, \quad (1)$$

とする。さらに、目的変数を家賃 t_n 、説明変数を専有面積 a_n として、線形回帰モデル $t_n = \beta_{0,j} + \beta_{1,j} a_n + \epsilon_n$ を

考えたときの、間取り j の専有面積 a_n に対する回帰係数を

$$\beta_{1j} = \frac{\sum_{n:s_n=j} (a_n - \bar{a}_j)(t_n - \bar{t}_j)}{\sum_{n:s_n=j} (a_n - \bar{a}_j)^2}, \quad (2)$$

とする。ここで、 $\bar{a}_j = \frac{1}{|\{n \in N: s_n=j\}|} \sum_{n:s_n=j} a_n$ および $\bar{t}_j = \frac{1}{|\{n \in N: s_n=j\}|} \sum_{n:s_n=j} t_n$ である。

2.2 多項分布型レジームスイッチングモデルによる可視化方法

一般に広く利用されている正規分布を仮定したレジームスイッチング手法 [2] に対し、間取りのような順序のないカテゴリのための同様の手法を構築する。 k 番目のレジームの開始価格を $T_k \in N$ 、 $\mathcal{T}_K = \{T_0, \dots, T_k, \dots, T_{K+1}\}$ をスイッチングタイムステップ集合とし、便宜上 $T_0 = 1$ 、 $T_{K+1} = N + 1$ とする。すなわち、 T_1, \dots, T_K は推定される個々のスイッチングタイムステップであり、 $T_k < T_{k+1}$ を満たすとする。そして、 N_k を k 番目のレジーム内のタイムステップ集合とし、各 $k \in \{0, \dots, K\}$ に対して $N_k = \{n \in N; T_k \leq n < T_{k+1}\}$ のように定義する。なお、 $N = N_0 \cup \dots \cup N_K$ である。

いま、各レジームの状態分布が J カテゴリの多項分布に従うと仮定する、 p_k を k 番目のレジームにおける多項分布の確率ベクトルとし、 \mathcal{P}_K はそれら確率ベクトルの集合、つまり $\mathcal{P}_K = \{p_0, \dots, p_K\}$ とすると、 \mathcal{T}_K が与えられたときの対数尤度関数は以下のように定義できる。

$$L(\mathcal{D}; \mathcal{P}_K, \mathcal{T}_K) = \sum_{k=0}^K \sum_{n \in N_k} \sum_{j=1}^J s_{n,j} \log p_{k,j}. \quad (3)$$

ここで、 $s_{n,j} = \{1 \text{ if } s_n = j, 0 \text{ otherwise.}$ はダミー変数である。各レジーム $k = 0, \dots, K$ と各状態 $j = 1, \dots, J$ に対する式 (3) の最尤推定量は $\hat{p}_{k,j} = \sum_{n \in N_k} s_{n,j} / |N_k|$ のように与えられる。これらの推定量を式 (3) に代入すると以下の式が導ける。

$$L(\mathcal{D}; \hat{\mathcal{P}}_K, \mathcal{T}_K) = \sum_{k=0}^K \sum_{n \in N_k} \sum_{j=1}^J s_{n,j} \log \hat{p}_{k,j}. \quad (4)$$

したがって、スイッチングタイムステップの検出問題は、式 (4) を最大化する \mathcal{T}_K の探索問題に帰着できるため、レジームスイッチングを考慮しないときの尤度からの改善度合いに基づき、貪欲法と局所探索法を組み合わせた方法で高速に解くことが可能となる [3]。したがって、カテゴリ s_n や系列 t_n は、主要駅からの距離や

Regime Switching Model Using Rental Property Prices and Attributes
†Hikaru SAITO †Kiyone KAMBE †Wataru SATO †Tetta NOSHIO
†Yuki YAMAGISHI §Masahiro HASHIMOTO
†Shizuoka Institute of Science and Technology
‡RIKEN
§Hosei University

専有面積，築年数等にも拡張が可能である．なお，本実験では貪欲法アルゴリズムの終了条件として最小記述長原理 (MDL) [4] を採用し，事前にレジーム数を設定することなく自動で終了させる．

3 評価実験とまとめ

国内最大級の賃貸情報サイトである SUUMO* における，藤枝駅，浜松駅，静岡駅の各駅から半径 3 km 圏内の物件情報を用いて，各種分析手法を適応する．藤枝駅，浜松駅，静岡駅のそれぞれの物件数 N は，871，2682，2535，である．

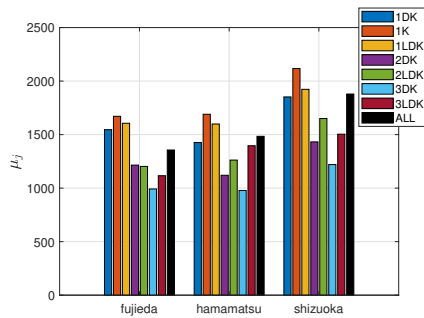


図 1: 各駅と各間取りの専有面積あたり家賃 μ_j

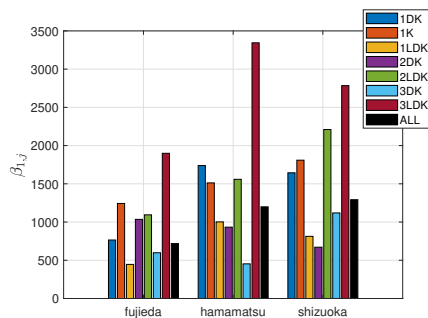


図 2: 各駅と各間取りの専有面積の回帰係数 $\beta_{1,j}$

図 1, 2 より，単純な可視化でも各駅の間取りごとの特徴を捉えることができるが，間取りが順序カテゴリではなく，サンプルサイズの違いもあるため，シームレスな比較考量は難しい．一方，提案レジームスイッチングによる可視化結果（図 3, 4, 5）では，横軸の家賃という系列に対し，どのように間取りのシェアが変化しているかが読み取りやすい．

参考文献

[1] Ti-Ching Peng and Chun-Chieh Wang. The Application of Machine Learning Approaches on Real-Time Apartment Prices in the Tokyo Metropolitan Area. *Social Science Japan Journal*, Vol. 25, No. 1, pp. 3–28, 08 2021.

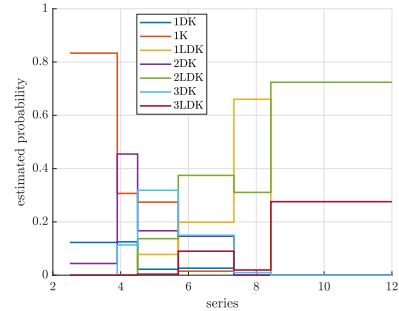


図 3: 藤枝駅の提案レジームスイッチング結果

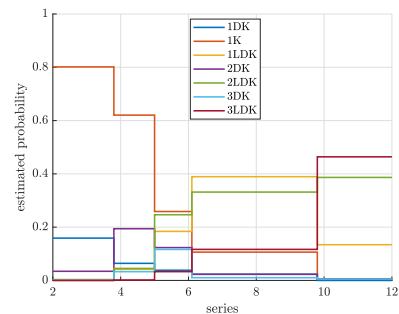


図 4: 浜松駅の提案レジームスイッチング結果

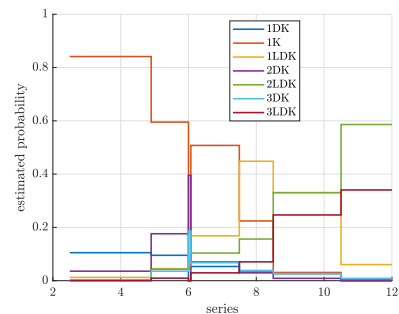


図 5: 静岡駅の提案レジームスイッチング結果

[2] Rebecca Killick, Paul Fearnhead, and I.A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, Vol. 107, pp. 1590–1598, 12 2012.

[3] Yuki Yamagishi and Kazumi Saito. Visualizing switching regimes based on multinomial distribution in buzz marketing sites. In *Foundations of Intelligent Systems - 23rd International Symposium, ISMIS 2017*, Vol. 10352 of *Lecture Notes in Computer Science*, pp. 385–395. Springer, 2017.

[4] J. Rissanen. Modeling by shortest data description. *Automatica*, Vol. 14, No. 5, pp. 465–471, September 1978.

*<https://suumo.jp/>