賃料予測残差に着目した特徴的部屋配置の抽出

Extraction of characteristic substructures of room layouts focusing on rent forecast residuals

尾崎 知伸 *
Tomonobu Ozaki

日本大学 文理学部
College of Humanities and Sciences, Nihon University

Abstract: The rent for rental property is determined by various factors such as area and age. In this paper, to extract interesting patterns strongly related with rent, we associate a class label to rental property based on estimated residual, and extract contrasting subgraph patterns considering the number of occurrences. Furthermore, we try to extract decision rules consisting of subgraph patterns by using a method in interpretability research of machine learning. The effectiveness of the proposed frameworks is evaluated in preliminary experiments using real world datasets.

1 はじめに

賃貸マンションの賃料は，部屋の広さや築年数，地域等の基本的な条件に加え，ペット可等の特記事由や間取り（部屋配置）等，様々な要因に左右される．賃料と強い関係を持つ部屋配置を特定することができれば，適正賃料の設定・判断に加え，新築・リフォーム時の参考材料としてそれらを利用することができ，貸し手と借り手の双方にとって大きな貢献が期待できる．

これまでに，賃料と部屋配置の関係に着目した分析がいくつか試みられている．例えば文献 [1] では，部屋配置を類型化した上で順序付けを行い，それらが賃料に与える影響を分析している．また文献 [2] では，複数の間取り図に共通して現れる部屋配置を属性とした回帰モデルを構築し，各部屋配置が賃料に与える影響を考察している．同様に文献 [3] でも，部屋配置を用いた回帰木・モデル木構築による分析が展開されている．一方文献 [4] では，部屋配置を属性とする回帰木アンサンブルからルールを導出することで，賃料に関して特徴的な部屋配置の抽出を行っている．

これらの既存研究と同様，本論文では賃料と関係の強い部屋配置を抽出することを目的とする．具体的には，グラフを用いて間取り図を表現し，実賃料と基本的な条件に基づく推定賃料との差（推定残差）に基づいてクラスを設定した上で，出現数を考慮した対比部分グラフを抽出する．さらに機械学習の解釈性研究 [5] における手法を援用し，部分グラフパターンに基づくルールを導出することで，賃料と関連の強い部屋配置の抽出を試みる．

以下に本論文の構成を示す．2 章では準備として，グラフ表現された間取り図集合におけるパターンと，各間取り図におけるパターンの出現数について議論する．3 章では，推定残差を用いたクラス設定について述べた後，閾値付き対比部分グラフを提案する．加えて，パターンとその出現数に関する決定リストの導出について述べる．4 章では，実験結果を示し考察を行う．最後に 5 章でまとめを行い，今後の課題を示す．

2 頻出部分グラフマイニング

2.1 頻出部分グラフ

既存研究 [2, 3, 4] と同様，本論文では，部屋やベランダ，玄関などを頂点，それらの繋がりを辺とみなすことで，各間取り図をラベル付き無向グラフとして表現する．グラフ $g = (V_g, E_g, \lambda_g)$ は，頂点集合 V_g と辺集合 $E_g \subseteq V_g \times V_g$，ラベル関数 $\lambda_g : V_g \cup E_g \to \mathcal{L}$（$\mathcal{L}$ はラベルの全体集合）の 3 項組で表現される．

2 つのグラフ $g = (V_g, E_g, \lambda_g)$ と $p = (V_p, E_p, \lambda_p)$ に対し，条件

1. $\forall u \in V_p\,[\lambda_p(u) = \lambda_g(f(u))]$
2. $\forall (u, v) \in E_p$
   $\exists (f(u), f(v)) \in E_g\ s.t.\ \lambda_p(u, v) = \lambda_g(f(u), f(v))$

を満たす単射関数 $f : V_p \to V_g$ が存在するとき，p を g の部分グラフと呼び，$p \sqsubseteq g$ と表記する．またグラフ

*連絡先：日本大学文理学部情報科学科
〒 156-8550 東京都世田谷区桜上水 3-25-40
E-mail: tozaki@chs.nihon-u.ac.jp

p とグラフ集合 $D = \{g_1, \cdots, g_n\}$ に対し，部分グラフとして p を持つ D 中のグラフの割合，すなわち

$$\mathrm{S}(p, D) = |\{g \in D \mid p \sqsubseteq g\}| \,/\, |D|$$

を p の支持度と定義する．

支持度に関する閾値 $\sigma\,(0 < \sigma \leq 1)$ に対し，条件 $\mathrm{S}(p, D) \geq \sigma$ を満たすグラフ p を，D における頻出部分グラフと呼ぶ．また，誤差に関するパラメタ $\delta \geq 0$ に対し，条件 $\forall q \sqsubseteq p\,(\,\mathrm{S}(q, D) > (1+\delta) \cdot \mathrm{S}(p, D)\,)$ を満たす頻出部分グラフを，頻出 δ フリー部分グラフと呼ぶ．D に対する頻出 δ フリー部分グラフの全体集合を

$$\mathcal{F}_D^{\sigma,\delta} = \left\{ p \,\middle|\, \begin{array}{l} \mathrm{S}(p, D) \geq \sigma \wedge \\ \forall q \sqsubseteq p\,(\,\mathrm{S}(q, D) > (1+\delta) \cdot \mathrm{S}(p, D)\,) \end{array} \right\}$$

と表記する．以降簡略化のため，頻出 δ フリー部分グラフをパターンと呼ぶ．

2.2 部分グラフパターンの支持度

一般に，部屋数が異なれば賃料が異なることからも明らかなように，特徴的なパターン（部屋配置）の抽出において，各物件におけるパターンの出現数は大きな意味を持つ．しかし，単一グラフ（間取り図）におけるパターンの支持度，すなわちパターンの出現集合から得られる出現の強さは自明ではない．

詳しくは後述するが，代表的な支持度である極大独立集合に基づく支持度（MIS）[6] や最制限頂点に基づく支持度（MRN）[7] は，出現の重複に関し強い制約を要求する．しかし部屋配置という対象の性質を考えた場合，重複に対する強い制約が必ずしも有効とは限らない．例えば MIS や MRN を用いた場合，キッチンに 1 つの部屋だけが繋がっているグラフ g^1 と，キッチンに 3 つの部屋が繋がっているグラフ g^3 に対し，(キッチン) $\longleftrightarrow$ (居室) というパターン p の支持度はすべて 1 となり，g^1 と g^3 を区別することができない．しかし，g^3 においてはキッチンへの導線が複数あり，また隣接した各部屋の利用方法にも複数の組み合わせが考えられることから，最も制約の強いキッチンだけを基準に支持度を考えることは必ずしも適切ではない．これらの議論を基に，近年，単一グラフにおけるパターンの新たな支持度が提案されている [4]．以下では形式的な準備の後，これらの各支持度を導入する．

(1) パターンの出現集合

グラフ g と p に対し，p と同型である g の部分グラフ，すなわち条件 $o \sqsubseteq g \wedge p \sqsubseteq o \wedge o \sqsubseteq p$ を満たす部分グラフ $o = (V_o, E_o, \lambda_o)$ を g における p の出現と呼ぶ．また g における p の出現の全体集合を

$$O(p, g) = \{o \mid o \sqsubseteq g, p \sqsubseteq o, o \sqsubseteq p\}$$

と表記する．

(2) 出現の有無に基づく支持度

支持度 $\mathrm{S}_{BIN}(p, g)$ は，グラフ g における p の出現の有無を表す．

$$\mathrm{S}_{BIN}(p, g) = \begin{cases} 0 & (\,O(p, g) = \emptyset\,) \\ 1 & (\,\text{otherwise}\,) \end{cases}$$

$\mathrm{S}_{BIN}(p, g)$ の値域は $\{0, 1\}$ であり，部屋配置という応用を考えた場合，識別能力（粒度）に関して問題があると考えられる．

(3) 極大独立集合に基づく支持度 [6]

支持度 $\mathrm{S}_{MIS}(p, g)$ は，頂点を共有せずに g 中に配置することのできる出現 o の最大数であり，

$$\mathrm{S}_{MIS}(p, g) = |MIS(p, g)|$$

と定義される．ここで $MIS(p, g)$ は，各出現 $o \in O(p, g)$ を頂点，（g における）頂点を共有する出現同士を結んだものを辺とするグラフ

$$\big(\,O(p, g),\ \{(o_j, o_k) \mid o_j, o_k \in O(p, g), V_{o_j} \cap V_{o_k} \neq \emptyset\},\ \lambda\,\big)$$

の極大独立集合を表す．

(4) 最制限頂点に基づく支持度 [7]

支持度 $\mathrm{S}_{MRN}(p, g)$ は，p 中の頂点 $v \in V_p$ の写像先集合（$\varphi_o(v)$ の適用結果）に対する最小要素数であり，

$$\mathrm{S}_{MRN}(p, g) = \min_{v \in V_p} |\{\varphi_o(v) \mid o \in O(p, g)\}|$$

と定義される．ここで $\varphi_o : V_p \to V_g$ は，出現 $o = (V_o, E_o, \lambda_o)$ に対し，条件

1. $\forall v \in V_p \Rightarrow \lambda_p(v) = \lambda_g(\varphi(v))$
2. $\forall (u, v) \in E_p \Rightarrow (\varphi(u), \varphi(v)) \in E_g$

を満たす頂点間の写像関数である．

$\mathrm{S}_{MIS}(p, g)$ 同様，$\mathrm{S}_{MRN}(p, g)$ も出現の重複を考慮した支持度である．$\mathrm{S}_{MIS}(p, g)$ は，出現 o の単位での重複を許さないのに対し，$\mathrm{S}_{MRN}(p, g)$ は，部分グラフ p の頂点単位で重複カウントを許容しない．

(4) 出現数に基づく支持度 [4]

支持度 $\mathrm{S}_{NOO}(p, g)$ は，出現 $o \in O(p, g)$ における辺集合 E_o の種類数であり，以下の様に定義される．

$$\mathrm{S}_{NOO}(p, g) = |\{E_o \mid (V_o, E_o, \lambda_o) \in O(p, g)\}|$$

$\mathrm{S}_{MIS}(p, g)$ や $\mathrm{S}_{MRN}(p, g)$ とは異なり，$\mathrm{S}_{NOO}(p, g)$ は完全な重複以外を許容する．これにより，出現の違いをより明確に表すことが可能であると考えられる．

(6) 頂点数の比に基づく支持度 [4]

支持度 $\mathrm{S}_{NOO}(p,g)$ は，出現の部分的な重複を考慮しないため，一ヶ所だけが異なるような出現を多数持つ部分グラフに対して不当に大きな値を与えてしまう可能性がある．この問題を軽減するために，支持度 $\mathrm{S}_{ROV}(p,g)$ では，出現に含まれる総頂点数を部分グラフの頂点数で除することで，正規化を行っている．

$$\mathrm{S}_{ROV}(p,g) = |\{v \in V_o \mid (V_o, E_o, \lambda_o) \in O(p,g)\}| \,/\, |V_p|$$

出現の重複に対して正規化の考え方を適用することで，重複を許容しない $\mathrm{S}_{MIS}(p,g)$ や $\mathrm{S}_{MRN}(p,g)$ と，重複を許す $\mathrm{S}_{NOO}(p,g)$ との中間的な性質を持つ支持度が実現されることが期待できる．

3 特徴的部屋配置の抽出

本論文では，後述するクラス情報を考慮することで，頻出 δ フリー部分グラフ集合 $\mathcal{F}_D^{\sigma,\delta}$ 中のパターンとその組み合わせを対象に，単一グラフに対する支持度 $\mathrm{S}_M(p,g)$ $(M \in \{BIN, MIS, MRN, NOO, ROV\})$ を考慮した特徴的なパターンの抽出を試みる．

3.1 回帰分析に基づくクラス付与

賃貸物件集合 T に対し，そこから導出される賃料回帰モデル f を用いてクラス付与を行う．賃貸物件 $t \in T$ の賃料を y_t と表記する．また，専有面積や築年数等，物件 t が持つ基本的な条件（属性）の集合を X_t と表記する．このとき，実賃料と回帰モデル f による推定賃料の差 $r_t = y_t - f(X_t)$ を推定残差と呼ぶ．

本論文では，賃貸物件集合 T に対し，推定残差 r_t が降順で上位 $\alpha^{\uparrow}$%以内である物件 t にクラス "↑" を，下位 $\alpha^{\downarrow}$%以内である物件 t にクラス "↓" を付与する．物件 t に付与されたクラスを c_t，T 中のクラス cls に属する物件の集合を $T^{cls} = \{t \in T \mid c_t = cls\}$ とそれぞれ表記する．同様に，物件 t の間取り図グラフを g_t，T^{cls} に属する物件の間取り図グラフの集合を $G_T^{cls} = \{g_t \mid t \in T^{cls}\}$ と表記する．

集合 $T^{\uparrow}$ に含まれる物件は，基本属性に基づく推定賃料より実賃料が高い物件であるのに対し，$T^{\downarrow}$ に含まれる物件は推定賃料よりも実賃料が安い．本論文では，間取りの観点から両者の違いを分析することで，賃料に影響のある部屋配置の抽出を試みる．

3.2 閾値付き対比部分グラフ

本節では，特定のデータベースにのみ頻出するパターンを抽出するため，対比集合 [8] のアイディアを部分グラフパターンと閾値の対へと適用した，閾値付き対比部分グラフを提案する．

グラフ集合 D に対し，パターン p と閾値 τ の対 $\langle p,\tau\rangle$ の支持度を，p の支持度が τ 以上となるグラフの割合

$$\mathrm{S}_M(\langle p,\tau\rangle, D) = |\{g \in D \mid \mathrm{S}_M(p,g) \geq \tau\}| \,/\, |D|$$

と定義する．

物件集合 T から得られるグラフ集合 $G_T = G_T^{\uparrow} \cup G_T^{\downarrow}$ を考える．このとき，支持度に関する閾値 $\rho\,(0 < \rho \leq 1)$ と増加率に関する閾値 $\theta(>1)$ に関し，下記の条件を満たすパターン・閾値対 $\langle p,\tau\rangle$ をクラス $C \in \{\uparrow,\downarrow\}$ に関する閾値付き対比部分グラフと呼ぶ．

$$\begin{cases} \mathrm{S}_M(\langle p,\tau\rangle, G_T^C) \geq \rho \\ \mathrm{S}_M(\langle p,\tau\rangle, G_T^C) > \mathrm{S}_M(\langle p,\tau\rangle, G_T \setminus G_T^C) \times \theta \end{cases}$$

クラス C に関する閾値付き対比部分グラフの全体集合を $\mathcal{C}_C$ と表記する．このとき，$\mathcal{C}_{\uparrow}$ に含まれるパターン・閾値対は，推定残差が大きい物件にのみ見られる特徴を表しており，賃料を増加させる部屋配置の特徴を捉えている可能性が考えられる．逆に $\mathcal{C}_{\downarrow}$ 中の対は，賃料低下の要因を探る手掛かりとして利用することが期待できる．

ところで，あるパターン p に対し，異なる複数の閾値 $\tau_1 \neq \tau_2$ との対 $\langle p,\tau_1\rangle$ と $\langle p,\tau_2\rangle$ が $\mathcal{C}_C$ に含まれる可能性がある．また同様に，パターンのみに着目した場合，包含関係 $p \sqsubset q$ を満たす対 $\langle p,\tau_p\rangle$ と $\langle q,\tau_q\rangle$ が共に $\mathcal{C}_C$ に含まれる可能性がある．これらの冗長な対を排除するため，本論文では，極小パターンに対する最小閾値を持つ対，すなわち条件

$$\begin{cases} \tau^* = \min_{\langle p,\tau\rangle \in \mathcal{C}_C} \tau \\ \forall \langle q,\tau\rangle \in \mathcal{C}_C \ (p \neq q \rightarrow p \not\sqsupset q) \end{cases}$$

を満たす対 $\langle p,\tau^*\rangle$ を最小閾値付き対比極小部分グラフと呼び，抽出・分析の対象とする．

最小閾値付き対比極小部分グラフは，グラフ構造とその出現数を同時に考慮することで，より細かい粒度での特徴的パターンの抽出を実現している．また今回，$\mathrm{S}_M(\langle p,\tau\rangle, D)$ の定義に "$\mathrm{S}_M(p,g) \geq \tau$" という閾値条件を用いているが，"$\mathrm{S}_M(p,g) < \tau$" や "$\tau_1 \leq \mathrm{S}_M(p,g) < \tau_2$" など，異なる条件を利用することも考えられる．

3.3 部分グラフに基づく決定リスト

本節では，クラス情報と頻出 δ フリー部分グラフ集合を用いて分類問題を構築し，機械学習の解釈性研究 [5] の成果を援用することで，複数のパターンから構成されるルールの抽出を試みる．具体的には，物件集合

T に対し，推定残差に基づくクラス値を目的変数，各パターンの出現数を説明変数とするデータベース

$$DB_{T,M} = \{\langle\{\mathrm{S}_M(p,t)\}, g_t\rangle \mid t \in T^{\uparrow} \cup T^{\downarrow}, p \in \mathcal{F}^{\sigma,\delta}_{\{g_t \mid t \in T\}}\}$$

を構築し，inTrees[9] を適用する．

inTrees は，ランダムフォレスト [10] などの分類木アンサンブルから決定リストを導出する手法である．決定木中の根から葉までの各パスを一つのルールとし，そこに現れる条件（パターン）を集計すると共に不要な条件の削除と複数ルールの集約を行う．その上で，集合被覆アルゴリズム [11] に基づき少数のルールを選択する．詳細は原著論文 [9] を参照されたい．

inTrees により $DB_{T,M}$ から得られる各ルールは

$$\mathrm{S}_M(p_1,t)\Diamond\tau_1, \wedge\cdots\wedge \mathrm{S}_M(p_n,t)\Diamond\tau_n \to \widehat{c_t} \quad (\Diamond \in \{\leq, >\})$$

の形式をしており，クラス分類に関して複数のパターンとその出現数に関する情報を含む．これらのルールを通じ，複数の条件を考慮した上で部屋配置の賃料に対する影響を考察することが可能となると考えられる．

4 実験

4.1 データセットと実験設定

実験には，株式会社 LIFULL が国立情報学研究所の協力により研究目的で提供している「LIFULL HOME'S データセット [1]」を利用した．具体的には，文献 [4] と同様，間取り図を確認しにくい物件や二階建て構造を持つ物件等を除いた東京 23 区内の 2 部屋マンション 400 件および 3 部屋マンション 561 件の合計 961 件を選定している．

基本的な属性として "徒歩距離"，"部屋面積"，"部屋階数"，"部屋の向き"，"部屋数"，"最寄り駅単位で集計した相場" の 6 属性を用い，XGBoost（以降 XB）[12]，Random Forest（以降 RF）[10]，LightGBM（以降 LG）[13] を用いて，それぞれ賃料推定モデル f を構築した．各モデルから得られる推定誤差の分布及び平均絶対誤差を表 1 に示す．なお今回の実験では，パラメタ $\alpha^{\uparrow} = \alpha^{\downarrow} = 25\%$を用いてクラス設定を行った．

物件 t に対する間取り図グラフ g_t は，文献 [14] で準備・使用されたデータを援用した．今回の実験では，居室やキッチン，リビング，収納など 11 種の頂点ラベルを採用し，辺ラベルは省略した．このデータに対し，パラメタ $\sigma = 0.1$，$\delta = 0.05$ を用い，11,770 件からなる頻出 δ フリー部分グラフの集合 $\mathcal{F}^{\sigma,\delta}_D$ を導出し，分析対象とした．得られたパターンの分布を表 2 に示す．

[1] https://www.nii.ac.jp/dsc/idr/lifull/homes.html

表 1: 推定誤差の分布

	XB	RF	LG
最小値	-44,762	-363,570	-729,876
第 1 四分位数	-5,146	-7,354	-4,658
中央値	-660	-842	-193
平均値	34	168	0
第 3 四分位数	4,394	6,296	3,646
最大値	63,631	285,806	401,582
平均絶対誤差	7,452	13,695	13,691

表 2: 頻出 δ フリー部分グラフの分布

	支持度	辺数	頂点数
最小値	10%	0	1
第 1 四分位数	11%	7	8
中央値	14%	8	9
平均値	16%	8.4	9.3
第 3 四分位数	18%	10	11
最大値	100%	16	17

4.2 閾値付き対比部分グラフの抽出

第一の実験として，パラメタ $\rho = 10/240$，$\theta = 5$ を用い，$\mathcal{F}^{\sigma,\delta}_D$ から最小閾値付き対比極小部分グラフを導出した．得られたパターン数を表 3 に示す．表における "非極小" 列は，閾値に関しては最小であるが部分グラフは極小ではない閾値付き対比部分グラフの数を表す．

結果より，支持度 S_{BIN} と S_{MRN} を用いた場合は少数の対比パターンしか得られず，S_{NOO} や S_{ROV} を用いた場合は多くのパターンが獲得できることが確認できる．これは S_{BIN} や S_{MRN} は，出現の重複に対する制限が強く，各間取り図グラフが持つ支持度のバリエーションが小さいことに起因すると推察される．このことから，S_{NOO} や S_{ROV} などの重複に寛容な支持度の利用が，分析において効果的であることが確認できる．

一方実験結果より，非極小な対比パターンに対し，極小なパターンの数が大幅に少ないことが確認できる．このことから，抽出対象を極小パターンに限定することは，分析コストの軽減に貢献するという点でも有益であると考えられる．

得られた推定残差の違いに着目すると，XB と RF ではクラス ↓ よりもクラス ↑ に関する対比パターンが多く，LG では逆にクラス ↑ よりもクラス ↓ に関する対比パターンが多く抽出されている．このことから，得られる対比パターンが，賃料推定に用いる回帰モデルの種類や精度，クラス設定に用いるパラメタに大きく

表 3: 閾値付き対比部分グラフ数

S_M	f	↑	↓	合計	非極小
BIN	XB	2	0	2	0
	RF	12	0	12	11
	LG	0	0	0	0
MIS	XB	21	2	23	771
	RF	71	2	73	1,652
	LG	19	40	59	1,162
MRN	XB	2	1	3	0
	RF	12	1	13	11
	LG	0	1	1	0
NOO	XB	24	2	26	777
	RF	76	2	78	1,678
	LG	19	40	59	1,169
ROV	XB	49	3	52	655
	RF	68	1	69	1,371
	LG	22	50	72	1,087

左右されることが伺える．今後は，これらの影響について理論的・実験的に検証を行う必要がある．

4.3 部分グラフに基づく決定リストの抽出

データベース $DB_{T,M}$ に対するランダムフォレストモデル（RF モデル）の精度及び inTrees によって得られたルール数と精度を表 4 に示す．なお RF モデル及び inTrees によるルールの獲得には，R 言語上の実装を利用した．また，RF モデル構築にはデフォルトパラメタを用い，inTrees によるルール獲得では，利用する分類木数を（RF モデルのデフォルト値である）500，ルール抽出のための最小支持度を 0.05 と設定している．

結果より，RF モデルの精度は必ずしも高くはものの，inTrees により得られたルールは少数ながらも RF モデルと同等以上の精度を達成していることが分かる．また，利用する支持度 S_M やクラス設定のための賃料推定モデル f の間で，大きな違いは確認できず，その意味で，安定した結果が得られていると判断できる．一方，得られたルール数に関しては，クラス分布が均等であることもあり，↑クラスに対するルール数が平均 6.0，↓クラスに対するルール数が平均 5.1 となり，大きな差は確認できない．

得られたルールの例を図 1 に示す．図中において各 p_x はパターン p_x の支持度 $S_{ROV}(p_x, t)$ を表す．結果より，多様なパターン及び閾値を用いたルールが構成されていることが分かる．また，一つのルールを構成するために最大 5 つのパターンと閾値の対が利用されており，組合せを用いた特徴的なパターンが抽出されていることが確認できる．

表 4: inTrees の結果

		RF の精度		ルール数		ルール精度	
S_M	f	↑	↓	↑	↓	↑	↓
BIN	XB	0.63	0.62	7	6	0.75	0.75
	RF	0.70	0.65	6	6	0.77	0.80
	LG	0.64	0.57	6	4	0.70	0.77
MIS	XB	0.60	0.62	7	6	0.81	0.71
	RF	0.69	0.64	6	6	0.77	0.81
	LG	0.65	0.59	5	5	0.68	0.79
MRN	XB	0.65	0.63	7	7	0.78	0.77
	RF	0.71	0.65	5	2	0.73	0.82
	LG	0.65	0.58	6	5	0.74	0.75
NOO	XB	0.65	0.60	6	4	0.69	0.77
	RF	0.71	0.65	5	5	0.74	0.77
	LG	0.67	0.57	7	6	0.74	0.75
ROV	XB	0.65	0.61	6	5	0.79	0.78
	RF	0.70	0.65	5	4	0.83	0.73
	LG	0.63	0.59	6	5	0.64	0.84

5 まとめと今後の課題

本論文では，賃料に強い関係を与える部屋配置やその組み合わせを抽出することを目的に，従来の対比パターンの拡張である閾値付き対比部分グラフを提案すると共に，部分グラフパターンとその出現数に関する決定リストの導出を行った．また初期的な実験により，提案手法により注目すべき部屋配置が獲得できる可能性を示した．

今後の課題としては，得られた結果の詳細な分析があげられる．また，より効果的に賃料への影響が強い部屋配置を特定するため，傾向スコア分析 [16] や回帰木アンサンブルからのルール抽出手法 [17] 等を併用することを予定している．

謝辞： 本研究では，株式会社 LIFULL が国立情報学研究所の協力により研究目的で提供している「LIFULL HOME'S データセット」を利用した．本研究の一部は，JSPS 科研費 17K00315 の助成を受けたものです．

参考文献

[1] 花里 俊廣，平野 雄介，佐々木 誠：首都圏で供給される民間分譲マンション $100m^2$ 超住戸の隣接グラ

IF	$p_{582} > 0.50 \ \wedge \ p_{2808} \leq 1.28 \ \wedge \ p_{8636} \leq 1.08$ THEN class $=\downarrow$
ELSE IF	$p_{2890} > 0.50 \ \wedge \ p_{7381} \leq 1.50$ THEN class $=\uparrow$
ELSE IF	$p_{1090} \leq 1.64 \ \wedge \ p_{4412} \leq 1.63 \ \wedge \ p_{5489} \leq 1.06 \ \wedge \ p_{7518} > 0.5 \ \wedge \ p_{10363} \leq 1.35$ THEN class $=\uparrow$
ELSE IF	$p_{760} \leq 0.50 \ \wedge \ p_{1279} \leq 1.42 \ \wedge \ p_{2888} \leq 1.05 \ \wedge \ p_{5708} > 1.05 \ \wedge \ p_{7511} \leq 1.28$ THEN class $=\downarrow$
ELSE IF	$p_{149} > 0.50 \ \wedge \ p_{7172} \leq 1.17 \ \wedge \ p_{11148} \leq 0.50$ THEN class $=\downarrow$
ELSE IF	$p_{420} > 0.50 \ \wedge \ p_{2994} \leq 1.08$ THEN class $=\uparrow$
ELSE IF	$p_{321} \leq 1.21 \ \wedge \ p_{2593} \leq 0.50 \ \wedge \ p_{3942} \leq 0.50$ THEN class $=\downarrow$
ELSE IF	$p_{2025} \leq 0.50 \ \wedge \ p_{2989} \leq 1.13$ THEN class $=\uparrow$
ELSE	class $=\uparrow$

図 1: 得られた決定リストの例：S_{ROV}，f=RF

フによる分析，日本建築学会計画系論文集 (591), pp.9–16, 2005.

[2] 瀧澤 重志，吉田 一馬，加藤 直樹： グラフマイニングを用いた室配置を考慮した賃料分析 ： 京都市郊外の 3LDK を中心とした賃貸マンションを対象として， 日本建築学会環境系論文集 (623), pp.139–146, 2008.

[3] 尾崎 知伸，小黒 淳斗：頻出部分グラフを用いた賃料分析，人工知能学会第 111 回知識ベースシステム研究会，pp.13–16, 2017.

[4] 長谷川 優也，尾崎 知伸：部屋配置とその出現数に着目した二段階賃料推定， 人工知能学会 第 21 回インタラクティブ情報アクセスと可視化マイニング研究会，pp.80–87, 2019.

[5] 原 聡：私のブックマーク，機械学習の解釈性，人工知能，Vol.33, No.3, pp.366–369, 2018.

[6] M. Kuramochi and G. Karypis : Finding frequent patterns in a large sparse graph, *Data Mining and Knowledge Discovery*, Vol.11, No.3, pp.243–271, 2005.

[7] B. Bringmann and S. Nijssen : What is Frequent in a Single Graph? *Proc.* of the 5th International Workshop on Mining and Learning with Graphs, pp.1–4, 2007.

[8] S. D. Bay and M. J. Pazzani : Detecting Change in Categorical Data: Mining Contrast Sets, *Proc.* of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.302–306, 1999.

[9] H. Deng : Interpreting Tree Ensembles with inTrees, International Journal of Data Science and Analytics, 2018.

[10] L. Breiman : Random Forests, *Machine Learning*, Vol.45, Issue 1, pp.5–32, 2001.

[11] J. Fürnkranz : Separate-and-Conquer Rule Learning, *Artificial Intelligence Review*, Vol.13, No.1, pp.3–54, 1999.

[12] T. Chen and C. Guestrin : XGBoost: A Scalable Tree Boosting System, *Proc.* of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785–794 , 2016.

[13] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu : LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Advances in Neural Information Processing Systems 30, 2017.

[14] K. Kamihori, T. Shimano, A. Oguro and T. Ozaki: Towards discovery of human's recognition mechanisms for complex structured images, Proc. of the Third International Workshop on Skill Science, pp.26–27, 2016.

[15] X. Yan and J. Han : gSpan : Graph-based Substructure Pattern Mining, *Proc.* of the 2002 IEEE International Conference on Data Mining, pp.721–724, 2002.

[16] 星野 崇宏： 調査観察データの統計科学一因果推論・選択バイアス・データ融合，岩波書店，2009.

[17] S. Hara and K. Hayashi : Making Tree Ensembles Interpretable: A Bayesian Model Selection Approach *Proc.* of the 21st International Conference on Artificial Intelligence and Statistics, pp.77–85, 2018.