

間取り図を用いた賃料予測モデルに関する一検討

服部 凌典¹ 岡本 一志¹ 柴田 淳司²

¹ 電気通信大学 大学院情報理工学研究科 情報学専攻

² 産業技術大学院大学 産業技術研究科 情報アーキテクチャ専攻

はじめに

大野喜久之輔：継続賃料鑑定評価を再考する,
都市住宅学, 住宅新報社, 2016.

賃貸物件の特徴

- ・ 同じ物件が存在しない
- ・ 間取り, 階数, 立地などの物件変数が賃料に影響を与える

賃貸物件の価格決定方法 [大野喜, 2016]

- ・ 積算法
- ・ 収益分析法
- ・ 賃貸事例比較法
 - ・ 対象物件の周囲の物件や類似する物件を参照して賃料を決定
 - ・ 決定支援にヘドニック・アプローチを採用

唐渡 広志：ヘドニック・アプローチを利用した不動産価格
指數の推定方法とその問題点, 都市住宅学, vol.2016, no.
92, pp.17-20, 2016.

ヘドニック・アプローチ

ある商品の価格をその商品の変数の価値に関する集合とし,
その商品価格の予測モデルを線形回帰で構築する技術[唐渡, 2016]

$$\hat{y}_i = \alpha_1 \underline{x_1} + \alpha_2 \underline{x_2} + \alpha_3 \underline{x_3} + \cdots + \alpha_n \underline{x_n} + \beta$$

賃料 占有面積 立地 階数 築年数

既存のヘドニック・アプローチは間取り図が考慮されていない

ヘドニック・アプローチを用いた関連研究

説明変数として使用されている変数のカテゴリ

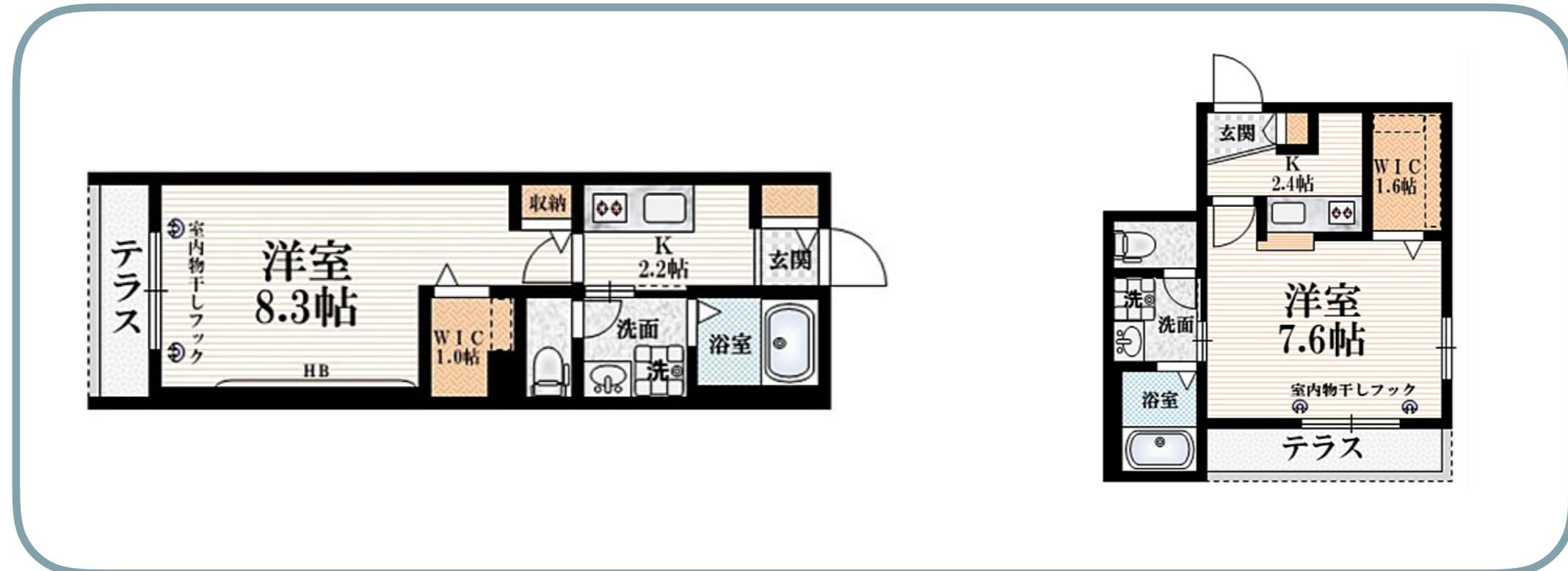
	建物構造	立地 /アクセス	周辺環境	売買契約	画像
Ram+ 2019	○	○	○	✗	✗
Philipp+ 2016	○	○	○	✗	✗
Jun+ 2017	○	○	○	○	✗
本研究	○	○	✗	○	○

- Ram P. Dahal and Robert K. Grala and Jason S. Gordon and Ian A. Munn and Daniel R. Petrolia and J. Reid Cummings :
A hedonic pricing method to estimate the value of waterfronts in the Gulf of Mexico, Urban Forestry & Urban Greening, vol.41, pp. 184-194, 2019.
- Dr. Philipp Deschermeier and Björn Seipelt :
A Hedonic Rent Index for Student Housing in Germany, Cologne Institute for economic research, pp.1–12, 2016.
- yung-Jin Jun, Hee-Jae Kim : Measuring the effect of greenbelt proximity on apartment rents in Seoul, Cities, vol. 62, 2017.

間取り図の特性

清田陽司, 山崎俊彦, 諏訪博彦, 清水弘：
エディトリアル：「不動産とAI」，人工知能,
vol.32, 2017.

間取り規格が同じであっても、部屋の形状が違うものが存在
希望の賃貸物件を探す際には間取り図を見る習慣がある[清田+, 2017]



同じ物件の同じ間取り規格で間取り図が違う例

研究内容

目的: 賃料予測における間取り図の影響を明らかにする

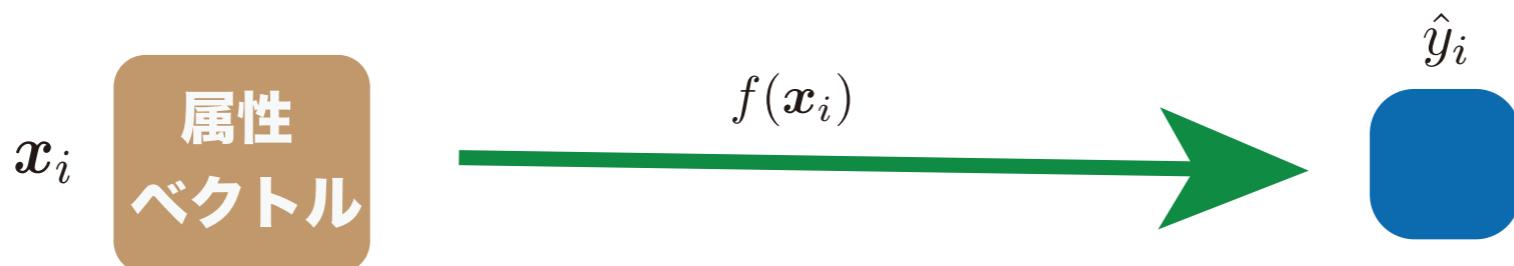
賃料の予測精度を比較

- 間取り図を考慮しない予測モデル (LR)
- 間取り図を考慮した予測モデル
 - A. 線形変換による特徴量抽出法 (PCA-LR)
 - B. 非線形変換による特徴抽出法 (VGG-LR)

LR (Linear-Regression)

一般的なヘドニック・アプローチ

- 予測式： $\hat{y}_i = f(\mathbf{x}_i) = \boldsymbol{\alpha}^T \mathbf{x}_i + \beta$
- 説明変数に間取り図を含まない
- scikit-learn (version : 0.20.2) のLinearRegressionクラス使用

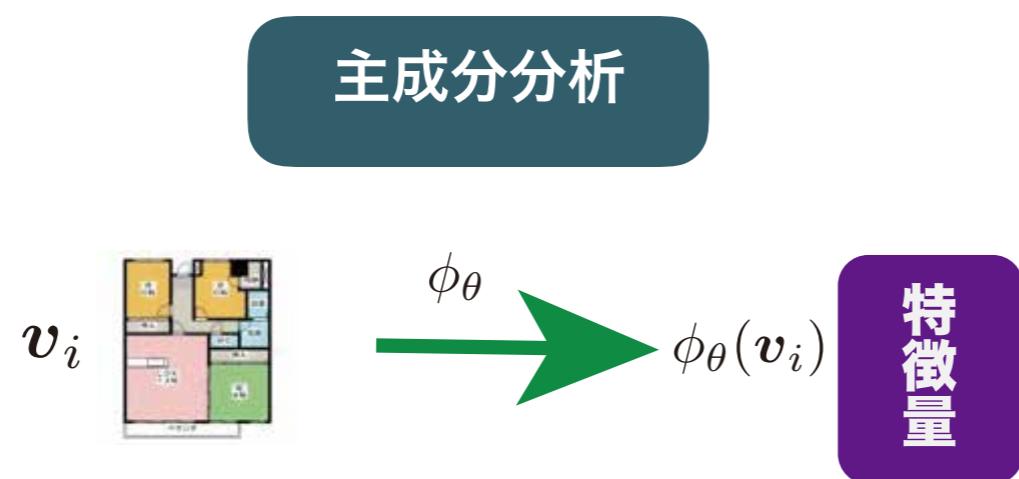


このLRと間取り図を考慮したPCA-LRとVGG-LRを比較する

PCA (Principal Component Analysis) -LR

LRの説明変数に間取り図を追加したモデル

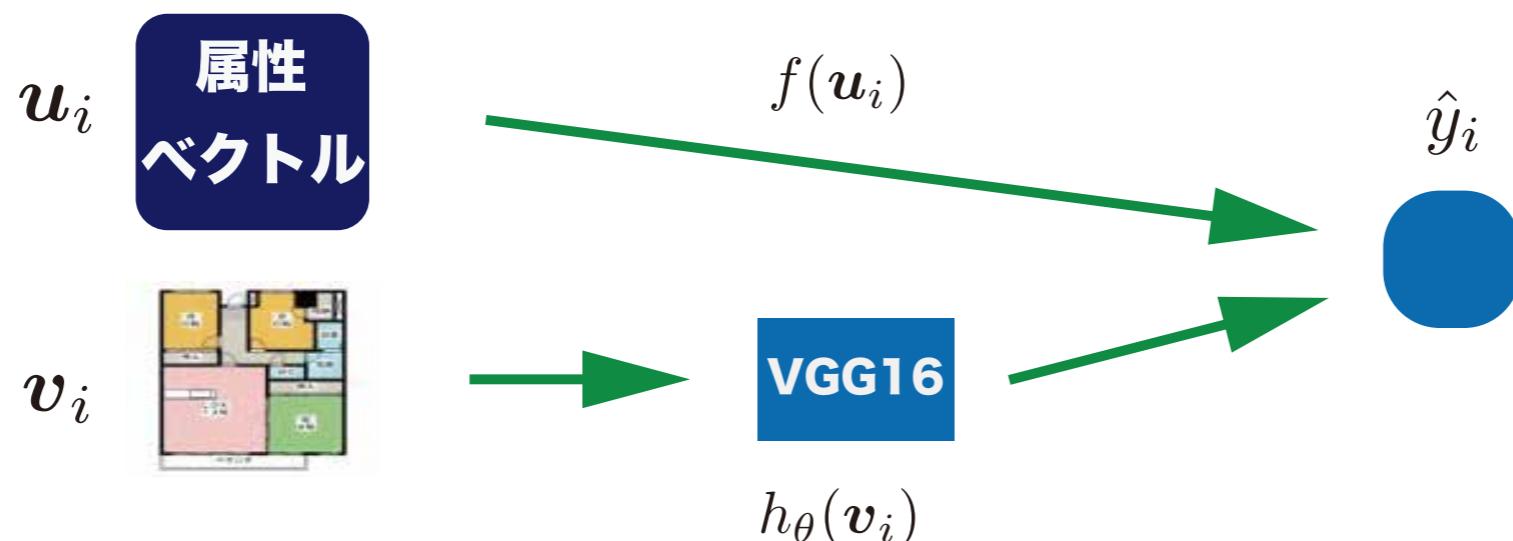
- 主成分分析による特徴抽出
- 抽出した間取り図の特徴量と
変数ベクトルを組み合わせ、線形回帰で賃料予測
- 間取り図の特徴量: 64, 128, 256, 512, 1024,
2048次元の6種類
- scikit-learn (version : 0.20.2)
のPCAクラス使用
- $\mathbf{x} = [\mathbf{u}, f_{\theta}(\mathbf{v})]$



VGG-LR

間取り図を非線変換で特徴抽出するモデル

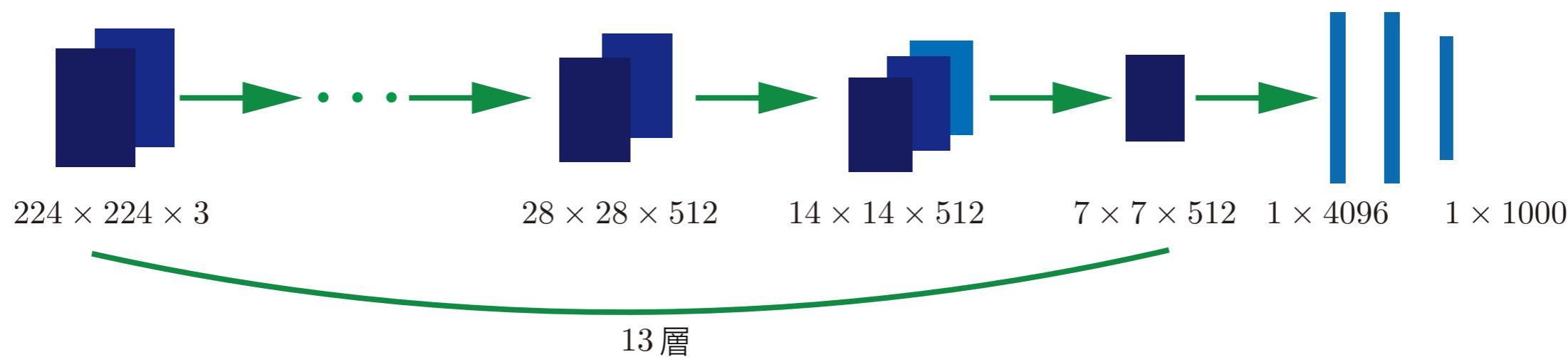
- ミニバッチ + 確率的勾配降下法を適用
- batch size: 100, epoch: 100
- 間取り図の特徴抽出: 64次元
- 活性化関数: ReLU
- Keras (version : 2.2.4) のFunction API
- $\mathbf{x} = [\mathbf{u}, h_{\theta}(\mathbf{v})]$



VGG16の構造

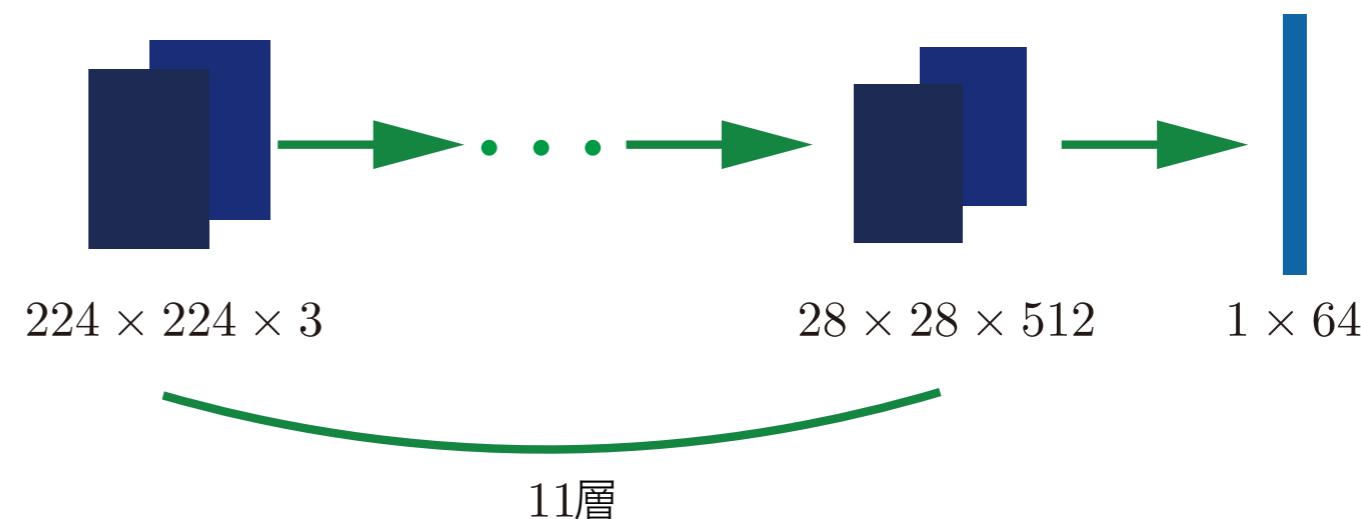
Karen Simonyan, and Andrew Zisserman : Very Deep Convolutional Networks for Large-Scale Image Recognition, Computing Research Repository, vol.abs/1409.1556, 2014.

VGG16 [Karen+, 2014]: オックスフォード大学のVGGチームによって開発された学習済みのニューラルネットワーク



使用するVGG16

- ・後ろの5層を除去
- ・64次元の全結合層追加
- ・後ろ3層のみ学習



実験環境

LIFULL HOME'Sデータセット（2015年9月時点）

- ・ 賃貸物件データ（70変数、533万件）
- ・ 120×120ピクセルの画像データ（8,300万ファイル）

使用する計算機（OS: Ubuntu 18.04.1）

- ・ CPU : Intel(R) Xeon(R) CPU E5-2650 v3 @ 2.30GHz
- ・ GPU : NVIDIA Corporation GM107GL [Quadro K2200] 4GB
NVIDIA Corporation GP102 [GeForce GTX 1080 Ti] 11GB
- ・ メモリ : 64GB

データセット

使用賃貸物件データ (14変数, 9万件)

- ・ 東京都のデータを使用 (\because 47都道府県で最もデータ量が多いため)
- ・ 目的変数：賃料 + 共益費
- ・ 説明変数：徒歩距離, 立地, 階数（地上）, 階数（地下）, 市町村区, 部屋階数, 契約期間, 建物構造, 駐車場料金, 新築・未入居フラグ, 占有面積, 築年数, 間取り図
- ・ 間取り図：グレースケール変換

データセット分割

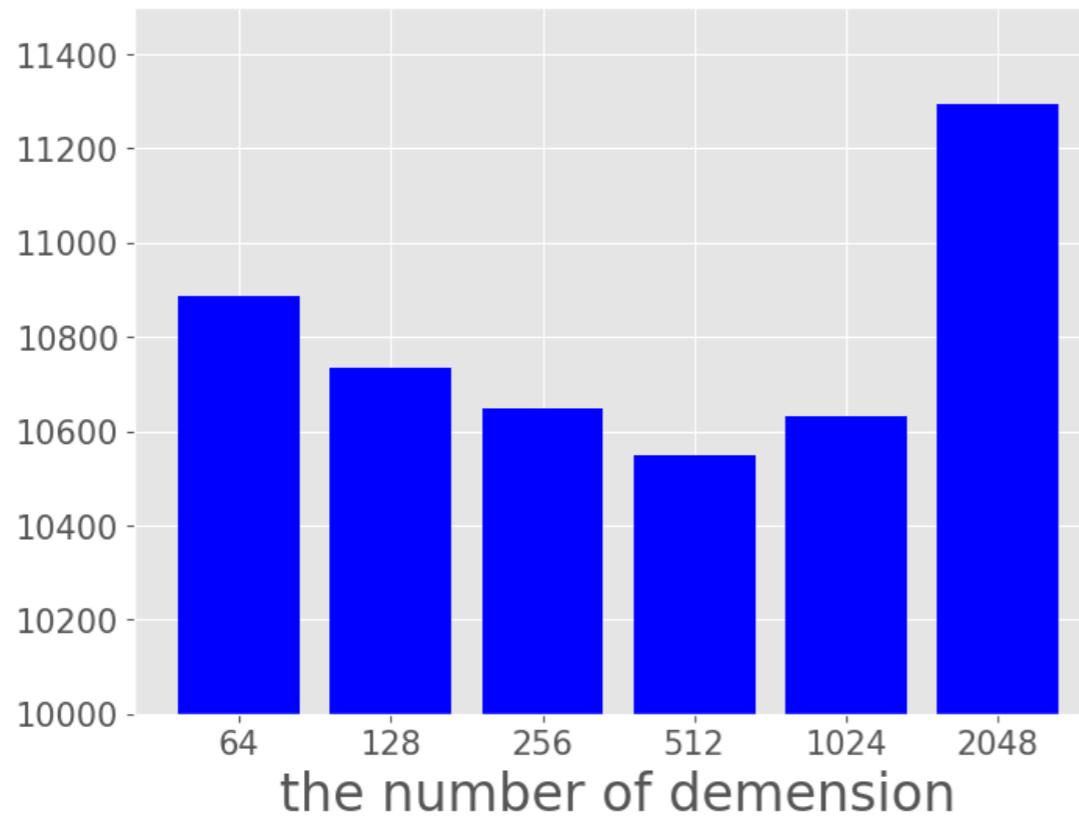
- ・ 使用賃貸物件データを間取り規格を基準にK, R, DK, LD, LDKの5種類に分割
- ・ LDはデータ数が35件のため使用しない
- ・ 20%をテストデータに, 80%を開発用データに分割

評価方法

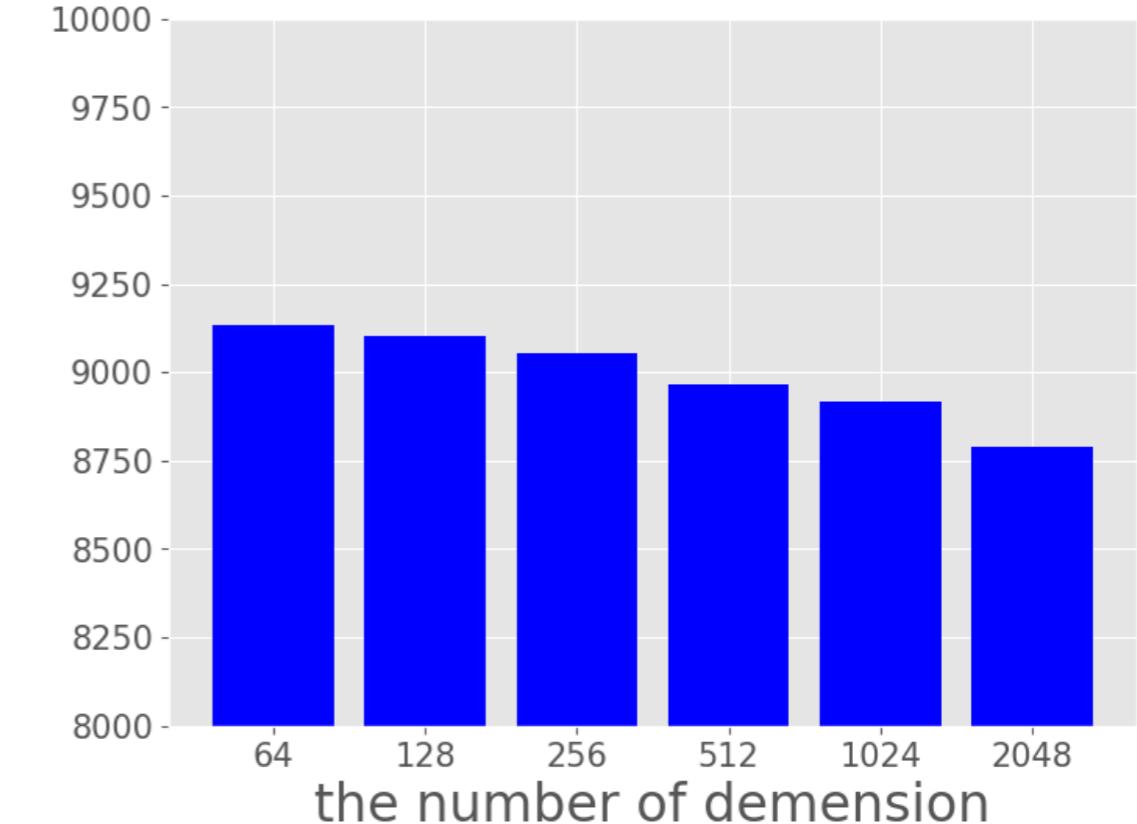
評価指標

- ・ 平方平均二乗誤差 (RMSE: Root Mean Square Error)
 - ・ ハイパーパラメータチューニング
 - ・ PCA-LRにおける適切な間取り図の次元数を探索
 - ・ 64, 128, 256, 512, 1024, 2048次元の6種類
 - ・ 開発用データに10-fold cross-validationを適用
 - ・ テストデータに対するLRとVGG-LR, PCA-LRの評価
- ・ 予測誤差の分析
 - ・ テストデータに対するLRとVGG-LR, PCA-LRの評価
 - ・ 累積密度関数
 - ・ 予測値の可視化

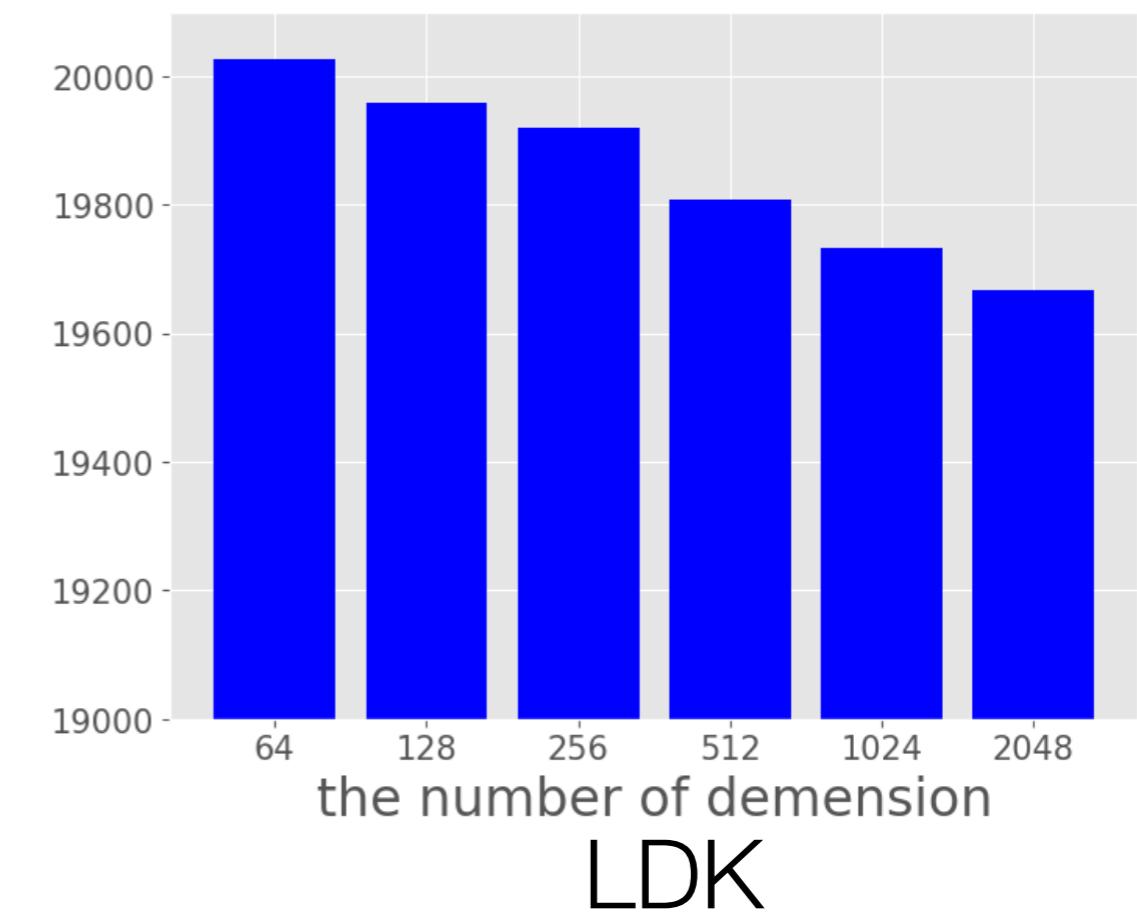
PCA-LRにおける適切な間取り図の次元数の探索結果



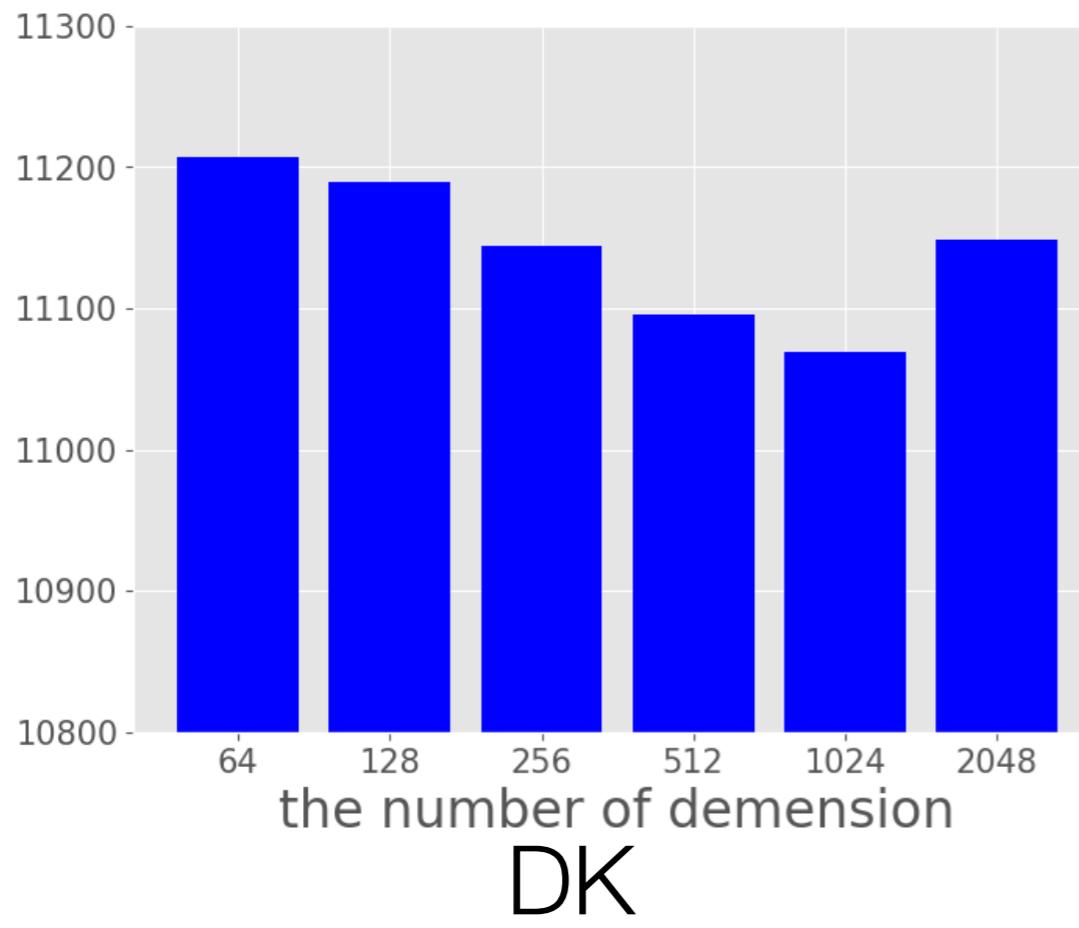
K



R



LDK



DK

RMSEによる評価と学習時間

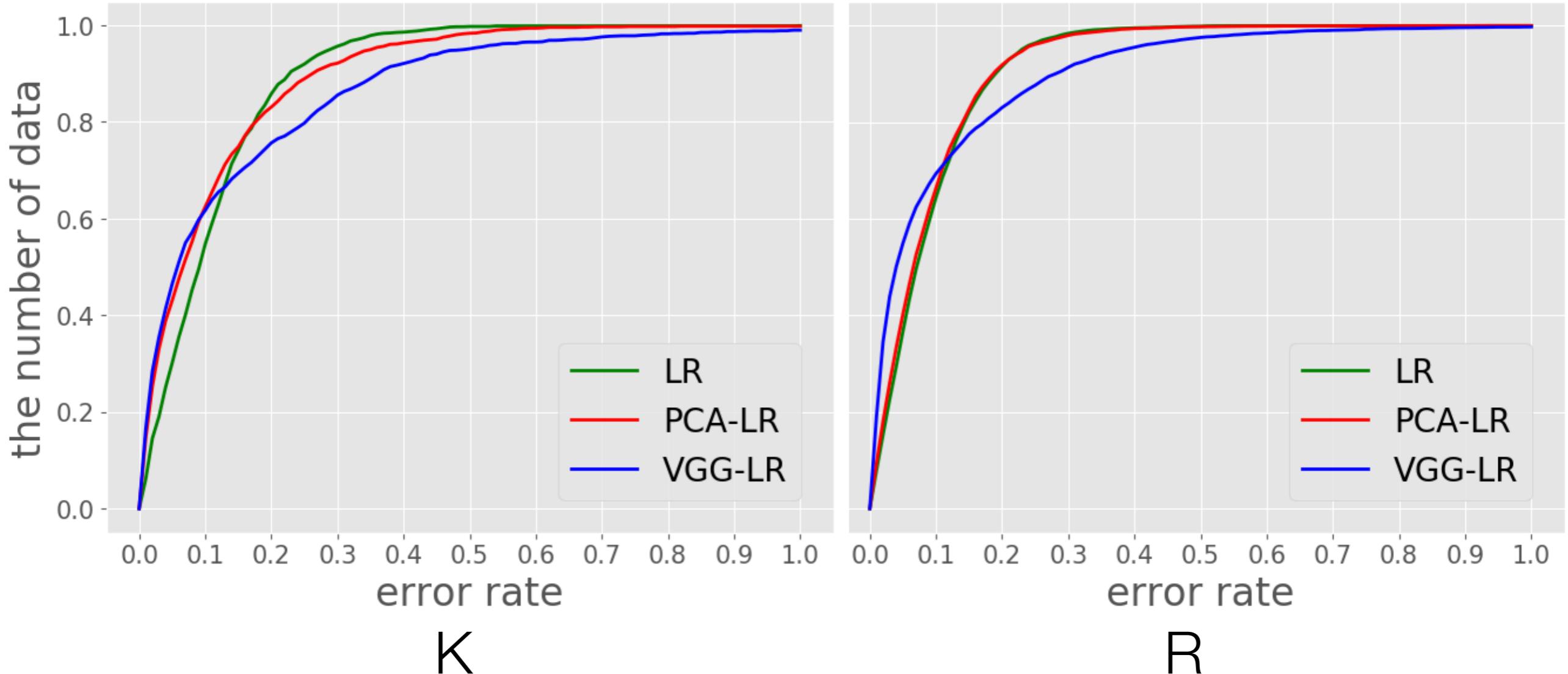
テストデータに対するRMSE[円]

	データ数	LR	PCA-LR	VGG-LR
K	1,712	11,602	11,689	16,862
R	7,525	9,217	8,727	12,541
DK	3,548	11,156	10,957	16,467
LDK	6,842	20,121	19,646	22,878

開発用データの学習時間[秒]

	データ数	LR	PCA-LR	VGG-LR
K	9,205	0.06	11	2,897
R	39,985	0.08	48	14,566
DK	18,975	0.05	23	6,471
LDK	36,595	0.07	45	13,026

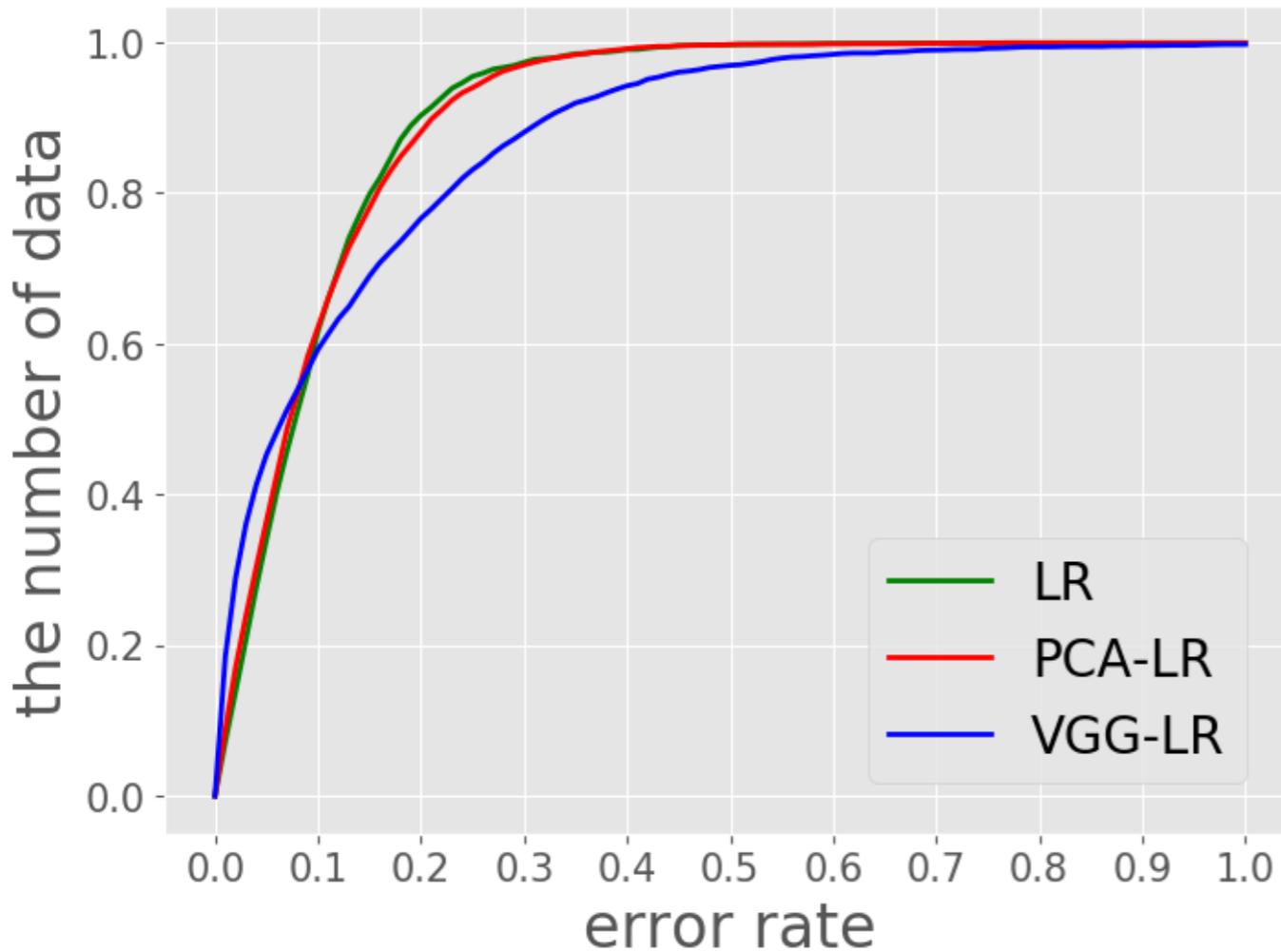
累積密度関数 (1/2)



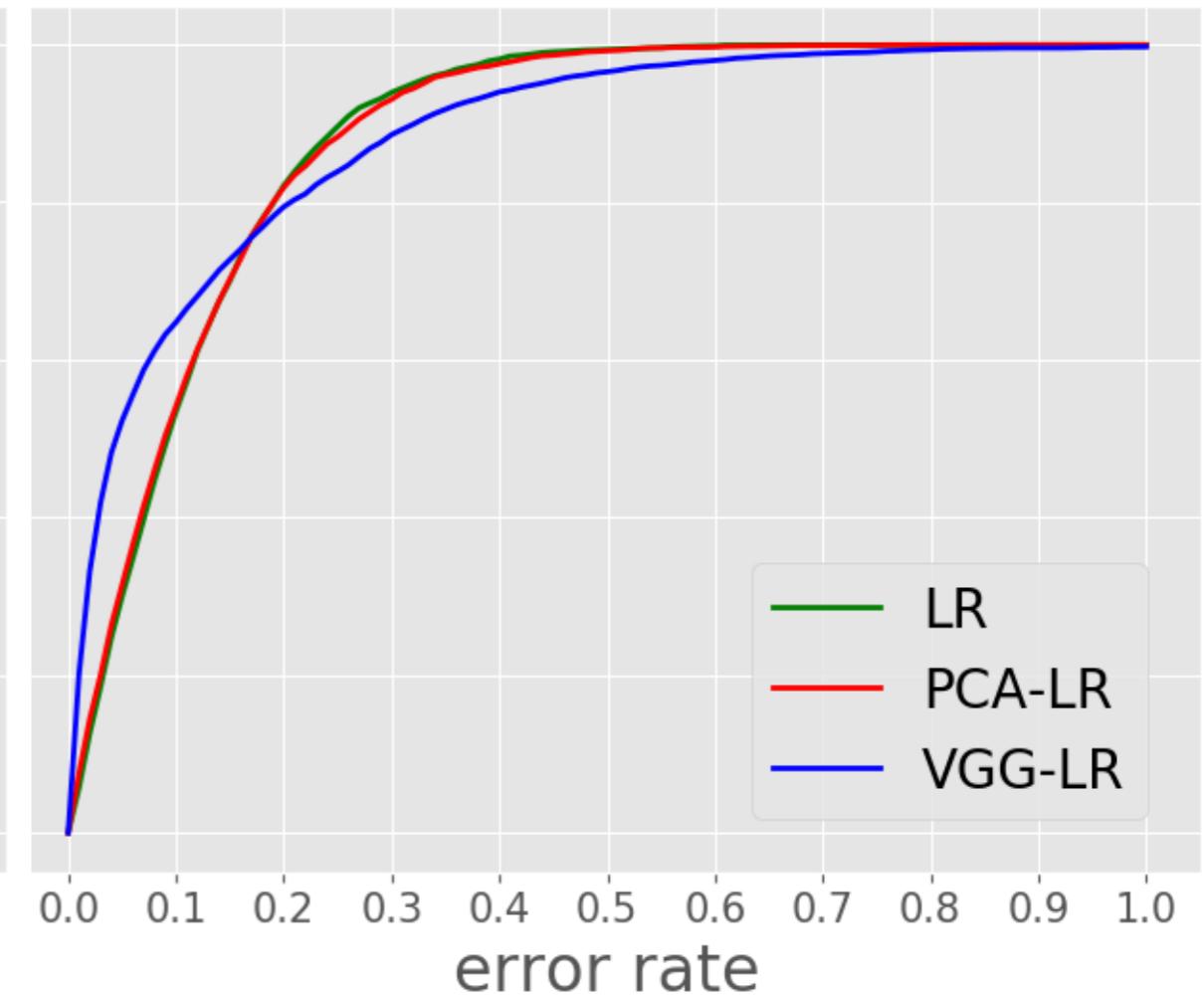
横軸 : error rate = $1 - \frac{\hat{y}}{y}$ y : 真値 \hat{y} : 予測値

縦軸 : errors rate 以下の誤差に収まっているデータ数の割合

累積密度関数 (2/2)

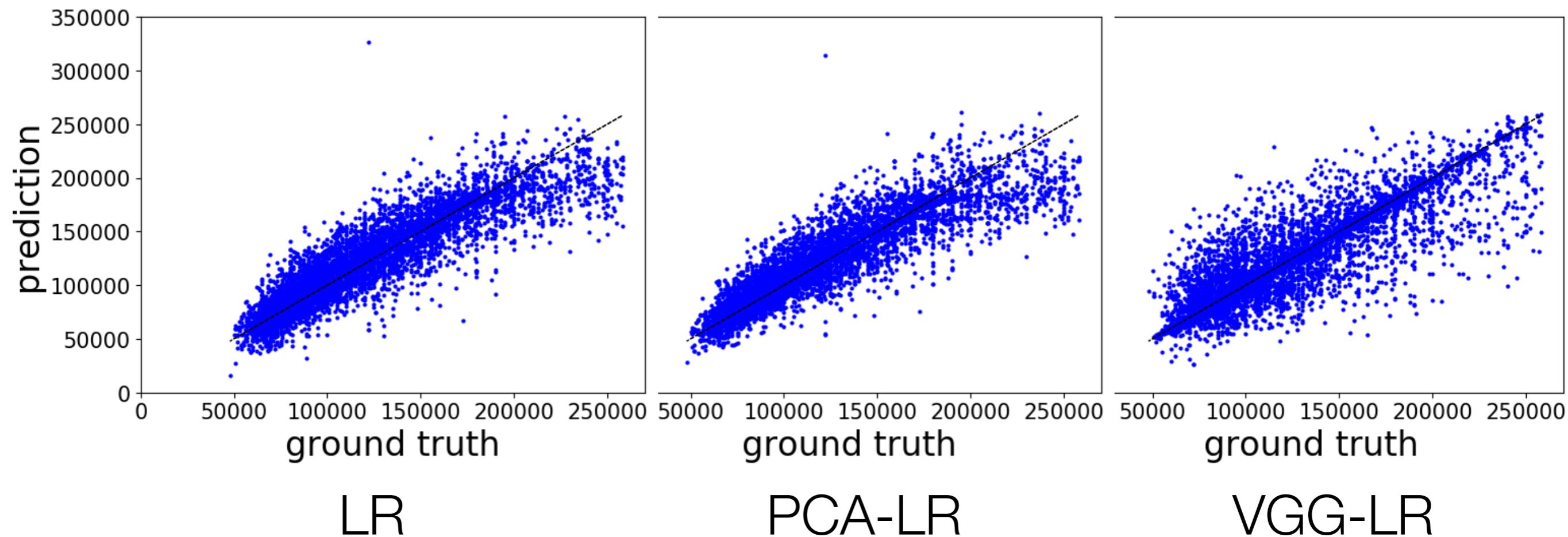


DK



LDK

LDKのテストデータに対する予測値の可視化



LIFULL HOME'Sデータセット外の物件予測

	間取り図	構造	立地	賃料	LR	PCA-LR	VGG-LR
研究室 メンバ 宅		木造	調布	74,000	66,695	68,701	58,561
自宅		RC	調布	65,000	61,361	62,545	83,695
指導 教員宅		鉄筋	墨田区	128,000	130,988	122,887	181,913
学生寮		RC	調布	47,700	56,931	49,716	64,526

おわりに

目的：賃料予測における間取り図の影響を明らかにする

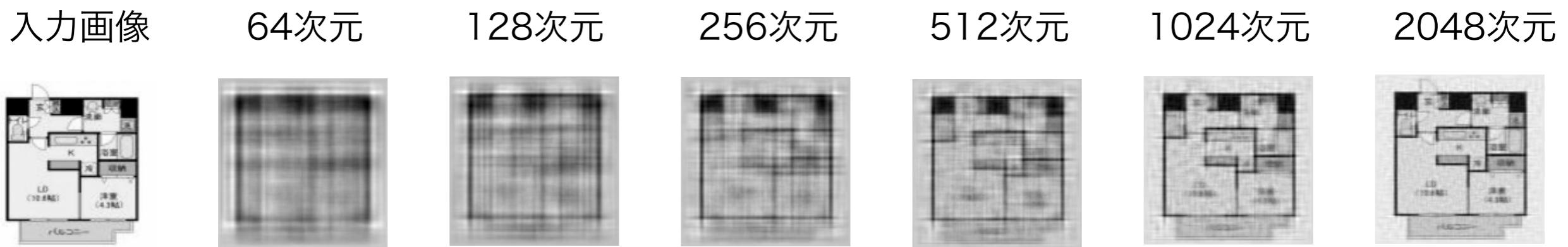
実験結果

間取り図を考慮することで予測精度があがる

- ・ 間取り図の特徴抽出に主成分分析を用いることでRMSEが向上する
 - ・ 主成分分析を使う際、その特徴量は1024, 2048次元の高次元が望ましい
 - ・ 1024, 2048次元の間取り図の特徴量で間取り図の持つ大部分の特徴量を抽出できる
- ・ 間取り図の特徴抽出にVGGを用いることで、的確に当てる物件が増える
 - ・ 予測精度のばらつきが大きくなる
 - ・ 計算コストが課題

PCAによる次元別の間取り図の特徴量

検証データ（LDK）に対する各次元数からの復元図



1024次元の特徴量で元の画像を認識できる程度に復元できている

テストデータに対する価格帯別予測誤差 (1/2)

間取りタイプ: K

	[0円, 5万円)	[5万円, 10万円)	[10万円, 15万円)	[15万円, 20万円)	[20万円, 25万円)	25万円以上						
データ数	279	954	421	54	3	1						
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値		
VGG-LR	11,434	5,917	9,093	4,417	9,375	3,130	22,311	7,785	53,707	23,504	90,673	90,673
PCA-LR	6,627	5,456	12,320	5,345	8,516	5,432	14,575	12,264	52,582	66,645	51,985	51,985
LR	6,106	5,348	12,920	6,993	10,317	8,147	18,045	17,847	54,364	61,570	49,106	49,106

間取りタイプ: R

	[0円, 5万円)	[5万円, 10万円)	[10万円, 15万円)	[15万円, 20万円)	[20万円, 25万円)	25万円以上						
データ数	1,073	5,436	945	65	6	0						
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値		
VGG-LR	6,839	2,317	6,677	3009	9,020	2,694	22,782	6,666	43,951	10,729	-	-
PCA-LR	5,140	3,935	5,863	4,577	8,508	6,256	24,556	22,291	54,623	56,182	-	-
LR	4,959	4,118	6,059	4,939	10,060	8,997	31,256	33,133	76,081	77,904	-	-

テストデータに対する価格帯別予測誤差 (2/2)

間取りタイプ: DK

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
データ数	91	2,598	760	98	1	0						
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG-LR	10,033	2,061	8,688	4,825	13,626	7,203	19,647	5,157	12,665	12,665	-	-
PCA-LR	6,377	4,690	7,398	5,719	9,708	7,987	15,883	14,587	69,914	69,914	-	-
LR	6,456	4,893	7,254	5,699	10,331	8,647	19,445	20,853	75,228	75,228	-	-

間取りタイプ: LDK

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
データ数	2	2,735	2,070	1,555	450	30						
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG-LR	47,045	47,045	11,404	3,853	14,858	7,953	12,641	3,870	23,011	7,323	47,122	47,218
PCA-LR	32,216	32,216	11,044	8,630	15,048	11,665	12,336	7,985	28,294	27,003	58,969	55,374
LR	20,233	20,233	10,219	8,163	15,133	11,925	13,705	9,635	33,436	33,540	61,898	65,562

PCA-LRのチューニング結果

RMSE	64	128	256	512	1024	2048
K	10,886	10,733	10,649	10,548	10,630	11,293
R	9,134	9,104	9,054	8,967	8,916	8,790
DK	11,207	11,189	11,144	11,096	11,069	11,148
LDK	20,027	19,958	19,919	19,807	19,733	19,668

LDKにおけるVGG16とPCA-LRで予測精度が高い 物件事例

テストデータ数の20.7%が両方のモデルで高い予測精度を示している

	間取り図	賃料	建物構造	築年数	部屋階数	立地
事例 1		168,000	RC	0	13	台東区
事例 2		78,000	軽量鉄骨	8	10	町田市
事例 3		237,000	SRC	17	4	千代田区

誤差密度による評価 (3/3)

各間取り規格のテストデータに対して予測誤差が真値の 5 %以下である割合[%]

	LR	PCA-LR	VGG-LR
K	16.0	24.8	20.5
R	36.7	40.2	54.9
DK	34.1	36.5	45.3
LDK	29.9	31.4	52.3

LDKにおけるVGG16のみで予測精度が高い 物件事例

テストデータ数の31.6%がVGG16のみで高い予測精度を示している

	間取り図	賃料	建物構造	築年数	部屋階数	立地
事例 1		146,000	木造	13	1	練馬区
事例 2		140,000	木造	8	1	世田谷区
事例 3		237,000	SRC	17	4	千代田区

LDKにおけるPCA-LRのみで予測精度が高い 物件事例

テストデータ数の10.1%がPCA-LRのみで高い予測精度を示している

	間取り図	賃料	建物構造	築年数	部屋階数	立地
事例 1		236,000	SRC	4	10	新宿区
事例 2		120,000	軽量鉄骨	4	1	府中市
事例 3		128,000	軽量鉄骨	0	2	小金井市

LDKにおける両方のモデルで予測精度が低い 物件事例

テストデータ数の36.9%が両方のモデルで低い予測精度を示している

	間取り図	賃料	建物構造	築年数	部屋階数	立地
事例 1		114,000	RC	20	2	江戸川区
事例 2		171,000	RC	0	2	品川区
事例 3		82,000	木造	28	2	三鷹市

LDKにおけるVGG16とPCA-LRで予測精度が高い 物件事例

テストデータ数の20.7%が両方のモデルで高い予測精度を示している

事例 1 8730617d91cf47293ef2a5930f6f



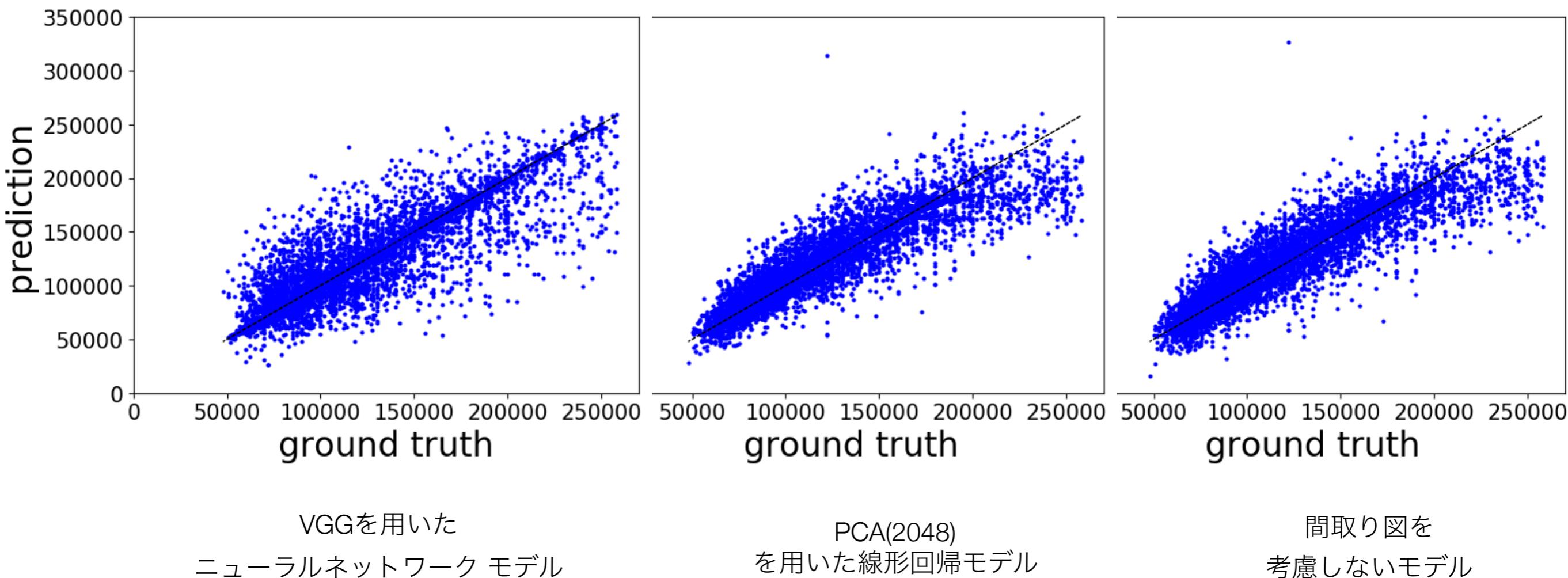
事例 2 e3e07e783efa1af04ff1a7fdabe2



事例 3 44d4d00c2608781fbe79df9c36d2



LDKのテストデータに対する予測精度



LDKにおけるVGG16のみで予測精度が高い 物件事例

テストデータ数の20.7%が両方のモデルで高い予測精度を示している

事例 1 2ba9eb7a25f5ea578d55dc42d7e9



事例 1

事例 2

1e182869a3e056912cb8686dc6ae



事例 3

634ccc3e55ba4986d1afae0ea1a4



LDKにおけるPCA-LRのみで予測精度が高い 物件事例

テストデータ数の20.7%が両方のモデルで高い予測精度を示している

事例 1

87687e55b997a09082d37539beb5



事例 2

6eaf4e251fb9ddb7dd00ceedb6f7



事例 3



8b3497c3d0579c24ea7c4f525316

LDKにおけるVGG16とPCA-LRで予測精度が低い 物件事例

テストデータ数の20.7%が両方のモデルで高い予測精度を示している

事例 1

634e29b8836b5fb29b854488f1f2

事例 2

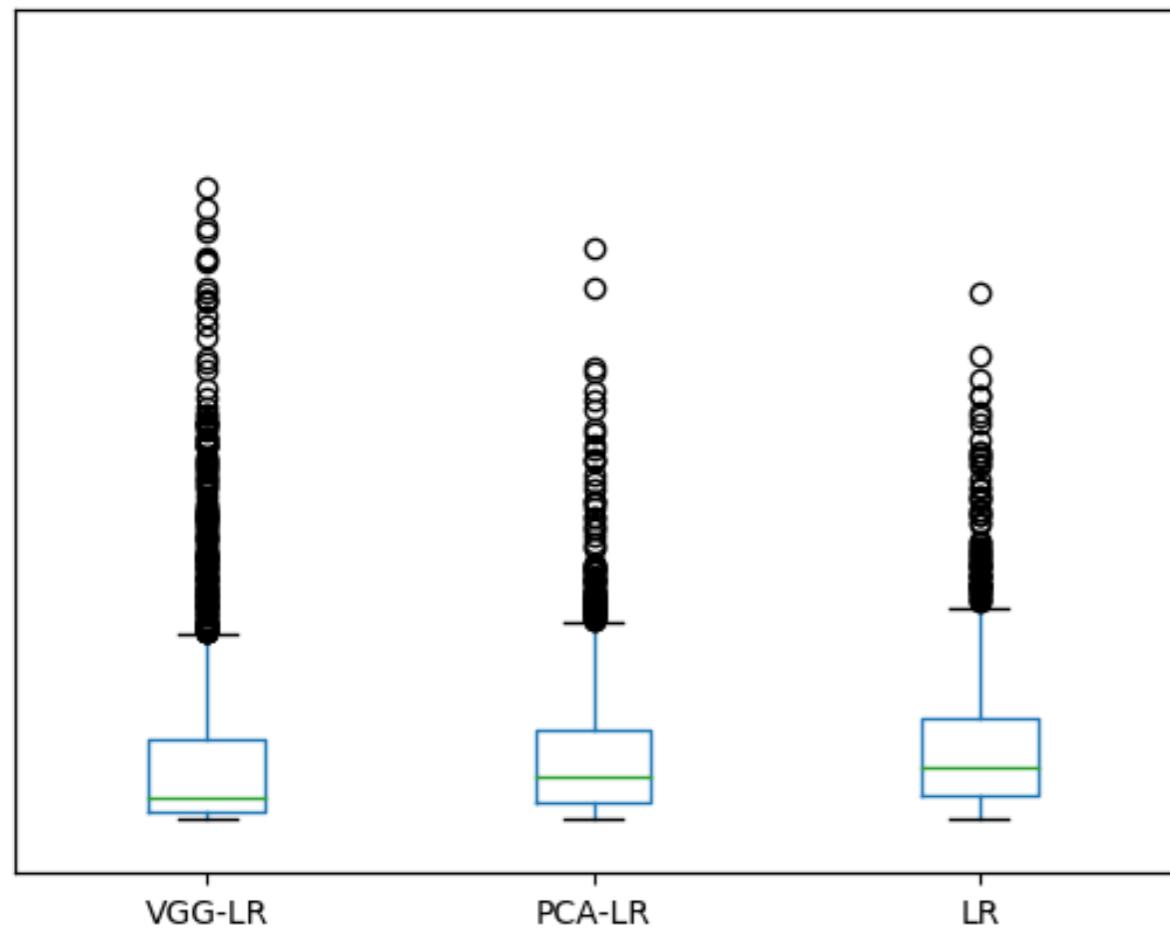
3305ee35c5ae9f6badc7d0a32ca7



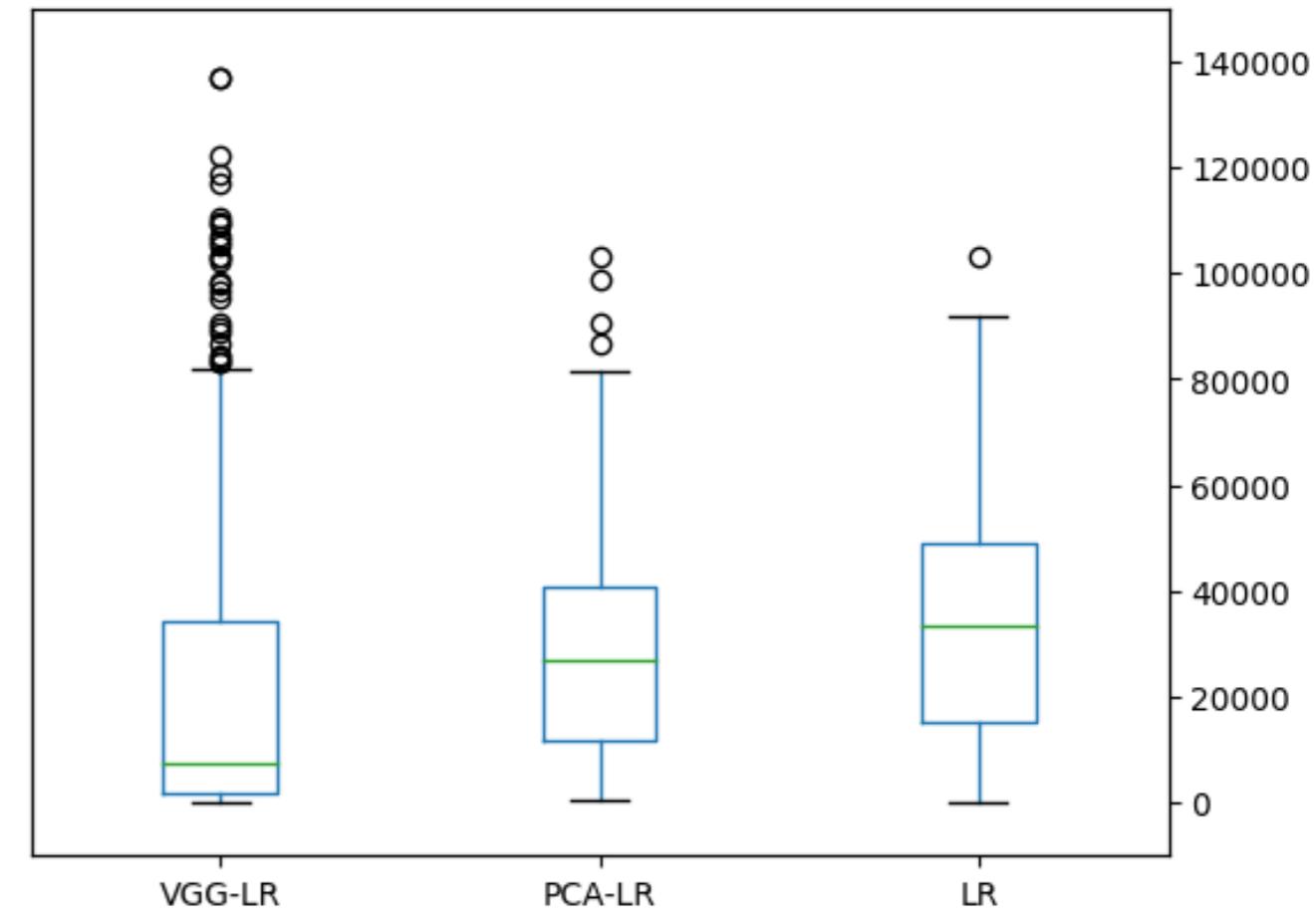
事例 3

2ebc9be5fce38eb1a78d2948def

LDKのテストデータに対する価格帯別予測精度

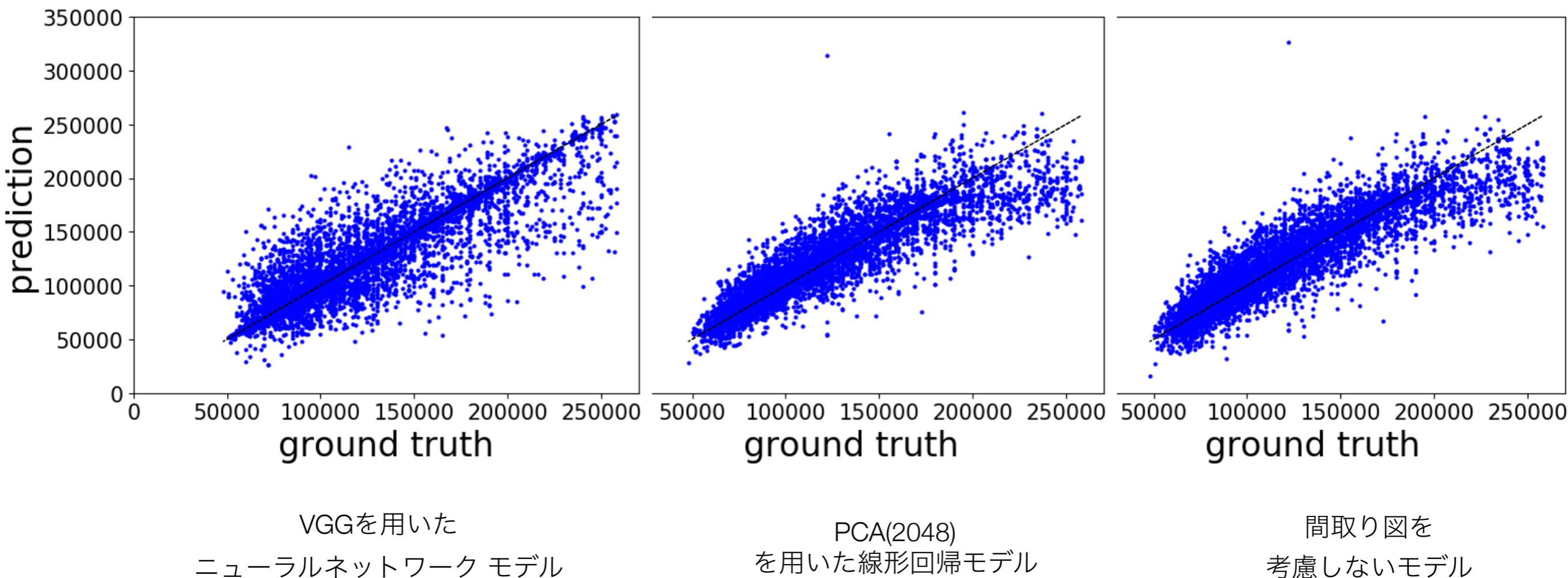


賃料: 15万円以上20万円未満

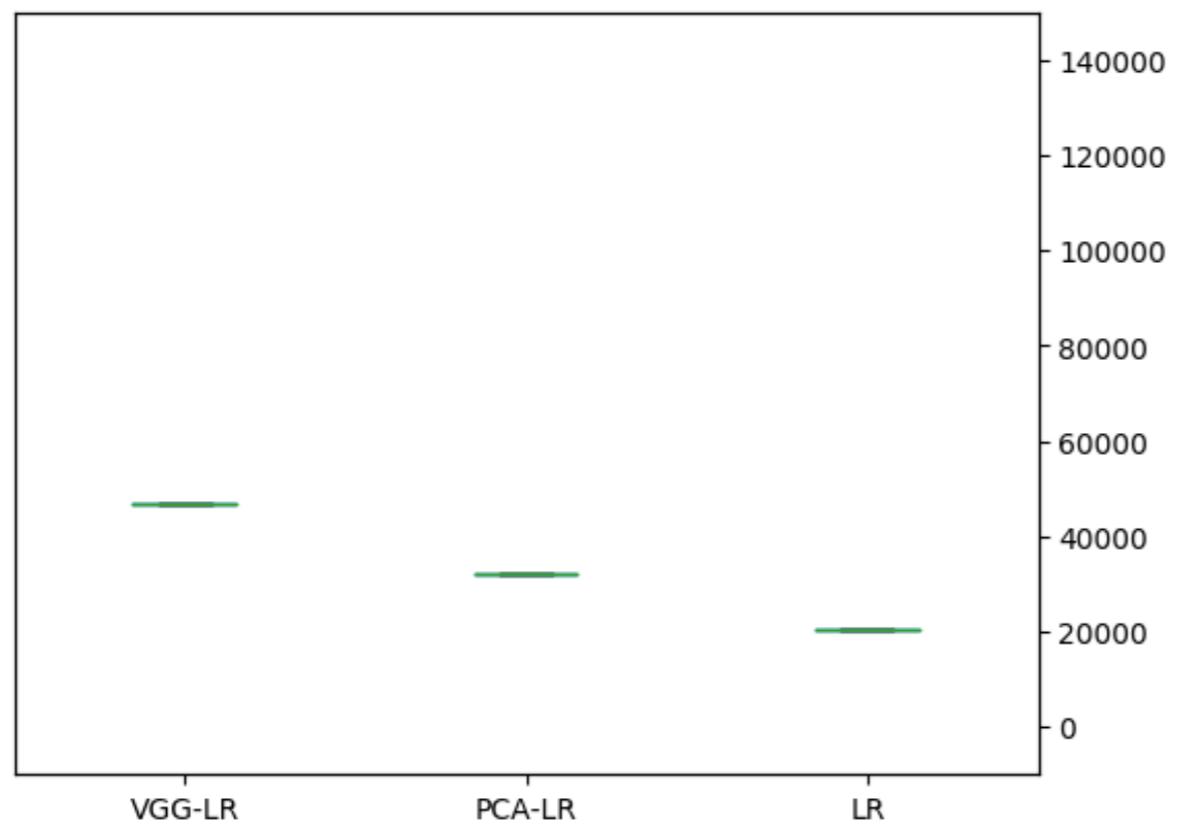


賃料: 20万円以上25万円未満

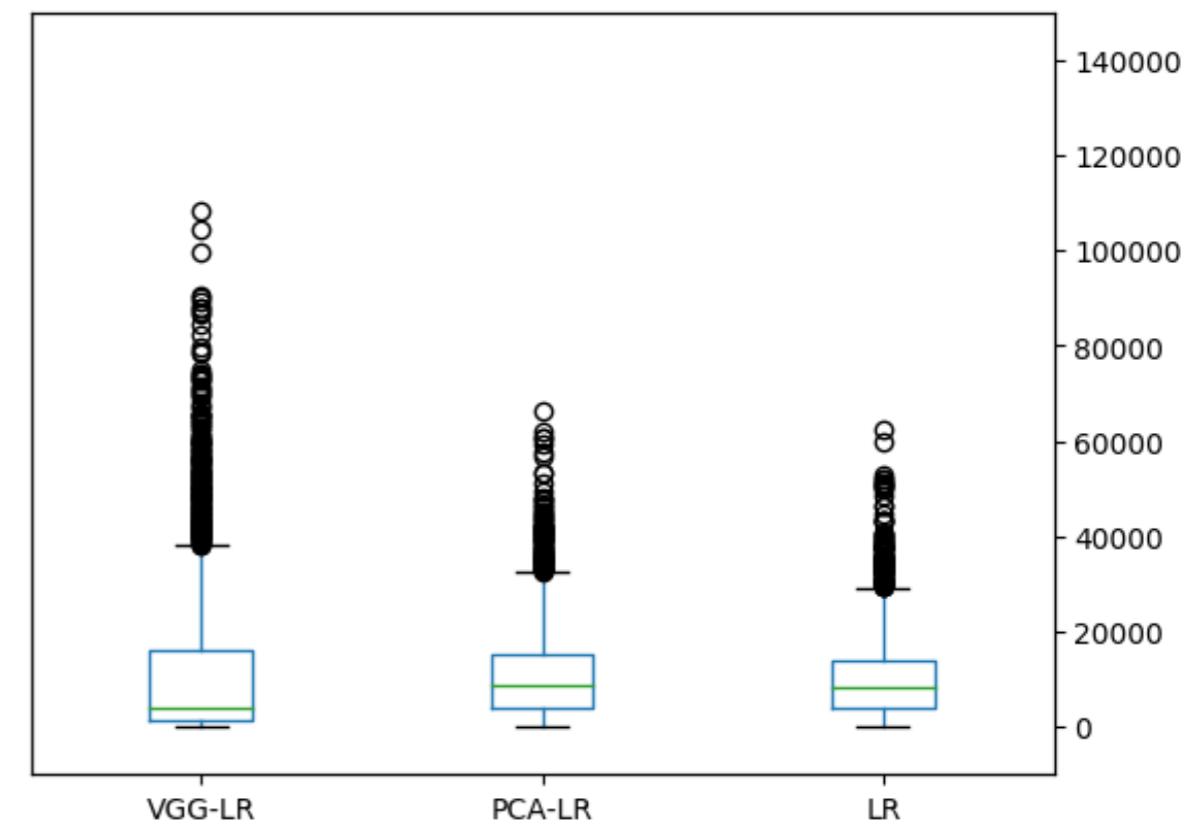
LDKのテストデータに対する予測精度



LDKのテストデータに対する価格帯別予測精度

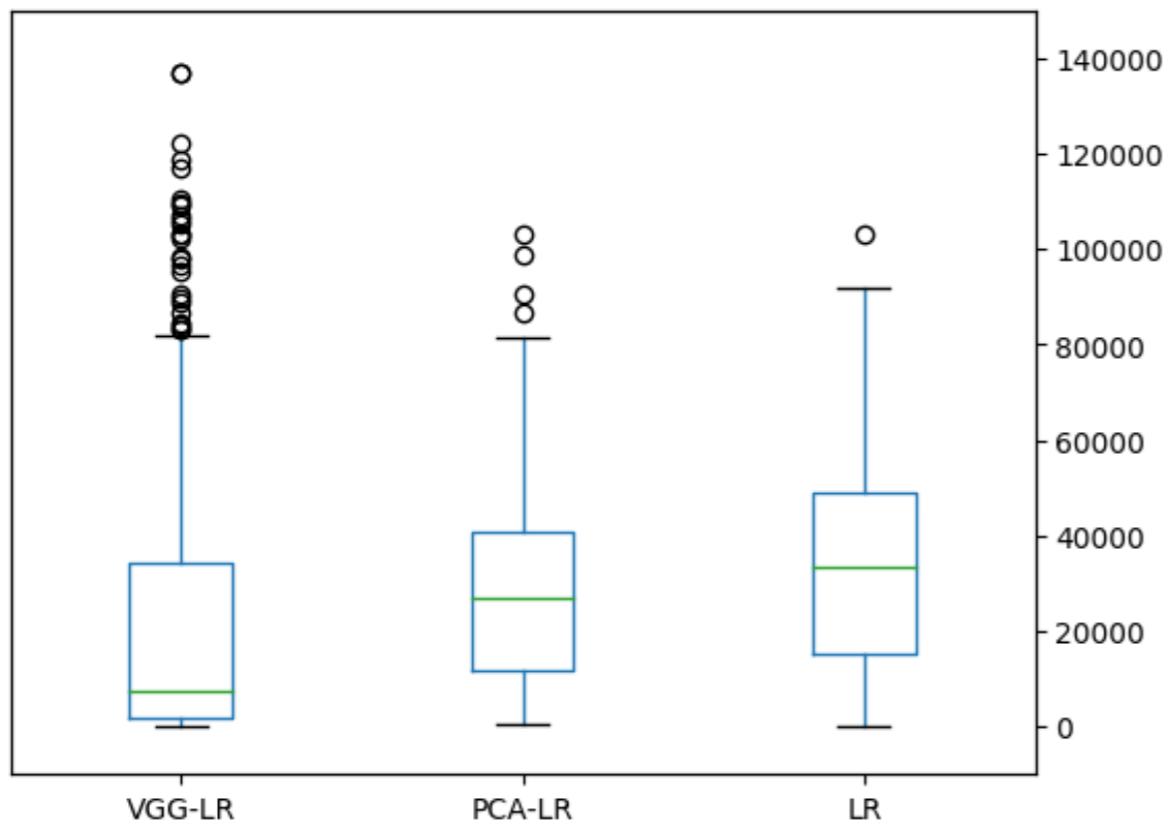


賃料: 5万円未満

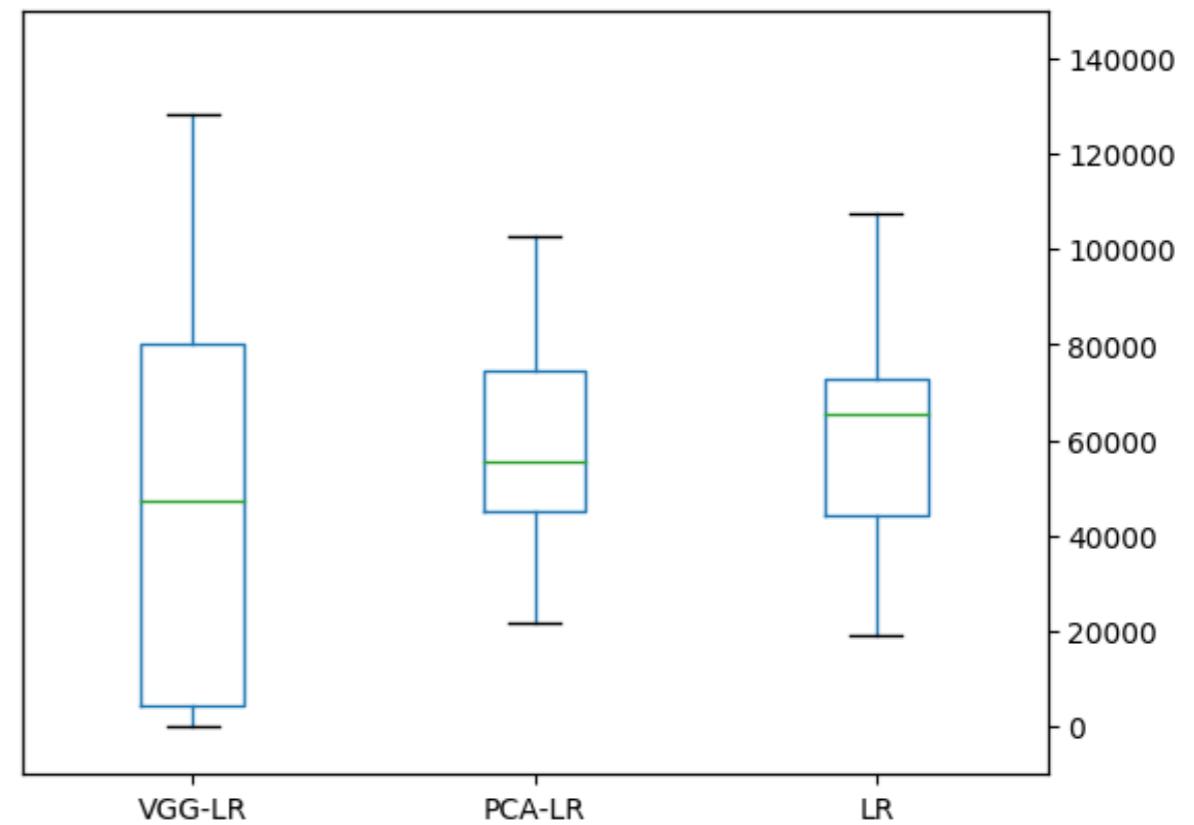


賃料: 5万円以上10万円未満

LDKのテストデータに対する価格帯別予測精度

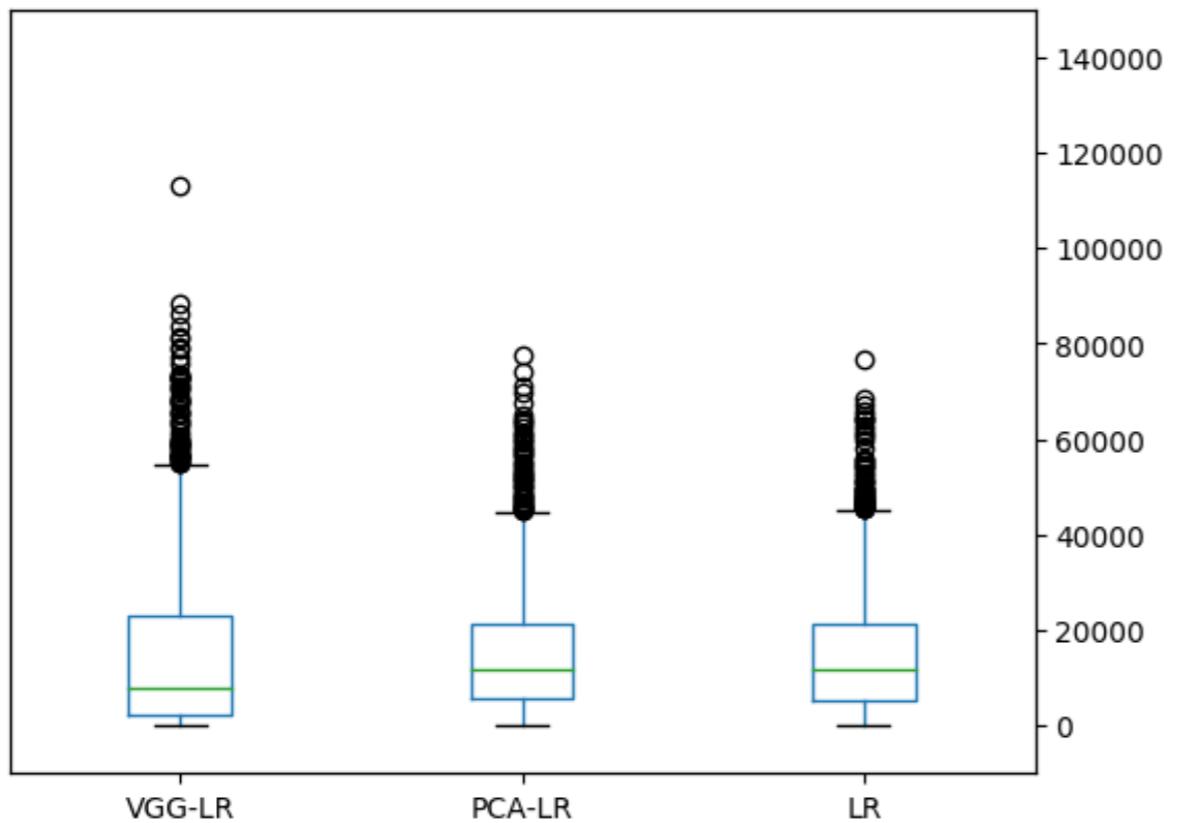


賃料: 20万円以上25万円未満



賃料: 25万円以上

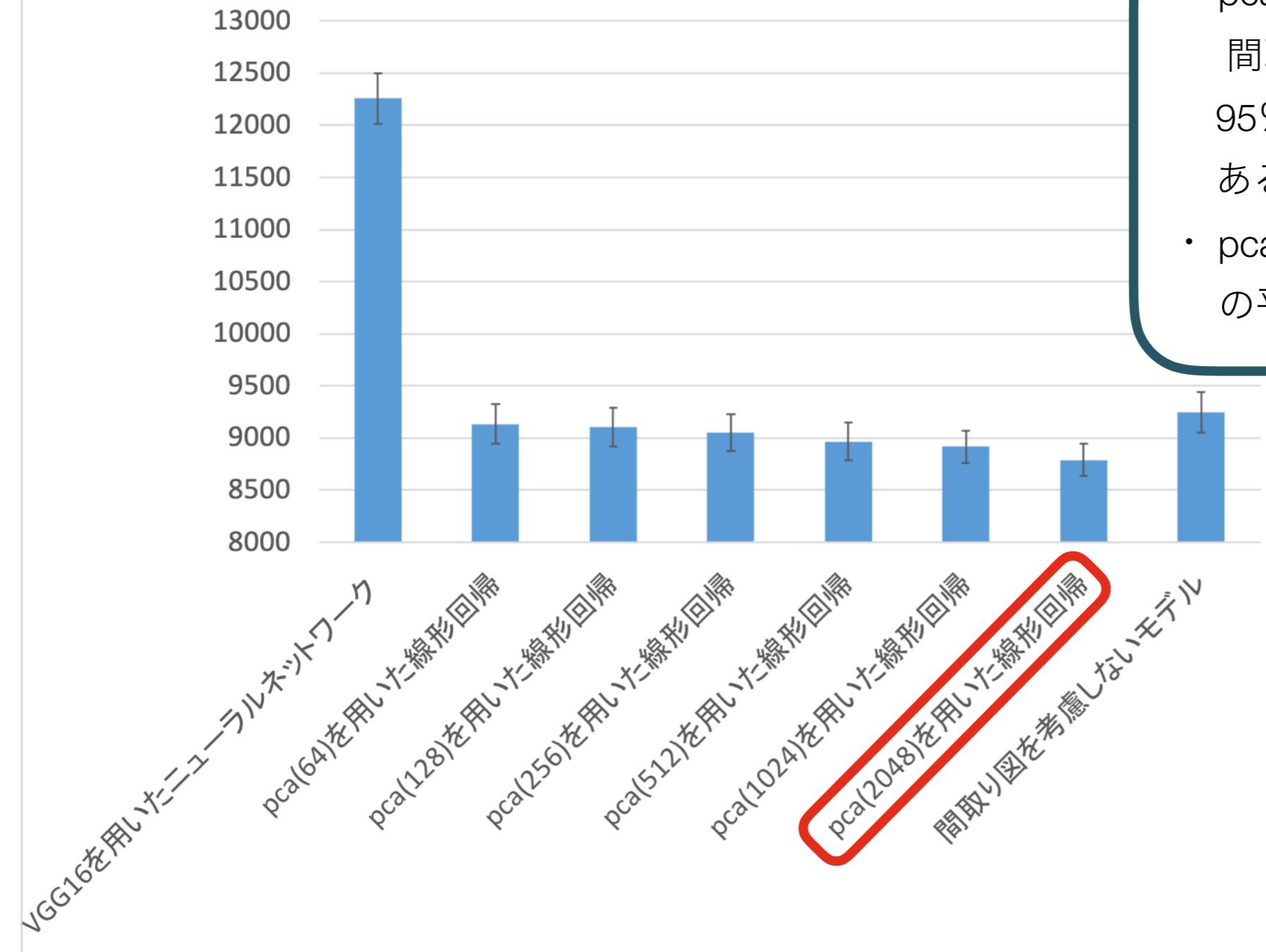
LDKのテストデータに対する価格帯別予測精度



賃料: 10万円以上15万円未満

賃料: 15万円以上20万円未満

Rの検証データに対する予測精度



- pca(2048)を用いた線形回帰モデルと間取り図を考慮しないモデルの95%信頼区間に重なりがなく優位性があると考えられる
- pca(2048)はpca(64)のモデルよりRMSEの平均が344.59改善

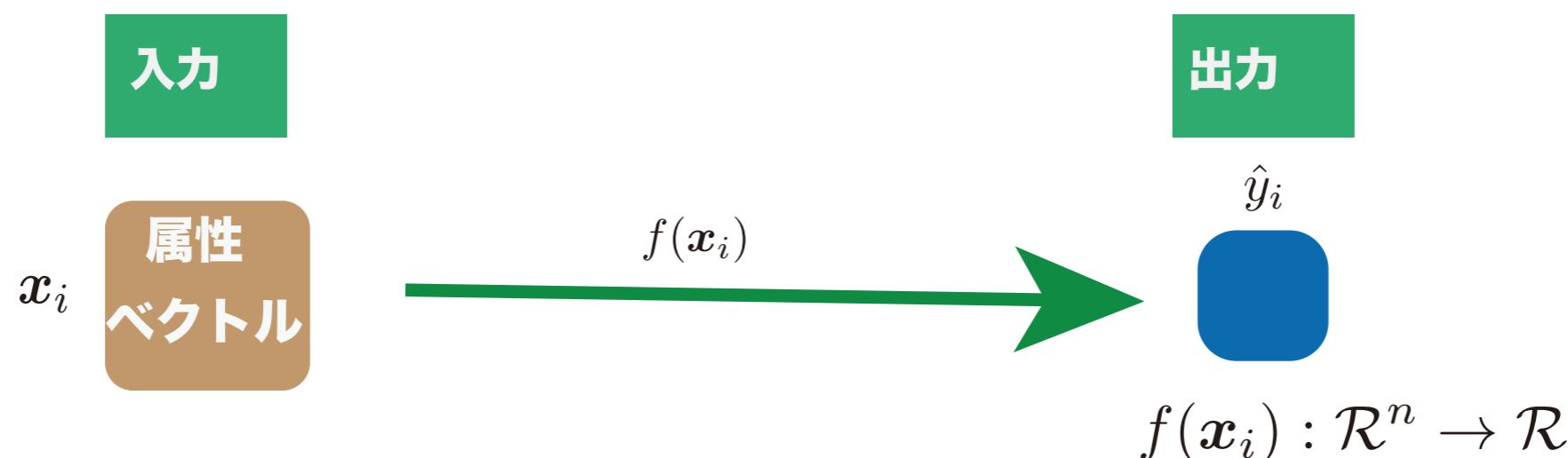
間取り図を考慮しないモデル

間取り図を考慮しないモデルにヘドニック・アプローチを採用

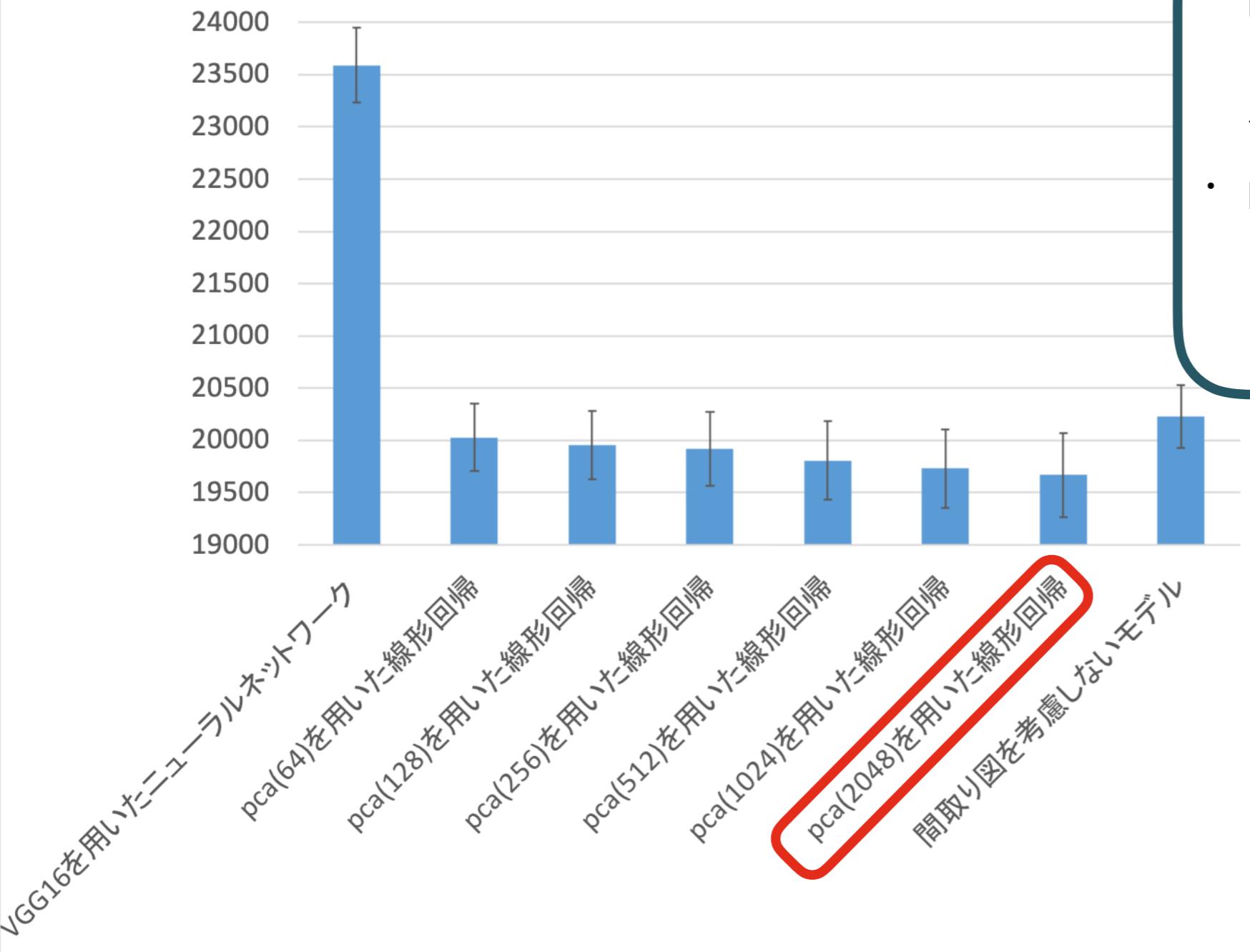
- $f(\mathbf{x}_i)$ で定義して α, β が最小となる ε を求める線形回帰モデル

$$\hat{y}_i = f(\mathbf{x}_i) = \boldsymbol{\alpha}^T \mathbf{x}_i + \beta$$

$$\varepsilon = \sum_{i=1}^m (y_i - \hat{y}_i)^2$$



LDKの検証データに対する予測精度



- pca(2048)を用いた線形回帰モデルと間取り図を考慮しないモデルの 95% 信頼区間に 145.61 の重なりあり
- pca(2048)を用いた線形回帰モデルと間取り図を考慮しないモデルの RMSE の平均の差は 559.13

おわりに

まとめ

- ・ 目的：賃料予測における間取り図の影響を明らかにする
- ・ 実験結果
 - ・ 間取り図を考慮することで予測精度があがる
 - ・ 間取り図の特徴抽出には主成分分析が良い
 - ・ 間取り図を考慮する際、その特徴量は1024, 2048次元の高次元が望ましい
 - ・ 1024, 2048次元の間取り図の特徴量で間取り図の持つ大部分の特徴量を抽出できている

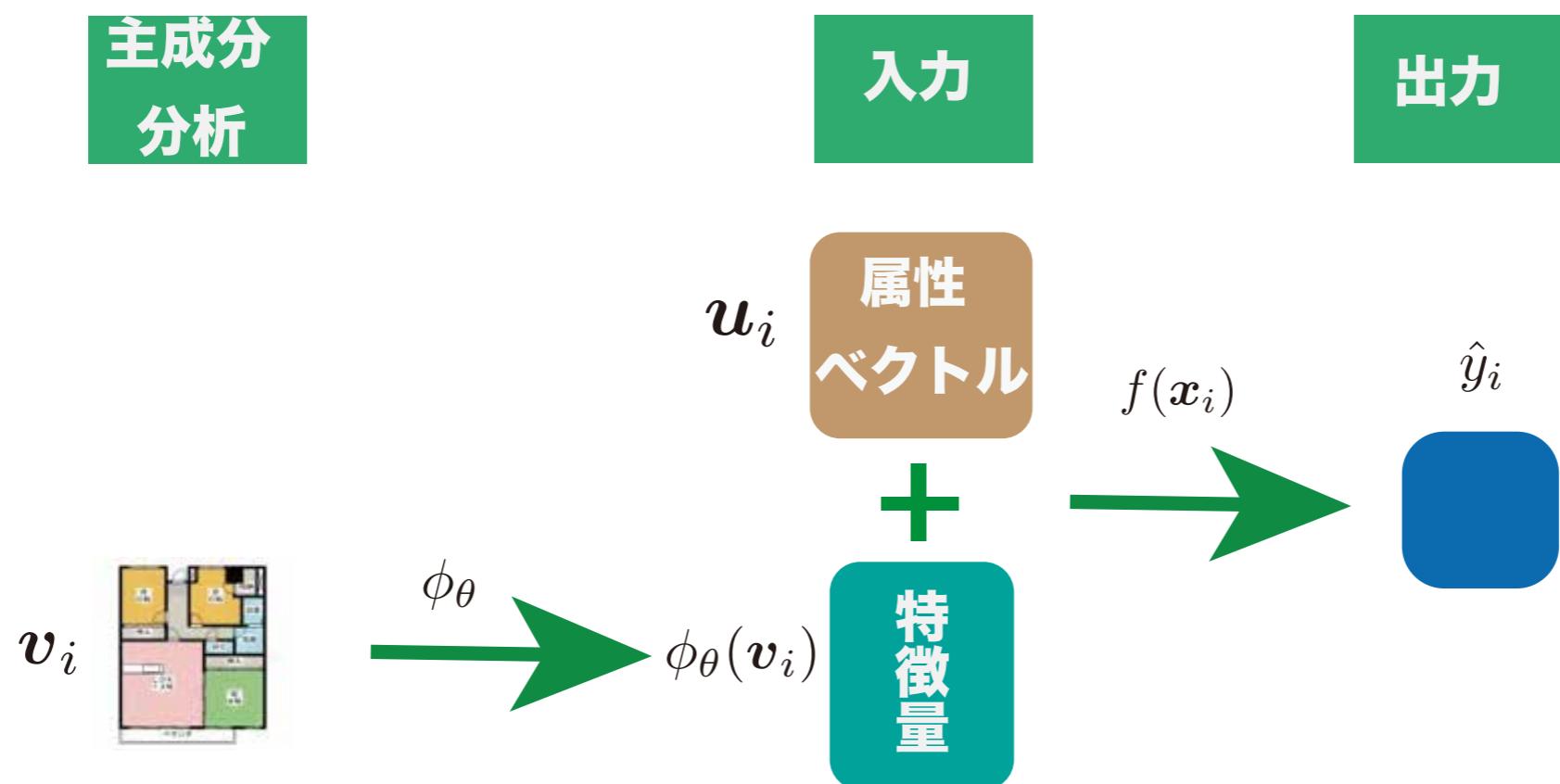
今後の方針

- ・ 欠損値の補完
 - ・ 尤もらしい値の代入
- ・ ニューラルネットワークモデルでの64次元よりも高い間取りの特徴抽出
 - ・ ResNetの採用

間取り図を考慮した予測モデル

PCA-LR

- ・ 主成分分析を用いて間取り図の特徴量を抽出
- ・ 抽出した間取り図の特徴量と変数ベクトルを組み合わせ、線形回帰で賃料予測
- ・ パラメータ θ と α, β を別々に学習



テストデータに対する価格帯別予測精度 (1/2)

間取りタイプ: K

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG	6,308.79	1,927.84	6,466.18	2,811.86	9,288.36	2,487.14	22,740.38	7,315.7	44,833.95	11,650.3	-	-
PCA	5,140.96	3,935.61	5,863.89	4,577.27	8,508.03	6,256.27	24,556.05	22,291.26	54,623.29	56,182.21	-	-
LR	4,959.35	4,118.17	6,059.36	4,939.87	10,060.15	8,997.77	31,256.09	33,133.38	76,081.01	77,904.07	-	-

間取りタイプ: R

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG	63,08.79	1,927.84	6,466.18	2,811.86	9,288.36	2,487.14	22,740.38	7,315.7	44,833.95	11,650.3	-	-
PCA	51,40.96	3,935.61	5,863.89	4,577.27	8,508.03	6,256.27	24,556.05	22,291.26	54,623.29	56,182.21	-	-
LR	49,59.35	4,118.17	6,059.36	4,939.87	10,060.15	8,997.77	31,256.09	33,133.38	76,081.01	77,904.07	-	-

テストデータに対する価格帯別予測精度 (2/2)

間取りタイプ: DK

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG	10,033.01	2,061.37	8,688.01	4,825.27	13,626.84	7,203.53	19,647.4	5,157.8	12,6665.64	12,6665.64	-	-
PCA	6,377.07	4,690.82	7,398.22	5,719.15	9,708.99	7,987.64	15,883.05	14,587.45	69,914.25	6,9914.25	-	-
LR	6,456.52	4,893.45	7,254.64	5,699.56	10,331.28	8,647.99	19,445.47	20,853.3	75,228.09	75,228.09	-	-

間取りタイプ: LDK

	[0円, 5万円)		[5万円, 10万円)		[10万円, 15万円)		[15万円, 20万円)		[20万円, 25万円)		25万円以上	
	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値	平均値	中央値
VGG	4,7045.09	4,7045.09	11,404.84	3,853.27	148,58.74	7,953.09	12,641.45	3,870.58	23,011.53	7,323.84	47,122.45	47,218.3
PCA	3,2216.32	3,2216.32	11,044.52	8,630.95	15,048.03	11,665.46	12,336.43	7,985.8	28,294.14	27,003.51	58,969.32	55,374.65
LR	2,0233.15	2,0233.15	10,219.48	8,163.42	15,133.73	11,925.57	13,705.06	9,635.61	33,436.41	33,540.97	61,898.01	65,562.94

テストデータの真値に対する誤差別予測精度

真値の5%内誤差

	VGG-LR	PCA-LR	LR
K	796(20.5%)	737(24.8%)	516(16.0%)
R	4,133(54.9%)	3,030(40.2%)	2,759(36.7%)
DK	1,608(45.3%)	1,297(36.5%)	1,209(34.1%)
LDK	3,579(52.3%)	2,153(31.4%)	2,052(29.9%)

真値の10%内誤差

	VGG-LR	PCA-LR	LR
K	16,862	11,602	11,602
R	12,541	9,217	9,217
DK	16,467	11,156	11,156
LDK	22,878	20,121	20,121

PCA-LRのチューニング結果

間取りタイプごとの最も予測精度の高いモデル

	PCAの次元数	RMSEの平均
K	512	10,548.18
R	2,048	8,789.96
DK	1,024	11,069.98
LDK	2,048	19,668.87

大野喜久之輔: 繼続賃料鑑定評価を再考する,
住宅新報社.

唐渡 広志: ヘドニック・アプローチを利用した不動
産価格指数の推定方法とその問題点, 都市住宅学,
vol.2016, no.92, pp.17-20, 2016.

はじめに

賃貸物件の特徴

- ・ 同じ物件が存在しない
- ・ 間取り, 階数, 立地などの物件変数が賃料に影響を与える

ヘドニック・アプローチ [唐渡, 2016]

ある商品の価格をその商品の変数の価値に関する集合とし,
その商品価格の予測モデルを線形回帰で構築する技術

$$\hat{y}_i = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \cdots + \alpha_n x_n + \beta$$

賃料 占有面積 立地 階数 築年数

間取り図を考慮していない

- 荒川, 諏訪, 小川, 荒川, 安本, 太田 : 暗黙知に基づく飲食店向け不動産賃料推定モデルの提案,
情報処理学会論文誌, vol.59, no.1, pp.33 - 42, 2018.
- 荒川, 諏訪, 小川, 荒川, 安本, 太田 : 暗黙知センシングに基づいた飲食店向き不動産店舗の賃料推定,
第31回人工知能学会全国大会, 2017.
- E. Ahmed, M. Moustafa: House price estimation from visual and textual features,
Proc. of 8th Int. Joint Conf. on Computational Intelligence, vol.3, pp.62-68, 2016.

研究背景

不動産業界を取り巻く課題 [荒川+ , 2018]

- ・ 経験豊富なベテラン営業マンの経験に大きく依存
- ・ 家賃の決定が営業マンが変われば大きく変わってしまう

→新人営業マンへのノウハウの伝達が難しい

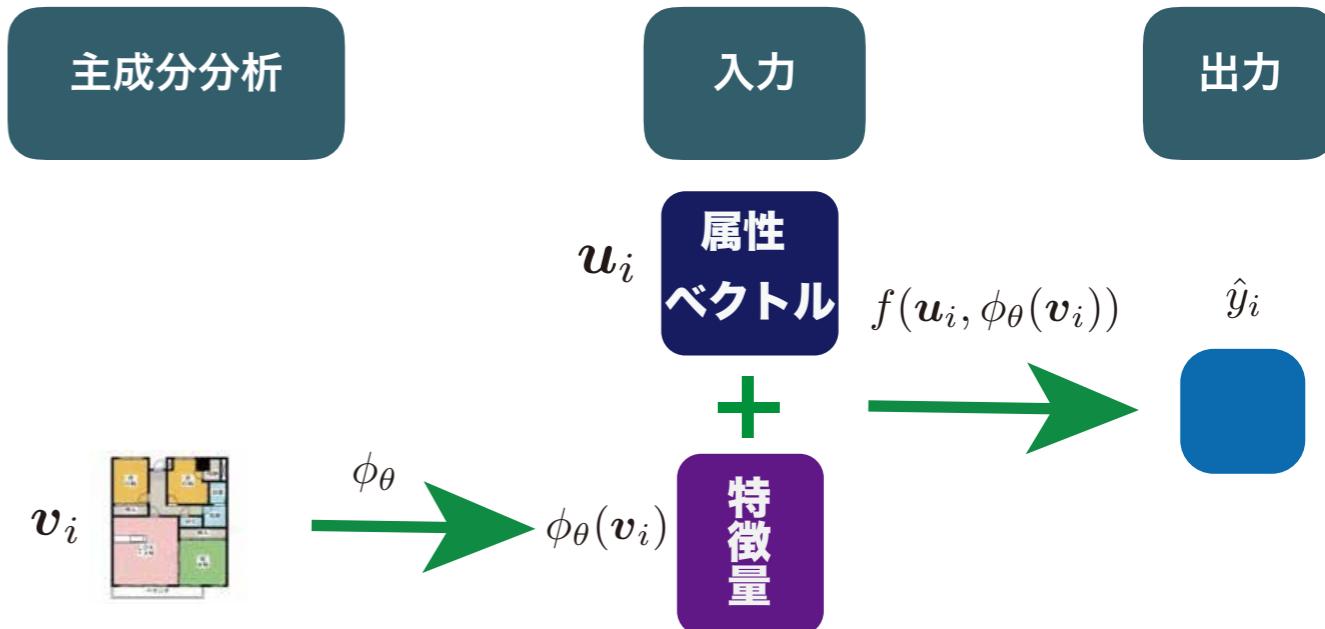
物件価格推定に対する技術的アプローチとその課題

- ・ 物件前交通量や視認性などの暗黙知を考慮した家賃予測 [荒川+ , 2017]
 - ・ 家賃の低い物件での賃料の予測精度が低い
- ・ 4種類の画像データを用いた物件価格の予測 [Eman+ , 2016]
 - ・ 物件ごとに画像として記録されている場所や枚数などが異なる
 - ・ 共通して入手できる画像を用いる必要がある

間取り図を考慮した予測モデル

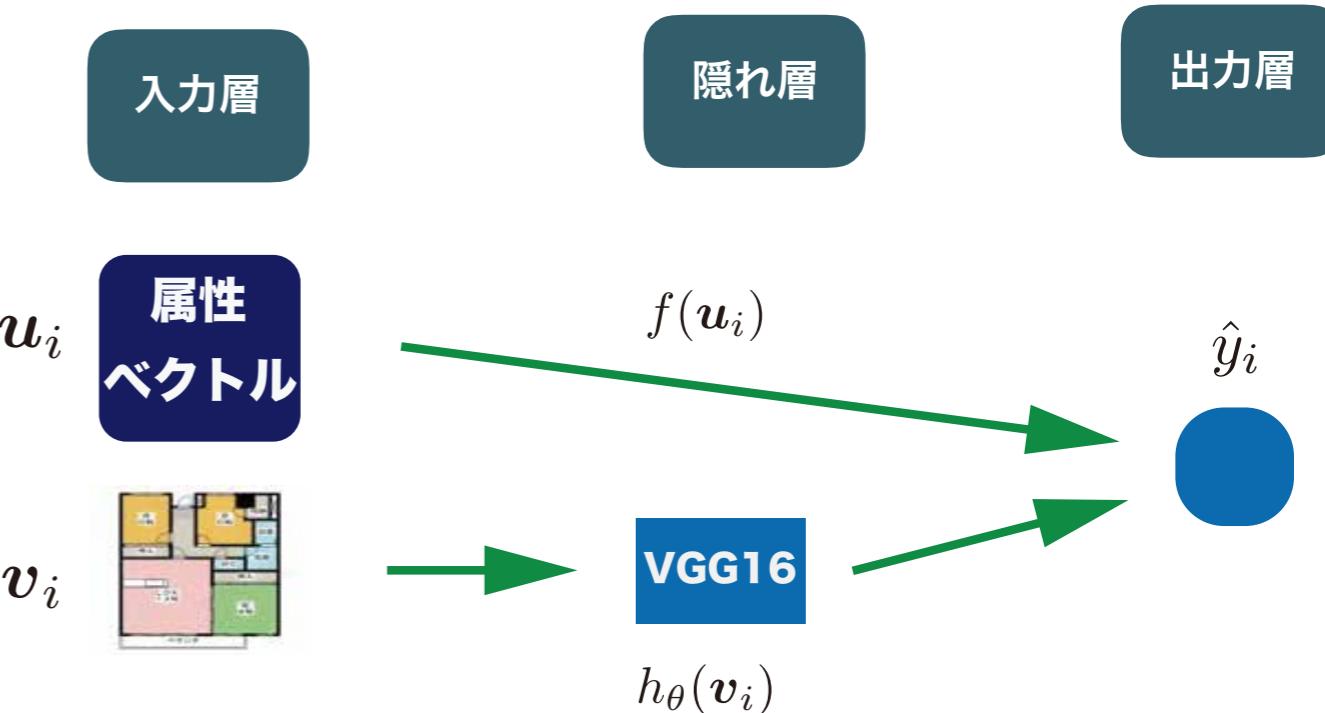
PCA-LR

- ・ 主成分分析による特徴抽出
- ・ 抽出した間取り図の特徴量と変数ベクトルを組み合わせ、線形回帰で賃料予測
- ・ 間取り図の特徴量: 64, 128, 256, 512, 1024, 2048次元の6種類



VGG-LR

- ・ ミニバッチ毎に確率的勾配降下法を適用
- ・ batch size: 100
- ・ epoch: 100
- ・ 活性化関数: ReLU
- ・ 間取り図の特徴抽出: 64次元



検証する3モデルの予測式

モデル名	予測式	パラメータ
LR	$\hat{y}_i = f(\mathbf{x}_i) = \boldsymbol{\alpha}^T \mathbf{x}_i + \beta$	$\boldsymbol{\alpha}, \beta$ を更新
PCA-LR	$\hat{y}_i = f(\mathbf{u}_i, \phi_{\theta}(\mathbf{v}_i))$ 間取り図	θ を更新した後, $\boldsymbol{\alpha}, \beta$ を更新
VGG-LR	$\hat{y}_i = h_{\theta}(\mathbf{v}_i)$ $\hat{y}_i = f(\mathbf{u}_i)$ 物件変数	$\boldsymbol{\alpha}, \beta, \theta$ を更新

予測精度の評価方法

使用モデル

- LR, VGG-LR, PCA-LR (予測精度の高いハイパーパラメータを使用)

予測精度の評価

- データセットの20%をテストデータに, 80%を開発用データに分割
- 間取りタイプごとに, 開発用データを学習に使用
- テストデータを予測

評価指標

- 平方平均二乗誤差 (RMSE: Root Mean Square Error)
- 予測誤差の密度

PCA-LRのチューニング結果

RMSE	64	128	256	512	1024	2048
K	10,886	10,733	10,649	10,548	10,630	11,293
R	9,134	9,104	9,054	8,967	8,916	8,790
DK	11,207	11,189	11,144	11,096	11,069	11,148
LDK	20,027	19,958	19,919	19,807	19,733	19,668

平方平均二乗誤差による評価

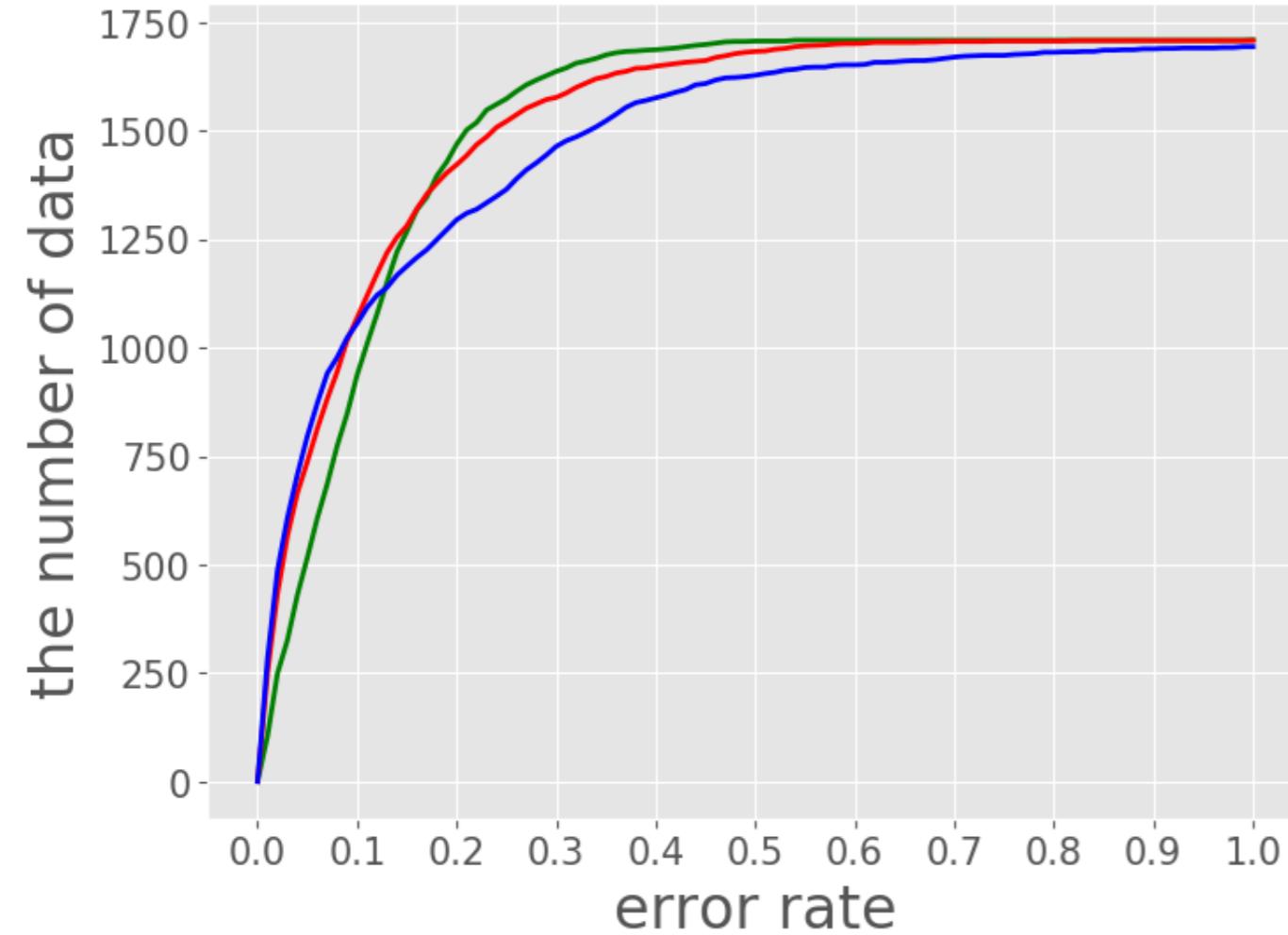
テストデータに対するRMSE[円]

	データ数	LR	PCA-LR	VGG-LR
K	[1,712]	11,602	11,689	16,862
R	[7,525]	9,217	8,727	12,541
DK	[3,548]	11,156	10,957	16,467
LDK	[6,842]	20,121	19,646	22,878

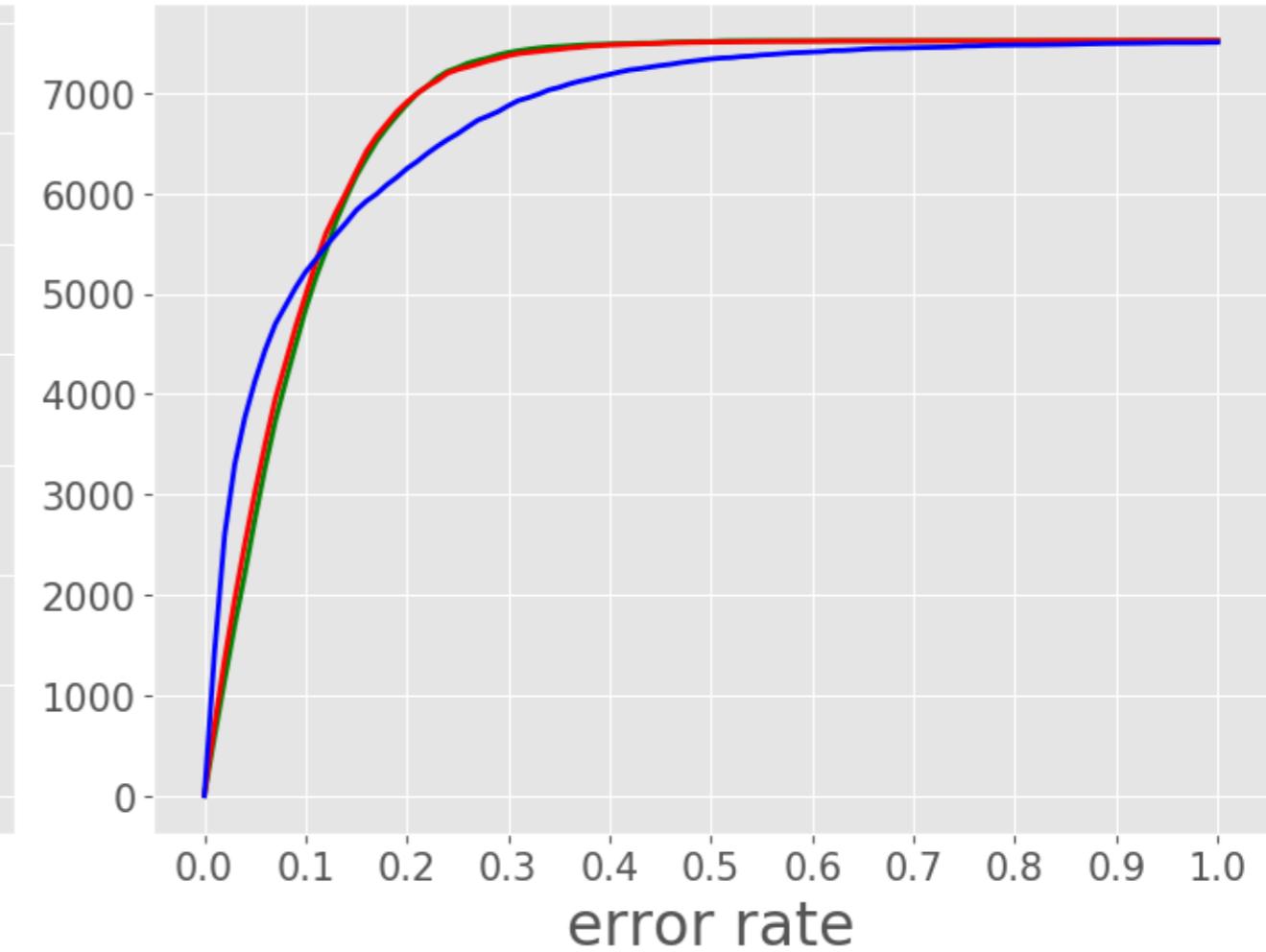
開発用データの学習時間[秒]

	データ数	LR	PCA-LR	VGG-LR
K	[9,205]	0.06	11	2,897
R	[39,985]	0.08	48	14,566
DK	[18,975]	0.05	23	6,471
LDK	[36,595]	0.07	45	13,026

予測誤差の密度による評価①



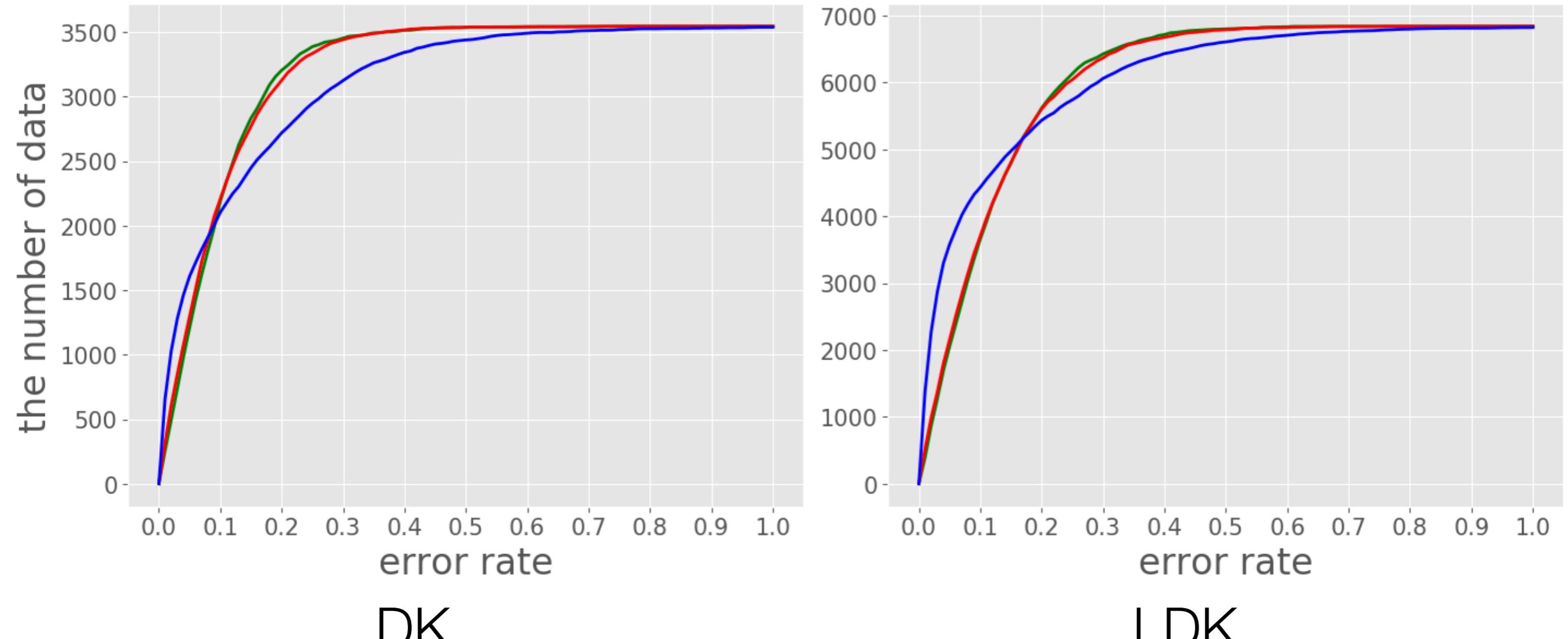
K



R

間取り規格LDKのテストデータに対する誤差関数

予測誤差の密度による評価②



間取り規格LDKのテストデータに対する誤差関数

