

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

Real estate rent estimation of restaurant properties based on tacit knowledge sensing

荒川 周造 *1
Shuzo Arakawa

諏訪 博彦 *1
Hirohiko Suwa

小川 祐樹 *2
Yuki Ogawa

荒川 豊 *1
Yutaka Arakawa

安本 慶一 *1
Keiichi Yasumoto

太田 敏澄
Toshizumi Ohta

*1 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

*2 立命館大学
Ritsumeikan University

Previously determining the rent of restaurants properties was based on the tacit knowledge, that is, intuition and experience gained by experienced salespeople. However, this business custom has problems (e.g., no evidence to the determined rent). Therefore, the transference of the knowledge and experience from the experienced salespeople to fresh one, is not effective. To make rent estimation more correctly and efficiently, we build the rent estimation concept model for restaurants properties. In addition, we discuss about the specific challenges in building the rent estimation system, that are (1) acquisition and indexing of the tacit knowledge and (2) construction of a rent estimation model. In this paper, we build rent estimation models based on some machine-learning methods. As the result of comparing, the case of using Random Forest regression algorithm achieved the highest accuracy that coefficient of determination was 0.738.

1. はじめに

近年、機械学習を活用した営業支援などが活発になされており、不動産分野においても住宅物件の価格推定に関する取り組み [1][2][3] がなされている。一方で、飲食店舗を対象とした不動産物件の賃料は、これまで豊富な経験を持つベテラン営業マンが培ってきた勘や経験からなる暗黙知に基いて決定されている。この手法では、賃料決定の根拠を示せないことから、人が変われば価格が変わるといった問題を抱えている。また、ベテラン営業マンから経験の浅い新人営業マンへの暗黙知の継承が課題となっている。しかしながら、賃料決定に根拠がないゆえに正しい知識伝達ができず、非効率な共同作業などの手法によって行われている現状がある。また、飲食店舗物件固有の属性が存在するため、従来からある住宅系の価格推定手法をそのまま活用することには問題がある。具体的な例として、物件選びの指標に、ガスの口径や排気設備、店舗の視認性や周辺の歩行者通行量などが考慮されており、立地と間取りを中心とした一般住宅とは選定基準が異なることが挙げられる。

そこで本研究では、SECI モデル [4] を基にした賃料推定概念モデル [5] を提案・構築している。SECI モデルは、知識は暗黙知を表出化して形式知にし、さらに連結化することで、概念として共有・伝承が可能となることを示したモデルである。この考え方に基くと、ベテラン営業マンの暗黙知は、表出化して形式知にし、連結化することで、共有・伝承が可能となる。

我々はベテラン営業マンの暗黙知を表出するために、飲食店向き不動産会社である株式会社 ABC 店舗のベテラン営業マンにインタビューを実施している。その結果、賃料を決める要素として、複数の具体的変数が回答され、これを営業マンが指標化できている顕在的情報とする。顕在的情報は変数の性質によって、物件固有の静的情報、物件とは独立して変化する動的情報に分類できる。また、営業マンは指標化できないが、物件の特徴を含む潜在的情報も賃料決定に影響していると考えられ

る。先行研究 [1][2][3] では、一般住宅物件を対象としており、静的情報が主に用いられているが、飲食店では動的情報や潜在的情報も重要であるため、本研究では 3 つの要素から賃料を推定する概念モデルを構築している。また、賃料推定概念モデルに基づき、賃料推定システムの構築を行う。システム構築における課題は、

- (1) 暗黙知の取得及び指標化
- (2) 機械学習に基づく賃料推定モデルの構築

である。

課題 (1) の実現に向けて、我々は株式会社 ABC 店舗の営業マンにインタビューを行い、賃料推定に関する要素を抽出している。その結果から、静的情報、動的情報、潜在的情報として用いる具体的な変数が決定された。静的情報は、ABC 店舗の HP 上で顧客に提供されている物件情報である。動的情報は、周辺地域の価値と物件の見つけやすさの変数である。周辺地域の価値は、地域毎の平均賃料を用いて指標化し。また、店舗の見つけやすさの指標として、店舗前の通行量と店舗の視認性について、営業マンがスコアリングしたデータを活用する。一方で、スコアリングによる指標化作業には個人による差異が生じることから、根拠あるシステムを構築するには望ましいとは言えない。そこで本研究では、通行量を機械的に取得可能な通行量センシングシステム [6] を独自に開発し、実験を通じてその有用性について検証している。さらに、潜在的情報は、営業マンが記述している物件キャッチコピーから、自然言語処理技術を用いて名詞・形容詞を抽出し、単語と賃料の関係に基いて指標化している [7]。

課題 (2) を実現するために、賃料推定モデルの実装・評価を行う。本稿では、ニューラルネットワーク、ランダムフォレスト、重回帰分析によって、推定に用いる回帰モデルを構築し、交差検証による比較を行う。データセットには、東京都内で契約が成立した実績のある 184 物件を対象としている。その結果、ランダムフォレストを用いて推定を行った際に、決定係数 R^2 が 0.738 となり、最も優れていることが確認された。

以下に本稿の構成について述べる。第 2 章では関連する研究について説明する。第 3 章では、賃料推定概念モデルと各要素の取得・指標化手法について述べる。第 4 章では機械学習に基づく賃料推定について述べる。第 5 章はまとめとする。

連絡先:

荒川 周造: 奈良先端科学技術大学院大学, 奈良県生駒市
高山町 8916-5, arakawa.shuzo.aj4@is.nasit.jp

諏訪 博彦: 奈良先端科学技術大学院大学, 奈良県生駒市
高山町 8916-5, h-suwa@is.nasit.jp

2. 関連研究

機械学習に基づく不動産価格推定に関する研究として、Victor[1]らは決定木とニューラルネットワークに基づいて不動産価格推定を行っている。研究対象は、米国キング郡の不動産物件であり、外れ値を取り除くために取引価格の上位・下位5%の物件を除外して推定を行い、価格帯ごとに分割して推定誤差を評価している。その結果、ニューラルネットワークを用いた方が良好な結果が得られることを明らかにしている。

また、Chih-Hung[2]らは、台湾での住宅選定に関わりの深い風水に着目し、風水を考慮した不動産価格推定に取り組んでいる。学習データセットには、多くの不動産価格推定に用いられている建物固有の属性に加え、風水におけるタブーを設定している。また、機械学習手法として、バックプロパゲーションニューラルネットワーク(BPN)とファジーニューラルネットワーク(FNN)、独自に開発したハイブリッド遺伝ベースのサポートベクター回帰(HGA-SVR)からなる複数のアルゴリズムで推定した結果、いずれの手法においても風水のタブーを考慮した方がより良好な推定結果となることを明らかにしている。また、それぞれのアルゴリズムの性能を評価し、HGA-SVR、FNN、BPNの順で優れていることを明らかにしており、最も優れたHGA-SVRの平均絶対パーセント誤差(MAPE)は4.79%を示している。さらに、BPNを用いて最も価格推定に影響のある属性が土地の大きさであること確認している。また、風水におけるタブーでは、トイレ、窓、家の外観、ドアの順に影響していることを明らかにしている。

さらに、Vincenza[3]らは、ターラント市(イタリア)における不動産価格に、交通システムと地域ごとの環境の質が深く関係していると考え、人工ニューラルネットワーク(ANN)を用いて不動産価格推定を行っている。データセットには、立地条件や建物の構造に加え、駅やバス停までの距離などの交通に関する属性や、SO₂、NO_x、NO、NO₂、CO、PM₁₀のそれぞれの値と最大値といった環境汚染に関する属性が含まれている。この研究でもデータセットの各属性の感度分析を行った結果、不動産価格に最も影響している要因として、ビーチへの近さや、ガレージおよびテラスの有無が求められた。環境汚染に関する項目で最も重要なものとしては、SO₂の最大値が挙げられ、全42ある属性の内の8番目であった。また、交通に関する属性も15番目付近に位置しており、影響があることを示している。さらには、環境汚染に関する属性を除いて学習したところ、大きな精度低下は見られなかったことから、不動産物件から工業地帯への距離を用いることで、環境汚染の属性を代用できることを明らかにしている。

以上3件の関連研究では、それぞれの地域に特化した手法を採用している点で共通している。地域ごとに賃料決定に影響のある属性に違いが見られるため、どの属性が重要であるかを吟味する必要がある。また、文献[2]では、風数的情報を用いており、これはある種潜在的な情報であるといえる。文献[3]においても、環境の質として、SO₂等の値をパラメータとして用いており、これは動的な情報に相当する。さらに、機械学習手法としては、多くの研究でニューラルネットワークを用いていることが確認できる。

3. 飲食店向き賃料推定概念モデル

3.1 SECIモデルに基づくモデル構築

従来の賃料決定手法では、ベテラン営業マンの勘や経験からなる暗黙知に依存しており、値付けを行う人が変われば賃料が変わるといった問題がある。また、新人営業マンへの知識伝

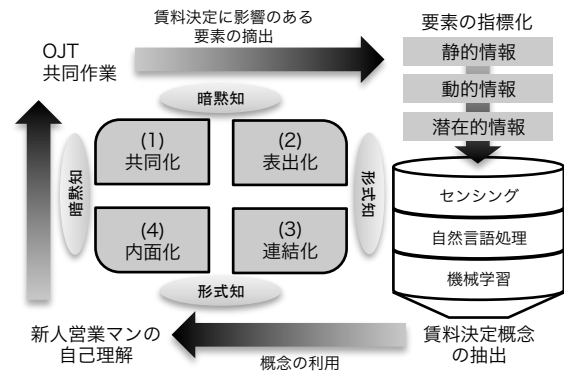


図1: 賃料推定概念モデル

承は、OJTや共同作業などの効率が悪い手法によって行われており、正しく知識伝承がされているかも不明である。暗黙知の伝承をより正しく効率的に実施するには、新人営業マンが賃料決定ルールを概念として取得し、自ら学習できるシステムがあると望ましい。システムを構築するためには、飲食店を対象とした概念モデルが必要となる。

このような暗黙知を伝承するための知識創造モデルとして、野中らのSECIモデル[4]がある。SECIモデルでは、暗黙知は言語化できない知識を意味し、形式知は言語化できる知識とされている。SECIモデルを図1の(1)~(4)に示す。野中らによれば、知識創造は、(1)共同化(暗黙知 暗黙知) (2)表出化(暗黙知 形式知) (3)連結化(形式知 形式知) (4)内面化(形式知 暗黙知) (1)共同化といったサイクルを繰り返すことによって可能となる。SECIモデルでは、知識は暗黙知を表出化して形式知にし、さらに連結化することで、概念として共有・伝承が可能となることを示している。言い換えると、ベテラン営業マンの暗黙知は、表出化して形式知にし、さらに連結化することで、共有・伝承が可能となる。

そこで、本研究では前述したSECIモデルに基づき、図1に示すような賃料推定概念モデルを構築する。まず、ベテラン営業マンに対するインタビュー調査によって、暗黙知を形式知として表出(化)する。ベテラン営業マンからは、賃料決定に影響を与える要素として、指標として取得可能な顕在的情報が回答された。顕在的情報は各変数の性質に基づき整理することで、物件固有の静的情報、物件に直接関連せずに変動する動的情報の2種類に分類できる。これらは、営業マン自身が言葉にできているゆえ形式知であるといえるが、それらの指標化にあたり営業マンの主観によるスコアリング値を活用しており、一部に個人による差異が生じる。この問題は、センシング技術を活用し、その特定の変数を機械的に取得することで解決する。また、営業マンが回答できなかったが、賃料決定に影響があると考えられる要素として、潜在的な情報が挙げられる。この要素は指標化がなされていないため、静的情報と動的情報に分類する判断もできない。潜在的な情報は、物件について記述されているテキストデータに内包されていると考え、自然言語処理技術を活用し指標化する。さらに、機械学習を用いて、指標化された要素と賃料との関係を連結(化)する。機械学習の結果として、賃料推定概念が抽出され、営業マンはそれらを共有することができるようになる。また、概念が抽出されることは、新人営業マンへの伝承を可能であることを示している。

3.2 要素の取得・指標化

インタビューによって回答された顕在的情報の各変数は、特性によって静的情報と動的情報に整理できる。また、営業マン

が回答できず指標化できなかった潜在的情報についても、指標化を行い推定に活用できるようにする。本章では、これらの要素を持つ各変数の取得と指標化について説明する。

まず、静的情報として、居抜きの有無、駅徒歩時間、階数、坪数が変数として回答された。居抜きとは、物件に付帯する什器（テーブル、カウンター、ガスレンジ等の設備）を意味し、その有無を 0・1 のフラグとして表現する。これらの変数は、株式会社 ABC 店舗が Web 上で公開している物件情報として提供されており、改めて指標化を行う必要はない。

次に、動的情報としては、物件周辺の価値を示す地域ポテンシャルと、店舗の視認性、店舗周辺の通行量が回答された。地域ポテンシャルは、最寄り駅の平均坪単価を算出し、物件の坪数を掛け合わせた駅推定賃料として定義し、指標化する。地域ポテンシャルを表現できる他の指標としては路線価が考えられるが、我々は駅推定賃料の方が効果的であることを既に確認している [7]。また、店舗の視認性と店舗周辺の通行量は、積を取ることで物件の見つけやすさを表す新たな指標として利用する。これら 2 つの変数の指標化には、2 名の営業マンが 5 段階でスコアリングした平均値を用いる。この手法では、個人による差異が生じる問題があるが、これは人に依存しないセンシングシステムの開発により解決できると考えられる。本研究では、実際に通行量を取得可能なセンサを開発し、実験によってその有用性を確認している [6]。一方で、視認性については間口面積や看板の大きさなどを活用すれば指標化できる可能性もあるが、周囲にある植込みや建造物などの要因を盛り込めないなどの問題があるため、さらなる検討が必要である。

最後に、潜在的情報の取得には、営業マンが記述している物件キャッチコピーを活用する。キャッチコピーには、物件の特徴や雰囲気などの多様な情報を含む一方で、指標化はされていない。そこで、キャッチコピー中に多く含まれる名詞と、それを修飾する形容詞を形態素解析によって抽出し、単語と賃料に与える影響を求めると、事前には潜在的情報のみで求めた推定賃料と、実際の賃料の比率を物件毎に求めておき、単語毎に平均値を求める平均推定賃料比率を定義している [5]。単語毎に算出された値より、賃料を上げる要因があるポジティブワードと、賃料を下げる要因があるネガティブワードとして抽出を行い、物件毎にそれらの積をとることで、ポジティブ指標・ネガティブ指標として賃料推定に活用する。また我々は、形容詞を活用することは、名詞のみの場合よりも精度が向上することについて確認している [7]。

これらの要素の賃料推定に与える影響としては、動的情報が基本価格の推定に優れ、潜在的情報は価格の調整に優れており、全ての要素を用いた場合が最も優れていることが明らかになっている [5][7]。ゆえに、次章では全ての要素を用いて機械学習手法の比較を行う。

4. 機械学習に基づく賃料推定モデル

4.1 賃料推定モデルの構築

賃料推定システムの実現のためには、機械学習に基づいた推定モデルの構築が必要である。本章では、賃料推定モデルを多層パーセプトロンのニューラルネットワーク（NN：Neural Network）、ランダムフォレスト（RF：Random Forest）、重回帰分析（LR：Linear Regression）の 3 種類の回帰モデルでそれぞれ構築し、比較を行う。モデル構築に用いる指標は、静的情報・動的情報・潜在的情報である。学習の入力として用いる特徴量は、居抜きフラグ、駅徒歩時間、階数、坪数、地域ポテンシャル、視認性×通行量、キャッチコピーのポジティブ指

表 1: 機械学習パラメータ

アルゴリズム	パラメータ	数値
NN	中間層数	2
	ニューロン数	(6,3)
	学習係数	0.01
	L2 正則化項	0.001
	活性化関数	ReLU
	最適化手法	L-BFGS
RF	学習回数	100,000
	木の数	18
	各決定木の特徴量数	3
	木の最大深さ	15

標、ネガティブ指標の合計 8 個である

定量的な精度評価は、決定係数 R^2 と推定値の平均二乗誤差（RMSE: Root Mean Squared Error）を比較することで行う。用いる決定係数の定義は (1) 式の通りである。このとき、 y_i は実績値、 \hat{y}_i は推定値、 \bar{y} は実績値の平均を示している。実績値は実際の賃料のことを示し、株式会社 ABC 店舗で実際に契約に至った賃料である。この値にはベテラン営業マンの暗黙知を含むため、真値にすることに疑問が残る可能性がある。しかしながら、不動産価格は需要と供給のマッチングによって決定されることから、本研究ではその原則に従いこれを真値とする。決定係数の最も良いスコア値は 1.0 である。また、RMSE は (2) 式によって求める。ここで、 N は全予測対象数を示している。RMSE は推定値が実績値からどれ程乖離しているかを示しており、モデルの推定精度の悪さを評価する指標である。よって、0 に近い程優れているといえる。

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (2)$$

学習における対象物件は、東京都内の飲食店用不動産物件である。このうち、推定に使用するデータセットとしては、60 万円以下の 184 物件を対象としている。ここでは、高額物件には特殊なものが多く、推定の外れ値となりえるため排除している。このときの平均賃料は 266,987 円であり、中央値は 250,000 円である。さらに、機械学習に基いて作成されるモデルの汎化性能を評価するために、3-fold 交差検証を行う。なお、NN と RF のモデル構築に用いるパラメータは、表 1 に示す通りである。NN の中間層数とニューロン数、RF の木の数と各決定木の特徴量数はグリッドサーチによって決定している。

4.2 機械学習手法の比較

各手法を交差検証することによって得られた決定係数 R^2 と RMSE の値を表 2 に示す。決定係数は 3-fold 交差検証により、3 回分のスコアが得られるため、それらの平均値と標準偏差で評価を行う。また、実際の賃料と推定賃料の分布を図 2～図 4 に示す。表 2 より、RF を用いた場合の決定係数の平均値が 0.738 と最も高く、ばらつきも小さいことが分かる。また RMSE 値も 52494 であり、平均賃料が約 26 万円の物件に対して実用的な精度で推定できているといえる。また、LR の決定係数は 0.716 であり、RMSE は 52982 となることから、RF に迫る精度で推定できていることが読み取れる。一方で、NN

では決定係数が 0.611, RMSE が 63397 となり, 他の手法よりは劣る結果となった。これらの結果より, 飲食店用不動産向けの賃料推定には弱学習器や単純な回帰モデルの方が向いている傾向があるといえる。また, 推定賃料の分布に着目する。図中の線は $\pm 30\%$ の範囲を示しており, NN では全体的にばらつきが大きく, RF では 10 万円台の低価格域の誤差が大きいことがわかる。また, LR では推定賃料が負になるものが存在することも明らかになった。

表 2: 機械学習手法の比較結果

アルゴリズム	決定係数 R^2	RMSE
NN	0.611 \pm 0.0863	62297
RF	0.738 \pm 0.0120	52494
LR	0.716 \pm 0.0626	52982

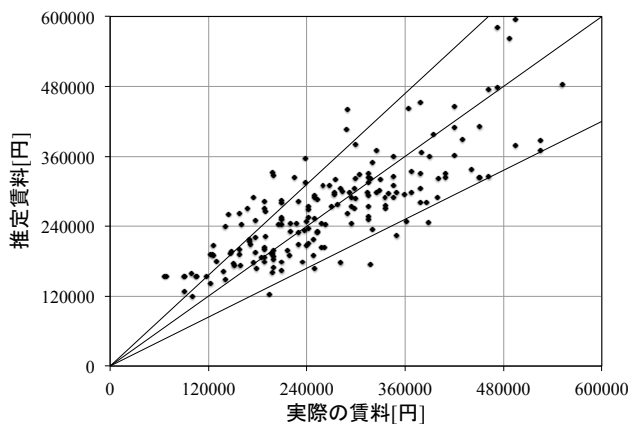


図 2: NN の推定賃料分布

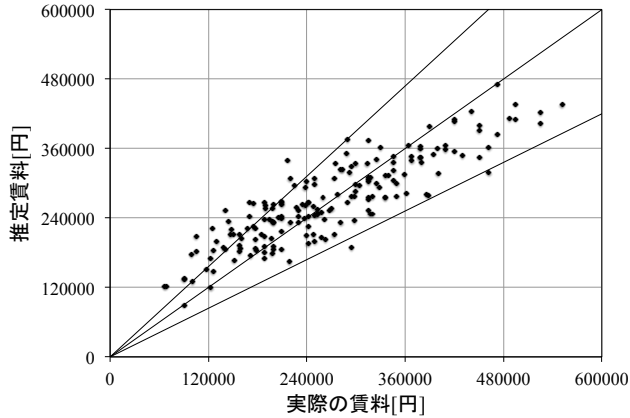


図 3: RF の推定賃料分布

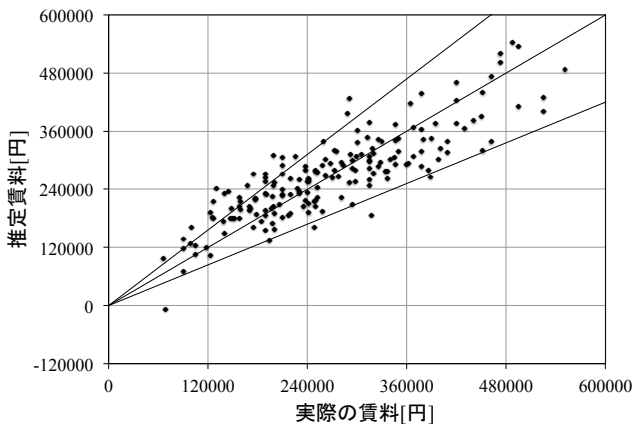


図 4: LR の推定賃料分布

5. おわりに

本稿では, 飲食店用不動産向けの賃料推定システムの実現に向けて, 機械学習に基づく賃料推定モデルの構築・検証を行った。推定モデルの構築のための回帰モデルとして, ニューラルネットワーク, ランダムフォレスト, 重回帰分析を用いて比較を行った。その結果, ランダムフォレストを用いたときに最も高いスコアが得られ, 決定係数 R^2 が 0.738, 平均二乗誤差 (RMSE) が 52494 となった。この値は, 物件の平均賃料が約 26 万円であるのに対し, 実用的な精度での推定ができていると考えられる。ただし, 10 万円前後の格安物件に対しては誤差の割合が未だ大きいことから, 低価格物件の精度向上が課題となる。

今後の課題としては, 構築した賃料推定システムを新人営業マンに活用してもらい, 推定モデルの改良に取り組むことなどが挙げられる。

謝辞

本研究の一部は科学研究費補助金 挑戦的萌芽研究 15K12161 の助成を受けたものである。また, 研究フィールドの提供にご協力いただいた株式会社 ABC 店舗の皆様へ感謝いたします。

参考文献

- [1] Victor Gan, Vaishali Agarwal, Ben Kim: DATA MINING ANALYSIS AND PREDICTIONS OF REAL ESTATE PRICES, Issues in Information System, Volume 16, Issue IV, pp.30-36 (2015)
- [2] Chih-Hung Wu, Chi-Hua Li, I-Ching Fang, Chih-Chia Hsu, Wei-Ting Lin, Chia-Hsiang Wu: HYBRID GENETIC-BASED SUPPORT VECTOR REGRESSION WITH FENG SHUI THEORY FOR APPRAISING REAL ESTATE PRICE, 2009 First Asian Conference on Intelligent Information and Database Systems, pp.295-300 (2009)
- [3] Vincenza Chiarazzo, Leonardo Caggiani, Mario Marinelli, Michele Ottomanelli: A Neural Network based model for real estate price estimation considering environmental quality of property location, Transportation Research Procedia 3, pp.810-817 (2014)
- [4] Nonaka, I., and Hirotaka, T.: The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford university press (1995)
- [5] 荒川周造, 諏訪博彦, 小川祐樹, 荒川豊, 安本慶一, 太田敏澄: 潜在的情報を用いた飲食店用不動産賃料推定モデル, 第 23 回社会情報システム学シンポジウム (2017)
- [6] 荒川周造, 諏訪博彦, 小川祐樹, 荒川豊, 安本慶一: 通行量センサを用いた飲食店用不動産賃料推定システムの提案, 社会システムと情報技術研究ウィーク in ルスツリゾート (2016)
- [7] 荒川周造, 諏訪博彦, 小川祐樹, 荒川豊, 安本慶一, 太田敏澄: 暗黙知に基づく飲食店用不動産物件の賃料推定システム, 社会システムと情報技術研究ウィーク in ルスツリゾート (2017)