

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

荒川 周造¹ 諏訪 博彦^{1,a)} 小川 祐樹² 荒川 豊¹ 安本 慶一¹ 太田 敏澄

受付日 2017年4月15日, 採録日 2017年10月3日

概要: 飲食店用不動産店舗の賃料は, ベテラン営業職員の勘や経験, 貸主の意向などの暗黙知によって決定されている。しかしながら, 従来手法では, 賃料を決定している具体的な根拠が示せない問題がある。また, ベテラン営業職員から経験の浅い新人営業職員への知識継承にも課題がある。本研究では, これらの問題を解決するために, ベテラン営業職員の暗黙知に基づいた賃料推定概念モデルの提案, および暗黙知の指標化と機械学習を用いた推定モデルを提案している。複数の機械学習手法に対して検証を行った結果, ランダムフォレストを用いたときが最も優れており, 決定係数が 0.738 と比較手法の中で最大になることを確認した。また, 平均二乗誤差 (RMSE) についても同様に確認した結果, 提案した 3 つの指標を用いてランダムフォレストを用いたときに RMSE が最小になったことから, 提案手法の有効性を確認した。

キーワード: 機械学習, データマイニング, 飲食店用不動産物件, 賃料推定, 通行量センシング, 自然言語処理

Proposal of Rent Estimation Model for Restaurant Properties based on Tacit Knowledges

SHUZO ARAKAWA¹ HIROHIKO SUWA^{1,a)} YUKI OGAWA² YUTAKA ARAKAWA¹ KEIICHI YASUMOTO¹
TOSHIZUMI OHTA

Received: April 15, 2017, Accepted: October 3, 2017

Abstract: 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.

Keywords: machine learning, data mining, real estate properties for restaurants, rent estimation, pedestrian traffic sensing, natural language processing

1. はじめに

近年, 機械学習の普及と発展にともない, 不動産分野においても, 住宅物件の価格推定 [1], [2], [3] などの営業支援が活発に取り組まれている。一方で, 飲食店を対象とした

不動産営業に着目すると, 経験豊富なベテラン営業職員が培ってきた勘や経験からなる暗黙知に依存している現状がある。この手法では, 賃料を決定している具体的な根拠が示せないため, 人が変われば価格が変わるといった問題が生じる。また, ベテラン営業職員から経験の浅い新人営業職員への知識継承が課題となっている。しかしながら, 賃料決定の根拠が不明確なためにスムーズな知識伝達ができず, OJT (On-the-Job Training) などに頼らざるをえない現状がある。また, 飲食店舗物件固有の属性が存在するた

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
² 立命館大学
Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan
^{a)} h-suwa@is.naist.jp

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ができない。具体例として、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。このような背景から、飲食店向け不動産物件の賃料はベテラン営業職員の暗黙知により決定されている現状がある。

そこで本研究では、飲食店向け不動産物件の賃料に影響する暗黙知に基づいた賃料推定モデルを提案し、実際の不動産会社で運用可能な賃料推定システムを構築することを目的とする。システムの基本となる賃料推定モデルは、野中らのSECIモデル[8]を基にして構築する。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。この考え方に基づく、ベテラン営業職員の暗黙知を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。

暗黙知を表出化する手法として、ノウハウ情報の記録やインタビューがある[4], [5], [6], [7]。本研究では、ベテラン営業職員の暗黙知を表出するために、飲食店向け不動産会社のベテラン営業職員にインタビューを実施している。インタビューの結果として、営業職員が形式知化（指標化）できる顕在的情報と形式知化（指標化）できない潜在的情報が存在することが確認された。また、顕在的情報は、物件固有の静的情報と、物件とは独立して変化する動的情報に分類できた。本研究ではこれら3つの構成要素（静的情報、動的情報、潜在的情報）を機械学習に用いることで賃料推定モデルの構築を行う。

賃料推定モデルを実現するにあたって、(1) 暗黙知の取得および指標化、(2) 機械学習に基づく賃料推定モデルの構築が課題である。

課題(1)の実現に向けて、我々は飲食店向け不動産会社の営業職員にインタビューを行い、賃料推定に関係する3つの構成要素を抽出している。静的情報は、飲食店向け不動産会社のHP上で顧客に提供されている物件情報であり、基本的に指標化されている。動的情報は、周辺地域の価値と物件の見つけやすさを指標化している。潜在的情報は、営業職員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単語と賃料の関係に基づいて指標化している。

課題(2)を実現するために、賃料推定モデルの実装・評価を行っている。本研究では、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っている。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。飲食店向け不動産

会社に確認したところ、この結果は実運用として賃料決定の参考になるとの回答を得ており、一定の精度が得られていると考える。

以下に本稿の構成について述べる。2章では関連研究について述べる。3章では、提案する賃料推定モデルについて述べる。4章では提案モデルの評価を行う。5章は考察とし、6章で結論を述べる。

2. 不動産価格推定に関する研究

機械学習を用いた不動産価格推定に関する研究として、Ganら[1]は決定木とニューラルネットワークに基づいて、米国キング郡の不動産物件を対象に推定を行っている。この研究では、取引価格の上位・下位5%の物件を除外して推定を行うことで、外れ値を取り除き、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場合に良好な結果が得られることを明らかにしている。

また、Wuら[2]は、台湾での住宅選定に影響があるといわれる風水に着目し、不動産価格推定に取り組んでいる。機械学習の特徴量には、一般的な建物固有の属性に加え、風水におけるタブーを変数として設けている。また、機械学習手法には、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を用いて比較を行っている。その結果、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、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説明し、最も優れたHGA-SVRの平均絶対パーセント誤差(MAPE)が4.79%となったことを示している。さらに、最も影響のある属性が土地の大きさであることを、BPNを用いて確認している。また、風水におけるタブーは、トイレ、窓、家の外観、ドアの順に影響があることを明らかにしている。

さらに、Chiarazzoら[3]は、ターラント市(イタリア)において、交通システムと地域ごとの環境の質が不動産価格に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)を用いて検証を行っている。特徴量には、立地条件や建物の構造に加え、交通に関する属性として駅やバス停までの距離などが、環境汚染に関する属性として、SO₂、NO_x、NO、NO₂、CO、PM₁₀の値と最大値がそれぞれ含まれている。この研究でも各属性の感度分析を行った結果、ビーチへの近さや、ガレージおよびテラスの有無が不動産価格に大きく影響していることが確認された。環境汚染に関する属性で最も重要なものとしては、SO₂の最大値があげられ、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外して検証したところ、大きな精度低下は見られず、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件の関連研究を含め実サービスとして展開されているIESHILLなどの賃料推定手法は、一般住宅を対象としているため、それらの手法をそのまま飲食店舗に適用することはできない。しかしながら、特徴量にどのような属性を活用しているかという点では参考になる。たとえば、文献[2]では、風水的情報を用いており、これはある種潜在的情報であるといえる。また、文献[3]においても、環境の質として、SO₂などの値を用いており、これは動的情報に相当している。本研究では、これらを参考にしながら、賃料推定モデルを構築する。

3. 飲食店向き賃料推定モデル

飲食店向き賃料は、まず物件所有者である貸主と飲食店向け不動産会社の営業職員との間で決定され、飲食店向け不動産会社のホームページへ掲載される。この時点では借主は決まっていないため、貸主と借主との価格交渉はできない。ホームページに掲載後、借主は掲載価格に基づいて検索や価格交渉を行うこととなる。そのため、掲載時点で賃料が安すぎた場合は、貸主は本来得られるはずだった利益を得られないことになり、掲載賃料が高すぎた場合は、借主が交渉（申込み）に至らず契約機会を逃すことになる。そのため、飲食店向け不動産会社の営業職員の決定（貸主に提案）する賃料は、適切である必要がある。

従来の賃料決定は、ベテラン営業職員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点が問題である。また、ベテラン営業職員から新人営業職員への知識伝承は、OJTや共同作業などの非効率な手法によって行われ、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するには、新人営業職員が賃料決定ルールを概念として習得し、自ら学習できるシステムがあると望ましい。本章では、システムを構築するために必要となる飲食店を対象とした賃料推定モデルの構築について述べる。

3.1 SECIモデルに基づくモデル構築

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[8]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、

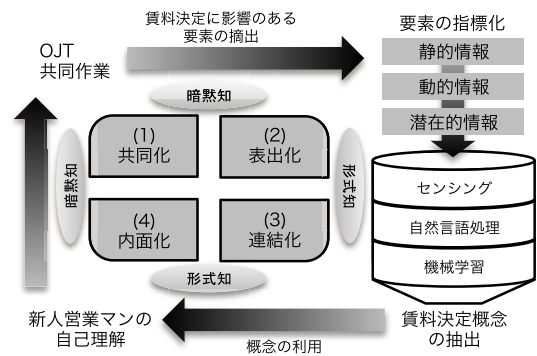


図1 SECIモデルと研究アプローチ

Fig. 1 SECI model and research approach.

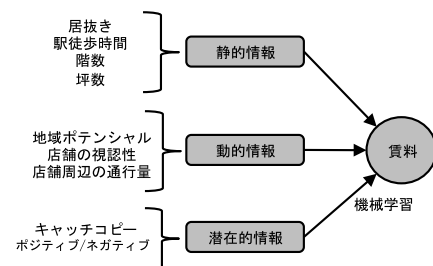


図2 賃料推定基本モデル

Fig. 2 Basic model of rent estimation.

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。この考えを援用すると、ベテラン営業職員の暗黙知は、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2に示すようなアプローチで賃料推定モデルを構築する。まず、ベテラン営業職員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。インタビューは、実際に物件の賃料を決定しているベテラン営業職員に対して行った。まず、「どのような要因が賃料に影響を与えますか」という質問を行った。営業職員からは、「立地が重要です」などの回答が得られたため、「立地とはどのような要因が含まれますか」と重ねて質問した。その結果、「場所そのものの良さでたとえば銀座や六本木などは良い立地です」や「人通りが多いところは良い立地です」、「人目に付きやすい場所です」などの回答が得られた。このように質問を重ねることで、「立地」という一言に集約されていた賃料の決定要因を詳細化していった。

また、「それ以外の要因で賃料に影響を与える要因はないですか」と、別の視点から賃料に影響を与える要因について回答を得るための質問を行った。その結果、「駅に近いと賃料は上がります」や「路面店は賃料が高くなります」などの回答が得られた。このように別の視点を促すことで、幅広い賃料の決定要因を探っていった。

さらに、「面積や築年数はどうですか」など、他に影響を与えそうな要因についても質問を行った。その結果、「当

然ですが、面積が広いと賃料は上がります」や「築年数は、新しければいいわけではなく、古くても前の借主が繁盛していれば逆に賃料は高くなることもあり一概にはいえません」などの回答が得られた。このように、営業職員からは出てこない要因についても確認することで、営業職員にとっては当たり前に含まれる要因（面積）や含まれない要因（築年数）を探っていった。

このように、1. 質問を重ねて回答を詳細化すること、2. 他の要因を促す質問をして回答を幅広くすること、3. 回答されない要因について確認する質問をすることにより、特徴量の取捨選択を行った。その結果、坪数、駅からの距離、階数、居抜き、通行量、視認性などの回答が得られた。これらの要因は、容易に指標化できる要因である。たとえば、坪数であれば広くなるほど、駅からの距離であれば近くなるほど賃料は高くなる。このように営業職員が容易に指標化できる要因を顕在的情報と呼ぶ。顕在的情報は、物件固有の静的情報と物件周辺の状況で変化する動的情報の2種類に分類できる。静的情報とは、面積、駅からの距離、階数などであり、年数が変化しても基本的に変わらない物件固有の情報である。一方、動的情報とは、地域ポテンシャルや通行量、視認性などであり、物件周辺の状況によって変化する要因である。たとえば、地域ポテンシャルや通行量は、地域の再開発や大型施設の出店/撤退により大きく変化する。また、外国人観光客の増加により浅草の人气が上昇するなど、人気となる地域は外部要因によって変化する。視認性についても、街の動線や周辺の建物の状況により変化するため、動的情報に分類している。

また、営業職員が容易に指標化できない要因を潜在的情報と呼ぶ。潜在的情報は、たとえば、「大通り沿い」や「希少物件^{*1}」など、その店舗特有の好（悪）要因が該当する。この潜在的情報を指標化する難しさは、単に大通り沿いであればいいわけではなく適度に交通量がありかつ停車ができることや、通行する車両から目に付くことなど、他の要因と複雑に関連しているところである。

このように、我々は抽出した要因について、営業職員が容易に指標化できる顕在的情報と容易には指標化できない潜在的情報に大きく2つに分類し、さらに、顕在的情報については物件固有の静的情報と物件周辺の状況により変化する動的情報として、3つの構成要素に分類している。我々の課題の1つは、これら3つの構成要素（静的情報、動的情報、潜在的情報）について、利用可能な指標として表出化することである。

また、2つ目の課題として、指標化された各構成要素と賃料との関係を概念として連結化する必要がある。図1は、インタビューに基づいて構築した賃料推定基本モデルである。本研究では、この基本モデルに基づいて機械学習を用

いることで、実営業で利用可能な賃料推定モデルを構築する。これにより各構成要素と賃料との関係を連結化する。これにより、賃料推定概念がモデル化され、営業職員は賃料推定概念を共有することが可能になる。概念がモデル化されることにより、新人営業職員への伝承が可能となる。

3.2 構成要素の指標化

インタビューによって回答された顕在的情報の各要因は、特性によって静的情報と動的情報に整理できる。また、営業職員が明確に説明できず指標化できなかった潜在的情報についても、指標化する必要がある。本節では、これらの構成要素が持つ具体的な要因の指標化について説明する。

3.2.1 静的情報

静的情報として、居抜きの有無、駅徒歩時間、階数、坪数が要因として回答された。居抜きとは、物件に付帯する什器（テーブル、カウンター、ガスレンジなどの設備）を意味し、その有無を0・1のフラグとして表現する。これらの要因は、飲食店向け不動産会社が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。

3.2.2 動的情報

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。なお、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ため、5章にて比較を行う。

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指標として定義する。積をとる理由は、たとえば視認性が良くても通行量のまったくない店舗は低評価と判断したいからである。視認性と通行量のスコアリングは、2名の営業職員が5段階で評価した平均値を用いる。営業職員による5段階評価は、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る。このことについては、6章において考察する。

3.2.3 潜在的情報

潜在的情報の取得には、営業職員が記述している物件キャッチコピーを活用する。具体例としては、「大通り沿い!」「希少な焼肉店居抜き!」「看板大きく角地で視認性良好!」などが。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。その手法としては、事前に顕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ごとに求めておき、続いて各単語ごとにそれらを含む物件を抽出し、比率の平均値を求める。本研究ではこの値を

^{*1} 新規物件が少ない地域の店舗、特別な設備（ガスの口径が多きいなど）がある店舗など

平均推定賃料比率として式 (1) のように定義する.

$$\text{平均推定賃料比率} = \frac{1}{n} \sum_{i=1}^n \frac{\text{各物件の推定賃料}}{\text{各物件の実際の賃料}} \quad (1)$$

n : キャッチコピーに対象の用語を含む物件の数

この平均推定賃料比率の値を, 用語のポジティブ・ネガティブの値として活用する. この値が 1 より小さい場合 (推定賃料 < 実際の賃料) は, 物件賃料が過小評価されており, 賃料を上げるポジティブ要因が潜在的情報として含まれている. 反対に 1 より大きい場合 (推定賃料 > 実際の賃料) は, 物件賃料が過大評価されている状況であり, 潜在的に賃料を下げるネガティブ要因が含まれている. 本研究では, 特徴のある用語を抽出するために, 値が 0.9 以下の用語をポジティブワードとして抽出し, 値が 1.1 以上の用語をネガティブワードとして抽出している.

物件ごとにポジティブ指標・ネガティブ指標を算出するために, 各物件が持つポジティブ/ネガティブワードの平均推定賃料比率の積として算出される. たとえば, キャッチコピーにポジティブワードとして「大通り」(値: 0.81) が含まれ, ネガティブワードとして「焼肉」(値: 1.31), 「角地」(値: 1.11) が含まれる場合, ポジティブ指標は 0.81, ネガティブ指標は 1.4511 となる. この処理を, すべての物件に対して行うことによって, 指標化されていなかったキャッチコピーが指標化される. すなわち, 営業職員が指標化できなかった物件の潜在的情報が指標化される.

3.3 賃料推定モデルの構築

指標化された各要因と賃料との関係を概念として連結化するために, 機械学習を用いた賃料推定モデルを構築する. 本研究では, 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 種類でモデル構築を試みる. 推定モデルの構築に用いる構成要素は, 静的情報・動的情報・潜在的情報である. 具体的特徴量としては, 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標, ネガティブ指標の合計 8 個を入力とする (図 2).

推定モデルの精度評価を行うには, 決定係数 R^2 と推定値の平均二乗誤差 (RMSE: Root Mean Squared Error) を用いる. 本研究で使用する決定係数の定義は式 (2) のとおりである. 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. ここで実績値とは, 実際の賃料のことであり, 飲食店向け不動産会社で実際に契約に至った賃料である. 賃料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 は式 (3) 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, すなわち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ことを示す指標である.

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

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

学習対象とする物件は, 東京都内にある 184 物件である. これらの物件は, 60 万円以下の飲食店向き不動産物件であり, 推定の外れ値となりうる高額な特殊物件はあらかじめ除外している. 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. また, 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 のモデル構築に用いるパラメータは, 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, 学習回数 100,000 である. また, RF のモデル構築に用いるパラメータは, 木の数 18, 特徴量数 3, 木の最大深さ 15 である.

4. 賃料推定モデルの評価

構築された賃料推定モデルを評価するために, 機械学習手法ごとの比較, 使用する構成要素ごとの比較を行う.

4.1 機械学習手法の比較

各手法 (NN, RF, LR) を検証した評価結果 (決定係数 R^2 と RMSE の値) を表 1 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を求めている. 表 1 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 52,494 であり, 平均賃料が約 26 万円の物件に対して約 2 割程度であり, 賃料の推定においては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 52,982 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 63,397 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と, NN では全体的にば

表 1 機械学習手法の比較結果

Table 1 Comparison results of machine learning methods.

アルゴリズム	決定係数 R^2	RMSE (円)
NN	0.611 ± 0.0863	62,297
RF	0.738 ± 0.0120	52,494
LR	0.716 ± 0.0626	52,982

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確認された。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。このことから、飲食店向け不動産物件の賃料推定には RF が適していると考えられる。そのため、以降の分析は RF を対象に行う。

4.2 構成要素が賃料に与える影響

推定に用いている各構成要素が、賃料に与える影響を検証するためそれぞれを組み合わせた推定モデルを作成し、3-fold 交差検証により比較を行う。学習アルゴリズムには RF を用い、データセットは表 2 に示す全 7 通りの組み合わせを用いる。モデル構築に用いるパラメータは、木の深さ=15 で固定とし、木の数・用いる特徴量の数をグリッドサーチによって決定している。

交差検証によって求められた結果を表 3 に示す。まず、最も賃料決定に影響している構成要素を抽出するため、それぞれの構成要素を単独で用いた場合で比較する。静的情報を用いた場合（データ 1）の決定係数は 0.250 であり、RMSE は 88,212 であった。また、動的情報を用いた場合（データ 2）の決定係数は 0.510、RMSE は 71,314 であった。さらに、潜在的情報を用いた場合（データ 3）の決定係数は -0.0445 であり、RMSE は 102,144 であった。よって、動的情報のみを用いた場合が最も高い決定係数が得られており、3 つの要素の中で最も精度が良いことが確認できる。ここで、静的情報と動的情報を用いた場合（データ 4）の決定係数は 0.638 であり、RMSE は 62,394 である。このことから、2 つの要素は組み合わせることで精度が向上することが確認された。

表 2 検証に用いるデータセット

Table 2 Data set using for validation.

番号	構成要素	木の数	特徴量の数
データ 1	静的情報	26	2
データ 2	動的情報	26	1
データ 3	潜在的情報	10	1
データ 4	静的・動的	18	3
データ 5	静的・潜在的	18	3
データ 6	動的・潜在的	29	3
データ 7	すべての構成要素	18	3

表 3 各要素の比較結果

Table 3 Comparison results of each factors.

データ番号	決定係数 $R^2 \pm \sigma$	RMSE (円)
データ 1	0.250 ± 0.0577	88,212
データ 2	0.510 ± 0.0445	71,314
データ 3	-0.0445 ± 0.202	102,144
データ 4	0.638 ± 0.0270	62,394
データ 5	0.447 ± 0.0692	75,190
データ 6	0.734 ± 0.00902	52,987
データ 7	0.738 ± 0.0120	52,494

一方で、潜在的情報は負の決定係数が出現し、賃料決定能力がないことが分かる。決定係数には複数の異なる定義が存在し、統計ソフトウェアによって使われる定義が異なることは、Eisenhauer ら [9] が説明している。また、Motulsky ら [10] は、残差平方和が全平方和を上回るような不適切なモデルを選択した場合に、決定係数が負になる可能性があることを説明している。このことから、潜在的情報単独では負の決定係数が出現するため、モデルが不適切であり、賃料を説明する能力がないことを示している。

しかしながら、静的情報と組み合わせた場合（データ 5）の決定係数は 0.447 であり、RMSE は 75,190 である。また、動的情報と組み合わせた場合（データ 6）の決定係数は 0.734 であり、RMSE は 52,987 である。よって、潜在的情報のみで推定を行うのは難しいが、静的情報と組み合わせるときに 0.447、動的情報とでは 0.734 となり、価格調整に優れていることが確認できる。

最後に、すべての情報を用いた場合（データ 7）の決定係数は 0.738 であり、RMSE は 52,494 である。この結果、決定係数を基準に考えると、賃料推定には、すべての情報を用いた場合が最も良い精度が得られることが確認できる。また、この結果について営業職員に確認した結果、「現場で十分参考になる」との回答を得ている。これらの物件は数万円単位の価格操作は日常的に行われており、数十万円オーダで決定される賃料に対して十分な精度であるといえる。

5. 考察

5.1 動的情報の指標化に関する考察

本研究では地域ポテンシャルの指標として、駅推定賃料を利用している。しかしながら、他の指標として路線化の利用も考えられる。そこで、路線価を利用して賃料推定を行い、駅推定賃料を活用した場合との比較を行う。

保有しているデータセットでは、路線価のデータがない物件も含まれるため、それらはあらかじめ除外した 169 物件で比較する。比較するのは動的情報のみと、すべての情報を活用した場合であり、推定に用いるパラメータは表 4 のとおりである。表 5 に推定結果を示す。駅推定賃料を使用した際の結果が、表 3 と異なるのは、モデル構築に活用している物件数が異なるためである。結果より、動的情報

表 4 地域ポテンシャル検証時の機械学習パラメータ

Table 4 Machine learning parameters for area potential testing.

種類	用いる要素	木の数	特徴量の数
駅推定	動的情報	17	2
路線価推定	動的情報	16	2
駅推定	すべての要素	18	3
路線価推定	すべての要素	18	3

表 5 地域ポテンシャルの比較結果

Table 5 Comparison results of area potentials.

種類	用いる要素	決定係数 $R^2 \pm \sigma$	RMSE 値
駅推定	動的情報	0.427 ± 0.0251	77,960
路線価推定	動的情報	0.172 ± 0.0835	92,750
駅推定	すべての要素	0.690 ± 0.0112	57,925
路線価推定	すべての要素	0.665 ± 0.0175	60,604

表 6 地域ポテンシャルと物件の見つけやすさの比較

Table 6 Comparison between area potential and property finding easiness.

変数	木の数	決定係数 $R^2 \pm \sigma$	RMSE 値
駅推定賃料	15	0.352 ± 0.0668	83,982
視認性 $\times$ 通行量	11	0.0354 ± 0.0528	100,389

のみを用いた場合と、すべての要素を用いた場合の両方において、駅推定賃料の方が賃料の決定に優れていることが分かる。

さらに、動的情報の中でも地域ポテンシャルと物件の見つけやすさの影響する度合いを検証するため、それぞれの変数のみで推定を行う。このときの結果を表 6 に示す。結果より、動的情報の中でも駅推定賃料は決定係数が高く、賃料の基本推定能力に優れていることが分かる。また、この値は、すべての静的情報を用いた場合よりも高く、物件の周辺価値は重要度が高いといえる。一方で、視認性 $\times$ 通行量は、単体では賃料決定能力が乏しい。しかしながら、駅推定賃料と組み合わせることで、価格調整の役割を担っていると考えられる。

5.2 動的情報の取得に関する考察

本研究では、動的情報の 1 つとして通行量を用いており、2 名の営業職員が 5 段階でスコアリングした平均値を用いているが、適切に判断するには経験が必要である。この問題の解決のために、通行量を取得可能なセンサを独自開発している [11]。

これまでの通行量センシングに関する研究として、Yan ら [12] は、スマートフォンの bluetooth を用いた方法を提案しているが、bluetooth 機能を有効にしていない歩行者や端末を所持していない歩行者などの検知が行えない。また、Pinelli ら [13] は、携帯電話の位置情報を用いる手法を提案しているが、携帯電話の位置情報は通信事業者以外では取得が難しいといった問題がある。Pane ら [14] や Ma ら [15] は、カメラを用いたシステムを提案しているが、物件オーナーの許可やプライバシー面でのハードルが高く、飲食店用不動産物件には適していない。設置面では電源線が必要となる場合がほとんどであり設置コストが高い。プライバシーに配慮したセンサを用いたアプローチとして、Wahl ら [16] や Zhu ら [17] は、受動型赤外線 (PIR: Passive Infrared Ray) センサを用いた手法を提案しているが、電源線の敷

表 7 各地点におけるセンサの反応回数

Table 7 Sensor counts in each places.

物件	17:00~23:59	0:00~13:00	合計	1 時間あたり
A(1.5)	58	145	203	10.15
B(2)	827	1,526	2,353	117.65
C(3)	2,677	5,375	8,052	402.6
D(4)	4,204	8,612	12,816	640.8

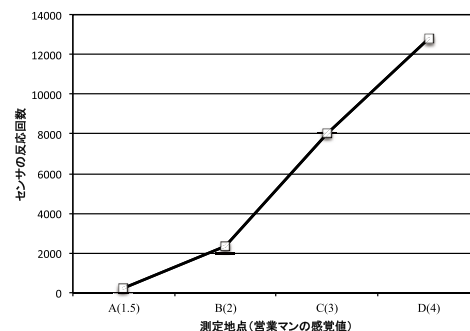


図 3 センシングデータと主観データの比較

Fig. 3 Comparison between sensor data and subjective data.

設については考慮していない。

これらをふまえ、本研究においては、PIR センサを用いたエナジーハーベスティングセンサを新たに開発することで、店舗への設置に適したシステム構築している。カメラと異なり個人の識別をしないためプライバシーに配慮でき、低消費電力であるためセンサを小型にできる特長がある。また、PHS を接続した通信用ゲートウェイを別で設けており、遠隔地でもデータ取得を容易にしている。

プロトタイプとして構築した通行量センシングシステムの有用性を確認するために、営業職員の感覚に基づく通行量と、実際に通行量センサで取得した値の比較を行った。検証には、営業職員の主観データが異なる 4 つの物件を抽出し、通行量センシング実験を実施した。ここで抽出した地点は、それぞれ A(1.5), B(2), C(3), D(4) とし、括弧内の値は主観データ値 (1~5: 0.5 刻み) を示している。

実験によって得られた結果を表 7 および図 3 に示す。グラフは地点 A, B, C, D それぞれにおけるセンサの反応回数を示しており、人の感覚に比例して増加していることが分かる。したがって、本システムのように賃料に影響を与える要素を機械的に取得可能なシステムを用いれば、構成要素の指標化を人に頼らずに自動化することが可能となる。一方で、視認性については間口面積や看板の大きさなどで指標化できる可能性もあるが、周囲にある植え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。

5.3 潜在的情報の指標化に関する考察

本研究では、潜在的情報の取得には、営業職員が記述している物件キャッチコピーを活用している。これまでも、

テキスト中から用語を抽出し、利用する取り組みは多くなされている。Shimoda ら [18] は、特定分野で世界的に高いシェアを持つ GNT (global nice top) 企業に共通する成功要因を理解するため、GNT 企業に関する自由記述データをテキストマイニングし、さまざまな産業の成功要因に関する情報を取得している。また、Chang ら [19] は、Twitter のつぶやきから、ユーザの位置を予測する問題に取り組んでいる。このほかにもテキスト中から用語を抽出して活用する研究は数多くなされている [20], [21], [22], [23], [24]。一方で、不動産の賃料推定にこうした手法を用いている前例はないことから、その点において本研究には独自性があるといえる。

抽出されたポジティブワード群の具体例としては、「黒字、バス停、会社、勝どき駅、すし…」などがある。まず、「黒字」からはその物件が儲かることを想像できる。また、「バス停」、「大通り」、「交差点」、「繁華街」といった、人通りが多く予想されるワードが含まれる。さらに、「会社」、「大学」、「オフィス街」、「高層ビル」からは、学生やサラリーマンが多く集まる場所では、高い賃料が見込めることを読み取れる。加えて、「勝どき駅」や「浅草」といった特定の場所では賃料を高め設定できることが分かる。また、「すし」、「ホルモン」といった料理店の居抜きを持つ場合も、高い賃料を見込めるといえる。

一方で、ネガティブワード群としては、「値下げ、駅前、環八通り、蕎麦、韓国料理…」などがある。ネガティブワードは、ネガティブな物件という意味ではなく、価格が抑えられた物件と解釈できる。まず、「値下げ」、「格安」、「キャンセル」といったワードからは、売れない/売れなかった物件であることが読み取れる。つまり、「売れなかったので値下げしたためお買い得である」という意味でキャッチコピーとして掲載される。キャッチコピーとしては、より早く契約を得るためのポジティブな行為であるが、賃料という観点では値段を押し下げるネガティブな要因が存在することを表している。

また、「駅前」、「商店街」といった一見よさそうな条件も含まれている。便利そうな場所にあるからといって、良い物件であるとは限らないことが読み取れる。また、近年では、シャッター通りになっているような商店街も珍しくなく、来客が見込めない可能性も考えられる。「環八通り」、「昭和通り」、「靖国通り」、「自由通り」からは、店舗が面している道路も賃料に影響することが分かる。これらには通りに面していることを推しているが、駅からは遠かったり不便な場所にある可能性がある。さらに、「蕎麦」、「韓国料理」、「手打ちうどん」、「ビストロ」、「タイ料理店」、「焼肉」、「小料理」、「焼鳥」、「ラーメン店」といった、特定の居抜きの種類を示す単語が多く見られる。これらについて、営業職員にインタビューを行ったところ、「確かに蕎麦や韓国料理の居抜き物件は売れにくい」といったコメントが

得られたことから、こうしたワードにも賃料に対して一定の影響力を持っていることが示唆された。

5.4 提案手法の優位性

評価の結果、提案手法は既存手法（ベテラン営業職員の勘や経験）と同レベルの性能で価格推定ができることが確認された。しかしながら、提案手法の優位性は、単にベテラン営業職員と同レベルの性能で価格推定ができることではなく、従来不明確であったベテラン営業職員の着眼点が、提案手法によって明らかになったことである。これは営業品質の向上という観点において、飲食店向け不動産会社や新人営業職員にとって画期的である。

5.3 節の最後でも述べたとおり、潜在的情報において特定の居抜きの種類を示す単語は、売れにくい物件の特徴を示しており、前借主の営業形態が結果として賃料に影響を与えることを示している。このことは、ベテラン営業職員にとっては暗黙知であったが、提案手法に基づくフィードバックにより形式知として表出化されている。このように、ベテラン営業職員の暗黙知を飲食店向け不動産会社において組織知として獲得できる点は、提案手法の優位性の 1 つである。

さらに、形式知として表出化されることにより、新人営業職員への伝承が可能となる。新人営業職員は、物件の立地や広さだけでなく、前借主の営業形態も考慮する必要があることが理解できる。加えて、「蕎麦」、「韓国料理」など具体的に賃料が安くなりやすい営業形態を知ることができる。このように、提案手法によりベテラン営業職員の暗黙知を新人営業職員へ継承できることは、提案手法の優位性の 1 つである。

5.5 賃料推定システムのプロトタイプ

賃料推定モデルに基いて、実際に新人営業職員への伝承を行うためには、推定モデルを営業現場で利用可能にするシステムが必要である。具体的には、推定モデルを新人営業職員が自身で活用し、各種条件に対する推定結果を学習することで、知識の伝承を行える狙いがある。また、システムを利用することによって、これまで業務とは別に収集していたスコアリング値（視認性、通行量）を自然に収集可能となる。これらを活用しモデルの学習・検証を繰り返すことによって、将来的に賃料推定モデルのさらなる精度向上が期待できる。さらに、実際に営業職員に活用してもらうことで、改善点の発見につながるが見込める。そこで、推定モデルを営業現場で利用可能にするためのシステム化を行った。構築した賃料推定システムの利用画面を図 4 に示す。

本システムは飲食店向け不動産会社内に設置しているローカルサーバに導入し、Web アプリケーションとして営業職員の各 PC から利用可能にしている。システムは

図 4 賃料推定システム利用画面

Fig. 4 Rent estimation system interface.

Python2.7で構築しており、機械学習ライブラリにはscikit-learnを、WebフレームワークにはFlaskをそれぞれ使用している。使い方としては、各特徴量を選択または入力し、推定ボタンを押下することで、推定賃料が上部に表示されるように構築している。また、推定賃料とは別に、地域ポテンシャルである「駅推定賃料」を下部に表示している。さらに、推定には活用していないが、物件の住所を入力すれば、データセット中の最寄り5つの物件をジオコーディングにより抽出し、下部に表示する機能も実装している。

これらの機能を新人営業職員が活用し、物件やパラメータを変えながら結果を見ることや、賃料に影響を与える要因（特徴量）を学習することによって、ベテラン営業職員の暗黙知の継承につながる。また、入力したこれらの情報は、アクセス元PCのIPアドレスと共に、CSVファイルに記録されるようにしており、入力者、新規物件ID、新規物件に対するスコアリング値の3つを紐付けて収集できる。これにより、これまでベテラン営業職員に依頼してスコアリングしてもらっていた視認性・通行量の主観データを自動的に収集できるようになり、今後の精度向上のためのシステムの更新に活用可能となる。また、スコアリング値は個人差があることが分かっており、スコアリングするにも経験が要求される。そのため、新人営業職員のスコアよりもベテラン営業職員のスコアを優先するなどの対応が必要となるが、入力者を特定することで対応可能である。

さらに、本システムが精度向上を繰り返した後に、既存物件に対して賃料推定を行えば、既存賃料の外れ値の探索にも活用できる。具体的にはオーナーの意向やこれまでの風習など、システムで考慮されていない要因によって、値付けされている物件の発見に活用できる。なお、本システム

は現在飲食店向け不動産会社内にて稼働中である。

6. 結論

本研究では、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52,494となった。これは平均賃料が約26万円の物件に対して約2割の誤差で推定できているという結果であり、この結果について、飲食店向け不動産会社に確認したところ、実運用として賃料決定の参考になるとの回答を得ており、一定の精度が得られていると考える。

また、各構成要素が賃料推定に与える影響についても検証を行った。その結果、動的情報がもっとも賃料推定に影響を与えている明らかとなった。さらに暗黙知として特に着目した潜在的情報は、それ自体では賃料推定能力は低いものの、顕在的情報（静的情報、動的情報）と組み合わせることで、優れた賃料推定能力を発揮することが明らかとなった。

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職員に活用してもらい、より暗黙知の継承に有効なモデルや利用方法を検討することが考えられる。また、更なる精度向上のため推定モデルの改良に取り組むことなどがあげられる。

参考文献

- [1] Gan, V., Agarwal, V. and Kim, B.: Data Mining Analysis and Predictions of Real Estate Prices, Issues in Information System, Vol.16, No.IV, pp.30-36 (2015).
- [2] Wu, C.-H., Li, C.-H., Fang, I.-C., Hsu, C.-C., Lin, W.-T. and Wu, C.-H.: Hybrid Genetic-based Support Vector Regression with Feng Shui Theory for Appraising Real Estate Price, 2009 1st Asian Conference on Intelligent Information and Database Systems, pp.295-300 (2009).
- [3] Chiarazzo, V., Caggiani, L., Marinelli, M. and Ottomanelli, M.: A Neural Network based model for real estate price estimation considering environmental quality of property location, Transportation Research Procedia, Vol.3, pp.810-817 (2014).
- [4] 菅谷克行, 上野恵美子: 熟練技術の伝承支援に向けて, 茨城大学人文学部人文コミュニケーション学科論集 7, pp.219-232 (2009).
- [5] 成子由則: モノづくりにおける知識・ノウハウの伝承, 情報管理, Vol.49, No.8, pp.439-448 (2006).
- [6] 橋本峻平, 関良明, 諏訪博彦: 短期的な世代交代のある組織における注記事項伝承システム LEAVES, 情報処理学会論文誌, Vol.52, No.1, pp.121-130 (2011).
- [7] 大野光太郎, 小川祐樹, 諏訪博彦, 太田敏澄: 東京消防庁における消防活動経験の伝承を支援するSNSの提案, 情報処理学会論文誌, Vol.54, No.1, pp.121-130 (2013).
- [8] Nonaka, I. and Takeuchi, H.: The knowledge-creating

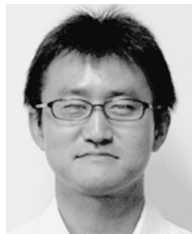
company: How Japanese companies create the dynamics of innovation, Oxford university press (1995).

- [9] Eisenhauer, J.G.: Regression through the Origin, Teaching Statistics, Vol.25, pp.76-80 (2003).
- [10] Motulsky, H. and Christopoulos, A.: Fitting models to biological data using linear and nonlinear regression, Oxford University Press (2004).
- [11] 荒川周造, 諏訪博彦, 小川祐樹, 荒川 豊, 安本慶一, 太田敏澄: 通行量センシングと機械学習に基づく飲食店用不動産賃料推定, マルチメディア, 分散, 協調とモバイル (DICOMO 2016) シンポジウム (2016).
- [12] Yan, Z. and Yang, J.: Emmanuel Munguia Tapia, Smart-phone bluetooth based social sensing, *Proc. 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, pp.95-98 (2013).
- [13] Pinelli, F., Di Lorenzo, G. and Calabrese, F.: Comparing Urban Sensing Applications Using Event and Network-Driven Mobile Phone Location Data, *2015 16th IEEE International Conference on Mobile Data Management*, pp.219-226 (2015).
- [14] Pane, C., Gasparini, M., Prati, A., Gualdi, G. and Cucchiara, R.: A People Counting System for Business Analytics, *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp.135-140 (2013).
- [15] Ma, H., Zeng, C. and Ling, C.X.: A Reliable People Counting System via Multiple Cameras, *ACM Trans. Intelligent Systems and Technology (TIST)*, Vol.3, No.2, pp.1-22 (2012).
- [16] Wahl, F., Milenkovic, M. and Amft, O.: A distributed PIR-based approach for estimating people count in office environments, *2012 IEEE 15th International Conference on Computational Science and Engineering*, pp.640-647 (2012).
- [17] Zhu, F., Yang, X., Gu, J. and Yang, R.: A New Method for People-Counting Based on Support Vector Machine, *2009 2nd International Conference on Intelligent Networks and Intelligent Systems*, pp.342-345 (2009).
- [18] Shimoda, A. and Yabuki, T.: Success Factors in Global Niche Top Companies: Analysis of Free Description Data Using Text Mining, *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp.828-833 (2016).
- [19] Chang, H.-W., Lee, D., Eltaher, M. and Lee, J.: @Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage, *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp.111-118 (2012).
- [20] 諏訪博彦, 梅原英一, 太田敏澄: インターネット株式掲示板の投稿内容分析に基づくファクターモデル構築の可能性, 人工知能学会論文誌, Vol.27, No.6, pp.376-383 (2012).
- [21] Suwa, H., Umehara, E. and Ohta, T.: Using the Factor Model to Analyze Internet BBS Messages and Stock Returns, *SICE2011* (2011).
- [22] Wysocoki, P.D.: Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards, Working paper, University of Michigan (1999).
- [23] Tumarkin, R. and Whitelaw, R.F.: News or Noise? Internet Postings and Stock Prices, *Financial Analysts Journal*, Vol.57, pp.41-51 (2001).
- [24] Antiweiler, W. and Frank, M.Z.: Internet Stock Message Boards and Stock Returns, University of British Columbia Working Paper (2002).



荒川 周造

2017 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在, 株式会社デンソーに勤務。



諏訪 博彦 (正会員)

2006 年電気通信大学大学院情報システム学研究科博士後期課程修了。博士(学術)。2014 年より奈良先端科学技術大学院大学助教。社会情報システム学に関する研究に従事。



小川 祐樹 (正会員)

2011 年電気通信大学大学院情報システム学研究科博士課程修了。博士(工学)。現在, 立命館大学情報理工学部助教。ソーシャルメディアにおける意見形成等の研究に従事。



荒川 豊 (正会員)

2006 年同大学院理工学研究科博士課程修了。博士(工学)。2013 年より奈良先端科学技術大学院大学准教授。ネットワークアプリケーション, ソーシャルマイニングに関する研究に従事。



安本 慶一 (正会員)

1995 年大阪大学大学院博士後期課程退学。博士(工学)。2011 年より奈良先端科学技術大学院大学教授。ユビキタスコンピューティングに関する研究に従事。



太田 敏澄

元電気通信大学教授。2016 年逝去。1977 年東京工業大学博士課程修了。博士(工学)。『社会情報システム学序説』や多くの論文を通じて, 社会情報システム学の発展に尽力した。