

# 幸福度予測モデルに基づく居住エリアレコメンドシステムの開発

鳥居 克哉<sup>†</sup>

東京大学大学院新領域創成科学研究科<sup>†</sup>

宮下 知也<sup>§</sup>

東京大学大学院新領域創成科学研究科<sup>§</sup>

青木 龍人<sup>‡</sup>

東京大学大学院工学系研究科<sup>‡</sup>

稗方 和夫<sup>¶</sup>

東京大学大学院新領域創成科学研究科<sup>¶</sup>

## 1 序論

近年メディアなどにより人々の「モノ消費」から「コト消費」への価値観の変化が注目されており、その影響は不動産業界においても例外ではない。消費者は新たな住まいへの転居を機に、日々の生活や体験にどのような変化があるかといった点に価値を見出し始めていると言える。

以上の背景を踏まえて本研究では、各地域に住んだ場合の将来的な幸福度の予測値と、その地域にて希望の物件が見つかる度合いを表す成約率との二軸における散布図を示すことで、住まい探しを行うユーザーにとってより良い居住エリアの選択を可能にするレコメンドシステムを提案する。

## 2 システムの構築

### 2.1 システムの概要

本研究で開発したシステムのプロット図を図1に示す。

最初に、ユーザーは複数の項目からなる趣向アンケートに回答する。続いて、回答内容のデータが幸福度予測モデルに渡される。幸福度予測モデルは「生活者総合調査2019」のデータを用いて学習したもので、各地域ごとにユーザーがその地域に住んだ場合の幸福度の予測値を配列として出力する。なお、今回対象とする地域は東京都23区とした。

最後に、上記幸福度の予測値と各地域における「希望物件の見つかりやすさ」を表す成約率（詳し

くは後述）との二軸による散布図を表示する。ユーザーはこの結果を基に、パレート解の中から住むのにより良い地域を判断することができる。

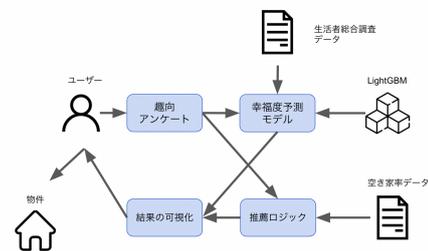


図1 システムのプロット図

### 2.2 幸福度予測モデル

今回はモデルの構築にあたって、人々の日々の生活や価値観に関する様々な項目についてアンケートをとった「生活者総合調査2019」による13856人分の調査回答データを用いる。13856人分のデータを9:1の割合で訓練データとテストデータに分け、モデルの学習および予測を行う。目的変数である幸福度は同調査内の質問項目の一つであり、実際の回答データを正解値として用いる。

#### 2.2.1 モデル概要

システムの利便性を考慮すると、説明変数の数ある程度限定する必要がある。今回は、まず同調査の調査項目から幸福度に寄与しそうな変数を51個抽出し、変数ごとの重要度を導出することができるLightGBM[1]による学習を行う。その後、重要度の大きいものから順に変数を選択する。

学習の結果、幸福度への寄与が特に大きいと考えられる変数として「通勤通学エリア」、「居住エリア」、「世帯年収」、「職業」、「住宅の広さ」、「所有資産」、「個人年収」、「同居の子ども有無」、「家族と過ごす時間をなるべく多くしたいか」、「休みの日には家族そろって出かけるか」、「休日にはそろって出か

Development of a residential area recommender system based on a happiness prediction model

<sup>†</sup> Katsuya Torii, The Graduate School of Frontier Sciences, The University of Tokyo

<sup>‡</sup> Ryuto Aoki, School of Engineering, The University of Tokyo

<sup>§</sup> Tomoya Miyashita, The Graduate School of Frontier Sciences, The University of Tokyo

<sup>¶</sup> Kazuo Hiekata, The Graduate School of Frontier Sciences, The University of Tokyo

けるような、仲の良い家族が望ましいか」が挙げられることが分かった。次項にて、変数をこれら 11 個に限定したモデルと全ての変数を用いたモデルについて、テストデータに対する予測精度の比較を行う。

### 2.2.2 モデルの精度評価

全ての変数を用いたモデルと変数を上述の 11 個に限定したモデルの予測精度を、平均二乗誤差 (RMSE) と決定係数 (R2) の観点から比較した (表 1)。なお、予測の精度を上げるため、ハイパーパラメータ自動最適化フレームワーク Optuna[2] を用いてモデルのハイパーパラメータチューニングを行った。また、ベースラインモデルとしてガウスカーネルを用いた Support Vector Regression (SVR) による予測精度も併せて示している。

表 1 より、全ての変数を用いた場合と変数を限定した場合を比較すると、前者の予測精度を大きく落とすことなく変数を限定できていることが分かる。

表 1 全ての変数を用いたモデルと変数を限定したモデルの予測精度比較

モデル	RMSE	R2
SVR(全変数)	2.006	0.178
LightGBM(全変数)	2.000	0.183
SVR(11 個の変数)	2.044	0.146
LightGBM(11 個の変数)	2.031	0.181

### 2.3 成約率の算出

各地域における、ユーザーが希望する間取りを有する物件の家賃が「希望する間取りの家賃相場額を平均、その 1/3 の値を標準偏差とする正規分布」に従うと仮定する。各地域について、ユーザーの希望家賃額がとる確率密度値を計算し、最小値が 0、最大値が 10 となるようにスケールリングしたものを成約率とした。ここで、家賃相場については物件情報サイト「SUUMO」が公開しているデータを用いた。

### 2.4 実装

ユーザーのアクセス性向上およびデータ分析モデルとの接続を考慮して、システムは Web アプリケーションの形で提供する。アンケート項目を入力し終わると、バックエンドから結果が引き渡され、画面上に Recommend 結果が表示される (図 2 を参照)。ユーザーは赤い点で示されるパレート解の中から居住エリアを選択する。

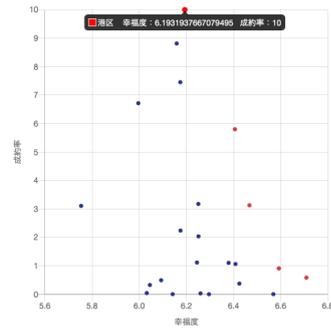


図 2 アプリケーションの Recommend 結果の一例

## 3 システムの評価

本システムの有効性を検証するため、テストユーザーに対してシステム利用に関する調査を行った。「居住地選びにおいて幸福度と成約率のどちらを重視するか」という質問に対してばらつきのある回答が得られ、本システムがユーザーの価値観に合わせた居住地選びの機会を提供できていると言える。

一方で、「入力した情報が幸福度にどう結びついているのかが不透明である」という意見が複数みられた。よって、システムの説明性という点においては改善の余地があると言える。

## 4 結論

本稿では幸福度と成約率の二軸により居住エリアを Recommend するシステムを提案し、テストユーザーによる調査を基に一定の有効性を示した。今後の課題としては上述の通り、より説明性の高いシステムの実現を目指すなどが挙げられる。

### 謝辞

本研究では株式会社 ADK マーケティング・ソリューションズからデータの提供を受けました。

### 参考文献

- [1] Guolin Ke, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, Vol. 30, pp. 3146–3154, 2017.
- [2] Takuya Akiba, et al. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.