

不動産ポータルサイトでのユーザの潜在的嗜好の抽出

Extraction of User Latent Preference for Real Estate Website

池上 顕真
Kenshin Ikegami

伊藤 博典
Hironori Ito

野村 眞平
Shimpei Nomura

*1 株式会社リクルート住まいカンパニー
Recruit Sumai Company Ltd.

*2 京都大学情報学研究科
Graduate School of Informatics
Kyoto University

*1 株式会社リクルート住まいカンパニー
Recruit Sumai Company Ltd.

As a current trend, in the consumption of durable consumer goods such as cars, real-estate properties, increasing number of buyers are making their purchasing decisions based on information on the Internet. Despite that, since real-estate contraction process is barely common or frequent, more users (consumers) than not suffer from a lack of adequate knowledge. Consequently, searching for a house on-line could become confusing. Renting a house, specifically, often comes time restriction, and from times requires compromise in search terms. To help, search terms and listing recommendations have to be more properly generated according to each user's needs and preferences. In this analysis, we propose a method of user preference extraction and provide insights on the results.

1. はじめに

近年、車の購入や、不動産契約などの耐久消費財の商取引において、ユーザがウェブ上の情報を元にして購買の意思決定を行う機会は増加している。一方で、不動産契約は頻繁に行う取引ではないため、消費者であるユーザが十分な知識を持っていないことも多い。その結果、知識が十分でないことにより適切な探し方できず、特に賃貸物件においては選定期間に制約があることが多いため、妥協した選定をせざるを得ない場合がある。この問題を解決するためには個人に応じた検索条件や、物件の推薦をより適切に行う必要がある。今回の分析では、これらを実施するためにユーザの潜在的嗜好の抽出を行ったのでこの結果について報告を行う。

2. SUUMO について

2.1 SUUMO に蓄積されているデータ

不動産情報ポータルサイトである SUUMO には様々なデータが蓄積されている。各企業から提供される物件情報や施工事例といった不動産データを始めとして、各不動産会社のデータ、ユーザが SUUMO の Web サイトやスマートフォンアプリケーションを利用した際の行動履歴データ、SUUMO 独自で収集を行っている街情報や、住みたい街ランキングに代表されるアンケートなどの調査データをはじめ、様々なデータが存在している。本稿では、上記のうち不動産データと行動履歴についてそれぞれ簡単に紹介する。

§ 不動産データ

本稿における不動産データとは、主に不動産会社から提供される物件情報や施工事例などのデータのことを指す。例えば物件の緯度経度や最寄り駅などの地理情報や、その物件が何階に所在し、どのくらいの広さを持ち、どんな設備を持っているかといったスペックの情報、どのような物件かを視覚的に確認できる画像情報、そして現在から過去にかけていくらか掲載されていたのか、という価格情報など多岐にわたった情報が蓄積されている。

§ ユーザの行動履歴

行動履歴とは、ユーザが SUUMO の Web サイトやスマートフォンアプリケーションを利用した際のアクセスログを指す。cookie 毎にどのページにいつアクセスし、どんな検索を行い、どんな物件を比較し、どの不動産会社に問い合わせたのか、などの情報が含まれている。

3. 分析事例

3.1 分析の背景

SUUMO の Web サイト上では、ユーザは家賃などの検索条件を複数設定しながら、物件の検索を行うことができる。このようにユーザが明示的に指定する検索条件に対して、検索条件には設定していないが、ユーザが暗黙的に重視している物件の条件があると仮定する。このような条件のことを本稿ではユーザの潜在的嗜好と呼ぶ。ユーザの潜在的嗜好の抽出を行うことができれば、レコメンドシステムなどへの応用が可能になる。

3.2 先行研究

土方(土方, 2005)は、Web サイトを利用しているユーザの嗜好抽出を行う方法として、2種類の手法があると述べている。1つはユーザに興味に関する情報を入力してもらう明示的な方法である。もう1つは、Web 閲覧時の挙動からユーザの興味に関する情報を取得する暗黙的な手法である。暗黙的な手法は、閲覧した情報から手がかりを利用してユーザが閲覧した情報への興味の有無を推定する手法である。SUUMO の Web サイトを利用しているユーザに当てはめると、閲覧した情報とは、ユーザが閲覧した物件のことを指し、手がかりとは物件を閲覧した時間などを指す。

明示的な手法を SUUMO の Web サイトの閲覧履歴に応用した例では、布施ら(布施, 2018)の分析がある。彼らは、ユーザが入力した複数検索条件を分散表現化することで、検索条件の可視化を行なった。

浅沼ら(浅沼, 2011)は暗黙的な手法によってユーザのレシピ閲覧履歴からレシピの推薦システムの構築を行なった。彼らが提案する手法では、閲覧したレシピに使用される食材に注目し、食材を素性とした TF-IDF 表現でレシピを表現することにより、推薦するレシピのスコアリングを行なった。

3.3 提案手法

本項では、SUUMO の Web サイトの行動履歴を用いて、ユーザの潜在的嗜好を抽出する手法について述べる。手法を考える上で、ユーザの物件閲覧行動には以下の傾向があると仮定した。

仮説 1: ユーザは物件に求める条件に対して内的優先順位を持つ

仮説 2: 複数の物件を閲覧していく中で、物件に求める条件の内的優先順位は変動する

仮説 3: 物件に求める条件の内的優先順位は物件を閲覧するごとに変動するが、その中でも常に上位に存在する条件が存在する

ユーザが仮説 1, 仮説 2 に基づいて物件の閲覧を行なった際に、仮説 3 に当てはまる物件の条件をユーザの潜在的嗜好とし、条件の抽出を行なった。具体的には以下の手順に基づいて潜在的嗜好を抽出した。

まず、ユーザが物件を閲覧するたびに、直近閲覧した N 件の物件を用いて、ユーザが物件に求める条件の内的優先順位を求める。 N 件の物件を 1 つの文書、集合内に含まれる物件の条件を単語とみなし、物件条件の TF-IDF 値を求める。 j 番目の物件を閲覧した際の、物件条件 c_i の TF-IDF 値は以下の式で求める。

$$TFIDF_{i,j} = \frac{n_{i,j}}{DF_i \sum_k n_{k,j}} \quad (1)$$

ただし、 $n_{i,j}$ は直近閲覧した N 件中、条件 c_i を満たしている物件の数、 DF_i は全物件中、条件 c_i を満たしている物件の数を表している。TF-IDF 値を元に、物件条件を降順にソートし、 j 番目の物件を閲覧した際の物件条件 c_i の内的優先順位 $rank_{i,j}$ を求める。

次に物件を閲覧するたびに求めた物件条件の内的優先順位から、常に上位に存在し、かつ順位の変動幅が小さい条件を抽出する。具体的には、以下の式を用いて、 j 番目の物件を閲覧した際の物件条件 c_i の平均順位 $rankave_{i,j}$ 、物件条件 c_i の平均順位変動 $rankdiff_{i,j}$ を求める。

$$rankave_{i,j} = \frac{\sum_{t=1}^j rank_{i,t}}{j} \quad (2)$$

$$rankdiff_{i,j} = \frac{\sum_{t=2}^j |rank_{i,t-1} - rank_{i,t}|}{j-1} \quad (3)$$

最後に平均順位 $rankave_{i,j}$ が上位 $X\%$ かつ平均順位変動 $rankdiff_{i,j}$ が上位 $Y\%$ の条件 c_i を抽出し、これを j 番目の物件を閲覧した際のユーザの潜在的嗜好とみなす。

3.4 分析準備

本節では実験に用いたデータ、評価指標において説明を行う。

関東地方の賃貸物件を閲覧しているユーザを分析の対象とした。期間 2017-01-01 ~ 2017-01-31 で関東地方の賃貸物件を閲覧しているユーザの内、CV (これ以後、CV は閲覧した物件の資料請求を行うことを指す) 以前に 20 件以上、100 件未満の物件を閲覧しているユーザの中からランダムに抽出した 100 人を実験対象とした。

手法の評価には、各ユーザから抽出した潜在的嗜好と CV した物件の条件との適合率、再現率、F-measure の平均を用いた。またベースラインとして、ユーザが閲覧した物件からランダムに抽出した物件の物件条件、ユーザが閲覧した物件の中で最も閲覧時間が長い物件の物件条件と、同じく CV した物件の条件との適合率、再現率、F-measure の平均を算出した。

3.5 分析結果

提案手法の定性評価と定量評価を示す。あるユーザの CV 直前までの閲覧履歴から抽出した潜在的嗜好と、その直後に CV した物件の物件条件を表 1 に示す。

表 1: 閲覧履歴から抽出した潜在的嗜好と、その直後に CV した物件の物件条件

提案手法により抽出した潜在的嗜好	CV した物件の物件条件
地域:練馬区	地域:練馬区
徒歩分数:10 分以内	築年数:築 30 年以上
設備:建物設備	家賃:6.8 万円
間取り:2K	間取り:2K
建物種別:マンション	設備:建物設備
	沿線:西武池袋線
	徒歩分数:徒歩 8 分
	専有面積:33 平米
	建物種別:マンション

表 1 より、提案手法によって抽出された潜在的嗜好は 5 条件中 4 条件が、CV した物件の物件条件と一致しており、適合率が高い。一方で、抽出された嗜好数は 5 条件と少ないため、再現率は低くなっている。これは CV した物件の物件条件の全てが必ずしもユーザが重要視している項目ではないために起こると考えられる。

ユーザ 100 人の閲覧履歴を用いて行った定量的評価の結果を表 2 に示す。

表 2: 提案手法とベースラインの定量評価

	適合率	再現率	F-measure
提案手法	0.75	0.21	0.32
ランダムな物件	0.31	0.04	0.07
閲覧時間の長い物件	0.65	0.11	0.18

表 2 より、提案手法の方が、ランダムな物件、閲覧時間の長い物件と比較した際に高い精度を示しており、ベースラインよりも高い精度でユーザの潜在的嗜好が抽出できていると言える。

4. まとめ

今回の実験では、ユーザが明示的に入力している検索条件を用いることなく、ユーザの閲覧履歴から一定の精度で潜在的嗜好を得られることを示した。今後は実験対象者を十分に増やすとともに、明示的に入力されるユーザの嗜好との比較を行うことにより、より適切な絞り込み条件のサジェストや物件推薦の精度向上を行っていきたい。

参考文献

- [土方 2005 年] 土方嘉徳: “嗜好抽出と情報推薦技術”, 情報処理学会論文誌, 土方嘉徳. “嗜好抽出と情報推薦技術.” 情報処理学会論文誌 47.4, 2005.
- [布施 2018 年] 布施拓馬, 關翼人, 村田昇, 杉浦太樹, 野村眞平: “検索条件のベクトル表現を用いた検索の興味の可視化”, 情報処理学会第 80 回全国大会, 2018.
- [高畑 2011 年] 高畑麻理, 上田真由美, 中島伸介: “食材に対する好き嫌いを考慮した料理レシピ推薦手法の提案.”, DEIM Forum. 2011.