

情報閲覧順序を用いた不動産物件推薦手法の提案

三條 知美[†] 岡 昂一郎^{††} 櫻井 彰人[‡]

株式会社ネクスト[†] 株式会社日本政策投資銀行^{††} 慶應義塾大学[‡]

1 背景

不動産ポータルサイトを訪れるユーザの多くは物件検索に不慣れであるため、よりユーザの嗜好に適った物件をなるべく多く推薦することが必要である。しかし、不動産ポータルサイトは、継続的に利用する性質のサイトではないため、個人属性の蓄積が難しく、個人の属性に従った推薦を行うことは難しい。

そこで、本論文では過去にユーザが物件を閲覧した順序を用いて物件に重みを付け、ある物件（基準物件）に対して関連の強い物件から順に物件を推薦する手法を提案する。

2 提案

過去にサイトを訪れたユーザが物件を閲覧した順序から、基準物件との関連の強さ（関連度）を集計し、基準物件に対し、関連度の強い物件の上位5件をユーザに推薦する。

関連度の計算方法次の通りである。まず、あるユーザが1セッション内に見た物件を閲覧順に並べる（以下、閲覧履歴）。その閲覧履歴から、閲覧順序が連続するいくつかの物件を抽出し、その中で一番初めに見られた物件を基準物件とする。基準物件の次以降に見られた物件に閲覧順序に従い、重み付きのポイントを付与する。同様の手順で、全ユーザのすべての閲覧履歴を集計する。

最後に、基準物件ごとに閲覧された物件のポイントを累計する。この累計済みポイントを、関連度と呼ぶ。関連度の高い順に基準物件に対する閲覧物件のランキングを作成する。そして、ランクの高い順に物件を推薦する。

連続する物件をいくつ抽出するか、については検証実験を行い、一番良い値を用いる。

この時、いくつ抽出するかは、“d”によって表記し、 $d \in \{1, 2, \dots, 10\}$ とする。

またポイントに関する重みの算出方法について

A research on a property recommendation method using the information browsing order.

[†] Tomomi SANJO, NEXT Co., Ltd.

^{††} Koichiro OKA, Development Bank of Japan Inc.

[‡] Akito SAKURAI, Keio University

でも検証実験を行い、一番良い値を用いる。この時、重み“β”は以下の3種類を用いる。

方法Ⅰ：重みなし

方法Ⅱ： $\beta = 1/d$

方法Ⅲ： $\beta = 1/(2^{d-1})$

式1, 2, 3の確率Pは実際の計算では $P \times \beta$ とする。

2.1 関連度計算手法

● 関連度計算手法1（非マルコフ型）

確率変数間の依存関係（条件付き独立性）を導入しないで、ポイント进行を定義する。

aを基準物件としたときの、物件cのポイントを次の式1で定義する。

$$\sum_{x_1 \dots x_n} P((x_1, \dots, x_n, c) | a) = \sum_{x_1 \dots x_n} P((a, x_1, \dots, x_n, c)) / P(a) \quad \dots \text{式1}$$

基準物件からの各物件のポイント例（重みなし）を図1に示す。

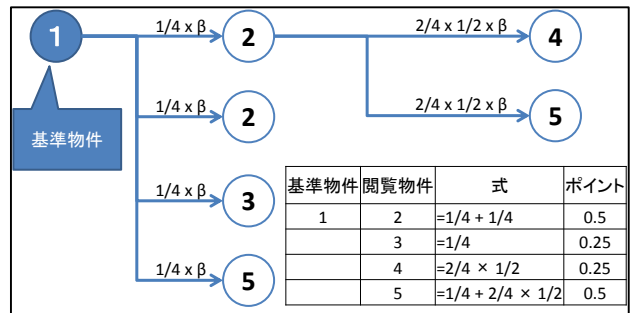


図1 関連度計算手法1

● 関連度計算手法2（1次マルコフ型）

いくつかの条件付き独立性を導入したモデルを考える。計算方法は式2の定義の通りで、各物件に与えられるポイントの例（重みなし）を図2に示す。

遷移確率はひとつ前の物件からの遷移確率のみ用いる。

$$\sum_{n, x_1, \dots, x_n} P((a, x_1, \dots, x_n, c)) / P(a) = \sum_{n, x_1, \dots, x_n} P(c | (a, x_n)) P(x_n | a) = P(c | a) + \sum_{x_1} P(c | (a, x_1)) P(x_1 | a) + \sum_{x_1, x_2} P(c | (a, x_2)) P(x_2 | a) + \dots \quad \dots \text{式2}$$

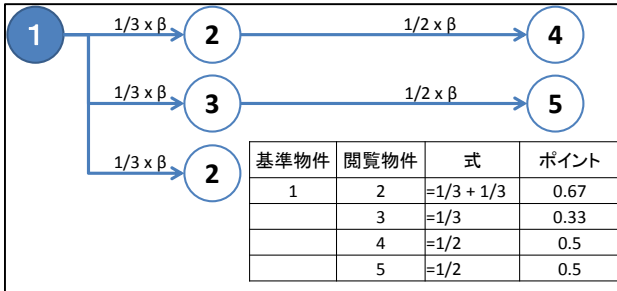


図2 関連度計算手法2

● 関連度計算手法3 (回数型)

さらに簡素化した手法を提案する。式2から、
 $P(x_1|a) = P(x_2|a) = \dots$
 $P(c|a, x_1) = P(c|a, x_2) = \dots$
 と近似する。すなわち

$$\text{左辺} \approx \frac{1}{z} \left(1 + \sum_{x_1} 1 \cdot 1 + \sum_{x_1, x_2} 1 \cdot 1 \cdot 1 + \dots \right) \quad \dots \text{式3}$$

を定義する。各物件に与えられるポイントの例 (重みなし) を図3に示す。

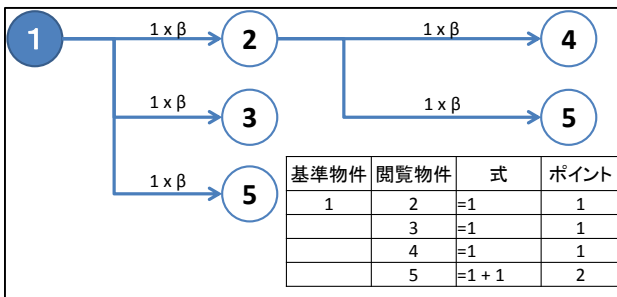


図3 関連度計算手法3

3 実験

不動産ポータルサイト HOME'S

(<http://www.homes.co.jp>) を訪れたユーザの閲覧履歴を用いて、2節の提案手法の関連度計算手法および、距離、重み算出方法に関する実験を行う。その結果得られた最適な組合せを用いて、他の既存手法と精度の比較実験を行った。

3.1 データ概要

検証実験に用いたのは、HOME'S に掲載された賃貸物件の閲覧履歴データである。同じ駅を最寄駅とする賃貸物件を閲覧したユーザの閲覧履歴を用いる。

用いたデータの詳細は表1の通りである。

表1 実験用データ

物件種別	期間	駅	
		駅名	物件数
賃貸	2015年 2月1日～7日	池袋	4,617
		中野	2,960
		吉祥寺	2,960

3.2 検証実験

3.1節で説明した検証用データを、駅毎に約1/2を学習用として関連度計算用に使い、残り1/2のデータをテストデータとする。得られた関連度によるランキングから推薦される物件と、テストデータの閲覧履歴の一致度で精度の比較を行った。

基準物件に対し、関連度の高い順に5件の物件を推薦する (推薦5物件)。テストデータに含まれる閲覧履歴の中で、基準物件の次に見た物件が、推薦5物件の中に含まれていた場合、正解数1とする。正解率 = 正解数 / 全推薦回数とする。

その結果、一番正解率が高い関連度計算手法は、関連度計算手法1 (非マルコフ型) であった。また、距離に関しては $d = 8$ 、重みづけ方法に関しては方法Ⅲの $\beta = 1/(2^{d-1})$ の正解率が一番高く、その正解率は3駅平均で約5.3%であった。

3.3 比較実験

3.2節の正解率を既存のランキング手法である RankingSVM により得られるランキングによって物件を推薦した場合、及び時系列データ予測に用いられる MarkovChain により得られた閲覧物件の予測データに従って物件を推薦した場合と比較を行った。

その結果、わずかではあるが提案手法の正解率が最も良かった (図4)。

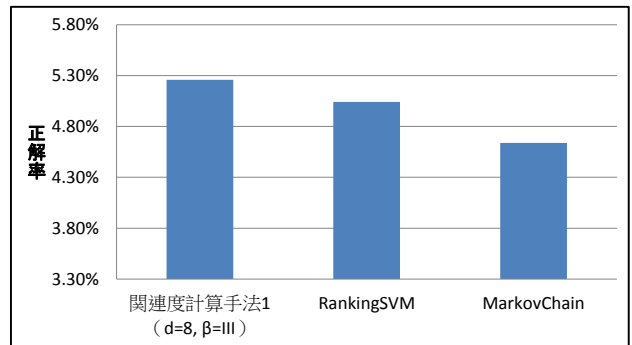


図4 既存手法との比較

4 まとめ

ユーザが物件を閲覧した順序を用いて、基準物件との関連度を計算して物件を推薦する手法を提案した。関連度の計算手法、閲覧抽出数 d 、および d に従った重み β の算出方法をいくつか用意し、検証実験により最適な組合せを導出した。

最適な組み合わせは関連度計算手法1 (非マルコフ型)、 $d = 8$ 、 $\beta = 1/(2^{d-1})$ であり、この組み合わせを用いた正解率は約5.3%であった。

また、既存のランキング手法と比較し、提案手法の正解率が最も高いことを確認した。