

2013年12月18日

株式会社リクルート住まいカンパニー

## 東京大学との共同研究 ビッグデータを活用した推薦アルゴリズムを新規に開発

ポータルサイト内での思考の発散と収束を促すことでCV率の向上を促進

株式会社リクルート住まいカンパニー(東京都千代田区/代表取締役社長：野口孝広)は、国立大学法人東京大学(東京都文京区/総長：濱田純一、以下「東京大学」)大学院工学系研究科総合研究機構松尾豊准教授の研究グループ、株式会社経営共創基盤(東京都千代田区/代表取締役CEO：富山和彦)の三社共同研究にて、住宅購入検討者向けの新しい推薦アルゴリズムを開発しましたのでお知らせいたします。

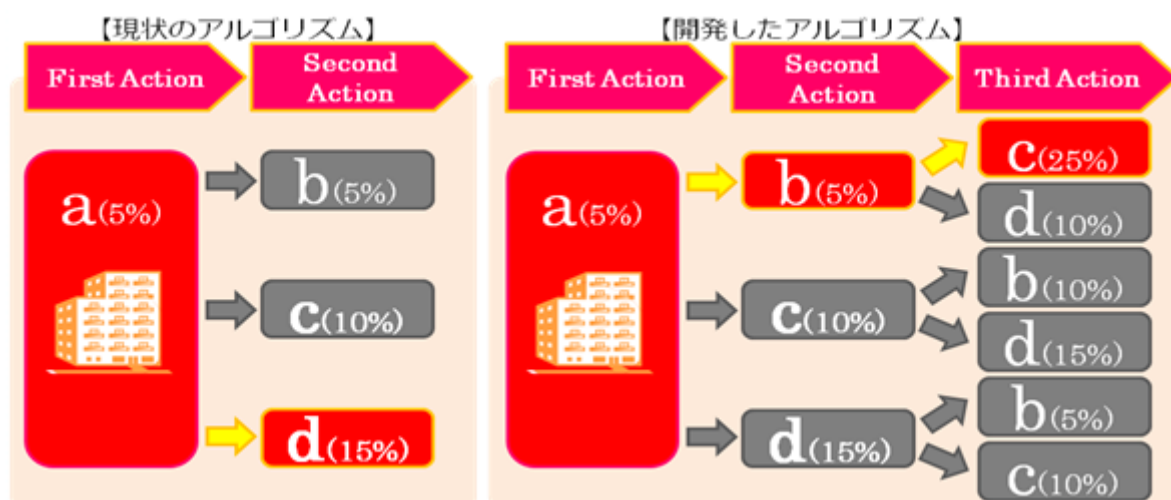
### 本研究の要点

不動産は購入金額が高く、検討期間が長い商品です。そのため購入検討者には、複数回にわたる不動産ポータルサイトへの訪問を通じ、欲しい商品イメージを固め、徐々に購入への決意を固めてもらうことが重要になります。本研究では、**1回目のサイト訪問時点で最も関心が高い物件を提示するよりも、あえて異なった他の物件の閲覧を通すことで最終的な購買意欲の促進につながることを明らかにし、その知見を基に新しい推薦アルゴリズムを開発しました。**

### 研究内容

不動産ポータルサイトにおいて、購入確度の高い検討者は、対象物件の詳細が記述してある資料をWEB上、または電話を通して請求する傾向があることが知られています。全購入検討者のうち、この資料請求を行う検討者数の割合をCV率(CV=コンバージョン)とします。今回、開発したアルゴリズム(下図イメージ)では、検討者の過去のサイト閲覧履歴を基に、次の訪問(Second Action)だけではなく、その次の訪問(Third Action)時の閲覧行動まで仮定した上でCV率が最大になるような物件を推薦するものとなっております。

【新しい推薦アルゴリズムのイメージ】



リクルート住まいカンパニーはこれからも、ひとりひとりにあった「まだ、ここにはない、出会い。」を届けていきます

【【本件に関するお問い合わせ先】】

株式会社リクルート住まいカンパニー 企画統括室 統括部 広報グループ

電話 03-6835-5290

<http://www.recruit-sumai.co.jp/support/press.html>

前頁の図は物件a~dの閲覧順序を表した例です。各ノード（構成要素）上の百分率は、そのノードまで閲覧した時点でのCV率を示しています。

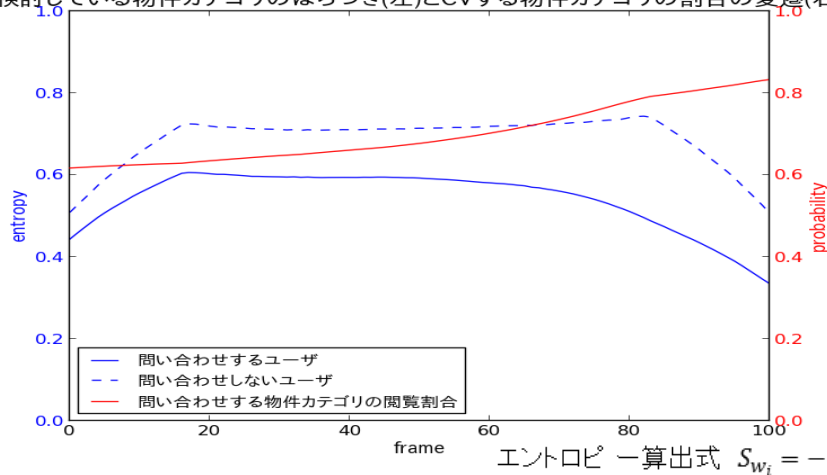
これまでの多くの推薦アルゴリズムの場合は、物件aを見た検討者に対しては、次（Second Action）に最もCVしやすいと考えられる物件dを提示していました。（図中【現状のアルゴリズム】を参照）。しかし、次の次（Third Action）まで物件を閲覧することを考慮した場合、Second Action時に物件dを推薦することは必ずしもCV率の上昇に貢献していません（図中【開発したアルゴリズム】を参照）。a→b→cと閲覧した場合が最もCV率が高くなる図の例では、Second Actionの時点では、物件bは閲覧してもCV率は低いのですが、その後のThird ActionでのCV率の上昇に影響を与えています。このことから、一見CV率の低い物件bが、検討者が購入意思を固める上で重要な働きをしていることがわかります。つまり、aを訪問した検討者にb→cという順序で閲覧するように推薦することで、購入検討を後押ししています。

このアルゴリズムの精度を検証するために、オフラインテストを実施した結果、単純にSecond Action時にCV率が最も高い物件を提示することに比べ、最大で約4倍CV率が上昇する物件の閲覧パターンがあることが判明しました。

## 【Appendix】

購入検討者の物件閲覧行動の特徴を表すために検討物件のばらつきを式(1)で定義します。地域・沿線・価格などから作成した物件カテゴリのばらつきとCV率の変遷を表した下図からは、CVする検討者ほど物件を絞り込んでいることがわかります。

検討している物件カテゴリのばらつき(左)とCVする物件カテゴリの割合の変遷(右)



$$\text{エントロピー ー算出式 } S_{w_i} = -\sum_{c=1}^{N_c} P_{w_i} \log_2 P_{w_i} \quad (1)$$

この結果から物件を推薦する場合、閲覧/CV履歴を利用して検討段階を捉え、興味の拡散と収斂を手助けし、最終的には収斂させることにより、将来のCV率が最も高まるような物件を提示する手法が有効になると考えられます。実際のアルゴリズムでは検討者カテゴリをi、閲覧物件カテゴリをj、推薦物件カテゴリをkのとしたときのCVRをpとすると

$$p_{ijk} = \frac{S_{ijk}}{T_{ijk}} \quad (T_{ijk} = 0 \text{ なら、} \hat{p}_{ijk} = 0, T_{ijk} : \text{PV/CVの総数}, S_{ijk} : \text{CVの総数}) \quad (2)$$

となります。しかし、実際には詳細に検討者をセグメント化するため信頼性のある  $p_{ijk}$  を得ることはできません。そこで直接  $p_{ijk}$  を求めるのではなく、以下のように算出することでこの問題に対応します。

$$\hat{p}_{ijk} = \hat{p}_{ijk}^{MLE} = f(p_{ijk}^1, p_{ijk}^2, \dots, p_{ijk}^M; \beta) \quad (3)$$

このとき、 $f(x; \beta) = \frac{1}{1 + \exp(-\beta^T x)}$  (シグモイド関数) Mは検討者、閲覧、推薦物件の組み合わせ数

とすると、pの推定はβの最適化問題となります。

(参考)

■論文「ユーザの成長を促進する情報推薦」, 人工知能学会全国大会(第27回)論文集, 2013  
<https://kaigi.org/isai/webprogram/2013/pdf/554.pdf>