

不動産仲介マーケティングのためのユーザ行動予測

Prediction of user behaviors for real estate brokerage marketing

大浜 毅美^{*1}
Takemi Ohama

^{*1} 株式会社 ietty
ietty Inc.

In this research, we built a model to predict how much each customer would make a contract from customer attributes and behavior. The validity of our model has been experimentally verified by using 11,823 customers' behavior data. As a result, it has been made clear that prediction is possible using only behavior and attributes for 30 minutes after customers' registration. In marketing, it is required to repeat the PDCA cycle by quickly, and this approach will contribute to the efficiency of such activities.

1. はじめに

インターネットによる商品販売のチャネルとして、チャットインターフェイスを利用した「対話型 e コマース」が注目を浴びている。従来の e コマースは希望の商品をカテゴリ選択や自由文による検索によって希望商品をユーザ自身が探して比較検討するのに対して、対話型 e コマースではユーザ(顧客)が商品の大きな希望条件を伝え、オペレータ(販売員)が顧客の希望に応じて商品の提案やさらなる質疑応答を行い、商品の選定をサポートする。これにより、従来の e コマースでは困難であった顧客と販売者のリアルタイムのコミュニケーション、すなわち「接客」という工程を取り入れることができるようになり、特に購買に至る意思決定に後押しが必要な中古自動車、保険商品や不動産などといった高単価・高付加価値の商品をネットで取り扱うことが可能になってきている。

特に居住目的の不動産売買および賃貸物件については、同質の財が存在せず、品質に関する情報が構造的に不足しがちであるという問題が指摘されており[清水 2016]、また契約交渉や重要事項説明といった法的側面も必要なことから、資格を持つ販売員による情報提供、すなわち接客が欠かせない。このためこれまでの e コマースサイトでは扱うことは著しく困難とされてきたが、チャットと対面による接客を組み合わせたビジネスモデルによる新しい試みがいくつか行われ始めている。

一方こうした対話型 e コマースを店舗での対面販売と比較すると、ネットを利用することにより広域のユーザをターゲットとすることができ、やりとりは一般的に文字(テキスト)のみによって行われるため、店舗の販売員が日常的に行っている顧客の態度や印象といった非言語の情報取得が困難であり、現状では接客技術という点で大きな不利を抱えている。

しかし、このテキストのみのやり取りであるという点を活かし、顧客の発言および行動のデータを適切に援用してデータ分析や機械学習の成果を適用できれば、上記の欠点を克服するだけでなく、より精緻な接客を展開できる可能性がある。

例えば顧客を獲得するマーケティングフェーズや来店直後の初動接客フェーズにおいては、さまざまな広告媒体や施策を組み合わせて最終的に成約に結び付く可能性の高い顧客をできる限り多く獲得・誘導することがその目的となるが、不動産賃貸や売買の場合、商談を始めてから契約に至るまで数週間以上

かかることも少なくなく、広告・施策の効果測定も同様の期間かかってしまうことになる。ここに機械学習による成約確率の予測値を早い段階で示せれば、実施中の媒体・施策の有効性をより早いサイクルで推定できるようになるため、効率的なサイクルでの施策変更が可能となると考えられる。

本稿ではこの議論を実証するため、属性情報と来店初期の行動から、来店者が最終的に商品を購入するに至るかどうかの予測モデルを構築し、マーケティングの広告効果測定に適用可能かどうかを検証する。

2. 対象データ

本研究ではオンライン不動産賃貸仲介店舗「ietty」における顧客データを対象として予測モデルを構築する[ietty 2017]。

ietty では、サイトに訪問した顧客に「来店カード」という形で賃貸仲介接客のための希望物件の情報(引越希望日、住みたい地域の最寄り駅、間取り、広さ、家賃等)や顧客の属性(在住の地域、性別、年齢等)を回答してもらい、この内容に従って DB を検索し該当物件を提案する。気に入る物件がない場合はルールベースのエンジンにより自動で条件を緩和したり、顧客とアドバイザー(チャットオペレーター)が相談して条件自体を変更したりといった接客を行いながら、より顧客に適した物件を探していく。提案された物件については「内見したい」「お気に入り」「興味なし」などの選択肢で評価できるほか、物件についての感想をコメントとして入力することが可能となっている。こうした希望の物件を探す活動の後、特に気になる物件を実際に訪問して確認(内見)し、顧客が納得すれば契約の申込を行う。

この手続き自体は一般の不動産賃貸仲介店舗で行われているそれと大きな違いはないが、ietty の場合内見以外の全てのやり取りが原則チャット上で行われるため、行動(発言)の内容や時刻が全てデータとして保存されているところに特徴がある。

今回はこの顧客の属性及び登録初期(30 分以内)の行動を特徴量とし、その顧客が登録後 2 ヶ月以内に契約の申込を行ったケースを正例、それ以外を負例としてモデル構築を行った。具体的な特徴量を表 1 に示す。

表1 クラスと特徴量

(クラス)成約結果	登録後2か月以内に契約申し込みを行ったか
ユーザ属性	性別,年齢,引越理由,同居人数
希望条件	家賃上限,広さ,引越希望時期
登録時の行動	来店カード記入にかかった時間,フリーコメントの長さ,利用デバイス
登録後の行動	初回のチャット発言までの経過時間,初回の物件評価までの経過時間

連絡先: 大浜毅美, 株式会社 ietty, takemi.ohama@ietty.co.jp

なお、ある期間のデータを抽出したところ、正例(申し込あり)が 619 件、負例が 33,212 件であった。

3. モデル構築

抽出されたデータは著しく不均衡であるため、多くの学習アルゴリズムではそのままモデル構築を行うと適切な識別境界を決定できずに分類性能が著しく低下することが知られている。

本研究ではこれを是正するためまず検証用のデータを無作為に 20%取り置いた後、それ以外を負例に対するランダムアンダーサンプリングと SMOTE による正例のオーバーサンプリングを組み合わせることでサンプル数を調整することとした[Chawla 2002]。ただしこうしたリサンプリングはデータを統計的に増殖させる手法であり過学習を招く恐れがあるため、後述の他のパラメータと組み合わせることで複数のパターンを実験している。

学習モデルは Random Forest を用いて無効な特徴量の取捨選択や分散化のレンジ調整を行った後、全特徴量を二値化したデータ(83次元)に対して RBF カーネルの SVM を適用し、複数の C および γ の組み合わせを試行した。

精度指標としては予測に適用するデータ自体も不均衡であることと、予測の目的が前述のようにマーケティング効果の識別にあることから、予測精度の平均(Average Precision: AP)と ROC 曲線の AUC の 2 種類を参照した[Davis 2006]。

表 2 に試行したリサンプリングおよび SVM パラメータのパターンを、表 3 に AP 上位 10 試行の結果を示す。

結果の傾向として、リサンプリングにより性能は向上するものの正例を 5 倍程度増殖させたところで頭打ちとなりそれ以上の改善は見られなくなる。むしろ計算時間の少ない小サンプルでも適切な SVM パラメータの調整を行うことにより遜色ない性能が確保できることが明らかになった。以降の実験では AP、AUC がいずれも高い値を示したサンプル数 4000(正負各 2000)、C=10、 $\gamma=0.001$ のパラメータで構築した学習モデルを採用する。

表2 パラメータパターン(計18試行)

リサンプリング後のデータ数	2000, 4000, 8000
SVMの誤差項ペナルティ(C)	1, 10, 100
RBFのカーネル係数(γ)	0.01, 0.001

表3 試行パラメータと評価結果(AP上位10)

AP	AUC	n	C	γ
0.180	0.871	4,000	100	0.001
0.179	0.878	4,000	10	0.001
0.178	0.877	8,000	100	0.001
0.177	0.879	8,000	10	0.001
0.176	0.879	2,000	100	0.001
0.175	0.877	2,000	10	0.01
0.172	0.878	4,000	10	0.01
0.172	0.874	4,000	1	0.01
0.166	0.875	8,000	1	0.01
0.166	0.868	2,000	10	0.001

4. ライブテスト

前章で構築された学習モデルを実際のビジネスに活用するため、ユーザの新規登録後 30 分経過後に成約確率の予測を行い、結果をチャットシステムに出力するチャットボットを開発して実地での検証を行った。

実験はモデルに利用したデータ期間以後から開始し、新規登録が 1 万件を超えるまで実施した。最終的に 11,823 件のデータを取得している。開発したチャットボットはまず 30 分毎に新規登録者の属性および登録から 30 分間の行動データを学習モデルと同様の手法で取得する。次にこのデータを構築済みの学習モデルに適用し「2 か月以内に成約する」クラスに属する確率とサイトに訪れるきっかけとなった広告手法コード(キャンペーンコード)をオペレータ向けチャットに出力する。

マーケティング担当者は通常同時に複数の媒体に広告を出稿するが、上記のチャットボットにより、どの媒体・手法がより契約に結びつきやすいかの指標とし、広告の費用対効果の推定値を算出して出稿量やコストの調整を試みる。

表 4 に獲得顧客数の多かった A から G までの 7 手法について、予測精度の平均値と、実際にその媒体から 2 か月以内にどの程度成約があったかの実測値を示す。予測精度と実際の成約率には 20 倍程度の乖離はあるものの、相関係数は 0.81 と高い値を示しており、広告戦略の最初期の効果推定という点では十分活用可能であると考えられる。なお広告手法 A のみ実際予測・実測値とも他の手法に比べて格段に高いが、登録企業や大学向けの手数料半額キャンペーンであり、一般の Web 広告とは異なりクチコミに近い特徴を持っているためである。

表4 広告手法別の成約推定値と成約率

媒体	予測精度平均	実際の成約率
広告手法A	0.718	0.237
広告手法B	0.523	0.026
広告手法C	0.464	0.026
広告手法D	0.442	0.028
広告手法E	0.439	0.018
広告手法F	0.274	0.010
広告手法G	0.215	0.007

($r=0.805$)

5. 結論

本研究では顧客の属性と行動から購買を予測するモデルを構築し、その効果をライブテストで検証した。この結果顧客の新規来店後 30 分間という短い時間の行動だけを用いても、識別力のある予測が可能であることが明らかになった。特にマーケティング分野ではその施策に対して素早く効果を推定して PDCA サイクルを繰り返すことが求められており、今回のようなアプローチが改善活動の効率化に寄与することが期待される。

また、今回は最初期の行動のみを用いたデータを利用したが、ユーザの滞在時間が長くなるほどその予測精度は向上すると期待されるため、今後はマーケティングだけでなく接客やシステム自体の改善に適用できるアプローチの探索を課題としたい。

参考文献

- [清水 2016] 清水 千弘: 透明で中立的な不動産流通市場の条件、土地総合研究 2016 年冬号、土地総合研究所, 2016.
- [Chawla 2002] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer: SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research 2002
- [Davis 2006] J. Davis and M. Goadrich: The relationship between precision-recall and ROC curves, In Proceedings of the 23rd International Conference on Machine Learning, 2006.
- [ietty 2017] <https://ietty.me> (2017.03.08)