

対面販売でのヒアリング内容と購入結果の食い違い予測

Discrepancy prediction between the results purchase and pre-listening of the face-to-face sales

内田 匠^{*1*2}
Takumi Uchida

吉田 健一^{*2}
Kenichi Yoshida

^{*1}株式会社リクルートコミュニケーションズ
Recruit Communications Co., Ltd.

^{*1*2}筑波大学大学院 ビジネス科学研究科
University of Tsukuba Graduate School of Business Sciences

It is difficult to interview real hope from customer through face-to-face consultation. For example, when customers try to purchase houses, even if they explain their hope to the sales staffs, they frequently purchase houses which are different from what they explain before. In other words, getting real hope of customer from face-to-face consultation is a difficult task. As the result, sales staff cannot offer satisfactory houses and lose their customers. In this study, we conducted a questionnaire on customers who bought house to see the relationship between pre-listening results from them and their actual purchases.

1. はじめに

本研究では顧客と対面販売員のコミュニケーション問題として、必ずしも購入される商品が顧客から事前にヒアリングした内容と同じものではない現象について分析を試みた。この現象は企業が適切な商品を顧客に提案する上で、問題となっている可能性が高い。本研究にあたり、1) 始めにこの問題が実在するかをアンケート調査で確認した。さらに、2) 何をヒアリングすれば購入内容を予測できるか、を機械学習の手法を用いて調査した。

1.1 「実際にこの問題が実在するか」の概要と結論

1年以内に分譲マンションを購入した人を対象にアンケート調査を行い、購入の際に事前に重視していた項目と実際に購入したマンションの情報を比較した。この解析結果から、事前の重視項目と実際の購入物件に直接的な関係がある場合とそうでない場合が確認できた。(図 1,2,3) 例えば、広さを重視すると回答した顧客が、実際に床面積の広い物件を購入する傾向は観測できなかった。一方、駅からの徒歩分数を重視すると回答した顧客は、そうでない顧客と比較した場合に、駅から近い物件を購入する傾向にあった。ただし、重視すると回答したとしても必ずしも駅から近い物件を購入するわけではなく、本研究が対象とする問題は実在すると考えられる。

1.2 「何をヒアリングすれば購入内容を予測できるか」の概要と結論

我々は価格、床面積、駅からの徒歩分数などは顧客の様々な事情や状況の組み合わせで決定され、有能な販売員は適切なヒアリングを行って顧客が満足する提案を行っていると仮説した。よって、対面販売でどのようなヒアリングをすればこれらが予測できるかを分析した。具体的には「都道府県、家族構成、物件選びで重視する項目、購入の目的、希望の生活」などをアンケートで取得し、実際に購入された物件の「価格、床面積、駅までの徒歩分数」が予測可能かを検討した。結果、線形回帰モデルや回帰木などの単純なモデルでは「都道府県、家族構成」のみを説明変数として利用する場合が最も予測精度がよく、それ以上の変数追加は精度を悪化させることが確認でき

表 1: アンケートの項目

変数分類	質問内容	選択肢など
目的変数	購入物件の価格	整数を入力。単位は万円
	購入物件の床面積	整数を入力。単位は平方 m
	購入物件の最寄り駅までの徒歩分数	整数を入力。単位は分
説明変数	購入物件の都道府県	[東京, 大阪, ...]
	家族構成	[配偶者, 子ども, 親] の有無
	物件の重視項目	[価格, 広さ, 駅までの距離, ...] など 18 項目から最大 5 つ
	購入目的	[家族のため, 仕事の都合, ...] など 24 項目から最大 3 つ
	希望する生活や暮らし	[ホームパーティ, ペット, ...] など 50 項目から最大 5 つ

た。一方で、Gradient Boosting Tree や多層ニューラルネットワークなどの複雑なモデルでは「物件選びで重視する項目、希望する生活や暮らし」などの複雑な情報を用いた方が精度が高いことも確認できた(表 2)。

2. 実施したアンケートについて

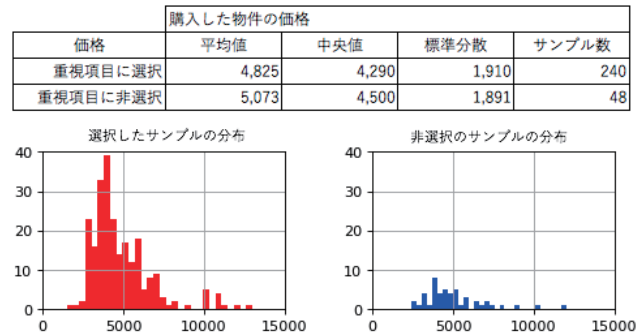
1年以内に新築分譲マンションを購入した 330 人に対してインターネット調査を行った。調査項目を表 1 に示す。

3. アンケートの解析結果

顧客が事前に重視していた内容と実際の購買結果にどの程度の差があるのかを確認するために、アンケート項目の「検討にあたって重視する項目」と「実際の購入結果」を比較した(表 1, 2, 3)。これらの結果から、価格と駅の近さについては事前の重視項目と実際の購入に若干の関係性が見られたが、明確な影響があるとは言い難い結果となった。また、物件の広さについては事前の重視項目に選択されていても、実際に床面積の広い物件が購入される傾向は確認できなかった。よって、新築マンションの販売現場において、事前のヒアリング内容と実際に購入される物件に違いがあることは確認できた。

連絡先: 内田 匠, 株式会社リクルートコミュニケーションズ,
〒104-0054 東京都中央区勝どき 1-13-1 イヌイビル・カチドキ

図 1: 重視項目に価格の有無と実際の購入価格の関係



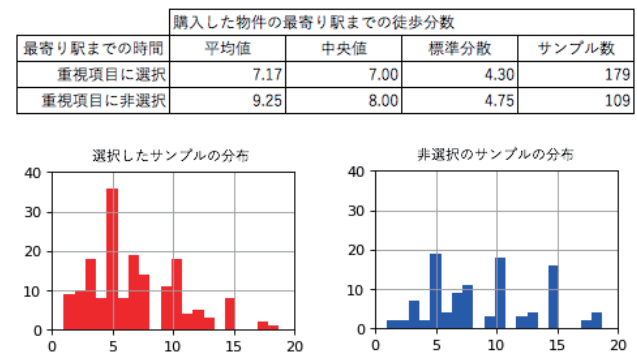
重視項目に選択したサンプルは、実際に安い物件を購入する傾向がある。しかし、サンプル数に大きな偏りがある。

図 2: 重視項目に広さの有無と実際の床面積の関係



重視項目に選択したサンプルが、床面積が広い物件を購入するとは限らない

図 3: 重視項目に駅までの距離の有無と実際の徒歩分数



重視項目に選択したサンプルは、駅までの徒歩分数の少ない物件を購入する傾向がある。ただし、大きな傾向とは言えない。

4. ヒアリング内容を変数とした購入物件の予測精度

対面販売において、ヒアリング内容と購入物件の食い違いという課題が実在することを確認した上で、どのようなヒアリングを行えば購入物件を特定できるか検討した。具体的にはアンケートで取得した顧客それぞれの希望や状況（都道府県、家族構成、物件選びで重視する項目、購入の目的、希望の生活）を説明変数として用いて、購入した物件のスペック情報（価格、床面積、駅までの徒歩分数）を学習し予測させた。検証方法は 10-fold の交差検証による予測誤差の絶対値平均（MAE）を確認した。この時、学習に用いる説明変数は対面販売で実際にヒアリングすることを想定して以下のように分類し、それぞれの説明変数群ごとの予測精度を確認した。

変数群 A 都道府県

変数群 B 都道府県 + 家族構成

変数群 C 都道府県 + 家族構成 + 物件の重視する項目

変数群 D 都道府県 + 家族構成 + 購入目的

変数群 E 都道府県 + 家族構成 + 希望する生活や暮らし

変数群 F 変数全て

上記の変数群ごとに、予測に用いる手法として、線形回帰モデル、回帰木 [Breiman 84], ランダムフォレスト [Breiman 01], GradientBoostingTree [Friedman 01], 多層ニューラルネットワーク [Hinton 89] を用いて実験した。目的変数である「価格、床面積、徒歩分数」はそれぞれ単位が違うことから、あらかじめ 0 から 1 に標準化した上で学習、予測を行なった。実装は機械学習ライブラリ [scikit-learn] を使い、調整パラメータは原則として該当ライブラリのデフォルト設定を利用した。

ただし、線形回帰モデルは L2 正則化を導入しその係数を 1 として、次元の呪いを抑制した。これは、変数群 F になるとかなりの多次元データになるがサンプル数が 330 程度しかないことから予測精度が著しく悪化することが明らかであったためである。また、多層ニューラルネットワークについては実装は [pytorch] を使い、ニューロンを 20 個の Relu を活性化関数とする 2 つの隠れ層で、各層が全て全結合で接続されたネットワークを構築した。

全体的な結果として、床面積については変数群 B の [都道府県, 家族構成] だけを用了場合が最も予測精度が高く、予測に用いるモデルもシンプルな線形回帰が最も良かった。床面積については、変数や予測手法を複雑にしても精度の改善は確認できなかった。一方で、価格については変数群 E の [都道府県, 家族構成, 希望の生活] を用いた GradientBoostingTree が最も精度が高く、こちらでは複雑なモデルで予測精度を改善することができた。また、最寄り駅までの徒歩分数については変数群 C の [都道府県, 家族構成, 重視項目] を用いた線形モデルが最も精度が高く、こちらは単純なモデルでかつ重視項目をヒアリングすることで精度が改善することを確認できた。

この結果の詳細は表 2 でまとめた。

参考文献

[scikit-learn] Lars Buitinck, et al.: API design for machine learning software: experiences from the scikit-learn project, 2013.

表 2: 10hold の交差検定による平均絶対誤差 (予測対象の変数は 0 から 1 にスケーリング)

予測手法 / 説明変数	A) 都道府県	B) A + 家族構成	C) B+重視項目	D) B+購入目的	E) B+希望の生活	F) 全変数
↓購入したマンションの価格の予測誤差						
01.LinearRegression	0.108	0.108	0.111	0.116	0.117	0.130
02.RegressionTree	0.109	0.118	0.144	0.146	0.134	0.167
03.RandomForest	0.109	0.114	0.117	0.118	0.118	0.118
04.GradientBoostingTree	0.110	0.112	0.114	0.111	*0.107	0.112
05.MultiNNet	0.116	0.121	0.117	0.118	0.115	0.117
↓購入したマンションの床面積の予測誤差						
01.LinearRegression	0.110	*0.105	0.111	0.107	0.116	0.122
02.RegressionTree	0.111	0.112	0.143	0.143	0.139	0.160
03.RandomForest	0.111	0.109	0.119	0.116	0.113	0.116
04.GradientBoostingTree	0.109	0.107	0.109	0.108	0.116	0.114
05.MultiNNet	0.116	0.109	0.105	0.106	0.110	0.109
↓購入したマンションの最寄駅からの徒歩分数の予測誤差						
01.LinearRegression	0.195	0.198	*0.191	0.203	0.208	0.213
02.RegressionTree	0.203	0.215	0.255	0.279	0.244	0.246
03.RandomForest	0.200	0.217	0.209	0.224	0.193	0.197
04.GradientBoostingTree	0.197	0.204	0.196	0.203	0.203	0.199
05.MultiNNet	0.196	0.195	0.193	0.194	0.192	0.193

[pytorch] Paszke, et al.: API Automatic differentiation in PyTorch.(2017)

[Breiman 84] L. Breiman, J. Friedman, R. Olshen, and C. Stone, “Classification and Regression Trees”, Wadsworth, Belmont, CA, 1984.

[Breiman 01] L. Breiman, “Random Forests”, Machine Learning, 45(1), 5-32, 2001.

[Hinton 89] Hinton, Geoffrey E. “Connectionist learning procedures.” Artificial intelligence 40.1 (1989): 185-234.

[Friedman 01] J. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, Vol. 29, No. 5, (2001).