

多変量解析を用いたヤフオク!の落札価格予測方法の提案

A Proposal of Prediction Method for Sold Price in Yahoo Auction Using Multivariate Analysis

1841034 片岡 泰斗

Taito KATAOKA

指導教員 秋葉 知昭

In recent years, the number of people using Internet auctions has been increasing. In this study, I analyzed the factors related to the sold price and constructed a sold price prediction model for three product categories. Based on these results, I evaluated the sold price prediction models.

1. 緒言

近年,インターネットの普及によってオンラインショップが必要不可欠になってきている.その中でもコロナ禍の巣ごもり需要なども相まってインターネットオークションの人気も高まってきており,市場も拡大しつつある. 国内のネットオークション最大手であるヤフーオークションは出品数は1500万点以上,ユーザー数は600万人を超えるほどの規模となっている.ネットオークションでは,落札時期,入札数など様々な要因で落札価格が変動する.落札価格の動きは経験的なもので予測されるものが多く,分析されたものは少ない[1][3].

そこで本研究では,人気の高い家電を対象にカテゴリをテレビ,冷蔵庫,カメラの3つに分け,2019年1月から2021年8月までの期間を対象に,落札価格に関係する要素について分析し,落札価格の予測モデル製作を行い,予測値と実際の落札価格を比較する.オークション開始前にある程度価格の予測が出来れば,オークション初心者でも冷静さを欠いて高額な商品を落札してしまうといった問題を回避できると考える.

2. 分析対象と使用する変数

本研究では,2019年1月から2021年8月のヤフオク!で落札された商品を対象とする.対象となる商品カテゴリは冷蔵庫,テレビ,カメラの3つである.目的変数として「落札価格」を選択した.説明変数として「入札数」「開始価格」「商品の状態」「送料」「出品者(ストア,個人)」「希望落札額」「オークションタイプ」「かんたん取引」「かんたん決済」「値下げ交渉」「注目のオークション」「税率」「景気動向指数」を候補とした.トレランスの異常値を考慮して最終的に選択された変数は「入札数」「開始価格」「商品の状態」「オークションタイプ」「かんたん取引」である.

3. 分析結果

2019年1月から2021年8月までの期間のデータ3500分の落札価格を対象に目的変数,その他の変数を説明変数とし,重回帰分析を行った.

分析の結果,冷蔵庫カテゴリが良好な結果となった.

表1 冷蔵庫カテゴリの重回帰分析結果

変数名	重回帰係数	寄与率R ²	R+ ²	R++ ²	AIC	標準偏回帰係数	標準偏回帰	トレランス	VIF	
目的変数名	落札価格	0.953	0.908	0.908	73476.998					
	残差平方和	残差自由度	残差標準偏差							
	604232.026	3492	8776.002							
vNo	説明変数名	残差平方和	変化量	分岐比	F値	(上側)	偏回帰係数	標準偏回帰	トレランス	VIF
0	定数項	80161869.135	32557636.309	343.1988	0.000	-44632.748				
4	入札数	59704060.770	12098827.850	33696.0825	0.000	1592.858	0.951	0.976	1.024	
5	開始価格	11439603.660	63835370.835	0.8288	0.366	0.028	0.005	0.993	1.007	
6	状態	76389464.700	28785231.874	9.4625	0.002	2071.861	0.016	0.984	1.016	
7	送料	47593318.834	-10913.992	0.0001	0.938	-				
8	消費税	49771111.754	97833121.072	5.1716	0.022	+				
10	希望落札額	07940513.201	39655719.625	1.8137	0.174	+				
11	オークション	06392825.762	41221807.064	4.4349	0.033	-				
12	かんたん取引	79042398.689	31439105.863	0.4082	0.530	-799.386	-0.003	0.981	1.019	
13	yahooかんたん	69478655.981	20974423.155	2.8678	0.086	-2225.787	-0.009	0.982	1.018	
14	値下げ交渉	72069977.861	75535354.965	8.7907	0.003	+				
15	注目のオーク	28730506.419	18973726.407	1.5437	0.211	-				
16	税率	91529343.447	56074889.379	9.8417	0.002	+				
17	日付	66603850.304	79000382.521	11.4470	0.001	+				
18	景気動向指数	68700382.925	21096130.089	0.2739	0.607	9.144	0.003	0.993	1.007	

寄与率R²が0.908と高い値をとった.標準偏回帰係数の値から入札数の影響力が強く,次に開始価格が強いということがわかった.またテレビカテゴリの寄与率R²は0.846,カメラカテゴリの寄与率R²は0.707と有意な値をとった.テレビ,カメラ両カテゴリとも入札数と状態の影響が強い結果となった.

4. モデル式の作成・評価

重回帰分析の結果をもとに,選択した説明変数を用いてモデル式を算出する.落札価格予測の算出方法を整理すると以下ようになる.落札価格の予測式を定式化するため,次の記号を定義する[2].

i : 予測するカテゴリ, j : 予測する時期

$C(i, j)$: カテゴリ*i*の時期*j*の落札価格

a_i : カテゴリ*i*の定数項

k : 説明変数

X_{ij} : カテゴリ*i*の時期*j*の変数ベクトル

X_{ijk} :カテゴリー*i*の時期*j*における説明変数*k*の値

B_i :カテゴリー*i*の偏回帰係数ベクトル

B_{ik} :カテゴリー*i*の変数*k*の偏回帰係数
 予想落札価格を式(1)で表すことができる。

$$C(i,j) = a_i + B_i^T X_{ijk} \quad (1)$$

得られた予測式をもとに落札価格を予測する。予測精度は相対誤差を参考にして評価する。相対誤差の絶対値の平均を MAD(平均絶対偏差)といい、その予測モデルがどのくらいの精度なのか測ることのできる指標である。MAD が小さければ精度の高い予測モデルと言える。算出した予測式の一部を抜粋して表 2, 表 3 に表す。

表 2 冷蔵庫カテゴリーの予測精度

年,月	予測値	平均実測値	残差	相対誤差	MAD
2019 第 1 四半期	42151.82	39902.85	2248.97	0.053354	0.007607
2019 第 2 四半期	37650.48	37614.49	35.99	0.00095	
2019 第 3 四半期	27637.11	27321.6	315.51	0.01141	
2019 第 4 四半期	22841.67	24080.66	-1238.99	-0.05424	
2020 第 1 四半期	28576.51	29773.38	-1196.87	-0.04188	
2020 第 2 四半期	38428.84	37993.49	435.34	0.01132	
2020 第 3 四半期	38395.45	37232.62	1162.82	0.03028	
2020 第 4 四半期	29359.79	29520.19	-160.40	-0.00546	
2021 第 1 四半期	24444.82	25221.03	-776.21	-0.03175	
2021 第 2 四半期	25643.77	26266.11	-622.32	-0.02427	
2021 第 3 四半期	23583.66	23222.33	361.32	0.13394	

冷蔵庫カテゴリーの MAD は 0.007607 と精度の良い結果が得られた。結果から、実際の商品を予測する場合に十分役立つ数値だと考えられる。

表 3 テレビカテゴリーの予測精度

年,月	予測値	実測値	残差	相対誤差	MAD
2019 第 1 四半期	268175.6	53589.26	214586.4	0.800171	0.77269
2019 第 2 四半期	268877.5	57245.41	211632.1	0.787095	
2019 第 3 四半期	267104.8	55306.23	211798.6	0.792942	
2019 第 4 四半期	270942	59108.51	211833.5	0.781841	
2020 第 1 四半期	269179.3	57452.08	211727.2	0.786566	
2020 第 2 四半期	298831.4	84521.91	214309.5	0.717159	
2020 第 3 四半期	280497.7	69991.86	210505.9	0.750473	
2020 第 4 四半期	276318.6	64360.06	211958.5	0.767084	
2021 第 1 四半期	269627.8	56249.57	213378.2	0.791381	
2021 第 2 四半期	270409.9	59662.54	210747.4	0.779363	
2021 第 3 四半期	278760.6	70940.2	207820.6	0.745516	

テレビカテゴリーの MAD は 0.772689 となった。結果から、実際の商品をこのモデル式を用いて予測した場合、77%の誤差が生まれるということである。例を出すと 1 万円の商品を買おうとした場合、17700 円から 2300 円の誤差が生じるということである。

カメラカテゴリーの MAD は 0.924692 となった。誤差がおおよそ 92%の予測式と、実際の商品を

予測するには使える段階には無い結果となった。原因としては、用意した変数では目的変数を説明しきれなかったことが原因として挙げられる。

3 カテゴリーとも説明変数間の種類が類似していたが、予測の結果を見ると冷蔵庫カテゴリー以外には予測に使える精度とは言えない結果になった。要因としては値段が大きく関係していると考えられる。データを取得する際に人気順に取得した結果、データに偏りが生じた可能性がある。データに偏りができないよう注意し、説明変数間の相関を考慮し、より多くの変数を用意できれば、より高い精度で予測できるのではないだろうか。

5. 結 言

本研究では、オークションの落札価格を予測するため、重回帰分析を使用してヤフオク!の落札価格に関係する要素について分析した。

ヤフオク!の複数のカテゴリーの落札価格予測モデルを作成し、実際に落札価格の予測を行った。結果としては 3 カテゴリー中 1 つしか十分な精度を得られなかったが、冷蔵庫カテゴリーの予測精度はオークションに入札する際の参考とするには十分な精度を得られたと考える。しかしながら予想していた落札価格が大きく外れたオークションも多く、説明変数では説明しきれなかった要素があったと考えられる。

本研究では分析する対象をヤフオク!の 3 カテゴリーにしたため、対象の幅が広く、価格帯が安定しなかった。冷蔵庫カテゴリーの結果から、単一の商品かメーカーやブランドを対象とすれば、より高い精度で価格の予測ができると推測できる。また四半期ごとではなくより短い間隔で予測値と実測値を比較することができれば更に予測精度が高められると考える。

今後の課題としては説明変数の追加、近年ユーザーが増えつつあるメルカや eBay などと言ったヤフオク!以外でも使えるモデルの作成、複数のカテゴリーでも使えるモデル式の作成が挙げられる。

文 献

- [1] 井上翔太：ヤフオク!における価格上昇率を考慮した落札価格予測方法の提案, 千葉工業大学 2016 年度卒業研究 (2016)
- [2] 中野優真：5ヶ国におけるうつ病症状数の要因分析と予測モデルの提案, 千葉工業大学 2020 年度卒業研究 (2020)
- [3] 黒澤聡,前川徹：インターネットオークションにおける入札者の行動分析,情報処理学会報告 Vol.2001,No.118,pp.7-14 (2001)