

ネットオークションにおける商品取引の特性分析

石塚 美加[†] 松田 和浩[†] 吾妻 晋[‡] 佐島 敬眞[‡] 石井 啓之[‡]

[†] 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所 〒180-8585 東京都武蔵野市緑町 3-9-11
[‡] 東海大学電子情報学部コミュニケーション工学科 〒259-1292 神奈川県平塚市北金目 1117

あらまし インターネットを利用したオークション（以下ネットオークション）は、欲しいものを安く手に入れられるという利点から、インターネット上の代表的なサービスのひとつとなっている。ネットオークションにおける商品取引では、匿名性や非対面性が高いため、ユーザが入手し得る情報には限りがある。従って、取引に関わるユーザ自身が商品取引の妥当性を簡単に判断できる方法が必要とされている。本報告では、このような背景を踏まえ、商品取引の妥当性を判断する指標を得るための最初のステップとして、ネットオークションで行われる商品取引の分析、モデル化を行う。特に、今回の分析では、情報の入手、分析が容易に行えるとともに、商品取引の特徴を最も的確に表すと考えられる落札価格の時系列変化に着目して分析を行う。

キーワード インターネットオークション、落札価格、時系列分析、正規分布

Statistic Analysis of the Highest Bids in the Internet Auctions

Mika ISHIZUKA[†], Kazuhiro MATSUDA[†],

Susumu AZUMA[‡], Yoshimasa SAJIMA[‡], and Hiroshi ISHII[‡]

[†] NTT Information Sharing Platform Laboratories, NTT Corporation

[‡] Dep. of Communications Engineering, School of Information Technology and Electronics, Tokai University

Abstract The Internet-auction is one of the most representative services in the Internet. Though users can get low-priced items through auctions, anonymity and indirect dealings cause uncertainty. Thus, it is important to give auction-users clues to judge the validity of the dealing. To this end, we analyze and model the Highest Bits in the Internet auctions. In this report, we focus on time series analysis of the Highest Bits since we consider time series of the Highest Bits stands for characteristics of dealings most appropriately.

Key words Internet auction, highest bits, time series analysis, normal distribution

1. はじめに

インターネット上で提供されるサービスが多様化すると共に、それらのサービスを利用するユーザ数も増加傾向にある。特に、インターネットを利用したオークション（以下ネットオークション）では、欲しいものを安く手に入れられる、不用品を適価で処分できるという利点からインターネット利用者全体の53.9%が利用する代表的なサービスのひとつとなっている [1]。携帯インターネットからの利用も可能になってきたことから、今後もネットオークションを利用するユーザは増えることが予想される。しかし、ネットオークションにおける商品取引では匿名性や非対面性が高いため、ユーザが入手し得る情報には限りがある。従って、取引に関わるユーザ自身が商品取引の妥当性を簡単に判断できる方法が必要とされている。

本報告では、このような背景を踏まえ、商品取引の妥当性を

判断する指標を得るために、ネットオークションで行われる商品取引の分析、モデル化を行う。商品取引の分析方法としては、特定のユーザの過去の取引行動を分析する方法と同一商品の取引の履歴を分析する方法とが考えられる。ユーザの過去の取引行動の分析は、出品者や入札者の信頼性を測るのに適しているが、一人一人の取引情報を過去に遡って分析するのは、多大な計算時間が必要であるだけでなく、情報の収集自身も困難である。そこで、今回の特性分析では、情報の入手、分析が容易に行える同一種類の商品の取引の履歴を分析する。また、分析項目としては、商品取引の特徴を最も的確に表すと考えられる落札価格の時間変化に着目する。ユーザの立場から落札価格の分析を見ると、取引が妥当であることと、なるべく安く買う（高く売る）の両方を満たす取引が抽出できるのが望ましい。よって、ユーザにとっては、平均的な落札価格特性よりも落札価格の分布の情報（特に、分布の裾の情報）が有益である。従って、今回

の報告では落札価格の分布のモデル化を、特に検討する。

以下に、本報告の構成を示す。まず、2.節で、準備として、本報告で用いる分析手法について説明する。3.節では、データの取得方法を述べる。更に、取得したデータの落札価格の時系列変化の特徴について考察する。4.節では、3.の考察に基づいた分析手順を示し、5.節で分析結果を述べる。最後に、6.で本報告のまとめを行う。

2. 準備

分析にあたっては、落札価格 (Y) そのものではなく、落札価格の対数 ($\ln(Y)$) を分析の対象とする。これは、価格が大きいほど、価格変動の幅も大きいと考えるのが自然なためである。以降の議論では、価格という言葉は特に断らない限り価格の対数を示す。

以下で、本報告で用いる分析手法を簡単に説明する。

2.1 モデルの選定 [2]

次式で表される一次の自己回帰モデルを考える。

$$X_{t+\delta t} = \rho X_t + \mu_t + \lambda_t d_t + \epsilon_t \quad (1)$$

ここで、 ϵ_t は独立、同一の分布に従う確率変数である。また、時刻 t は取引件数の累積を表し、 d_t は落札日を表す。

このモデルは、 μ_t, λ_t が時間によらず一定値 μ, λ を取り、特性方程式の根 (ϕ) が絶対値で 1 より小さい場合に定常であるという。モデルが非定常となる場合で、 $|\phi| > 1$ (発散)、 $\phi = -1$ (振動) は時系列データとしては分析の意味をなさないの、通常 $\phi = 1$ の場合のみを問題にする。モデルの定常性を検定するために、式 (1) を次のように変形する。

$$X_{t+\delta t} - X_t = \delta X_t + \mu + \lambda d_t + \epsilon_t \quad (2)$$

$$\rho = 1 + \delta \quad (3)$$

回帰分析により、式 (2) 中の δ, μ, λ を決め、 $\delta = 0$ の仮説を検定することは、定常性の仮説を検定することを意味する。 $\delta = 0$ の検定は、 $\hat{\tau} = \hat{\delta}/s_{\hat{\delta}}$ ($\hat{\delta}$: δ の推定値、 $s_{\hat{\delta}}$: $\hat{\delta}$ の標準誤差) を τ 分布の下側 $\alpha\%$ の臨界値と比べることによって行う。 $\hat{\tau}$ 値が臨界値よりも小さい場合、仮説: $\delta = 0$ (非定常であるという仮説) を棄却する。

モデルが非定常の場合は、さらに階差をとって定常性の検定を行う、つまり、 $X_{t+1} - X_t$ の定常性を同様の手順で検定する。この手順を繰り返すことにより、モデルの次数を決める。

2.2 正規分布性の検定 [3]

正規分布性の検定としては、Jarque-Bera 検定、D'Agostino の D 検定の二つを用いる。

以下、それぞれについて説明する。

- Jarque-Bera 検定:

歪度 ($\sqrt{b_1}$)、尖度 (b_2) により、以下の量 (JB) を計算する。

$$JB = \frac{nb_1}{6} + \frac{nb_2^2}{24} \quad (4)$$

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \quad (5)$$

$$b_2 = \frac{m_4}{m_2^2} \quad (6)$$

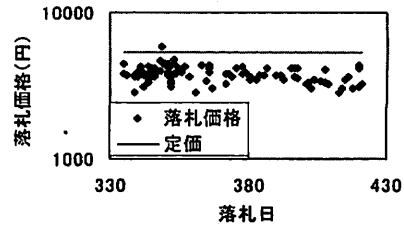


図1 商品 ID1 の落札価格

ここで m_k は、平均のまわりの k 次モーメントを表す。

歪度、尖度は、正規分布で 0 になる。従って、 JB が 0 に近いほど、正規分布であるという仮説に整合的だといえる。 JB テストの臨界点としては、[3] の数表に記載された値を用いる。

- D'Agostino の D 検定: 式 (7) ~ (9) から計算される Y の 0 からの乖離度によって、正規分布であるという仮説の整合性を判断する。 D テストの臨界点についても、[3] の数表に記載された値を用いる。

$$Y = \frac{\sqrt{\pi}(D - 0.28209479)}{0.02998598} \quad (7)$$

$$D = \frac{\sum_{i=1}^n c_i X_{(i)}}{n^{3/2} \{ \sum_{i=1}^n (X_i - \bar{X})^2 \}} \quad (8)$$

$$c_i = i - \frac{1}{2}(n+1) \quad (9)$$

$X_{(i)}$ は、 X_1, \dots, X_n を大きさの順に並べ替えた変数である (つまり、 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$)。

3. 落札価格の時系列変化

3.1 データの取得方法

オークション統計サイト [4] から、商品名を指定して過去三ヶ月間の落札価格、落札日情報を取得した。ここで、同一の日に落札された取引間の時間的順序関係の情報は、取得できないことに注意する。

平均的に一日 1 件以上の落札があるという意味で、分析対象は落札件数 90 件以上の商品に限定し、7 種類の商品の落札価格 (対数値) の分析を行った。データ取得の際には、明らかに落札価格の高低に影響を与えるキーワードを含む商品^(注1) (特異取引) は取り除いた。

3.2 落札価格の時系列変化についての考察

図 1~7 に、落札価格の時系列変化を示す。横軸は、平成 17 年 1 月 1 日からの経過日を表す。同グラフには、以下の特性が見られる。

- 落札価格が定価を越える取引が見られる。更に、定価以上で落札される取引とそれ以外の取引の落札価格が明らかに離れている (商品 ID1, 2, 4, 6, 7)。これは、一般的には商品そのものの価値を越えた価格で落札されることはないと考えられるものの、商品が品薄等の特別な事情がある場合に、定価を越えた落札価格でも商品を手入れしたいと考える利用者が、一部にい

(注1): ジャンク、キズあり、訳あり、海外版、全巻新刊、全巻帯付、おまけ付き

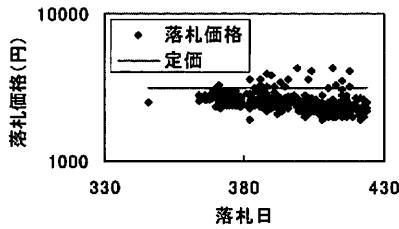


図2 商品ID2の落札価格

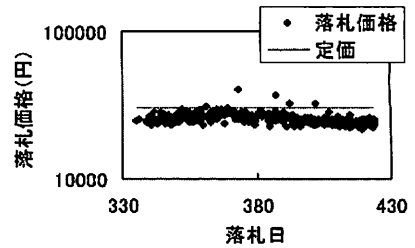


図6 商品ID6の落札価格

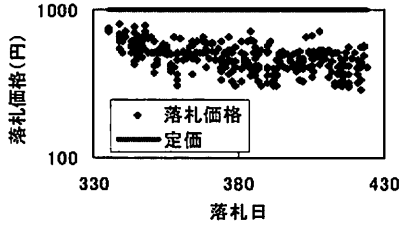


図3 商品ID3の落札価格

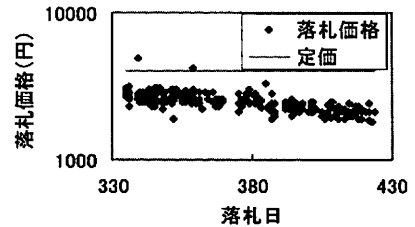


図7 商品ID7の落札価格

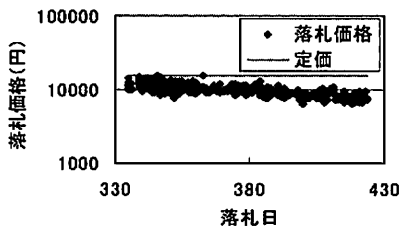


図4 商品ID4の落札価格

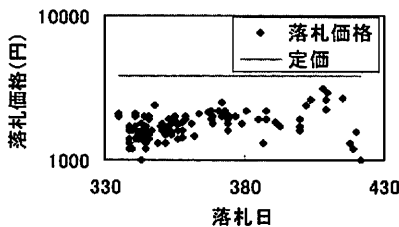


図5 商品ID5の落札価格

るためだと考えられる。定価以上で落札される取引とそれ以外の取引が離れていることから、定価を越えた落札価格でも商品を購入するユーザーによる取引は、特異取引であると考えられる。

● 落札価格の傾向（増加、減少の傾き、価格変動の幅）が変わる時点がある。これは、季節による価格変動、メディアが取り上げた等による一時的な価格の急騰など、外部的な要因によると考えられる（商品ID4, 5, 6）。

4. 分析方法

前節の考察に基づき、以下の手順で取得したデータの分析を行う。

● ステップ1：定価以上の落札価格で落札された取引を特異取引として取り除く。

● ステップ2：一ヶ月の落札件数を N とすると、一日の平均落札件数 n は、 $n = \lceil N/90 \rceil$ となる。 $\delta t = 2n$ （2日相当の意味）とする。回帰分析により、式(2)のパラメータ δ 、 μ 、 λ を決め、 $\hat{\tau}$ を求める。2.節の手順により、モデルの次数を決める。

● ステップ3：モデルから計算される価格と実際の価格の差（残差）の分布について、正規分布性を検定する。選択したモデル、パラメータが妥当であれば、残差は i.i.d. となるので、サンプル数が十分に多い場合、残差の分布は正規分布に従うはずである。残差が正規分布となる場合には、分析を終了する。そうでない場合は、ステップ4に行く。

● ステップ4：落札価格の時間変化を表すグラフから、落札価格の傾向が変わる時点 t_1 を見つける。測定期間を二つの時間区間 $[t_s, t_1]$ 、 $(t_1, t_e]$ (t_s : 測定開始日, t_e : 測定終了日) に分けて、(ステップ2)、(ステップ3) を繰り返す。残差の正規分布性が確認されるまで、時間区間を更に細かく分割していく。

5. 分析結果

前節の手順に従って、商品の落札価格（対数値）の分析を行った。

5.1 モデルの選定と適合性

(ステップ2)において求めた、各々の商品の $\hat{\tau}$ 値を図8に示す。全ての商品で $\hat{\tau}$ は9.5%臨界値を下回り、定常性の仮説が支持された。

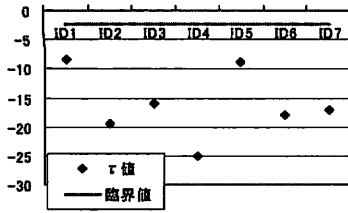


図 8 落札価格の対数に関する τ 値

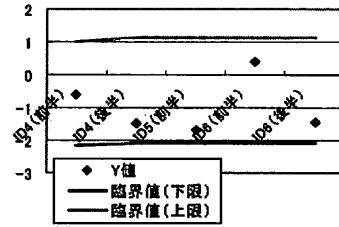


図 12 残差の Y 値 (分割)

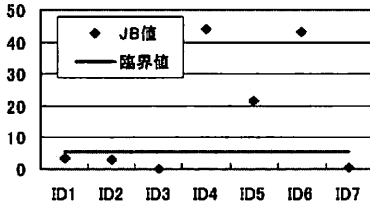


図 9 残差に関する JB 値

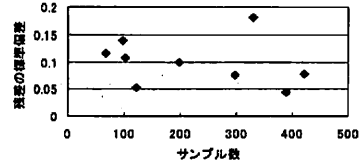


図 13 残差の分散とサンプル数の関係

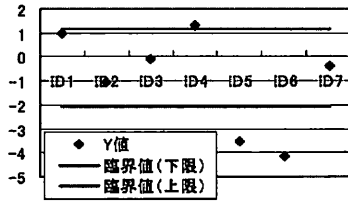


図 10 残差に関する Y 値

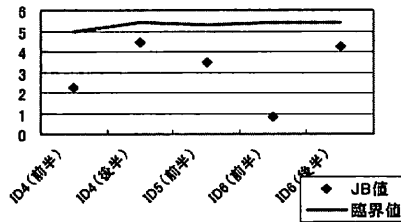


図 11 残差の JB 値 (分割)

次に、残差の正規分布性を検定したところ、7品目中4品目 (ID1, 2, 3, 7) が正規分布に適合した。各々の商品について、残差分布の JB 値 (式 (4) より計算)、Y 値 (式 (7) より計算) を、図 9, 10 に示す。

次に、残差が正規分布とならない3品目 (ID4, 5, 6) について、(ステップ4) を実行した結果を示す。全ての品目で、測定区間を2つの区間に分割することによって、残差が正規分布となった。なお、商品 ID5 については、分割後の後半区間 [396, 450] のデータ数が30に満たなかったため、前半区間の結果のみ記載する^(注2)。

以上の結果より、特異取引を除いた落札価格の時間変化は、区分的に線形なトレンドモデルの組み合わせによって記述できるといえる。

次に、モデルの精度とサンプル数の関係を見るために、残差 ϵ の分散とサンプル数の関係を図 13 に示す。サンプル数が多いと、分散が減少傾向にあることが確認できる。今回の分析における最小のサンプル数は68件であり、実際、残差の正規分布性を示す目的のためだけであれば、サンプル数は30件程度以上で十分と考えられる。しかし、サンプル数を増やすことにより、残差のばらつきを減らすことができる、つまり精度の高いモデル化ができるので、測定区間は極力長くとりサンプル数を多くするほうが望ましいといえる。

6. まとめ

本報告では、ネットオークションにおける落札価格の分析、モデル化を行った。分析の結果、品質の高低を表すキーワードを含む取引、定価以上の取引を除いた落札価格の時系列変化は、区分的に線形なトレンドモデルとして記述できることがわかった。トレンドが線形で記述できる区間長は外部的な要因で決まると考えられ、変化後のトレンドを表すパラメータも不連続に変化する。落札価格の変動を長期にわたって予測することは難しい。しかし、今回の分析法方法は、現在、および数日後の落札価格の分布の予測には十分適用と考えられ、利用者が落札価格の上限、下限の目安をつける手段に応用することができる。今後の課題としては、特異データの抽出トレンドの変化点の抽出を自動で行うことが挙げられる。

文 献

- [1] 総務省, "情報通信白書平成 17 年度版", <http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/h17/pdf/index.html/>
- [2] 山本 拓著, "経済の時系列分析", 創文社, 1998.
- [3] 森谷 千鳳彦著, "ブラックショールズモデル", 東洋経済新報社, 2000.
- [4] オークション統計ページ (仮), <http://www.aucfan.com/>

(注2) : 後半区間の急激な価格の下落は、新版の発売開始による (図 5)