

正規化時系列のクラスタリングによるオンラインオークション 落札価格予測

Forecasting Online Auction Final Price with Normalized Time Series Clustering

横谷 拓也[†]
Takuya Yokotani[†]

黄 宏軒[‡]
Hung-Hsuan Huang[‡]

川越 恭二^{††}
Kyoji Kawagoe^{††}

1. はじめに

近年, インターネットの普及に伴い電子商取引が発達している. 代表的な電子商取引の一つにインターネットを用いた競売であるインターネットオークションがある. インターネットオークションでは多くのユーザが自由に価格を決めるため, 同じ商品でも落札価格にバラつきがあることが挙げられる.

例えば, 図1はインターネットオークションで, ある商品の一年間における落札価格のリストである. 図1の通り, (Jan-10) から (Aug-10) までで最低落札価格が28,001円, 最高落札価格が67,000円と, 落札価格に大きなバラつきが見られる. これは, 例えば, 同じ商品AとBにおいて, 商品Aが開催中だった時は入札者が多く, 商品Bが開催中だった時には入札者が少ないなど, それぞれの商品の様々な状況によって落札価格にバラつきが生じたと考えられる. この落札価格のバラつきにより, ユーザは最終落札価格を予測するのが困難である.

現在, 過去に行われた取引データを用いて商品分析や落札価格予測を行うWebサービスが存在する. しかし, 時間経過によって入札状況が変化することを考慮していないため, 予測価格の精度に問題がある.

そこで, 本論文では, 正規化時系列のクラスタリングによる落札価格予測手法の提案を行う. 本手法では, 過去に取引された入札時系列データについて, 時間軸と価格軸で正規化を行う. さらに, 正規化した時系列データをクラスタリングし, 落札価格を推定する際に, 現時点までの入札時系列データと類似した状況のクラスタを随時比較する. 本手法により, 動的に変化する入札状況に対してリアルタイムに落札予測価格を算出することが可能となる.

| Jul-10 | Jun-10 | May-10 | Apr-10 | Mar-10 | Feb-10 |
|--------|--------|--------|--------|--------|--------|
| 60000円 | 39300円 | 40000円 | 45000円 | 36100円 | 57000円 |
| 37000円 | 43499円 | 41000円 | | 45800円 | 40300円 |
| | 49800円 | 61000円 | | 49990円 | 40000円 |
| | 41611円 | 48000円 | | | |
| | | 61005円 | | | |
| | | 36000円 | | | |
| | | 47000円 | | | |
| | | | | | |
| Jan-10 | Dec-09 | Nov-09 | Oct-10 | Sep-10 | Aug-10 |
| 40000円 | 38500円 | 65000円 | 47000円 | 58000円 | 66900円 |
| 63000円 | 58000円 | 45000円 | 59800円 | 57500円 | 52300円 |
| 39000円 | 50100円 | 46500円 | 63000円 | 52500円 | 59000円 |
| 28001円 | 53000円 | 48700円 | | 63000円 | 66900円 |
| 45500円 | 45000円 | 45000円 | | 55000円 | 56980円 |
| 58800円 | 40000円 | | | 63000円 | 58500円 |
| | | | | 52500円 | 67000円 |
| | | | | 56980円 | |

図1: ある商品の落札価格

2. 正規化時系列のクラスタリングによる落札価格予測

2.1. 基本的考え方

本論文では, 予測落札価格を算出するために, 過去に取引された時系列データとリアルタイムの時系列データの双方を考慮する. 予測落札価格を算出するために, 過去の時系列データを入札履歴の特徴別にクラスタリングを行う.

例えば, 入札開始すぐに価格が跳ね上がってその後は緩やかに価格が伸びる特徴や, 経過時間に比例して順調に価格が伸びる特徴, 入札終了間際に価格が跳ね上がる特徴などが挙げられる. これらの特徴別にクラスタリングを行うことで, 類似した特徴をもつ商品同士の時系列データを分類することが出来る. 予測したい商品のリアルタイムの時系列データが現在どのような特徴をもって推移しているか, その特徴がどのクラスタに分類されるかを判別する. 予測する商品の特徴と, 分類されたクラスタは類似した特徴をもっているため, 予測する商品の今後の価格推移は, クラスタの入札終了までの価格推移と同じように推移する可能性が高いと考えられる. このことを用いて予測落札価格を算出する.

2.2. 予測落札価格の算出

提案手法の処理手順を図2に示す. 図2の処理手

[†]立命館大学大学院理工学研究科

[†]yokotani@coms.ics.ritsumeimei.ac.jp

[‡]立命館大学情報理工学部

[‡]huang@fc.ritsumeimei.ac.jp

^{††}立命館大学情報理工学部

^{††}kawagoe@is.ritsumeimei.ac.jp

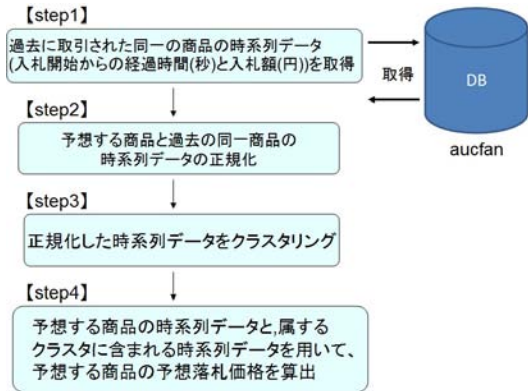


図 2: 予測落札価格の算出

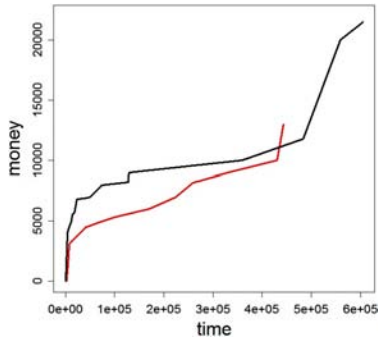


図 3: 正規化前 1

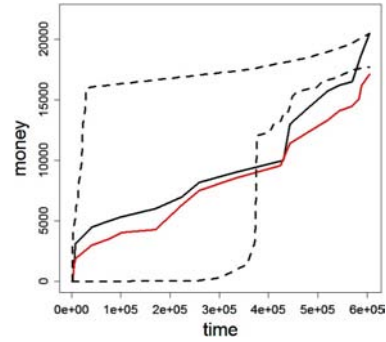


図 4: 正規化前 2

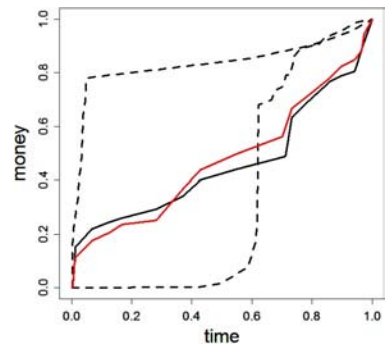


図 5: 正規化後

順を説明する。(step1) 落札価格を予測する商品の、過去に取引された同一の商品の時系列データを取得する。ここで、時系列データとは、入札開始からの経過時間(秒)と入札額(円)の2つの要素からなる入札履歴と定義する。(step2) 予測する商品のリアルタイムの時系列データと過去の時系列データの正規化を行う。(step3) 正規化した過去の時系列データを k-means 法によって特徴別にクラスタリングを行う。(step4) リアルタイムの時系列データと、その時系列データが属するクラスタに含まれる時系列データを用いて、予測落札価格の算出を行う。

2.2.1. 正規化

価格を予測したい時系列データは入札開始からの経過時間が一定ではないため、過去の時系列データと比べると図3のように X 軸(入札開始からの経過時間)において差が生じてしまう。さらに、図4の2つの実線は一見同じ特徴の曲線であるが、Y 軸(入札価格)に差が生じている。このように X 軸と Y 軸に差が生じることによって、そのままクラスタリングを行うと別のクラスタに分類されてしまう恐れがある。そのことを防ぐため、過去の時系列データとリアルタイムの時系列データの正規化を行う。各時系列データを X 軸(入札開始からの経過時間(秒)), Y 軸(入札価格(円))ともに最小値0, 最大値1に正規化する。正規化後の状態を図5に示す。

正規化前の時系列データ D_i は入札履歴のことであり、その入札履歴 $[d_{i1}, \dots, d_{in}]$ は、式(1)に示すように

データの入札開始からの経過時間 $x_j (j = 1, \dots, n)$ と x_j において入札された入札額 y_j の2つの要素から成る2次元ベクトルのシーケンスである。ここで、 n は入札件数である。

$$D_i = \langle d_{i1}, d_{i2}, \dots, d_{in} \rangle, d_{ij} = \begin{bmatrix} x_j \\ y_j \end{bmatrix}, (j = 1, \dots, n) \quad (1)$$

また、正規化前の時系列データ D_i から正規化したデータ S_i は同様に式(2)で表現することができる。

$$S_i = \langle s_{i1}, s_{i2}, \dots, s_{in} \rangle, s_{ij} = \begin{bmatrix} w_j \\ z_j \end{bmatrix}, (j = 1, \dots, n) \quad (2)$$

ここで、 w_j は正規化前の時系列データ D_i の入札開始からの経過時間 x_j を正規化した値であり、 z_j は x_j において入札された入札額 y_j を正規化した値である。それぞれ最小値0, 最大値1に正規化するため、 w_j, z_j をそれぞれ次式で算出する。

$$w_j = \frac{x_j}{\max(x_j)}, z_j = \frac{y_j}{\max(y_j)} \quad (3)$$

したがって、時系列データ数 $M(D_1, \dots, D_M)$ 個を式(3)によって正規化することにより、図5のように曲線の特徴に依ったクラスタリングを行うことが可能となる。

2.2.2. k-means 法によるクラスタリング

本研究では、非階層型クラスタリング手法である K-means 法を用いて時系列データの特徴に依ったクラスタリングを行う。クラスタリングするにあたり、正規化された2つの時系列データ S_i, S_k 間のユークリッド距離を算出する。ユークリッド距離を、時系列データ S_i の

入札履歴 s_{ij} と時系列データ S_k の入札履歴 S_{kj} を用いて式 (4) で算出する。

$$d(S_i, S_k) = \sqrt{\sum_{j=1}^n (s_{ij} - s_{kj})^2} \quad (4)$$

図6はクラスタリングした結果の例をクラスター別に示したものである。

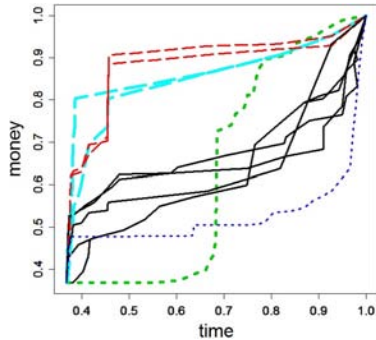


図6: クラスタリングの結果

2.2.3. スプライン補間

入札時刻は時系列データ毎に異なるため、クラスタリングに必要な距離を算出することができない問題が生じる。そこで、時系列データ S_i の入札件数がある一定値以下 (例えば5件以下) の場合、スプライン補間によって入札数の補間を行う。

2.2.4. 予測落札価格の算出

クラスタリングによって得られた結果より、リアルタイムの時系列データが属しているクラスターを判別する。そのクラスター内に存在する m 個の過去の時系列データ D_m の落札価格の平均を式 (5) で算出し、予測落札価格として出力する。

$$= \frac{\sum_{k=1}^m x_j k}{m} \quad (5)$$

3. 評価実験

3.1. 実験方法

提案手法の有効性を示すために実際に取引中の商品の落札価格の予測を行った。そして、予測落札価格が、実際の落札価格とどこまで近い値をとれたかを表す予測精度 $A(\%)$ を式 (6) で算出する。また、既存 Web サービスである Rakuboz[4] での予測価格と同一商品の平均落札価格の2つの値を同様の方法で求め、提案手法との比較を行う。

$$A = \left(1 - \frac{|\text{予測落札価格} - \text{実際の落札価格}|}{\text{実際の落札価格}}\right) \cdot 100 \quad (6)$$

表1: 実験で用いたデータ

| 商品 | 最高～最低落札価格 | 平均落札価格 | 標準偏差 |
|-----|----------------|--------|--------|
| 商品1 | 99.800～66.000 | 87.634 | 10.173 |
| 商品2 | 105.000～40.500 | 56.003 | 16.652 |
| 商品3 | 70.001～34.800 | 59.880 | 7.143 |
| 商品4 | 25.800～15.000 | 19.005 | 3.196 |
| 商品5 | 13.500～5.000 | 8.725 | 2.423 |

表2: 実験で使ったデータ件数

| 商品 | 過去の件数 | 予測した件数 |
|-----|-------|--------|
| 商品1 | 75件 | 13件 |
| 商品2 | 185件 | 40件 |
| 商品3 | 94件 | 26件 |
| 商品4 | 62件 | 10件 |
| 商品5 | 80件 | 11件 |

3.2. 実験条件

実験を行う上での条件を、以下のように定める。

- (i) Yahoo!オークション [5] で取引されている商品を扱う。
- (ii) 同一の商品で過去の時系列データが多く存在している商品を対象とする (60件以上を設定)。更に落札価格にバラつきが存在する商品の落札価格を予測する。
- (iii) 入札開始から落札までの時間の70%の時間が経過している商品を対象として、その落札価格の予測を行う。

上記の条件を考慮して表1,2に示す5種類の商品を対象として実験を行った。予測する商品数は合計100件である。

3.3. 実験結果

実験による商品毎の予測精度の平均値を表3に、標準偏差を表4に示す。

表3: 予測精度の平均 (単位:%)

| 商品 | 提案手法 | Rakuboz | 平均落札価格 |
|-----|-------|---------|--------|
| 商品1 | 89.93 | 63.40 | 88.92 |
| 商品2 | 83.10 | 58.01 | 79.07 |
| 商品3 | 85.41 | 63.98 | 87.4 |
| 商品4 | 88.83 | 52.23 | 85.46 |
| 商品5 | 84.70 | 61.20 | 75.19 |
| 全体 | 85.53 | 60.02 | 82.73 |

なお、Rakubozでは予測落札価格を出力する際、ERRORを出力することがあった。その場合、予測精度を0%で計算した。

実験結果より、提案手法で高い評価を得ることが出来た。提案手法の予測精度が85%を超えており、これは、実

表 4: 予測精度の標準偏差 (単位:%)

| 商品 | 提案手法 | Rakuboz | 平均落札価格 |
|------|-------|---------|--------|
| 商品 1 | 7.42 | 32.89 | 7.97 |
| 商品 2 | 10.53 | 35.97 | 13.41 |
| 商品 3 | 18.43 | 33.28 | 5.06 |
| 商品 4 | 5.04 | 36.24 | 5.38 |
| 商品 5 | 6.66 | 37.49 | 12.87 |
| 全体 | 12.38 | 35.29 | 11.45 |

際の落札価格の ± 1.5 割の値段以内に予測落札価格を算出できていることになる。本実験は落札価格がバラついている商品を対象としているため、高い予測精度を得たといえる。

Rakuboz では本手法と比べて予測精度の標準偏差が大きな値となり、予測落札価格の値に不安定な場面が多々見られた。その結果、提案手法との結果に差が出たと思われる。

平均落札価格では、平均予測精度が 80% を超え、標準偏差も 11.45 と、全体的に安定した結果が出ている。しかし、提案手法が平均落札価格より全体的に上回る結果となった。これは、本提案手法が類似した特徴を持つ過去のデータだけに絞り込み、そのクラス内で平均値を出して予測落札価格として算出しており、提案手法の有効性を示すことが出来た。しかし、商品 3 の標準偏差が 18.43 となり、平均落札価格による結果と比べて下回る結果となった。商品 3 における結果を分析して改善することが今後の課題である。

さらに、落札価格を予測する手法の 1 つに回帰式を用いるものがある [1]。この手法 (4.1 節の式 (7)) と本提案手法を比較するため、4.2 節で挙げた条件 (i) ~ (iii) と同様に、商品 2 の合計 40 件の商品で評価実験を行った。その結果を表 5 に示す。

表 5: 提案手法と回帰式の評価実験

| | 提案手法 | 回帰式 | 平均落札価格 |
|------|-------|-------|--------|
| 平均 | 83.10 | 83.22 | 79.07 |
| 標準偏差 | 10.53 | 18.83 | 13.41 |

表 5 に示すように、提案手法が平均で回帰式による手法と同程度の精度を得た。また、標準偏差で上回る結果となった。回帰式は 99% を超えるなど非常に高い予測精度となる場合がしばしば見られた。しかし、38% や 45% など低い予測精度となる場合もあり、不安定な結果となった。表 5 により、安定性の面で本提案手法の有効性を示すことが出来た。

4. 先行研究および Web サービス

4.1. 回帰式による価格分析

黒澤ら [1] は実際のインターネット上のオークションから入札履歴や落札価格の動向などの入札行動に関するデータを収集し、分析した。分析の結果、価格変動様式

を式 (7) で表すことが可能だとした。

$$P = \alpha_2 T^{\alpha_1} \quad (7)$$

α_1, α_2 はオークション変動係数、 P および T はそれぞれ正規化した価格と時間を表す。また、稀なケースで取引開始直後にある程度の価格上昇が起き、その後は大きな価格上昇が生じない傾向のある商品が存在することを示し、その場合は価格変動様式を式 (8) で表す方が適切だとした。

$$P' = \beta_1 \log T' + \beta_2 \quad (8)$$

β_1, β_2 はオークション変動係数、 P' および T' はそれぞれ正規化した価格と時間を表す。

4.2. aucfan・aucDATA

aucfan[2]・aucDATA[3] は、過去に行われた膨大な量の取引データを蓄積しており、そのデータから様々な視点で商品を分析することが可能な Web サービスである。ユーザは多岐に渡る分析要素から目的別に条件指定し、商品の分析を行う。その分析によって商品の相場を得られる。しかし、過去のデータのみを扱っており、落札価格を予測する商品が時間経過によって入札状況が変化することを考慮していない。そのため、これから起こりうる様々な状況に過去の取引データのみで対応することが困難であるという問題がある。

4.3. Rakuboz

Rakuboz[4] は、Yahoo!オークションにおいて商品の予測落札価格を得ることが可能な Web サービスである。しかし、現在の入札価格よりも低い結果が出るなど、リアルタイムの情報を考慮していない。

5. おわりに

本論文では正規化時系列によるオンラインオークション落札価格予測の手法の提案を行った。提案手法によって落札価格にバラつきがある商品においても適切な落札価格を予測することが可能となった。また、評価実験によって本提案手法の予測精度が平均 85.33% となり、Rakuboz の平均 60.02% と平均落札価格 82.73% と比較して本提案手法が有効であることを示した。

参考文献

- [1] 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, 情報処理学会研究報告, 電子化知的財産・社会基盤, 2001(118), p7-14
- [2] aucfan, <http://aucfan.com/>
- [3] aucDATA, <http://data.aucfan.com/>
- [4] Rakuboz, <http://www.rakuboz.com/>
- [5] Yahoo オークション, <http://auctions.yahoo.co.jp/>
- [6] 元田浩, 山口高平, 津本周作, 沼尾正行: データマイニングの基礎, オーム社出版
- [7] Jank W, Shmueli G, Wang S: Dynamic, real-time forecasting of online auctions via functional models, ACM SIGKDD(2006), pp580-585