

ネットオークションにおける不正検出に関する考察

Shill Bidders detection for online auction

吉田 剛^{*1} 松井 藤五郎^{*2} 大和田 勇人^{*2}
Tsuayoshi Yoshida Tohgorou Matsui Hayato Ohwada

^{*1} 東京理科大学 大学院 理工学研究科 経営工学専攻

Department of Industrial Administration, Graduate School of Science and Technology, Tokyo University of Science

^{*2} 同理工学部経営工学科

Department of Industrial Administration, Faculty of Science and Technology

Recently many people are using online auction. Since the market has been expanded quickly, many problems have been spread. Unfortunately the fundamental measure are still required. This paper propose a method to detect shill bidders for online auction. It detects out layers with One-Class SVM. Then it transform the results into the form of decision tree using C4.5. We confirm that we can understand the resulting rule as the rule classify shill bidders.

1. はじめに

近年、インターネットの普及により、多くの人々がネットオークションに参加するようになってきた。ネットオークションとは、販売目的でネット上のサイトに出された商品を、最も良い購入条件(より高額)を提示した買い手に売却するために、各々の買い手が提示できる購入条件を競う取引形態を示す。しかしながら、急速に市場が拡大してきた事から、数々の問題点を生んでいる[1]。問題点の例としては価格の吊り上げ行為や、本物と区別の付きにくい偽ブランド品の存在などがある。

この様なネットオークションを利用する上でのリスクを極力減らし安全に利用するためには、不正な利用者を取引の相手として避ける手段の実現が望まれる[2]。そのため本研究では、[1],[2]の相補的な研究として、不正行為である価格の吊り上げ行為を検出する方法を検討した。

初めに、ネットオークション内の情報をシェルスクリプトを利用して取得し、得られたファイルから Perl を用いて本文内の情報を抽出した。抽出した情報を One Class Support Vector Machine(SVM)[3]へ入力し異常行動者を判別し、判別の結果を用いて C4.5[4]によって異常行動者に関する分類木を作成した。作成された分類木を分析することで、異常行動者に関する分類木が吊り上げ行為のルールとして理解できることを確認した。

以下、2章でネットオークションについて述べ、3章で吊り上げ行為の検出手法について述べる。4章で実験結果を述べ、5章を結論とする。

2. ネットオークションについて

ネットオークションとは、インターネットなどの通信サービス上で行われるオークションのことで「オンラインオークション」や「インターネットオークション」とも呼ばれている。出品者はWEBサイト上に、商品の名称や写真、状態、入札期限、配送方法、支払方法などの情報を掲載し、入札者が現れるのを待つ。期限内に最も高値を提示した入札者が商品を落札し、出品者と電子メールなどを使って連絡を取り合い、商品と代金を交換する。

連絡先: 吉田剛, 東京理科大学理工学研究科経営工学専攻
大和田研究室, 千葉県野田市山崎 2641, 04(7124)1501,
yoshida @ ohwada-lab.net

2.1 オークションサイト

本論分では実際のオークションサイト[5]を対象に分析を試みる。分析対象のオークションサイトは、出品者は必ずクレジットカードの番号などの個人情報をおオークションサイト側に教えるシステムになっているため、他のオークションサイトと比べて安全性は高いと思われる。また、利用者の数も多く一日の取引数も多いだけでなく、長年運営されていて利用者の信頼度も高い。しかし、入札者はメールアドレスを複数用意することで複数のIDを作成することが可能であるため、これを悪用した問題点が残っている。なお商品取引に関する情報は、出品アイテム自体の情報の他、出品者に対する過去の評価と出品アイテムへの入札履歴の情報があり、後者からは入札者に対する過去の評価の情報として整理することが出来る。

2.2 吊り上げ行為

本研究で取り扱うオークションサイトにおいては、自動入札というシステムを利用した吊り上げ行為の問題が大きい。吊り上げ行為とは、出品者Aが自分の出品しているアイテムを入札者 B が入札している時に、

- ① 出品者Aの仲間Cが入札者Bよりも高い価格で入札する。
- ② 再度、入札者BがCよりも高い価格で入札し直す。

上記の①②を繰り返すことで出品アイテムの落札価格を上げて、不当な利益を得るといふものである。

自動入札システムは、自分の支払える上限の金額を入札しておく、誰かが入札した時にシステムが自動的に入札し返してくれる仕組みである。本来予算内であれば、安く落札できる仕組みであるが、これを不正に利用することで吊り上げ行為の問題が大きくなる。

3. 提案手法

吊り上げ行為を行っている利用者は、オークションに参加している一般の利用者とは違った異常行動をしているものと思われる。異常なデータを判別するための研究としては、One Class SVM を使った小野田らの[3]が代表的なものである。そこで本研究でも One Class SVM の利用を考えた。

しかし、One Class SVM の結果は可読性がないため、どのようにルールが作られているかを知ることが出来ない。そこで、C4.5を用いて One Class SVM の結果を分析し、異常行動者の

分類木の作成を試みた。作成された分類木を分析し、吊り上げ行為者のルールとして理解できることを確認した。具体的には、図1に示すように以下の①～⑤のステップでデータを分析した。

- ①:オークションサイトから現在出品されているアイテムとその入札履歴を取得した。出品アイテムのページからは出品者の過去の評価がされている評価のページを、入札履歴からは入札者に対する過去の評価がされている評価のページを取得した。更に、取得した出品者及び入札者の評価のページより、過去に出品したアイテム及び落札したアイテムの入札履歴のページを取得した。
- ②:①より、入札者ごとの情報として「入札者 ID、総落札数、平均上乗せ額(%), 前の人との平均入札時間差、一番落札した出品者 ID、その落札した割合、二番目に落札した出品者 ID、その落札した割合、悪いと評価された回数、のべ評価、同じ人からの評価を含まない評価、総入札回数、上乗せ額が100%を越えた割合、100%丁度の割合、100%を下回った割合」の15属性を取得した。
- ③:②の情報に元、One Class SVMによって異常行動者(吊り上げ行為)の抽出を行った。
- ④:C4.5を用いて異常行動者の分類木を作成した。
- ⑤:C4.5 で作成された分類木を分析し、属性の取捨選択を繰り返した。そうすることで、より良い分類木の作成を試みた。

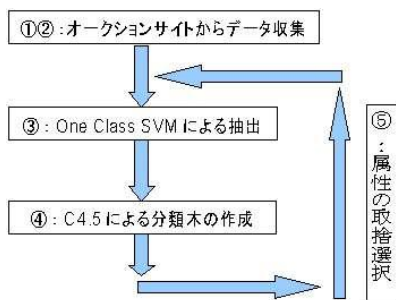


図1:提案手法の概要

4. 結果

オークションサイト内の全ての出品アイテムに対して実験を行うと、データが膨大な量となってしまったため、出品アイテム数が5,000件程のカテゴリーである昆虫生体(4,515件)を対象に実験を行った。2008年12月6日にシェルスクリプトを用いて、59,949人の入札者の情報を集めた(アイテム数67,244件:過去取引を含む)。集めた情報の中で総入札の回数が5回未満のものについては、分析を行うにはデータが不十分で、ノイズとなる可能性があると考え、総入札数が5回以上の入札者(11,265人)に関してのデータのみを以下の分析の対象とした。

なお、入札者ID及び取引相手のID(数値)を属性として与えてしまうと、誤った分類をしてしまう可能性が高くなるため、それを除いた12の属性を用いて実験した。また、One Class SVMで検出させる異常入札者の割合は0.001~3%に設定した。12の属性全てを用いてOne Class SVMで検出させると、異常入札者の割合が0.01%程度より小さい場合に、異常行動者の分類木として理解可能なものが作成できた。また、ほぼ同じ集団を集めていた。しかし、異常行動者の割合を0.5%程度より大きくすると、どういった入札を行った人を分類したのかが理解しにくい分類木が作成された。これは、不要な属性を用いた無理な分類が行われたと考えた。そこで、異常行動者として理解できる分類木を作っている属性を中心に属性の取捨選択を繰り返し、最終的に8属性を用いてOne Class SVMで異常行動者の判別を行い、C4.5で作成した分類木が、図2である。

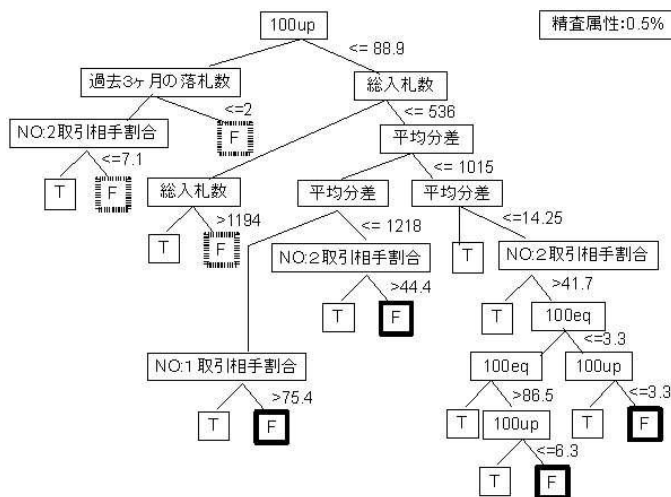


図2:精査した属性を用いた分類木

破線で囲まれた leaf node Fへ辿り着く枝は資金に余裕のある入札者、もしくはオークションを良く利用している入札者であると思われるが、太線で囲まれた leaf node Fへ辿り着く枝は、そこに辿り着いた入札者の情報を見た結果、異常行動者すなわち吊り上げ行為を行っていると思われる不正な入札者であると確認できた。

また、以上の方法によって作成された分類木及び吊り上げ行為者はデータを見ることで容易に判別が可能である。異常行動者の割合を1%とした時に、C4.5で吊り上げ行為者と識別した入札者の内、実際に吊り上げ行為を行っていた入札者は約75%であった。また、精査した属性を用いて異常行動者を分類すると、異常行動者の割合を0.3~2%程に変化させても、吊り上げ行為を行っている集団を分類している枝に集まる人数はそれほど変化がなく、資金に余裕のある入札者もしくはオークションをよく利用している入札者の人数が増えた(その枝にいくためのルールが緩くなった)。

以上のことより、提案手法によって吊り上げ行為を検出することが出来ることを確認した。

5. 結論

本研究では、ネットオークションから抽出した入札者に関する情報をOne Class Support Vector Machineの入力として与え異常行動者を判別し、C4.5によって異常行動の分類木を作成した。C4.5で作成された分類木を分析することで、その異常行動の分類木が、吊り上げ行為を識別できることを確認した。

参考文献

- [1] 横尾真. インターネットオークションの理論と応用. 人口知能学会誌,15(2):404-411.
- [2] 松尾徳朗. 架空名義入札者発見に基づく不正入札に頑健なオークションメカニズム. The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 2006
- [3] 小野田崇. One Class SVMに基づく水力発電所軸受異常振動の予兆発見. The 18th Annual Conference of the Japanese Society for Artificial Intelligence, 2004
- [4] Weka :<http://www.cs.waikato.ac.nz/ml/weka/index.html>
- [5] ビッドーズ :<http://www.bidders.co.jp/>