

電子商取引におけるユーザの関係に関する クラスタリング手法の提案

A suggestion of clustering methods about user relationships on e-commerce

小林真雄*¹ 伊藤孝行*²
Masao KOBAYASHI Takayuki ITO

*¹名古屋工業大学大学院産業戦略工学専攻
Master of Techno-Business Administration Nagoya Institute of Technology

*²名古屋工業大学 情報工学科/産業戦略工学専攻
Department of Computer Science / Master of Techno-Business Administration Nagoya Institute of Technology

Due to the recent widespread use of the Internet, many people easily access to the Internet Auctions as e-commerce trading. At the same time, the network structure like WWW has become huge and the large structure is analyzed on a grand scale. This paper proposes clustering methods on user relationships on Internet auctions. Our algorithm called "Auction Network Trust (ANT)" employs HITS's techniques and Internet auction's data. At this stage, we successfully implemented a crawler for Internet auction sites and compared our algorithm to a reputation value of Internet auctions with several approaches such as user rankings. Furthermore, our works possess a network analyzing system on a more large trading network that predicts which buyers and sellers are active and did better behaviors. Our experiments show many behavior of the Internet auctions and that ANT presents different scores from HITS on WWW and rankings by the reputation values.

1. はじめに

昨今、インターネット上における商取引が盛んであり、オンラインショッピングやネットオークションは代表的なものである。Yahoo!JAPAN の統計 [1] によると、インターネット利用者の中で、特にインターネットオークションの利用経験のある者は全体の 70% 以上である。本統計は他の再流通消費活動、つまり古本屋やフリーマーケットといった市場に比べて、インターネットオークションが最も利用されていることを示している。また、利用者数自体も年々増加している。国内最大手のインターネットオークションサイトである Yahoo!オークション*¹ の 2008 年 2 月の月次報告 (2008 年 3 月 10 日発表) では、サイト内有料サービスであるオークションの利用権限を持つプレミアム会員数が 680 万 ID にのぼる。一方で、近年、詐欺などの不正行為がメディアで報じられている。オークションサイト側はオークションの監視、データの解析を強化することで不審者を排除する働きかけを行っている。しかし、不正行為は防ぎきれない。一方、学術的な背景として、ネットワーク構造の研究が盛んに行われている。交友関係や企業間協定など、社会的ネットワークの持つ性質が数学的モデル [2] によって提唱されてから、複雑ネットワークという考え方が多くの分野でとり取り扱われてきた。特に近年、インターネットの普及により WWW といった新しいネットワーク構造が Web 上で拡大し、解析が盛んに行われている。

以上のような観点から、インターネットオークションをネットワーク構造として捉え、解析評価することでユーザ間の関係における新たな知識発見を支援することを主題とする。本論文では、オークションネットワークを評価する為のアルゴリズムとして Auction Network Trust (以下 ANT) を提案する。ANT をオークションネットワークに適用することで、個々の

連絡先: 小林真雄, 名古屋工業大学大学院 産業戦略工学専攻, 愛知県名古屋市昭和区御器所町, TEL:052-735-7968, FAX:052-735-7407, kobayashi@itolab.mta.nitech.ac.jp

*¹ <http://auctions.yahoo.co.jp/>

ユーザの出品者、また落札者としてのスコアを算出する。本手法の評価実験を実データに対して行った。実験結果で分かったところは主に以下の 3 点である。

- 評価値に依存しない取引評価指標をネットワーク構造を用いて算出した。
- Web における HITS の問題点を利用し、有益な情報として利用できた。
- 評価値差を用いたスコアリングによって優良ユーザを発見した

本論文の構成として、第 2 章では関連研究を述べる。第 3 章では本論文での提案手法である ANT に関して述べる。第 4 章では、実際のオークションサイトからデータを収集する為のクローラプログラムについて述べる。第 5 章では、ANT をクローラで収集した実データを用いて評価実験を行った。第 6 章ではまとめと今後の課題を述べる。

2. 関連研究

本章では、本研究に関する関連研究として、インターネットオークションを取り扱った既存研究とネットワーク評価に関する研究の 2 種類の視点から述べる。

2.1 インターネットオークションに関する研究

インターネットオークションに関する研究は様々な主題が置かれているが、特に不正者の発見や支援システムの開発があげられる。前章でも述べたように、インターネットオークションでは詐欺行為が問題になっており、オークションに内包されているデータから詐欺者を同定しようという研究が多い。代表的なものとして、オークションでの評価時間に着目したコミュニティ抽出 [4] や、オークション内の取引関係から確率推論を用いて特異なパターンを抽出し、詐欺者を同定する研究 [5] などがある。一方、システム実装に関して、オークションの入札支援を複数のエージェント間の協調によって行うもの [6] もある。

2.2 ネットワーク評価法

ネットワーク構造を解析するにあたり、他の分野で使われている評価法について述べる。近年では特に Web 上でのネットワークを解析するのが盛んであり、検索エンジンのランキングアルゴリズムとしてもネットワーク評価の手法が使われている。PageRank[7] は検索エンジン Google で使われているランキングアルゴリズムで、Web ページのリンク構造を評価している。また別の Web ページ評価アルゴリズムに HITS[8] がある。HITS はネットワークを構成するノードの情報がリンクを通じネットワーク全体に伝播することで、ページの評価を行う。HITS は多数のページからリンクを得る良質なページ「authority」と良質なリンク集ページ「hub」という概念を持つのが特徴である。また、HITS にはいくつかの問題点と、多くの改良法が提案されている [3]。本論文におけるインターネットオークションの評価法 Auction Network Trust については次章で詳細に述べるが、PageRank や HITS と同様に、ノードとリンクで構成されるグラフを隣接行列、または隣接リストを用いて計算、評価することが共通的な手法である。

3. Auction Network Trust

3.1 ANT の概要

本章では、ネットワーク解析の観点から、インターネットオークションを捉えるために Auction Network Trust (以下 ANT) を提案する。ANT では、先のネットワーク評価法で述べた HITS 要素に基づき、インターネットオークション独特の特徴を反映させることで、オークションユーザをスコアリングするものである。詳細な設定は後に述べるが、ANT では、多くのインターネットオークションサイトで採用されている評価システムによる評価値に依存しないスコアリングを目指す。オークションでの個々の取引では多くの情報が内包されており、他者との関係によって構成される情報も多い。ANT では以上のような情報の伝播によって得られる信頼性のうち、以下の4つのオークションでの要素を取り扱う。

1. 取引終了時刻と評価時刻の差
2. 取引時の落札者の評価
3. 同二者間での取引回数
4. 取引者間の評価値差

以上を隣接行列要素の定義に付加してスコアを算出する。ネットワーク解析の前提として、インターネットオークションにおけるネットワーク要素の扱いを以下の表 1 で示す。

表 1: ネットワーク要素の意味

ネットワーク要素	インターネットオークションでの意味
ノード	オークションユーザ
パス (A B)	有向な取引関係 (A:出品者, B:落札者)

3.2 ANT アルゴリズム

オークションでの取引関係を表す隣接行列を A とする。ユーザ i からユーザ j に商品が売られたとき $A_{ij} = 1$ とし、取引が存在しないとき、 $A_{ij} = 0$ となる。 A の要素定義式 (1) は、ANT の基本となる式であり、Web における HITS の隣接行列決定に用いられるものと同じである。

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

$$a_{ij} = \begin{cases} 1 & (\text{出品者 } i, \text{落札者 } j \text{ という取引があるとき}) \\ 0 & (\text{取引がないとき}) \end{cases} \quad (1)$$

ユーザ間の隣接行列 A に対して次の反復計算を行うことでネットワーク全体における、個々のユーザが出品者としてどの程度重要なのか、また落札者としてどの程度重要なのかをベクトルで出力する。落札スコア $X_i = (x_1, x_2, \dots, x_n)$ 、出品スコア $Y_i = (y_1, y_2, \dots, y_n)$ を n 次元ベクトルとし、 $X_0 = (1, 1, \dots, 1)$ 、 $Y_0 = (1, 1, \dots, 1)$ と初期化する。

$$X_{i+1} \leftarrow A^T Y_i \quad (2)$$

$$Y_{i+1} \leftarrow A X_{i+1} \quad (3)$$

$$X_{i+1} \leftarrow \frac{X_{i+1}}{\sum_n X_{i+1}}, Y_{i+1} \leftarrow \frac{Y_{i+1}}{\sum_n Y_{i+1}} \quad (4)$$

式 (2), (3) の更新ルールは各スコアが収束するまで繰り返される。式 (4) では正規化を行っている。上記の式の適用は隣接行列に関連する固有値計算に帰結する。計算時間の関係から、反復計算の終了条件は最大固有値の近似を用いている。計算によって得られるスコアが表すユーザ間の関係は、良い出品者ほど良い落札者に商品を売り、またより落札者ほど良い出品者から商品を買っている、ということを用いて定義している。

各取引の終了時刻と評価時刻との差を用いる隣接行列の定義式は式 (5) である。落札者 j の出品者 i に対する評価時刻を R_{ij} 、 i と j との取引終了時刻を T_{ij} とする。時刻の差を用いる理由はインターネットオークションにおいて、取引終了後の手際の良さで相手に対する評価が決まることが多いからである。

$$a_{ij} = \begin{cases} R_{ij} - T_{ij} & (\text{出品者 } i, \text{落札者 } j \text{ という取引があるとき}) \\ 0 & (\text{取引がないとき}) \end{cases} \quad (5)$$

取引時の落札者の評価を用いる隣接行列の定義式は式 (6) である。落札者のみの評価を用いる理由は、取引において落札者の出品者に対する評価を重視したためである。また出品落札両者の評価を計算に用いるとネットワークでの情報の伝播が打ち消し合ってしまうため、期待する値が得られない。

$$a_{ij} = \begin{cases} -3 & (j \text{ が非常に悪いと評価したとき}) \\ -2 & (j \text{ が悪いと評価したとき}) \\ 0 & (\text{取引がないとき}) \\ 1 & (j \text{ がどちらでもないと評価したとき}) \\ 2 & (j \text{ が良いと評価したとき}) \\ 3 & (j \text{ が非常に良いと評価したとき}) \end{cases} \quad (6)$$

同二者間での取引回数を用いる隣接行列の定義式は式 (7) である。取引回数を用いる理由は、インターネットオークションにおける複数回取引は当事者間の信頼度が高いことを示しているからである。

$$a_{ij} = \begin{cases} N_{ij} & (N_{ij}: \text{出品者 } i, \text{落札者 } j \text{ の時の取引回数}) \\ 0 & (\text{取引がないとき}) \end{cases} \quad (7)$$

取引者間の評価値差を用いる隣接行列の定義式は式 (8) である。 R_i は出品者 i の評価値、 R_j は落札者 j の評価値を表す。評価値差を用いる理由は、取引において特に落札者は相手の評価値を考慮しているからである。両者の評価値が高く、且

つその差が低い取引に高い重みを付加する．

$$a_{ij} = \begin{cases} \frac{R_i + R_j}{2|R_i - R_j| + 1} & (\text{出品者 } i, \text{ 落札者 } j \text{ の取引があるとき}) \\ 0 & (\text{取引がないとき}) \end{cases} \quad (8)$$

4. データ収集プログラム

本章では、実際のオークションサイトからデータを収集する巡回プログラム（以下クローラ）について述べる．本論文では、国内最大の規模を誇る Yahoo!オークションに準じてデータ収集を行った．クローラは Java で実装し、データベースに MySQL を使用した．クローラは探索基点となる Yahoo!オークション ID からデータ収集を開始し、取得した HTML を解析、ネットワーク構成に必要な要素など様々なパラメータを取得する．下記の図 4. にクローラのご概念図、表 2 に実際に Yahoo!オークションから個人のデータとして取得してきているものを MySQL のテーブルに格納している形で示す．

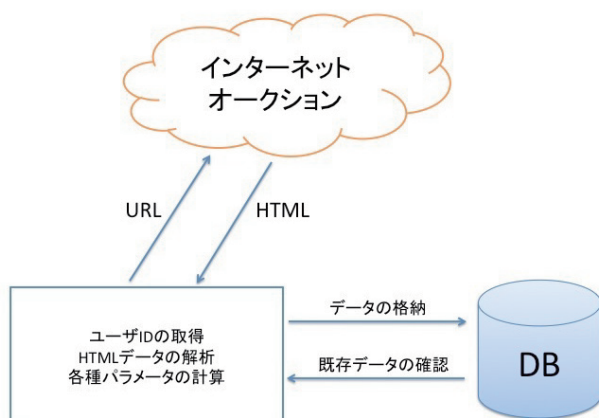


図 1: クローラのご概念図

表 2: ユーザ個人テーブルの構成

Field	データ型	説明
name	varchar	取引相手の ID
rep	int	取引相手の評価値
account	varchar	取引相手のアカウント情報
stance	char	取引相手の取引主体
rate	char	取引相手のその取引での評価
ttime	datetime	その取引の終了時刻
rtime	datetime	取引相手の評価時刻
cancel	varchar	取引キャンセル時の情報

5. 評価実験

実際にインターネットオークションサイトからクローラを用いてデータを収集し、ANT によるネットワークの評価を行った．以下に、実験環境と実験設定を示す．

- ノード（ユーザ）数：3,764
- リンク（取引）数：7,474(複数取引パスも含む)
- 計算機：iMac MacOSX 10.5.2 Intel Core 2 Duo 1G

- データベース：MySQL5.0
- 言語：Java(JDK1.5)

下図 2 は、今回収集したオークションデータを複雑ネットワーク可視化のオープンソースプラットフォームである Cytoscape で画像出力したものである．レイアウトに Inverted Self-Organizing Map Layout を採用している．

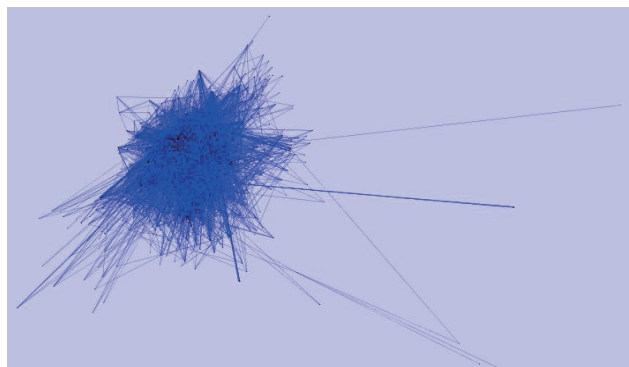


図 2: オークションネットワーク

今回の実験におけるノード数、またリンク数は Yahoo!オークションの規模からすれば全体を把握しているとは言えない．そこで本論文で提案した ANT が既存のオークション信頼指標である評価値と、どのような関係にあるのかについて検証する．今回取得したユーザデータに対して、前章で述べた隣接行列設定毎に ANT を算出する．式 (1) (5) (6) (7)、および (8) で計算される ANT をそれぞれ ANT(HITS) (従来の HITS)、t-ANT、r-ANT、m-ANT、d-ANT とする．実験における個々のユーザの持つ具体的なパラメータとしては (各 ANT の出品落札スコア、各スコアでの偏差値、自身の評価値) となる．上記の 5 つの ANT を計測し、出力される各スコアや偏差値について、評価値との比較のもと以下の項目に付いて検証する．

実験 1 平均以上の評価値を持つユーザ群の各スコアの偏差値の平均値

実験 2 平均以下の評価値を持つユーザ群の各スコアの偏差値の平均値

実験 3 各 ANT における両スコアの合計の上位 10 ユーザ中の優良ユーザ数

実験 1 の目的は、高い評価値を持つユーザの得た各スコアを見ることで、ANT スコアが評価値が高いユーザをどのように評価しているかを検証する．なお、取得ユーザの平均評価値は 1021.24 であり、各スコアの偏差値 0 がスコアの平均となる．表 3 中の [出品]、[落札] は各スコアを表す．実験 2 の目的は、実験 1 とは反対に低い評価値と ANT スコアとの関係を見る．実験 3 の目的は、実際にスコアをランキング指標として捉え、人の目（著者）でオークション内ページを確認し優良であるか判断した．

表 3 で注目すべき点の一つは、従来の HITS である ANT(HITS) と r-ANT が全実験においてほぼ同じ値を示した．このような結果になった原因として取引において互いの評価が異なることが少ないことをが考えられる．ルール適用時の正規化によって重みが吸収され、無評価の取引として計算され

表 3: ANT と評価値の実験結果

ANT	実験 1 [出品]	実験 1 [落札]	実験 2 [出品]	実験 2 [落札]	実験 3
ANT(HITS)	0.0707	-2.7339	-0.0127	0.4925	5
t-ANT	2.7739	-0.4870	-0.4997	0.08773	4
r-ANT	0.0704	-2.6899	-0.0127	0.4846	5
m-ANT	1.1674	0.6914	-0.2103	-0.1246	9
d-ANT	0.9713	1.001	-0.1750	-0.1803	10

ている。

実験 1 について, m-ANT と d-ANT は評価値の高さに依存して高いスコアを示した。一方で, 取引の立場によっては, スコアが低くなった ANT スコアもあった。以上の不一致は, 全ての要素が必ずしも評価値に依存するとは限らないことは言っている。

実験 2 では, 実験 1 で見られる性質が表れている。また, 出品と落札のスコアの差を実験 1 でのスコアの差と比較してみると, 等量にはなっていないことが分かる。これは ANT によるスコアリングでユーザを大別できることを意味する。また, ANT(HITS) と m-ANT を比較すると, 出品と落札で値の差に違いがある。この値の差は, ネットワークを通じて評価を伝播したことによるものだと考えられる。HITS での問題点であるホスト間相互補完関係を ANT では有益な信頼関係として評価している。

実験 3 では, 実際にスコアが高かったユーザの Yahoo!オークションでの評価ページを見て, 優良ユーザであるかを判断した。主な判断材料として, ユーザの自己紹介やプロフィールの記載, 自身の評価値に対して悪評価の比率, 悪評価があった場合に適切な対処ができていないか, などを用いた。結果として, d-ANT のスコア上位者については, オークション利用時にユーザが知り得る情報で信頼できると判断できることが分かった。以下に実験によって分かったことを述べる。

- t-ANT, および r-ANT によるスコアリングで得られた出品スコアと落札スコアにおいて評価値に依存しない値が得られた。
- Web における HITS の問題点であるホスト間相互補完関係を, m-ANT では取引回数をユーザスコアに反映させ, 良評価と悪評価を分化させたスコアリングを行えた。
- d-ANT によるスコアリングの上位 10 ユーザの持つ実際のオークションサイト内ページは全て優良ユーザと判断できた。

6. まとめ

本論文では, インターネットオークションサイトをネットワーク構造として捉え, 従来のオークションユーザの信頼指標である評価値とは違う指標として, ネットワーク構造を評価する Auction Network Trust を提案した。Yahoo!オークションの実データを用い, 4 つのオークション要素を取り入れてた ANT によって評価実験を行った。実験結果で分かった点として以下のようにまとめた。

- 評価値に依存しない取引評価指標をネットワーク構造を用いて算出した。
- Web における HITS の問題点を利用し, 有益な情報として利用できた。

- 評価値差を用いたスコアリングによって優良ユーザを発見した

今後の課題として第一にオークション要素の選択が上げられる。本論文では, 評価時間差, 落札者評価, 複数取引回数, および評価値差の 4 項目を扱ったが, インターネットオークションでは, 従来のオークションやショッピングサイトとは異なる多くの要素を内包している。オークション要素を如何にアルゴリズムに取り入れていくことが重要である。第二に, 扱うデータの規模の問題がある。本論文で, Yahoo!オークションからクローラを用いて取得したデータは, オークション全体から見ればごく一部である。大規模なインターネットオークションを扱うために, 効率よくデータを収集するためのクローラの改良が課題である。

参考文献

- [1] ネット生活予測 No.004 新しいライフスタイル「再流通消費」到来の兆し <http://docs.yahoo.co.jp/info/report/s>
- [2] Watts, D.J., and Strogatz, S.H., "Collective dynamics of small-world networks", Nature 393, pp. 440-442 (1998)
- [3] 手塚 友, 浅野泰仁, 西関隆夫, "現在の Web における HITS について", 信学技報, vol. 105, no. 679, COMP2005-62, pp. 37-44, 2006
- [4] 平手勇宇, 相吉澤明, 翁松齡, 井奥雄一, 木戸冬子, 山名早人, "インターネットオークションにおける不正行為者の発見支援," 日本データベース学会 Letters, Vol.5, No.2, pp. 77-80, 2006.
- [5] Shashank Pandit, Duen Horng Chau, Samuel Wang, and Christos Faloutsos NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks Proceedings of the 16th international conference on World Wide Web (WWW'07). May 8-12, 2007. Banff, Alberta, Canada. pp. 201-210
- [6] 伊藤孝行, 服部宏充, 新谷虎松, "エージェント間の協調的入札機構に基づく複数オークション入札支援システム BiddingBot," 人工知能学会誌, vol.17, no.3, pp.247-258, May. 2002.
- [7] S. Brin, L. Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine. WWW7 / Computer Networks 30(1-7): 107-117 (1998)
- [8] J.Kleinberg, Authoritative sources in a hyperlinked environment," Journal of the ACM, Vol.46, No.5, 1999