

# ネットワーク構造解析による 電子商取引システムユーザの特性分化

○河内佑美, 森本浩之, 吉井伸一郎

北海道大学大学院 情報科学研究科

**概要:** 本研究では電子商取引システム, 特にインターネットオークションに焦点を当て, ユーザのビヘイビアの特徴を抽出する. オークションユーザをノード, 商品取引を行った出品者と落札者の関係を出品者から落札者への有向リンクと考えネットワークを構成し, ネットワーク構造を解析する. 解析手法として HITS アルゴリズムを適用し, 各ユーザの評価値をユーザ間の関係性というネットワークを通じて再帰的に定量化する. 結果として出品専門ユーザのビヘイビアに特徴的なパターンが現れ, 特定のパターンを持つ落札者との取引関係によって“ブランド”と“ロイヤリティ”の関係にあるユーザが発見された.

## Differentiation of User Behavior in an E-commerce System using Network Structural Analysis

○Yuumi Kawachi, Hiroyuki Morimoto and Shinichiro Yoshii

Graduate School of Information Science and Technology, Hokkaido University

**Abstract:** This paper presents an approach to detect singularity (i.e. specific characteristics, idiosyncrasies, and unified patterns) in user behavior from network structures in Internet auctions. We can see the flow of goods from a seller to a buyer as a directed link and a user as a node, which comprises the auction network. One user behavior is related to user behaviors who conduct transactions with the users. Given this idea, our approach is that the evaluations of users are evaluated by other users' evaluation through the directed links. It is thought that the analyzing features of auction networks should lead to detecting specific characteristics in user behavior. As a result of calculation experiments using the HITS algorithm, we found the users who have relationships like brand and brand loyalty.

### 1. はじめに

インターネットや World Wide Web が普及し日常生活に浸透するにつれ, 人々はビジネスをインターネット上で行うようになってきた. ある消費者は商品, 製品またはデジタルコンテンツな

どをある業者が運営するウェブサイトを通して購入を行い、またある消費者はインターネットオークションを通して商品を売買している。このような電子商取引において市場促進のために消費者へのパーソナライゼーションやレコメンデーションが必要とされている。例えば、Amazon.com や eBay などの多くのサイトではすでにレコメンデーションシステムを採用している[1][2]。サイトを利用した消費者の購入商品、さらには消費者自身の性別、年齢など大量の顧客データを収集し解析することで消費者と商品間などの新たな関係性を導き出し、売上促進を目的とするデータマイニング手法である[3][4]。データ収集の方法としてはユーザ登録による情報、アンケート記入によるもの、過去の購入履歴や Web サイトで閲覧した商品履歴によるものなど様々である。データマイニング手法は収集した大量のデータを細分化し統計処理を行うことで人間の直感に頼らず市場の傾向をつかむものである。

一方、大規模で複雑に見えるデータについて、データを構成する要素とその要素間の関係性に焦点を当てネットワーク的に捉えることにより今まで明らかではなかった事象が現れてきた。例えば、スケールフリー構造の発見により、ネットワークの成長現象、脆弱性や頑健性を説明できるようになった[5][6][7]。またスモールワールドネットワークの概念[8][9]によって世界の膨大な人口に対する知人関係ネットワークの半径の短さが説明できる。近年、このように要素とそれらの関係つまりノードとリンクというネットワーク的視点からの研究が盛んに行われ多くの重要な知見が得られている。

そこで、本論文では電子商取引サイトにおけるユーザ間の関係性についてのネットワーク構造からユーザの特徴的なビヘイビア特性を導くことを目的とする。データマイニングのように情報を細分化するのではなく、ボトムアップ的にネットワークとして捉え構造を解析することでユーザ間の新たな関係性の発見を行い将来的にパーソナライゼーションやレコメンデーション技術の一要素としての提案を目指す。

## 2. インターネットオークションのネットワーク

電子商取引の一つであるインターネットオークションに焦点を当てる。本研究では日本最大であるオークションサイト、Yahoo! JAPAN オークション (<http://auctions.yahoo.co.jp/>) において出品者、落札者となるユーザをノード、一つの商品取引について出品者から落札者への商品の流れを一本の有向リンクとして捉えオークションのネットワークを構成する(図1)。クローリングシステムを構築し上記サイトからダウンロードした HTML データを解析することで、出品者、落札者の情報を得ることができる。以下に、取得したデータを示す。

取得ユーザ ID (ノード) 数 495,900 について、それらユーザ間での商品取引(出品者から落札者への有向リンク) 数は 748,540 であった。ユーザは行われる各オークションにおいて出品者にも落札者にもなりうる。オークションネットワークの一部をネットワーク描画ツール Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>) を用いて図2に示す。ここで、図の点は各ユーザを示し、リンクは出品者から落札者への商品の流れを示す。図の中央に位置するユーザはリンクでつながった他のユーザと商品の取引を行っており、さらにそれらのユーザはリンクでつながった他

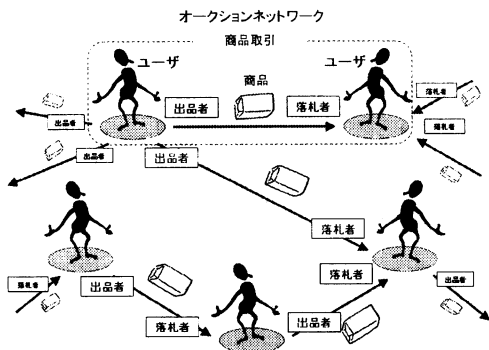


図 1. オークションネットワーク

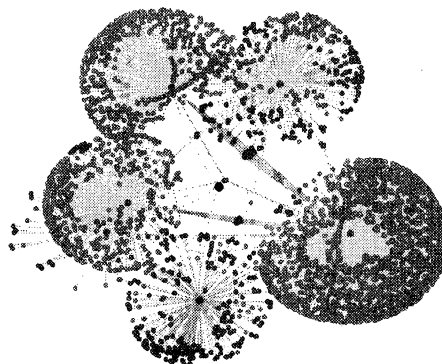


図 2. ネットワーク構造

のユーザと取引を行ったことを意味している。つまり図 2 はあるユーザを中心としたときの二次の隣接ネットワークを示したものである。

### 3. ネットワーク構造解析

ノードとリンクから構成されるネットワークの構造を解析する手法が多く提案されている。リンクの接続関係から各ノードに評価値を与える手法と、ノードをコミュニティに分割する手法に大別される。本論文では、インターネットオークションにおけるユーザの特徴抽出を行うことが目的であるので、前者の評価値を与える手法を適用する。また、特に Web サイトや Web ページとハイパーリンクで構成される World Wide Web のネットワーク構造の解析手法に焦点を当てると、次のような既存の提案手法が挙げられる。まず、代表的な検索エンジン Google の基礎技術となっている PageRank である[10][11]。多くのリンクが張られた Web ページは“よい”ページ，“よい”ページからリンクを張られたページもまた“よい”ページであるとの概念のもと、入リンクを考慮した各 Web ページの重要度ランキングを評価値として与える。PageRank はリンクを張ったページから張られたページへの有向リンクを通じて評価値が伝播し再帰的に評価を行う手法である。

また、HITS (Hyperlinked-Induced Topic Search) アルゴリズム[12]はあるトピックについて検索エンジンによって検索された Web ページ集合におけるハブとオーソリティを発見するアルゴリズムとして提案された。ハブとは多くのページへリンクを張っているページのことであり、オーソリティとは多くのページからリンクを張られているページのことである。一つのページの入リンクと出リンクを考慮し、ハブ度とオーソリティ度という二つの値を評価値として各ページに与える。評価の手法は PageRank と同様、有向リンクを通じて再帰的に評価値を伝播させるものである。本論文では、出品者、落札者の両者になりうるオークションサイトのユーザを評価したいので、入リンク、出リンクの両方を考慮したアルゴリズムである HITS アルゴリズムを適用してユーザの評価を行う。次に HITS アルゴリズムの詳細を示す。

### 3.1. HITS アルゴリズム

ネットワークの隣接行列を  $A$  とする。このとき Web ページ  $i$  から  $j$  へのハイパーリンクが存在するとき  $A_{ij} = 1$  であり、存在しないとき  $A_{ij} = 0$  である。また全ページに対し、 $a$  をオーソリティ度のベクトル、 $h$  をハブ度のベクトルとする。各ページ  $i$  は以下に示す式(1)、(2)によってハブ度とオーソリティ度が与えられる。ここで、 $i \rightarrow j$  はページ  $i$  から  $j$  へのハイパーリンクを示す。

$$a_i = \sum_{j \rightarrow i} h_j \quad (1)$$

$$h_i = \sum_{i \rightarrow j} a_j \quad (2)$$

さらに、以下にアルゴリズムの手順を示す。

$a = [1,1,1,\dots,1]$      $h = [1,1,1,\dots,1]$     (初期値)

Do {

$a = A^T h$ ;     $h = A a$ ;

$a = a / \|a\|$ ;     $h = h / \|h\|$ ;    (正規化)

} while ( $a$  と  $h$  が固定点  $a^*$  と  $h^*$  に収束していないとき)

$a^* = a$ ;     $h^* = h$ ;

Return  $a^*$ ,  $h^*$

$a^*$  と  $h^*$  は  $A^T A$  と  $AA^T$  の固有ベクトルに対応する。またハブ度とオーソリティ度はこのアルゴリズムによって  $a^*$ 、 $h^*$  の安定状態へ収束することが証明されている。ハブ度の高いページは多くのオーソリティ度の高いページへリンクを張り、またオーソリティ度の高いページは多くのハブ度の高いページからリンクを張られていることになる。

## 4. ユーザのビヘイビア特性

前述した取得ユーザ ID (ノード) と商品取引 (有向リンク) について、オークションネットワークを構成し、HITS アルゴリズムによってネットワーク構造を解析する。また HITS アルゴリズムにおいてハブ度とオーソリティ度として定義される評価値について、本論文では取引商品の流れを示すリンクの向きを照らし合わせ、ハブ度を「出品度」、オーソリティ度を「落札度」として定義する。まず、各ユーザのオークション参加回数 (出品回数または落札回数) に対する出品度と落札度の 3 次元分布を示し特徴的な傾向に焦点を当て、さらにその落札度と出品度に対するユーザビヘイビアの詳細について議論を行う。

### 4.1. ユーザのオークション参加数と評価値

HITS アルゴリズムを適用した計算実験の結果を図 3、4 に示す。図 3 のそれぞれの軸は、出品

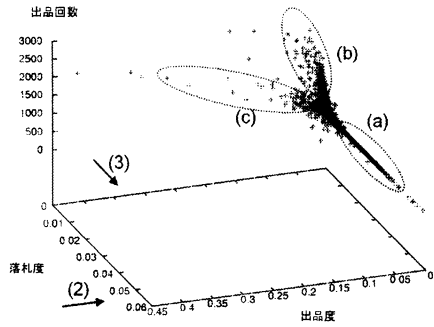


図 3. 出品度，落札度，出品回数の分布

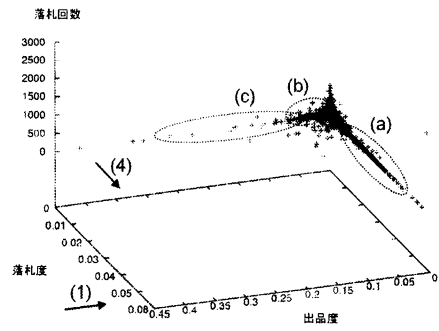


図 4. 出品度，落札度，出品回数の分布

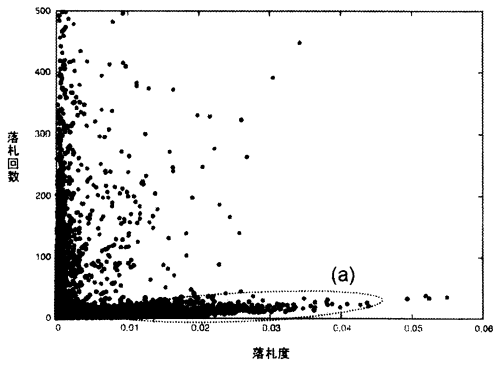


図 5. 落札度，落札回数の分布（視点(1)）

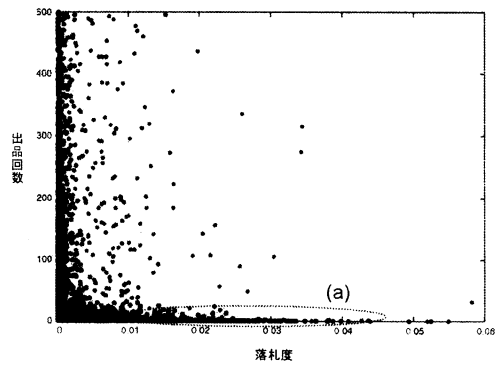


図 6. 落札度，出品回数の分布（視点(2)）

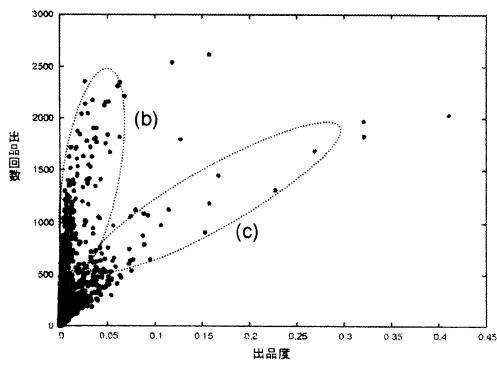


図 7. 出品度，出品回数の分布（視点(3)）

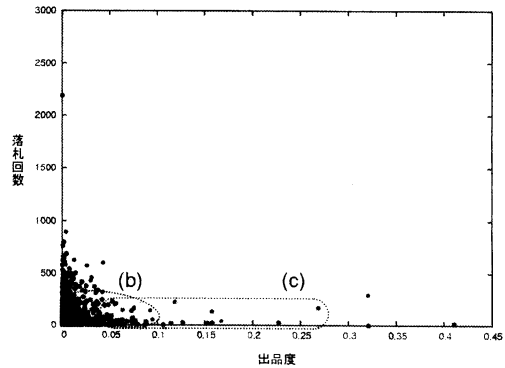


図 8. 出品度，落札回数の分布（視点(4)）

度、落札度、出品回数を示し、また図4では、出品度、落札度、落札回数を示す。図でのプロット点は一ユーザを示している。図3から分かるように、分布はおおよそ3つの方向(a), (b), (c)に分岐していると考えられる。図4でのインデックス(a), (b), (c)は図3でのインデックス(a), (b), (c)に対応している。さらに詳しく分布を調べるため、図3, 4に示す視点(1), (2), (3), (4)から2次元データとして考察を行う。

## 4.2. ユーザの落札度

図4の(1)から見たグラフを図5に示す。また図3の(2)から見たグラフを図6に示す。両図の横軸は落札度であり、縦軸はそれぞれ落札回数と出品回数を示している。図5でのグループ(a)近傍に属するユーザは他のユーザに比べ落札回数が比較的少ないにも関わらず、落札度は高くなっている。また図6に示すように(a)に属するユーザはほとんど出品をしておらず、回数が少ないながらも落札専門のユーザであることが分かる。これらのユーザの落札度が高い理由としては、高い出品度を持つユーザから落札する傾向があるといえる。つまり、出品者や出品商品を注意深く吟味し、オークションに対する警戒心を持ちながら確実に安心できるユーザから落札しているのではないかと考えられる。一方で、図5, 6での落札度が低いユーザは、落札だけではなく出品も数多く行っている傾向にある。

## 4.3. ユーザの出品度

次に、出品度について詳しく調べる。図3の(3)から見たグラフを図7、図4の(4)から見たグラフを図8に示す。両図とも横軸は出品度であり、縦軸はそれぞれ出品回数、落札回数である。図7でのグループ(b)(c)は出品回数が増加するにつれ出品度も増加する傾向にある。また、これらのグループのように多数の出品を行うユーザは図8から分かるように落札回数が少なく、出品専門のユーザであるといえる。しかしながら、グループ(b)と(c)ではたとえ出品回数がほぼ同数であったとしても、出品度は出品回数が増加するにつれ大きく異なっていくことがわかる。これらの差の原因として、どのような落札者と取引をおこなってきたかに依存すると考えられる。つまりHITS アルゴリズムを考慮すると、出品度の高いユーザは落札度の高いユーザとの関わっており、グループ(a)と(c)との関係性が強いといえる。グループ(b)の出品者は様々な落札者と取引を行っている一方で、グループ(c)の出品者はいわば、落札者グループ(a)という“ロイヤリティ”を持つ“ブランド”ユーザであるのではないかと考えられる。

オークションネットワークの構造解析によって出品専門ユーザのビヘイビアにおける特徴的なパターンと落札ユーザ間の関係性が現れてきた。特に、多くの出品を行っているユーザには比較的少数の落札を専門に行うユーザとの取引によって“ロイヤリティ”と“ブランド”の関係にあるユーザと、そうではない出品専門ユーザに分化することが分かった。これらのパターンは各ユーザの取引履歴データを詳細に解析するだけでは現れてはこない特徴であるといえる。

## 5. まとめ

本論文では、インターネットオークションにおける出品者と落札者の関係をネットワークとして構成し、その構造を解析することによってユーザのビヘイビアにおける特徴的なパターンを導き出すアプローチを示した。各ユーザの出品、落札に関する評価値をネットワーク構造でのリンクとその方向を通して再帰的に計算することで、直接取引を行っていないユーザに関する評価値情報がそれぞれのユーザに蓄積される。データマイニングにおいてはユーザが直接取引を行ったデータに関して統計処理を行い傾向やパターンを発見する手法である。一方本研究では、あるユーザのビヘイビアの評価値には、直接取引を行っていないユーザに関しても、直接取引を行ったユーザとの関係性を通じてそれらユーザのビヘイビアが反映されていることになる。

本論文では出品、落札の関係性にも焦点を当てたが、今後の課題としてはどのような種類の商品を取引したかというラベルをネットワークのリンクに付加することによって、ユーザの嗜好を加味したビヘイビアの特徴抽出が行えるのではないかと考えている。さらには、コミュニティ分析[13]、信頼度推定[14]などネットワーク解析を総合的に行うことで、実システムへの適応を目指す。

### 【参考文献】

- [1] J. B. Schafer, J. A. Konstan, and J. Riedi. Recommender systems in e-commerce. In *ACM Conference on Electronic Commerce*, pages 158-166, 1999.
- [2] J. B. Schafer, J. A. Konstan, and J. Riedl. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5(1/2):115-153, 2001.
- [3] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Application of dimensionality reduction in recommender systems - a case study, 2000.
- [4] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Analysis of recommendation algorithms for e-commerce. In *ACM Conference on Electronic Commerce*, pages 158-167, 2000.
- [5] A.-L. Barabási. *LINKED: The New Science of Networks*. Perseus Publishing, Cambridge, 2002.
- [6] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509-512, 1999.
- [7] H. J. Réka Albert and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378-382, 2000.
- [8] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440-442, 1998.
- [9] D. J. Watts. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, 1999.
- [10] T. Haveliwala. Efficient computation of pagerank. Technical Report 1999-31, 1999.

- [11] R. M. Lawrence Page, Sergey Brin and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.
- [12] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604-632, 1999.
- [13] J. Ando and S. Yoshii. Analyzing weblog space with topological structure. In *Japan-Australia Workshop on Intelligent and Evolutionary Systems*, 2005.
- [14] H. Morimoto and S. Yoshii. Estimation of information trustability on the internet. In *Information and Technology Letters*, pages 259-260, 2005.