

商品市場分析の新手法 — べき乗分布とネットワークによる分析 —

井 庭 崇[†] 吉田 真理子[†] 伊 藤 諭 志^{††}

本論文では、商品市場における販売量-順位分布がべき乗分布になるという特徴を考慮した、新しい市場分析の手法を提案する。提案する手法は、「販売量-順位分布にもとづく顧客クラスタリング手法」と「選択ネットワークによる商品間関係の分析手法」の2つである。前者が「商品を購入した顧客同士の関係性」に注目しているのに対し、後者は「購入された商品同士の関係性」に注目する。どちらの手法も、ニューラルネットワークの考え方に関係しているという特徴がある。

New Methods for Analyzing Commodities Market based on Power-Law Distribution and Complex Network

TAKASHI IBA,[†] MARIKO YOSHIDA[†] and SATOSHI ITOH^{††}

In this paper, we propose new methods for analyzing the market in which the relation between sale and rank follows power law. The proposed methods consist of (1) the method for clustering customers based on the relation between sale and rank, and (2) the method for analyzing the relation among commodities with customers' choice network. The former focuses on the relation among the customers, and the latter does on the relation among the commodities. Both methods are related to the concept of neural networks.

1. はじめに

先行研究^{1),2)}において私たちは、日本全国における書籍販売の実データを解析し、月間でみても年間でみても販売量-順位の分布が「べき乗分布」であることを明らかにした。同様に、商品の販売量-順位の分布がべき乗分布（ロングテール）になるということは、他の種類の商品においても発見されている³⁾。

本論文では、このような商品市場の特徴を考慮した新しい市場分析の手法を提案する。提案するのは、「販売量-順位分布にもとづく顧客クラスタリング手法」と「選択ネットワークによる商品間関係の分析手法」の2つである。これらの手法は、ともにeコマースの普及によって把握できるようになった購買履歴データを活用する点で共通しているが、データを分析する際の焦点が異なっている。前者が「商品を購入した顧客同士の関係性」に注目しているのに対し、後者は「購入

された商品同士の関係性」に注目する。また、前者は販売量-順位分布の詳細な情報を捨象することによって浮かび上がってくるのに対し、後者は販売量-順位分布からさらに詳細な情報に立ち入ることによって新しいリアリティが獲得できる。

このように、これらの手法は方向性が異なるにもかかわらず、本論文で示すように、実はどちらもニューラルネットワークの考え方に大きく関係している。以下では、これらの手法について順に説明していく。

2. 販売量-順位分布にもとづく顧客クラスタリング手法の提案

まず最初に、販売量-順位分布にもとづく顧客クラスタリング手法について提案する。この手法では、それぞれの顧客が選択（購入）した商品が販売数-順位分布のどの位置にあるのかという傾向をもとに分類するクラスタリング手法である。販売数-順位分布上で区間分けを行い、顧客がどの区間の商品をどれだけの数選択したかを把握し、その情報を用いて分類する。選択した商品の関係性の情報のみを用いて分類することで、顧客の年齢、性別、職業などの属性を用いない顧客クラスタリングが可能となる。

[†] 慶應義塾大学 総合政策学部
Faculty of Policy Management, Keio University
^{††} 慶應義塾大学 政策・メディア研究科
Graduate School of Media and Governance, Keio University

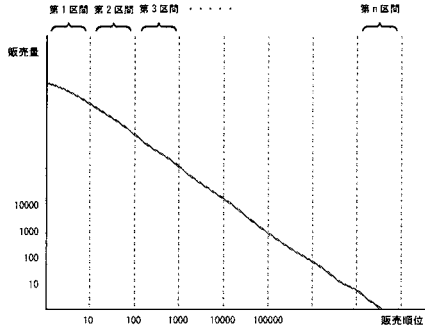


図1 販売量-順位分布における区間分け

2.1 販売量-順位分布における区間分け

具体的に順を追って見ていくと、まず y 軸を販売量、 x 軸をその順位とした両対数グラフ上で、 n 区間の区間分けを行う (図1)。ここで、区間数は、目的に応じて必要となる精度を考慮して決定する。次に、対象となる顧客が、各区間の商品をどれだけの数選択しているかの割合を算出する。そこで得られた割合の集合が、この顧客の特徴を集約したデータとなる。

例えば、ある顧客が第1区間で5つ、第2区間で3つ、第3区間で2つ商品を選択し、それ以外の区間では選択しなかったとすると、(0.5, 0.3, 0.2, 0.0, 0.0, 0.0, 0.0) という集合が得られる (図2上)。同様に、第1区間で2つ、第2区間で3つ、第3区間から第7区間でそれぞれ1つずつ商品を選択した場合には、(0.2, 0.3, 0.1, 0.1, 0.1, 0.1, 0.0) という集合が得られる (図2下)。このような集約データを用いて、顧客のクラスタリングを行う。クラスタリングは商品のジャンル別に行う場合もあれば、複数のジャンルを統合したかたちで行うこともできる。

2.2 自己組織化マップによるクラスタリング

上述の方法で得られた集約データのクラスタリングには、Teuvo Kohonen⁴⁾ によって提案された「自己組織化マップ」(Self-Organizing Map: SOM) を用いる (図3)。自己組織化マップとは、多次元データを2次元平面に写像する技術であり、教師信号を必要としないニューラルネットワークの一種と捉えられる。クラスタリングの結果、類似したものが互いに近くなり、類似していないものが遠くなるようにマップ上に配置される。また、出現頻度の高いものは低いものよりも、マップ上で面積が広くなるように表現される。

このように、自己組織化マップは入力パターンから規則性を抽出して、その特徴をわかりやすく表現する

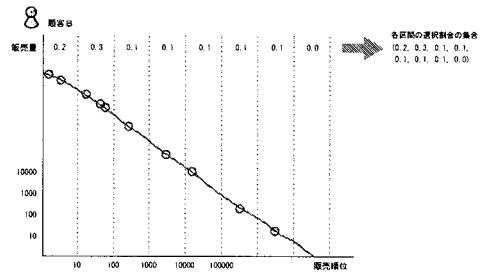
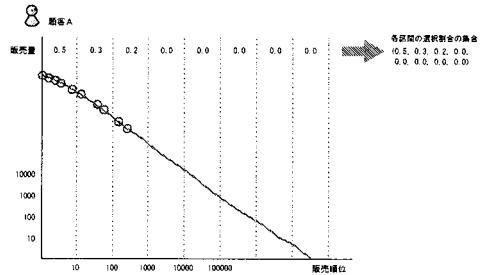


図2 販売量-順位分布における傾向の把握

ことができる。Kohonen⁴⁾ は、次元の大きなデータであっても、その特徴を浮かび上がらせるには2次元で充分であることを示している⁵⁾。また、自己組織化マップでは、大規模なデータであっても短時間でクラスタリングできることが知られている。これらのことから、非常に多くの顧客を対象とした市場分析の場合においても、クラスタリングおよびセグメンテーションのために有効な手法だといえる。

3. 選択ネットワークによる商品間関係の分析手法の提案

次に、選択ネットワークによる商品間関係の分析では、同一顧客によって選択 (購入) された商品間にリンクを張ることで、商品間の関係をネットワークとして可視化・解析する。このように、選択されたもの同士を結ぶネットワークを、本論文では「選択ネットワーク」(choice network) と呼ぶことにしたい⁶⁾。商品の選択ネットワークを可視化・解析することで、市場全体における商品の見取り図を得ることができる。

⁵⁾ ただし、2次元に写像するとき、Kohonen⁴⁾ が提案したような長方形のマップに写像すると、マップの端の部分で歪みが生じてしまうという問題も指摘されている⁵⁾。そこで、トラス型や球面型の自己組織化マップが提案されており⁵⁾、本手法においてもトラス型などを用いることが望ましい。

⁶⁾ 本論文の「選択ネットワーク」を、商品間関係の分析ではない別の事例に適用したものに、論文⁶⁾がある

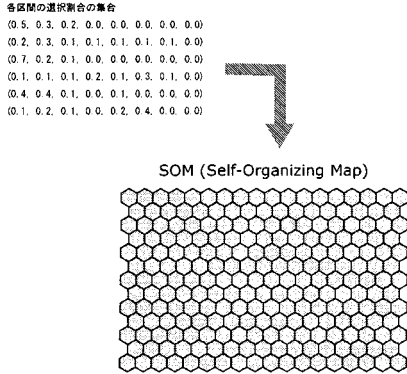


図 3 自己組織化マップによるクラスタリング

3.1 商品の選択ネットワークの構成

商品の選択ネットワークにおいては、各商品がノードになる。ネットワークは、以下の手順で作成する。ある2つの商品に着目したとき、同一顧客によってその2つの商品が選択されている場合には、その2つの商品のノード間にリンクを張る(図4上)。ある顧客が3つの商品を選択した場合には、それらの3つの商品に、そのすべての組み合わせの分だけリンクが張られる(図4下)。リンクを張る際には、何人がその組み合わせを選択したのかに応じて重みをつけ、太く表示する。

なお、商品をたった一つしか買わない顧客の選択は、この選択ネットワーク上には表現されない。しかし、通常はそのような顧客は稀であると考えられるし、ここで分析したいのは商品同士の関係性であるため、そのような例外は取り上げなくても問題はない。

ここで、販売量-順位分布との関係を確認しておきたい。販売量の多い商品、つまり販売量-順位分布のヘッド部分にある商品は、原理上、商品の選択ネットワークにおいてはハブとなる。ここでいうハブは、厳密にいうならば、単なる所有リンク数にもとづくものではなく、重みを考慮して計算されたものである。このようなネットワークの解析を行うことで、スケールフリーネットワークとべき乗分布の関係性⁷⁾の知見を活用して、販売量-順位分布がべき乗分布となる原理を解明できる可能性がある⁸⁾。

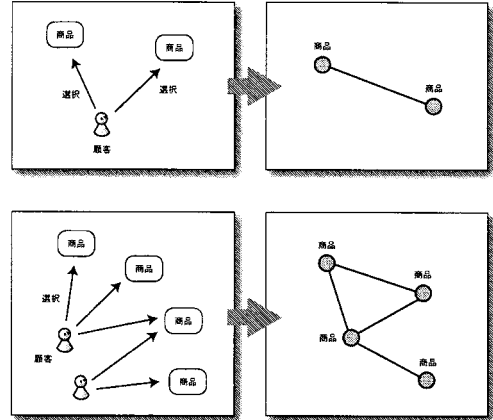


図 4 選択ネットワークの形成原理

3.2 商品の選択ネットワークの解析

商品間の関係性をネットワークとして捉えることで、社会ネットワーク分析 (Social Network Analysis) における手法⁹⁾¹⁰⁾やネットワーク科学 (Network Science) における手法¹¹⁾を活用することが可能になる。ここでは、「クラスター分析」、「ノード間の距離による分析」、「コミュニティ抽出」の3つのアプローチを紹介する。

クラスター分析

商品の選択ネットワークにおいては、顧客が3つ以上の商品を選択している場合には、原理上、クラスターが形成される。このような仕組みで複数の顧客の選択を描いていくと、多くの顧客が同じような選択をしている場合には、重みの値が大きい強固なクラスターが形成される。このように、ノード間のリンクにおいて、ある程度の重み以上のクラスターに注目することで、顧客の選択において頻度の高い組み合わせを理解することができる。

ノード間の距離の分析

商品の選択ネットワークにおいては、任意の二つの商品同士は、ネットワーク上での距離をもっている。このようにノード間の距離に注目することで、商品同士の近さ/遠さを把握できる。例えば、直接リンクで結ばれているノードAとノードBは、顧客が実際に選択した商品同士の組み合わせであることを示している。一方、ノードBを介した2本のリンクでつながっているノードAとノードCは、この組み合わせで選択されたことはないが、どちらもノードBと関係している(Bと組み合わせで選択されている)ことがわかる。このように、ノード間の距離から商品間関係性

* 厳密にいうと、商品の選択ネットワークは、「顧客」と「商品」をそれぞれノードとした二部グラフから算出される。つまり、ある顧客が商品を購入した場合に、購入した「顧客」と購入された「商品」の間にリンクを張る。このような顧客-商品の二部グラフから、商品-商品間のネットワークが算出できる。

とその近さ／遠さを把握することができる。この距離は、リンクの重みを考慮した上で計算することになる。

コミュニティ抽出

商品の選択ネットワークに対して「コミュニティ抽出」の手法を適用すると、組み合わせて選択された商品群が抽出できる。このように捉えられた商品群は、商品に関する一種の「フォークソノミー」だといえる。つまり、商品の供給側によって規定された分類ではなく、需要側からみた新たな商品分類である。

3.3 メタファーとしてのニューラルネットワーク

商品の選択ネットワークは、その形成原理を考えると、実はニューラルネットワークと同型であることがわかる。ニューラルネットワークでは、2つのニューロンが同時（一定時間内）に発火した場合に、その間のシナプス結合の重みが増すという仕組みになっている（ヘップ則¹²⁾）。商品の選択ネットワークでは、2つの商品が同一顧客によって選択された場合に、その間のリンクの重みが増すという仕組みになっており、ヘップ則と同じ仕組みであることがわかる。

この同型性から、生物の脳や人工ニューラルネットワークの研究で得られている知見を活用することができるようになる。例えば、ニューラルネットワークが重みの集合によって「記憶」を保持するように、商品の選択ネットワークにおいても、商品選択の重みの集合で、いわば「市場の記憶」を保持していると捉えることができる。また、ニューラルネットワークがニューロン間の相互作用によって思考するように、商品の選択ネットワークの商品間の相互作用によって、市場の思考のようなものが実現できるかもしれない。

このように、商品の選択ネットワークの特性や可能性を考える上で、ニューラルネットワークに関する知見を参考にできるという利点は大きい。今後は、ネットワーク科学の分野ではこれまであまり活発に研究されてこなかった「重みつきネットワーク」の解析手法の開発や、そのダイナミクスの理解についても研究を進めていくことが期待される。

4. おわりに

本論文では、商品市場における販売量-順位分布がベキ乗分布になるという特徴を考慮した新しい市場分析の手法を提案してきた。今後は、これらの手法を用いた具体的なデータ解析に取り組んでいくとともに、重みつきネットワークの解析手法の開発についても取り組んでいきたい。

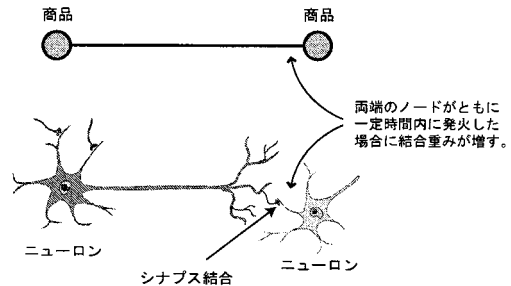


図5 選択ネットワークとニューラルネットワークの同型性

謝 辞

本研究の構想段階でアドバイスをいただいた慶應義塾大学総合政策学部の桑原武夫先生に感謝したい。

参 考 文 献

- 1) 井庭崇, 深見嘉明, 斉藤優: 書籍販売市場における隠れた法則性, 情報処理学会論文誌: 数理モデル化と応用, Vol. 17 (2007).
- 2) 井庭崇, 深見嘉明, 吉田真理子, 山下耕平, 斉藤優: 書籍販売市場における上位タイトルの売上分析, 情報処理学会第64回数理モデル化と問題解決研究会 (2007).
- 3) クリス・アンダーソン: ロングテール: 「売れない商品」を宝の山に変える新戦略, 早川書房 (2006).
- 4) T. コホネン: 自己組織化マップ (改訂版), シュプリンガー・フェアラーク東京 (2005).
- 5) 徳高平蔵, 大北正昭, 藤村喜久郎 (編): 自己組織化マップとその応用, シュプリンガー・ジャパン (2007).
- 6) 井庭崇, 湯村洋平, 若松孝次, 古市奏文: プロジェクト推進のパターン・ランゲージとその評価, 日本ソフトウェア科学会ネットワークが創発する知能研究会 & 情報処理学会 数理モデル化と問題解決研究会合同ワークショップ (2007).
- 7) アルバート=ラズロ=バラバシ: 新ネットワーク思考: 世界のしくみを読み解く, NHK 出版 (2002).
- 8) 伊藤諭志, 井庭崇: 商品ネットワークの成長モデル: 市場の秩序形成の探究に向けて, 情報処理学会第64回数理モデル化と問題解決研究会 (2007).
- 9) 安田雪: 実践ネットワーク分析: 関係を解く理論と技法, 新曜社 (2001).
- 10) 金光淳: 社会ネットワーク分析の基礎: 社会的関係資本論にむけて, 勁草書房 (2003).
- 11) 増田直紀, 今野紀雄: 複雑ネットワークの科学, 産業図書 (2005).
- 12) Hebb, D.: The organization of behavior, *Neurocomputing* (Anderson, J. and Rosefeld, E.(eds.)), MIT Press, chapter 4 (1949).