

# ECサイトにおける顧客の閲覧履歴を利用した商品ランキング生成法

武政 孝師, 後藤 順哉

## 1. はじめに

### 1.1 背景

インターネットを基盤としたさまざまなサービスがビジネスの主役となった現在、ビジネスにおけるビッグデータの有効活用が注目されている。例えば、幅広い商品を販売する EC (electronic commerce) サイトであれば、その顧客が当該サイト内をどのように閲覧し、最終的にどの商品を購入するに至ったか（あるいは至らなかったか）という、インターネットを介在しない従前のビジネスでは観察するのが困難だったデータが容易に得られる。一方、その膨大さ、無定形さゆえにどう集約・分析し活用していくのかは大きな課題となっている。

本論文では、EC サイトにおける顧客の商品閲覧履歴から商品のお勧めランキングを生成する方法を提案する。経営科学系研究部会連合協議会が主催する平成 24 年度データ解析コンペティションにおいて、株式会社リクルート・グループより提供された、同社の EC サイトであるポンパレに関する 2011 年 7 月 1 日から 2012 年 6 月 30 日までの情報を用いている。与えられた情報の規模概要を表 1 に示す。

表 1 提供データの概要

会員数	22,873
商品数	19,723
購買件数	171,599
閲覧履歴数	2,889,404

たけまさ あつし  
株式会社日立公共システム  
〒135-8633 東京都江東区東陽 2-4-18  
ごとう じゅんや  
中央大学理工学部経営システム工学科  
〒112-8551 東京都文京区春日 1-13-27  
受付 13.7.23 採択 14.5.1

### 1.2 動機と目的

多数の商品を扱う EC サイトでは、登録した会員に対して商品情報を e メール (メルマガ; メール・マガジン) にて送付していることが少なくない。例えば、さまざまなサービスのクーポンを扱う EC サイトであるポンパレでは、登録会員に対して毎日メルマガが配信されている。メルマガには値引きされている商品や新しく発売された商品のほか、売れ筋ランキングなど、さまざまな情報が掲載される。中でも、商品を売上高によって順位づけし、売れ筋商品を知らせてくれるメルマガは顧客にとって興味深い情報源である。一方で、当該ランキングには、“連日ほぼ同じ商品がメルマガに掲載される”、“メルマガに掲載されている商品が、すでに売り切れている”といった特徴が観察された。顧客に対して商品の情報を提供し、購買を促すことが目的であるメルマガにとって、同じ商品を何度も薦めることは好ましくない。特に、過去の販売実績を単純に追従するだけのランキングでは、顧客の閲覧行動から読み取ることのできる、潜在的な需要に対する情報を見逃し、ひいては顧客に対して有益な (意外性・説得力・多様性のある) “気づき” を提供することが難しい。また、売れている商品の購買を促すということは、ともすると売り切れている商品を薦めることにもつながるため、機会損失を生じさせやすくしているとも言える。このような点から顧客、企業の双方にとって、売上を基準に生成したランキングが必ずしも好ましいとは限らない。そこで、本研究では、上記の問題を解消する商品ランキング生成法を提案する。提案する手法は、Colley[1] が大学アメフトチームのランキングを題材として提案した手法 (Colley 法) をもとに、閲覧の少ない商品に対する評価にペナルティを加えるという修正を施したものである。

### 1.3 提案手法の特徴

ランキング生成において、重要な点の 1 つは何を基準とするかである。本コンペティションにおいて提供

された情報には“顧客情報”，“商品情報”，“購買履歴”，“閲覧履歴”などがある．中でも“閲覧履歴”はサンプル数が大きく，かつ各サンプルにはある顧客がいつ，どの商品をどのような順番で閲覧し，最終的にどの商品を購入したか（または閲覧のみで，購買には至らなかった）という情報が含まれている．閲覧履歴から，顧客が閲覧した中で最終的に購入された商品と購入されなかった商品間の選好関係を読み取ることができる．

そこで1つの閲覧履歴（1セッションと呼ぶ）において商品が“選ばれた”か“閲覧されたが選ばれなかった”かという部分的な情報を，すべての顧客について集計することでグラフを作成し，このグラフから商品の選ばれやすさを定量化し，多くの顧客に選ばれている商品ランキングの生成法を提案する．

顧客の閲覧行動，購買行動は日々異なるため，提案手法により毎日異なるグラフ，ひいては異なるランキングが生成され，顧客により多くの商品情報を提供することができると考えられる．また，提案手法では原則，選ばれやすさを観測された比較に基づく頻度により評価するため，購入された数が少ない商品でも高い評価を得られる可能性がある．このことは，連日同じ商品がランキング上位になる可能性を低め，より多くの商品を顧客に提示する機会の拡大につながると考えられる．

実際，提供された情報をもとに提案手法によりランキングを生成したところ，毎回異なるランキングが得られやすくなり，情報量を増幅させる効果があることがわかった．さらに，提供された情報をもとに算出した売れ筋ランキングに比べて，機会損失が生じる可能性を低める効果も期待できるという結果が得られた．

#### 1.4 関連研究

順位づける対象を点，対象間の一対比較関係を枝として，グラフを生成し，それをもとにランキングを生成する手法がさまざまな状況で提案されている．

その一例として，論文の評価 [2] が挙げられる．[2]では，多くの学生が卒業論文を発表する中で，教員が一部の学生の発表を評価する状況をグラフで表現している．さらに，生成されたグラフに対して，公理的アプローチ [3] を拡張したランキング手法を提案している．また，映画やレストランに対する口コミによる評価 [4]，検索エンジンにおける Web ページの順位づけ [5] もグラフを介したランキング手法と言える．

本論文で提案する手法は，閲覧履歴からグラフを作成し，そのグラフをもとに商品ランキングを生成する．

閲覧履歴に関しては，協調フィルタリング [6] やキーワードの抽出 [7] といった嗜好抽出や情報推薦 [8] のための研究が行われている．これらの研究は，閲覧履歴から顧客にとって有益な情報を抽出し，提供することを試みている点で本論文と類似している．

EC サイトにおける顧客の閲覧履歴を，その顧客の選好に基づく商品間の比較の結果と解釈し，商品の選ばれやすさを評価するためのグラフを生成するアイデアが本論文における貢献の1つである．以下，2節では Colley 法を紹介し，3節で改善点を説明する．4節では，売れ筋ランキングとの比較を通して目的に沿った手法であるかを検証し，5節では本研究の成果をまとめる．

## 2. 比較グラフと Colley 法

### 2.1 閲覧履歴から比較グラフへ

本節では，閲覧履歴から接続行列  $X$  を生成する手順を提示する．接続行列  $X$  から生成される有向グラフを比較グラフと呼び，商品に対する顧客の選好を視覚的に表現する．

接続行列  $X$  の生成に用いる閲覧履歴には，閲覧のみのセッションや購買のみのセッションは含めず，1回以上購買のあるセッションを用いる．購買のあるセッションに対して，接続行列  $X$  を生成する手順を例示する．ここで，以下のような商品 A, B, C, D, E に関する閲覧履歴があるとしよう．これらの閲覧履歴は，項番 1. から 5. ごとにセッションが異なり，矢印で示す順に顧客が商品閲覧したことを意味している．また括弧によって，ある顧客が括弧内の商品を購入したことを表すものとする．

1.  $A \rightarrow (B)$
2.  $A \rightarrow B \rightarrow (C)$
3.  $D \rightarrow A \rightarrow B \rightarrow (C)$
4.  $A \rightarrow B \rightarrow (A)$
5.  $C \rightarrow D \rightarrow (E) \rightarrow B \rightarrow E \rightarrow (A) \rightarrow D$

項番 1. は，ある顧客が商品 A を閲覧した後，商品 B を閲覧し，購入したことを意味している．この事実を“商品 B は商品 A と比較されて選ばれた”と捉え，下記に示す接続行列  $X$  の 1 列目のように，比較された商品 A を  $-1$ ，選ばれた商品 B を  $1$  とする．

項番 2. のように商品 C を購入する前に商品 A, B を閲覧していた場合，接続行列  $X$  の 2, 3 列目のように値を設定する．しかし，顧客が常に商品を比較し続けているとは限らないので，項番 3. のように 4 商品を

閲覧している場合、購買前の直近の2商品A, Bについて接続行列  $\mathbf{X}$  の4, 5列目のように値を設定する。

また、項番4.のように購買前の直近の2商品に購買した商品が含まれる場合は、接続行列  $\mathbf{X}$  の6列目のように比較された商品Bのみを考慮する。

さらに、項番5.のように1度購買した後に閲覧または購買された商品の閲覧履歴(B → E → (A) → D)は接続行列  $\mathbf{X}$  に含まない。これは、異なる2つの商品を連続して購買した場合など比較関係を想定しづらい場合が存在するためである。初回の購買までの履歴を考慮して7, 8列目のように値を設定する。

$$\mathbf{X} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} & \begin{pmatrix} -1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{matrix}$$

例示したように接続行列  $\mathbf{X}$  の生成には購買した履歴とその直前の履歴を用いる。提供された閲覧履歴2,889,404件のうち、生成に用いた履歴は108,905件であった。この接続行列  $\mathbf{X}$  より、図1に示すような点数  $n = 5$ 、枝数  $m = 8$  のグラフ  $(V, E)$  を得る。ここで、 $V$  はグラフの点集合、 $E$  は枝集合を表す。また、点数  $n$  と枝数  $m$  はそれぞれ接続行列  $\mathbf{X}$  の行数と列数に対応している。

このようにして生成されたグラフ  $(V, E)$  を比較グラフと呼ぶ。また、以下のように記号を定義する。

- $w_i$  : 点  $i$  の入次数 ( $i \in V$ )。
- $l_i$  : 点  $i$  の出次数 ( $i \in V$ )。
- $m_i$  : 点  $i$  の次数 ( $i \in V$ )、i.e.,  $m_i := w_i + l_i$ 。

ここで、 $m = \frac{1}{2} \sum_{i \in V} m_i$  の関係が成り立つことに注意しておこう。

比較グラフにおいて、点は商品、枝はある顧客の選好を意味している。つまり、入次数の大きい点ほど顧客に選ばれている商品であり、出次数の大きい点ほど閲覧されたが選ばれなかった商品と言える。

また、接続行列  $\mathbf{X}$  において行方向に1と-1の

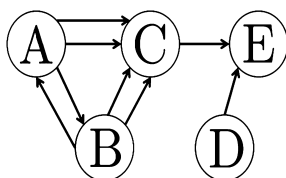


図1 接続行列  $\mathbf{X}$  から生成された比較グラフ

個数をそれぞれ数え上げることで、入次数ベクトル  $\mathbf{w} := (w_1, w_2, \dots, w_n)^T$  と出次数ベクトル  $\mathbf{l} := (l_1, l_2, \dots, l_n)^T$  が得られる。

## 2.2 Colley法の概要

次に、提案手法の基礎となるColley法[1]について、その含意と算出手法について概説する。ここに紹介する手法はColley[1]が大学アメフトチームのランキングを題材として提案したものである。

スポーツのランキングに用いられる一般的な指標として、勝率が挙げられる。いま、比較グラフにおける点をチーム、枝を2チーム間の対戦結果だとしよう。すると、次数  $m_i$  はチーム  $i$  の試合数、入次数  $w_i$ 、出次数  $l_i$  はそれぞれ勝ち試合数、負け試合数に対応する。さらに、 $r_i$  を勝率  $w_i/m_i$  によって定めるとしよう。 $i$  が1試合した結果、負けたとすると  $r_i = w_i/m_i = 0/1 = 0$  となる。このとき、 $r_i$  を将来の対戦における勝つ見込みに当てはめる場合には、勝率  $r_i = 0$  は  $i$  が勝つ可能性が100%ないことを意味し、 $i$  を評価するための指標として非現実的な解釈を与える。

そこで、勝率  $w_i/m_i$  を(1)式のように修正する。

$$r_i = \frac{1 + w_i}{2 + m_i}. \quad (1)$$

(1)式のように修正することで、0勝1敗のチーム  $i$  の評価値  $r_i$  は  $1/3$  となり、確率として勝ちも負けもありうる状態であることを表現できている。また、(1)式は試合のない評価対象に対して、 $1/2$  という評価を与えている。これは評価対象が、勝率として  $1/2$  という期待値(事前確率)を持っていることを意味する。

次に、 $w_i$  に注目し、対戦相手の評価値を考慮して、評価を決定することを考えよう。一対比較が部分的にしか不可能な状況において、実際の対戦結果に基づくだけでは有益な情報が得られない可能性が高い。そこで、評価の高い相手  $j$  と対戦して勝てば、評価対象  $i$  は対戦相手  $j$  の評価に応じて高く評価されるべきと考え、(1)式の  $w_i$  を対戦相手の評価値を考慮して決定することを考えよう。

上記のアイデアを評価値  $r_i$  に反映するために  $w_i$  に関する以下の恒等式に注目しよう：

$$w_i = \frac{w_i - l_i}{2} + \frac{m_i}{2} = \underbrace{\frac{w_i - l_i}{2}}_{(a)} + \underbrace{\sum_{j=1}^{m_i} \frac{1}{2}}_{(b)}. \quad (2)$$

(2)式は、勝ち試合数を、試合結果(a)と対戦相手の勝率の期待値(b)の和とみなしていると解釈する。こ

ここで、第2項 (b) に注目して“対戦相手の勝率の期待値  $1/2$ ”を“対戦相手の評価値”に置き換えることを考えよう。すなわち、左辺の勝ち試合数  $w_i$  の解釈を、実際の勝敗数  $w_i, l_i$  と相手の評価値によって

$$\hat{w}_i := \frac{w_i - l_i}{2} + \sum_{j=1}^{m_i} r_{ij},$$

とする。ここで、 $r_{ij}$  は  $i$  が  $j$  番目に対戦した相手の評価値とする。 $\hat{w}_i$  を (1) 式の勝ち試合数  $w_i$  に代入し、対戦相手の評価値を考慮した評価値  $r_i$  の関係式

$$r_i = \frac{1 + \frac{w_i - l_i}{2} + \sum_{j=1}^{m_i} r_{ij}}{2 + m_i}, \quad i \in V, \quad (3)$$

を得る。これを整理すると、

$$(2 + m_i)r_i - \sum_{j=1}^{m_i} r_{ij} = 1 + \frac{w_i - l_i}{2}, \quad i \in V,$$

となる。

ここで、 $i$  と  $k \in V \setminus \{i\}$  の対戦回数  $u_{ik}$  に対し、 $\sum_{j=1}^{m_i} r_{ij} = \sum_{k \in V \setminus \{i\}} u_{ik} r_k$  が成り立つことに注意すると、 $r_i$  の値は次の連立方程式の解として得られる。

$$(2 + m_i)r_i - \sum_{k \in V \setminus \{i\}} u_{ik} r_k = 1 + \frac{w_i - l_i}{2}, \quad i \in V. \quad (4)$$

これより、評価値  $\mathbf{r} := (r_1, r_2, \dots, r_n)^T$  を求めるには連立1次方程式を解けばよいことがわかる。

一方、対戦関係から導かれる接続行列  $\mathbf{X}$  に対し、 $\mathbf{X}\mathbf{X}^T$  の第  $(i, k)$  成分は  $-u_{ik}$ 、第  $(i, i)$  成分は  $m_i$  となる。この観察より、(4) 式は  $\mathbf{C} := \mathbf{X}\mathbf{X}^T + 2\mathbf{I}$  で定義される行列と、 $\mathbf{b}_i := 1 + \frac{w_i - l_i}{2}$  を第  $i$  要素とするベクトル  $\mathbf{b}$  に対して定義される連立方程式

$$\mathbf{C}\mathbf{r} = \mathbf{b} \quad (5)$$

として書けることが示される。ただし、 $\mathbf{I}$  は単位行列である。

なお、行列  $\mathbf{C}$  (Colley 行列と呼ばれる) は、定義から直ちに実対称正定値であることがわかる。したがって、 $\mathbf{C}$  は正則であり、 $\mathbf{r}$  は常に一意に求まる。

上記で示したように Colley 法では、評価値の算出に対戦相手の評価を考慮しているため、一対比較が部分的に得られる状況に対して適用することができる。

そのような状況の1つである比較グラフにおいて、Colley 法を適用することを考えよう。スポーツにお

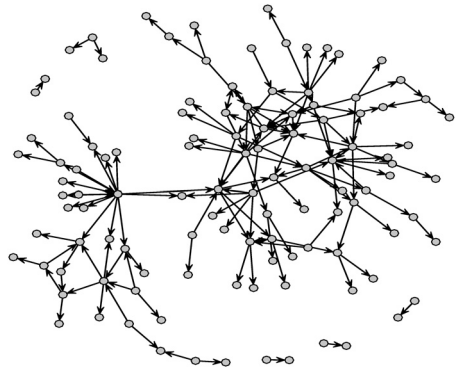


図2 ある1日分の閲覧履歴から生成した比較グラフ<sup>1</sup>  
(点数: 113; 枝数: 159)

る勝敗をECサイトにおける商品間の選好関係に置き換え、Colley 法により算出される評価値  $r_i$  は、商品  $i$  の“選ばれやすさ”を表している。(5) 式を解いて評価値  $\mathbf{r}$  を求め、これを降順に並べ替えることによって商品ランキングが得られる。

### 3. 評価値算出法の改良

Colley 法の特筆すべき特徴として、評価対象全部の総当たりの比較値を前提とせず、部分的な一対比較に基づき生成した有向グラフから評価値を生成している点が挙げられる。

比較されている回数が少ない対象 (次数が小さい点) にも順位をつけている点は、リーグ戦のように、どの評価対象 (チーム) も一様に一対比較 (試合) を経る状況下では問題になりにくい。しかし、図2に示すように商品によって比較された回数に著しくばらつきのある、本論文で扱うような場合には、優先して修正すべき事項と考えられる。

そこで、閲覧者数の少ない商品に対してペナルティを加えることで、多くの顧客に閲覧されている商品の評価を相対的に高める方法 (Colley+法) を提案する。

商品  $i$  の閲覧者数を  $v_i$  で表し、当該商品が比較されたセッション数で定義する。2節の閲覧履歴の例にあてはめると、接続行列  $\mathbf{X}$  を生成するために用いた閲覧履歴において商品が閲覧されたセッション数が閲覧者数であり、表2に示す値となる。

Colley 法より求まるランキングを表3に示す。閲覧者数が1である商品Eが1位となっているが、これは1人の顧客に選ばれた商品が1位となっていることを

<sup>1</sup> データ中に並行枝は存在するが、描写が技術的に困難なため省略している。

表 2 閲覧者数の算出

商品	閲覧履歴の項番	閲覧者数
A	1., 2., 3., 4.	4
B	1., 2., 3., 4.	4
C	2., 3., 5.	3
D	5.	1
E	5.	1

表 3 Colley 法によるランキング結果

順位	1	2	3	4	4
商品	E	C	D	A	B
評価値	0.769	0.654	0.423	0.327	0.327
閲覧者数	1	3	1	4	4

意味する。このように、商品ランキングとして Colley 法を採用した場合、閲覧者数の少ない商品が上位を独占し、“多くの顧客に選ばれている”とは言いがたいランキングが生成される。

提案手法である Colley+法は、(3)式に閲覧者数の関数である  $f(v_i)$  を付け加えた(6)式によって評価値  $r_i$  を算出する。

$$r_i = \frac{1 + f(v_i) + \hat{w}_i}{2 + m_i}. \quad (6)$$

ここで、関数  $f(\cdot)$  は、閲覧者数の少ない商品に対してペナルティを課すために、以下の2点を満たすように決定される。

- 単調増加関数である。
- $\lim_{v \rightarrow \infty} f(v) = 0$ .

なお、評価値は生成した比較グラフに対してのみ算出するため、 $f$  の定義域は1以上の整数とする。 $f$  としてどのような関数型を採用するかは議論の余地があるが、本論文では(7)式を採用した。

$$f(v) := -\frac{1}{2v^2}. \quad (7)$$

Colley+法によるランキングを算出するには、前節の導出過程をなぞることで、(4)式右辺に  $f(v_i)$  を加えればよいことがわかる。すなわち、 $b_i^+ := 1 + f(v_i) + \frac{w_i - l_i}{2}$  ( $i \in V$ ) を第  $i$  成分に持つ  $n$  次元ベクトル  $\mathbf{b}^+$  とし、(5)式の  $\mathbf{b}$  を  $\mathbf{b}^+$  に置き換えた連立方程式  $\mathbf{C}\mathbf{r} = \mathbf{b}^+$  を解けばよい。

2節の例に対し、Colley+法により導かれるランキングを表4に示す。

表 4 Colley+法によるランキング結果

順位	1	2	3	3	4
商品	C	E	A	B	D
評価値	0.597	0.572	0.291	0.291	0.191
閲覧者数	3	1	4	4	1

表 5 閲覧履歴 (365 セット) の代表値

	閲覧履歴数	セッション数
平均	297.68	106.84
中央値	293.00	106.00
標準偏差	97.41	34.81
最大値	738.00	275.00
最小値	77.00	24.00

表 6 点数と枝数 (365 セット) の代表値

	点数	枝数
平均	138.09	153.79
中央値	139.00	152.00
標準偏差	26.47	51.08
最大値	210.00	402.00
最小値	57.00	34.00

表3と表4を比較すると、Colley法では1位となっていた商品EがColley+法では2位となり、相対的に閲覧者数の多い商品A, B, CがColley法のランキングに比べ、順位を上げていることがわかる。Colley+法の特徴は、閲覧者数が少ない商品に対して評価値を差し引いている点にある。相対的に閲覧者数の多い商品を重視することで、多くの顧客に選ばれている商品ランキングを生成できる。

以下に、提案手法 Colley+法による商品ランキングの算出手順をまとめる。

1. 閲覧履歴より  $\mathbf{X}$ ,  $\mathbf{v}$  を生成する。
2.  $\mathbf{X}$ ,  $\mathbf{v}$  より  $\mathbf{C}$ ,  $\mathbf{b}^+$  を求める。
3.  $\mathbf{r} = \mathbf{C}^{-1}\mathbf{b}^+$  より商品ランキングを求める。

#### 4. 売れ筋ランキングとの比較

本節では、提供された情報をもとに算出した売れ筋ランキング、Colley法によるランキング、Colley+法によるランキングを比較する。閲覧のみのセッションと購買のみのセッションを除いた閲覧履歴を用い、ある日の15時から翌日の15時までを1セットとし、1年分365セットのランキングを生成した。表5, 6に

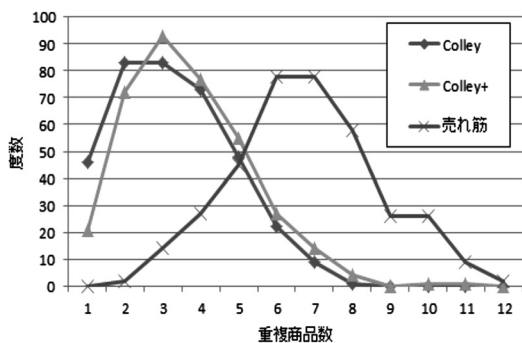


図3 重複商品数の分布

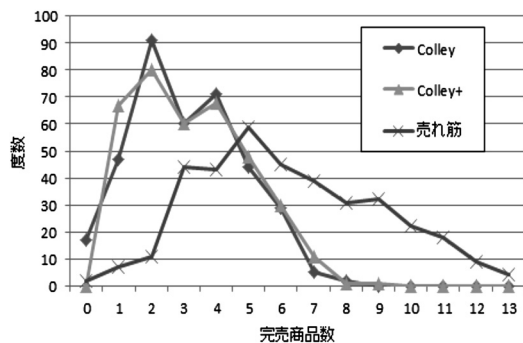


図4 完売商品数の分布

表7 重複商品数の分布の代表値

	Colley	Colley+	売れ筋
平均	3.25	3.65	6.78
中央値	3.00	3.00	7.00
標準偏差	1.60	1.67	1.92
最大値	8.00	11.00	12.00
最小値	0.00	0.00	2.00

表8 完売商品数の分布の代表値

	Colley	Colley+	売れ筋
平均	3.11	3.22	6.24
中央値	3.00	3.00	6.00
標準偏差	1.71	1.79	2.77
最大値	8.00	9.00	13.00
最小値	0.00	0.00	0.00

計算に用いた情報の規模概要を示す。実際のメルマガでは、売れ筋ランキング上位 20 位が配信されていることから、各ランキングの上位 20 位について重複商品数、完売商品数、閲覧者数を比較する。また、ここで算出した売れ筋ランキングは、提供された購買履歴をもとに売上高によって順位づけたものである。

まず、売れ筋と提案手法のランキング比較のための尺度として、重複商品数と完売商品数について説明する。

重複商品数とは、ある日の閲覧履歴をもとに生成したランキングの上位 20 位と、その日の前日のランキング上位 20 位の間で重複している商品の数である。より多くの商品情報を顧客に届けたいという意味では、重複商品数は少ないほうが好ましい。

図3、表7に重複商品数の比較結果を示す。図3より、Colley法、提案手法の分布が売れ筋ランキングの分布に比べて、左に寄っていることが読み取れる。また、表7に示したように基本統計量からも分布に差異があることがわかる。

これらの結果は、売れ筋ランキングに比べ Colley法、提案手法のほうが、前日のランキングに登場した商品と重複する商品が少なく、より情報量の多いランキングを生成できることを意味している。

もう1つの比較尺度である完売商品数は、生成したランキングの上位 20 位において、販売終了時点で完

売している商品数である。

メルマガでお勧めした商品がメルマガ配信直後に完売した場合、その時点以降、メルマガを見て商品を購入しようとした顧客は購入できず、機会損失となる。このような状況は、売れ筋ランキングのように、売れている商品を薦める場合に生じやすい。

機会損失を避けるためには、販売期間中に完売商品数が少ないことが望ましい。提案手法は選ばれやすさを基準にしたランキング手法であり、そのため、選ばれやすい商品ほど高く評価される。つまり、提案手法は売れ筋ランキングに比べ機会損失が生じる可能性を低めるランキング手法であると考えられる。

そこで、完売商品数を用いて、提案手法ではどの程度機会損失が生じる可能性を低めているか評価する。しかし、提供された情報からは商品が結果的に完売したかどうかはわかるが、完売した時点を読み取ることができない。そのため、完売時点の情報を考慮してランキングを生成することができないが、誰がいつ何を買ったかという購買情報とその購買に対応する閲覧履歴を用いることで、同じ条件下で売れ筋ランキングと提案手法によるランキングを算出し、完売商品数を評価する。

完売商品数が多いランキングほど、“メルマガに掲載されている商品がすでに売り切れている”という事象

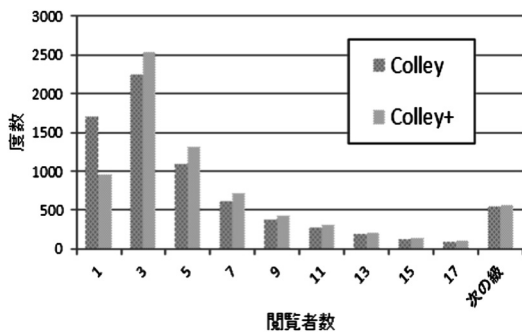


図5 観覧者数の度数分布

が生じやすい。この事象による機会損失を回避するという意味では、完売商品数は少ないほうが好ましい。

図4, 表8に示す結果から読み取れることとして、Colley法, 提案手法では、生成した365セットのランキングにおいて半数の機会で完売商品数が、3商品以下に限られており、多くても8または9商品に抑えられる。一方で、売れ筋ランキングでは、多くの場合、6商品が販売終了時点で完売しており、8商品以上完売することも稀ではない。

以上の結果から、Colley法, 提案手法で生成されたランキングは、売れ筋ランキングに比べて機会損失が生じにくいと言える。

これまで示した結果から、Colley法, 提案手法は売れ筋ランキングと比較して、重複商品数が少ないためにより多くの商品を提示できる手法であり、また、機会損失の可能性を低めるランキングを提示できることが言えた。

次に、Colley法, 提案手法で上位に位置づけられた商品は、“多くの顧客に選ばれた商品”と言えるのか、各手法のランキング365セット×上位20商品の観覧者数を集計することで検証する。

そもそも提案手法における関数  $f(v_i)$  の役割は、より多くの顧客に比較された商品を上位に位置づけるためのものであった。そこで、結果がそれを反映し、Colley法と比較して改善されているか確認するため、図5に観覧者数の度数分布を示す。縦軸の度数は、365セット×20商品中で横軸の観覧者数に対応する商品数を意味する。

図5より、Colley法に比べて提案手法のほうが観覧者数の多い商品を上位に位置づけていることがわかる。特に、観覧者数1の度数は、Colley法, Colley+法でそれぞれ1705, 952となっており、提案手法において観覧者数の少ない商品が上位に組み込まれにくくなっていることがわかる。

## 5. まとめ

本論文ではWebサイトの観覧履歴から顧客の嗜好を抽出した比較グラフを作成し、それに改良したColley法(Colley+法)を適用し商品ランキングを算出する方法を提案した。当該手法の利点としては

- 観覧履歴から商品間の嗜好をランキングに取り入れることが可能
- 最終的に  $n$  元連立線型方程式の求解に帰着といった特徴がある。また、本提案手法は、部分的な対象間のみ比較結果がある場合のランキング(例[2, 4])に広く適用可能である点を強調しておきたい。

一方で、比較グラフを生成する際、購買された商品と直近の何件の履歴を比較グラフに反映するかといった点や、Colley+法で用いる関数のより妥当な関数型の検討については今後の課題である。また、提案手法が実務で応用された場合、重複商品数、完売商品数を集計し、売れ筋ランキングと比較することで提案手法の有効性を検証することができる。

**謝辞** 平成24年度データ解析コンペティションのフリー学生部門において、経営科学系研究部会連合協議会および株式会社リクルート・グループよりデータを貸与していただきました。また、2名の査読者に有益なコメントをいただきました。深くお礼申し上げます。

## 参考文献

- [1] Colley, W. N., Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained, <http://www.colleyrankings.com/matrate.pdf> (2013年11月確認)。
- [2] Ryuo, S., Yamamoto, Y., "Ranking by Relational Power Based on Digraphs," *Journal of the Operations Research Society of Japan*, **52**, 245–262, 2009.
- [3] Brink, R., Gilles, R., "Measuring Domination in Directed Networks," *Social Networks*, **22**, 141–157, 2000.
- [4] Jiang, X., Lim, L., Yao, Y., Ye, Y., "Statistical Ranking and Combinatorial Hodge Theory," *Mathematical Programming*, **127**, 203–244, 2011.
- [5] Brin, S., Motwani, R., Page, L., Winograd, T., "The PageRank Citation Ranking: Bringing Order to the Web," Technical Report, Stanford InfoLab, 1999.
- [6] 高須賀清隆, 丸山一貴, 寺田実, "観覧履歴を利用した協調フィルタリングによるWebページ推薦とその評価," 電子情報通信学会技術研究報告, **107**, 115–120, 2007.
- [7] 松尾豊, 福田隼人, 石塚満, "ユーザ個人の観覧履歴からのキーワード抽出によるブラウジング支援," 人工知能学会論文誌, **18**, 203–211, 2003.
- [8] 土方嘉徳, "嗜好抽出と情報推薦技術," 情報処理, **48**, 957–965, 2007.