

レビュー順序グラフに基づく購買行動パターンの分析

稲福 和史[†] 伏見 卓恭^{††} 佐藤 哲司^{††}

[†] 筑波大学情報学群知識情報・図書館学類 〒 305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学図書館情報メディア研究科 〒 305-8550 茨城県つくば市春日 1-2

E-mail: [†]{inafuku,fushimi,satoh}@ce.slis.tsukuba.ac.jp

あらまし 近年、オンラインショッピングの拡大に伴い膨大な顧客購買データが利用できるようになっている。本研究では、モチーフ分析を用いて購買履歴から購買行動パターンの抽出に取り組んでいる。これらのパターンは、幾つかの特徴的な局所構造の組み合わせにより、異なる性質を有すると考えられる。本稿では、モチーフ分析の結果に対し、クラスタリングを適用することで購買行動パターンの抽出が行えることを示す。また同時に、主成分分析を適用することで、購買行動パターン間の関係を決定する要因となる局所構造の抽出が行えることを示す。実運用されているオンラインショッピングのレビューデータを購買データとみなし、購買行動パターンを分析したので報告する。

キーワード E-コマース, 購買行動, 有向グラフ, モチーフパターン

1. はじめに

オンラインショッピングの拡大に伴い、膨大な顧客購買データが利用できるようになっている。このデータを用いてユーザの購買行動の推定や商品推薦の研究が盛んに行なわれている。[1] [2] [3] [7] [9]。オンラインショッピングを行うユーザは、商品を購入後にレビュー記事の投稿も日常的に行っている。このため、レビューの投稿順序はユーザの購買順序を表していると考えられる。本研究ではこれまで、オンラインショッピングにおけるレビューの投稿順序を購買順序とみなした上で、有向グラフを構築し、モチーフ分析による購買行動パターンの抽出に取り組んできた。商品をノードとし、レビューの投稿順序に沿ってエッジを付与することで、購買順序を示す有向グラフが構築できる。ユーザー一人一人の購買順序をマージすることで、多くのユーザによって購入される商品間の関係について、購買される順番に着目した分析が可能となる。本稿では、これまでに抽出した購買行動パターンについて主成分分析を軸に、より詳細に分析を行っている。本稿の構成は次の通りである。関連研究を第2章、購買パターンの抽出手法を第3章、実験の結果及び考察を第4章及び第5章、最後にまとめを第6章に示す。

2. 関連研究

本研究において核となるのは、アイテムからなるネットワークのモチーフ分析である。本章では、モチーフ分析、アイテムベースのネットワーク分析及びネットワークベースのアイテム推薦に関する関連研究について説明する。

Salihoglu [6] は無向ネットワークを3ノード及び4ノードのモチーフパターン出現頻度ベクトルで表現し、道路ネットワーク、ASルーティングネットワーク、共購買ネットワークの構造はモチーフパターン出現頻度ベクトルに反映されることを明らかにしている。Srivastava [8] はモチーフパターンの出現頻度とコミュニティ検出によってAmazonの共購買ネットワークを分析している。著者はネットワーク全体に対するモチーフ分析を

行ったのち、ネットワーク全体をコミュニティに分割、ノードとして扱い新たなネットワークを構築、モチーフ分析を行った。これにより、さらに俯瞰的な視点での分析が可能になることを示した。また、ネットワーク全体及びコミュニティをノードとしたネットワークにおいて、同様のモチーフパターン出現頻度が3ノード及び4ノードの両方から得られたことを明らかにしている。

Zhou はアイテムとユーザから構築される二部グラフを用いたネットワークベースのアイテム推薦を提案している [13]。また、人気の高いアイテムが極端に高いスコアを獲得することを防ぐため、スコアの平均化プロセスを追加した [12]。しかし、この方法は人気の低いアイテムに過度のスコア補正を与えてしまうため、改善手法が提案されている [4], [5]。

また、Yu [11] によって多数のネットワークベースのアイテム推薦手法が収集・比較されているが、上述の手法を含め、これらは全ては無向グラフに基づくものである。本研究では、連続した購買行動に着目し商品をノードとした有向グラフを対象とする。

3. 購買パターンの抽出手法

本研究では、レビューデータから構築するレビュー順序グラフ (PHG^(注1)) のモチーフベクトルに基づき、購買行動パターンを抽出、分析する。ユーザー集合及びアイテム集合をそれぞれ U 及び I と定義する。ユーザ u がアイテム i のレビューを時刻 t に投稿したとき、レビュー r を $r = (u, i, t)$ と表す。ユーザ u が N 件のレビューを投稿した際のレビュー群を $R(u) = [r_1, r_2, \dots, r_{N_u}]$ と表す。 r_n は r_{n+1} より1つ前に投稿されたレビューであることを示す。ここで、ユーザ u にレビューされたアイテム群を $I(u) = [i_1, i_2, \dots, i_{N_u}]$ と定義し、 $I(u)$ において、連続する要素をペアにした集合を構築する。

(注1) : レビューの投稿順序がユーザの購買順序を表しているとしたことから、レビュー順序グラフを Purchased History Graph (PHG) と称する。

$$SI(u) = \{(i_1, i_2), (i_2, i_3), \dots, (i_{N_u-1}, i_{N_u})\} \subset I(u) \times I(u).$$

得られた SI を基に PHG を構築し分析する手順を以下に示す.

3.1 レビュー順序グラフの構築

本節では、各ユーザのレビューデータから有向グラフを構築する. 2名のユーザが合計で5種類、6個のアイテムを購入した際に構築される PHG の模式図を図1に示す. アイテムをノードと定義すると、ノード集合 V はアイテム集合 I と等しいため、 $V = I$ である. ここで、 $SI(u)$ のノードペア (i, j) 間に i から j への有向エッジを追加すると、エッジ集合 E は次のように表せる.

$$E = \left\{ (i, j) \in \bigcup_{u \in U} SI(u) \right\} \quad (1)$$

この $G = (V, E)$ で表される有向グラフを PHG^(注2) とする.

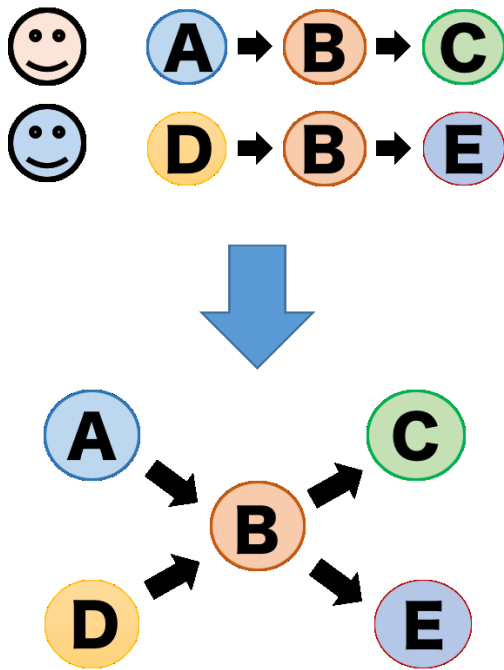


図1 レビュー順序グラフ

3.2 弱連結成分分解

3.1で構築した PHG に対し、弱連結成分分解を行う. 図2のように、PHG $G(V, E)$ において、ノード u とノード v が u から v ないし v から u に到達しうるとき、 u と v は同じ弱連結成分 (WCC) に属する. ここで、WCC 数を M とし、第 m 番目の WCC に属するノードを $V^{(m)}$ 及び $E^{(m)}$ と定義すると、各 WCC グラフは $G^{(m)} = (V^{(m)}, E^{(m)})$ で示される.

3.3 モチーフベクトルの構築

本節では WCC にモチーフ分析を適用する. 3 ノードからなるモチーフパターン (MP) は 13 種類あり (図3^(注3)), WCC

(注2): 典型的な購買行動パターンの抽出するために、 (i, j) ペアの出現頻度が 10 以上のものに対しエッジを追加している.

(注3): 各ノードに振られた番号は、各モチーフパターンにおいてノードをその構造に基づき分類した際の 30 種類のノードロール (Role) を示す. これにより、ノードの役割に着目した分析が可能になる.

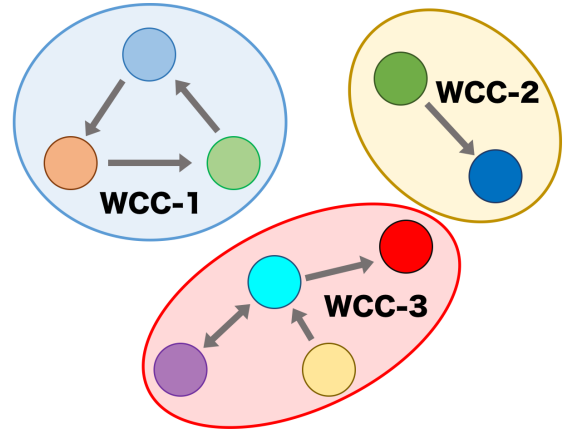


図2 弱連結成分分解

$G^{(m)}$ にモチーフ分析を行うことで、モチーフ分布 $\mathbf{x}_m$ が算出できる. $\mathbf{x}_m$ は各 MP の出現頻度を表しており、13 次元ベクトルとなる. このベクトルを正規化 (L_1 norm) し MP 出現確率ベクトルを取得する. これをモチーフベクトル^(注4) と称する.

グラフの構造をモチーフベクトルを用いて表現することで、より直感的に WCC の特徴を捉えることができる. 例えば、モチーフパターン 1 (MP1) は 1 種のメイン商品から多種のオプション商品への購買行動を示していると考えられる. 具体的には「iPhone と iPhone ケース」のような関係である. MP2 は購買される順序が決まっている商品を示しており、「コミックの 1 巻, 2 巻, 3 巻」というようなケースが考えられる. また、MP13 は購買の順序性がない「服」などの購買行動を示していると考えられる. 基本となる購買行動パターンは 1 つの MP で表せることが望ましい. しかし、実際の購買行動パターンは複数の MP の組み合わせ (混合モデル) になると考えられる. そのため、WCC の構造的特徴をモチーフベクトルを用いて表現する.

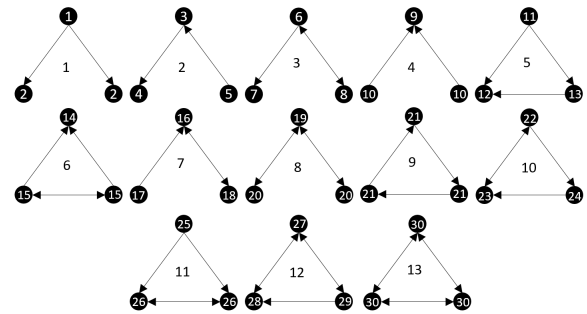


図3 3 ノードモチーフパターン及びノードロール

3.4 k -means クラタリング

3.3 で構築したモチーフベクトル $\mathbf{x}$ に k -means クラスタリングを適用し、WCC を K 個のクラスタに分ける. WCC $G^{(m)}$ と $G^{(n)}$ の距離は次のように求める.

$$d(\mathbf{x}_m, \mathbf{x}_n) = \|\mathbf{x}_m - \mathbf{x}_n\|_{L_2}^2. \quad (2)$$

(注4): モチーフパターン出現頻度の累積度からより典型的な購買行動パターンを抽出するため、モチーフ分布ベクトルを確率ベクトルへと変換している.

各クラスはモチーフベクトルに共通の傾向を持つ WCC によって構成される。すなわち、一つのクラスは一つの購買行動パターンに対応すると考えられる。ここでは、各クラスにおけるセントロイドのモチーフベクトルを代表値として抽出、購買行動パターンとしてどのような意味を持つのか分析する。クラス数の決定は GAP 統計量 G_k [10] を用いた。 G_k を最小とする K が求めるクラス数となる。

3.5 主成分分析

3.3 で構築したモチーフベクトルに対し、主成分分析を行う。上位の主成分の固有ベクトルを求め、さらに上位及び下位の要素を観察することで、購買行動において強い影響力を持つ MP を抽出・分析できる。

具体的には、各 WCC $G^{(m)}$ の 13 次元モチーフベクトル $\mathbf{x}_m$ を集めた $M \times 13$ の行列 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]^T$ から 13×13 の共分散行列

$$\Sigma = (\mathbf{X} - \mu)^T (\mathbf{X} - \mu)$$

を計算し、 Σ の固有ベクトル群 $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{13}]^T$ を求める。ここで、 $\mu = [1, \dots, 1]^T [\mu_1, \dots, \mu_{13}]$ は、各モチーフパターンの出現頻度の平均を要素とした $M \times 13$ の行列である。値が大きい H 個の固有値 $\lambda_1, \dots, \lambda_H$ と対応する固有ベクトル $\mathbf{v}_1, \dots, \mathbf{v}_H$ を抽出する。

4. 実験

4.1 データセット

本研究では、Nii が研究用に公開している楽天市場のレビューデータ (注5) を使用した。6,500 万レビューからなるデータセットから、投稿者が一意に判別できるレビューを抽出した。その際、購入した事が確認できない、あるいは投稿日時が欠落しているレビューを除外した。以上の処理を行い、2,445,084 ユーザによる 17,794,337 レビューを評価データセットとした。

4.2 PHG の統計分析

4.1 のデータから構築した 10,266 ノード、13,939 エッジの PHG を図 4 に示す。PHG は少数の大きなサイズの連結成分と多数の小さなサイズの連結成分から構成されている。また、複数箇所に類似した構造が観察されることから、一定の構造パターンが存在すると考えられる。PHG 全体のモチーフベクトルは図 5 に示すように MP3, MP7, MP8 が高頻度で出現している。これは、図 4 からわかるように、1 つの中心的な商品 (コアノード (注6)) とそれを取り囲むように存在する周囲の商品によって構成される、いわば花びら型のような巨大な構造が存在するためであると考えられる。このため、モチーフベクトルの分布は巨大な花びら型を持つ弱連結成分に全体が引っ張られていると考えられ、PHG 全体からの購買行動パターンの抽出を行うことは難しい。

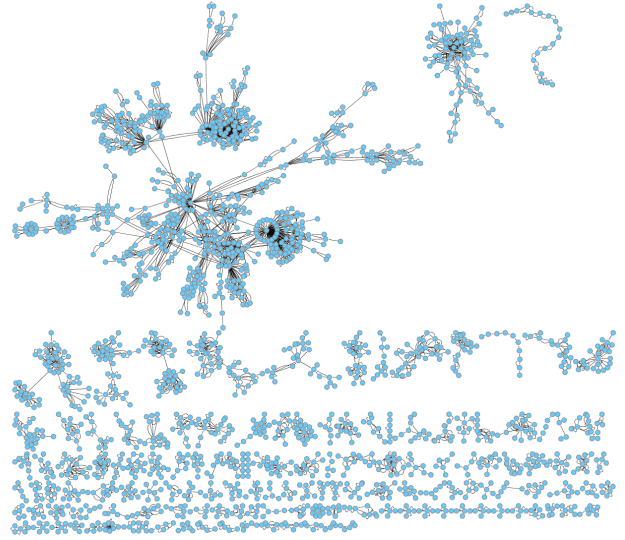


図 4 PHG 可視化結果

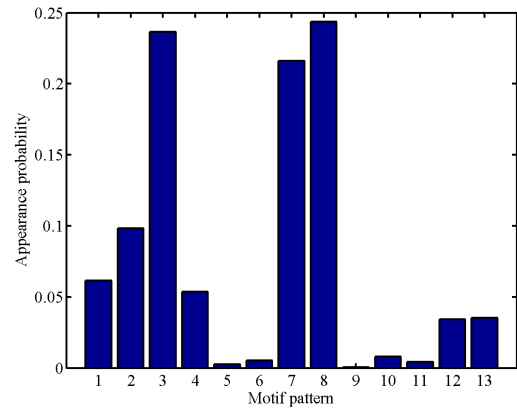


図 5 PHG モチーフベクトル

4.3 弱連結成分分解

4.2 で構築した PHG に対し弱連結成分分解を行った。この際、WCC のサイズが 5 以上のものを実験対象とした。結果、PHG は 141 WCC に分解され、2,094 ノード及び 5,399 エッジからなるグラフが構築された。図 6 は WCC のサイズ分布を示している。べき乗則の存在が確認できることから、4.2 でも記したように、少数の大規模な WCC と多数の小規模な WCC から成るグラフだといえる。

4.4 モチーフ分析及び k -means クラスタリング

4.3 で構築した WCC 群に対し、それぞれモチーフ分析を行い 141 個のモチーフベクトルを取得した。これに対し、PHG と同様に正規化 (L_1 norm) し、 k -means クラスタリングを行った。GAP 統計量 G_k によりクラス数 K は 8 とした。表 1 は、クラス別の WCC 数を示す。また、図 9 は、各クラスのセントロイドベクトルを示す。

• Cluster-1

モチーフ分布に着目すると、もっとも多く出現するモチーフパターンが MP2、続いて MP3, MP7 である。これらはいずれも 2 ノード間の関係から構築されるパターンであり、特に MP2 はコアノードを経由する一方通行型である。このパターンにおい

(注5) : <http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html>

(注6) : コアノードとは、5.2 に示すように、各ノードに対してモチーフパターンに基づく 30 種類のノードロールを数え上げた際に Role1,3,6,9,16,19 の比率が大きいノードと定義できる。

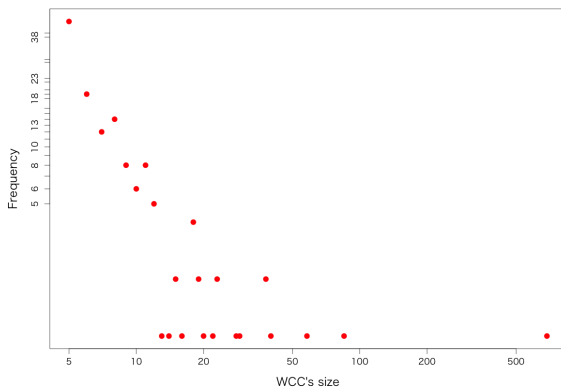


図 6 WCC サイズ分布

表 1 クラスタ別 WCC 数

Cluster-1	Cluster-2	Cluster-3	Cluster-4
17	8	14	38
Cluster-5	Cluster-6	Cluster-7	Cluster-8
28	11	19	6

てユーザは、図 7 のようにコアノードを経由して購買行動を次々と発展させているといえる。これを踏まえると、Cluster-1 は「中心経由発展型」のであると考えられる。

● Cluster-2

本クラスタでは、MP7 が最も多く観察され、第二位に MP2、第三位に MP4 が続く。これらの MP はいずれも 1 ノードへ集中する傾向を持つ。このことから、Cluster-2 はコアノードとその周辺ノードによって主に構成され、コアノードへ集中するパターンであると考えられる。具体的には、図 8 のように多種多様なメイン商品群とそれらに共通して使用出来るオプション商品の組み合わせが想定される。

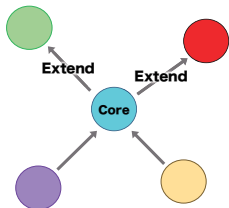


図 7 中心経由発展型 模式図

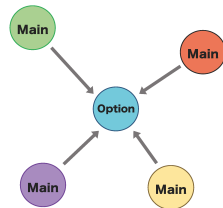


図 8 コアノード集中型 模式図

● Cluster-3

Cluster-3 は MP8,MP3 が極端に多く出現するクラスタであり、いずれの MP も 2 ノード間の関係からなる。Cluster-2 と同様にコアノードと周辺ノードによって構成されるクラスタであるが、その関係性が双方向エッジであるという点で異なり、コア商品と周辺商品の前後を問わず購買が行われる。またコアノードは始点にも終端にもなることから、コアノードが購買行動を強く促す役割を持つといえる。

● Cluster-4

Cluster4 は全クラスタの中で最大のクラスタであり、非常に複雑なモチーフ分布を示した。WCC 毎のノード数・エッジ数の

標準偏差はそれぞれ 116 と 363 と他のクラスタと比較して大きな値を示した。cluster4 は最大連結成分 (LWCC) を有しており、これが標準偏差に大きな影響をもたらしていると考えられる。そのため、LWCC を除外してノード数・エッジ数の標準偏差を算出したところ、17 と 50 であった。結果として、LWCC の影響を差し引いてもほかのクラスタに比べ大きな標準偏差を示した。以上から、Cluster4 は様々な WCC を含む複雑型であると考えられる。

● Cluster-5

Cluster-5 は主に MP3 と MP1 から構成されるクラスタであり、いずれの MP もコアノードから周辺ノードへの有向エッジが多い、放出型 MP である。そのため、Cluster-5 では、コアノードが購買行動の起点となる。また、さらなる連続購買へと発展せず、短いシーケンスで完結する傾向にある。Cluster-5 は 1 種類のメイン商品と多種多様なオプション商品によって成り立っているクラスタであると考えられる。

● Cluster-6

Cluster-6 は MP2 が極端に多く、続いて MP1,MP4 が出現するクラスタである。これらの MP はいずれも片方向エッジから構成される MP であり、商品の購入順が明確に決まっていることを示す。具体的には、定期的に発売される人気アーティストの CD や、コミックのシリーズ単行本などで観察されるパターンである。

● Cluster-7

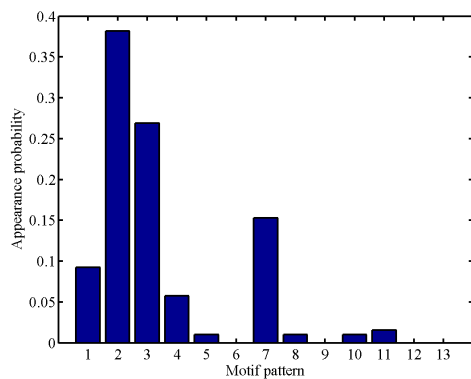
Cluster-7 は MP8 及び MP7 が極端に多いクラスタである。これらの MP はコアノードと周辺ノードの双方向エッジで構成される MP であり、Cluster-2 及び Cluster-5 のハイブリッドであるともいえる。つまり、1 種類のメイン商品と多種多様なオプション商品ないし多種多様なオプション商品と 1 種のオプション商品から構成されるパターンであると考えられる。加えて、コア商品と周辺商品が同時に購入された場合そのレビューの前後はユーザによって異なることから、双方向エッジは同時購入される商品にも観察されることが考えられる。

● Cluster-8

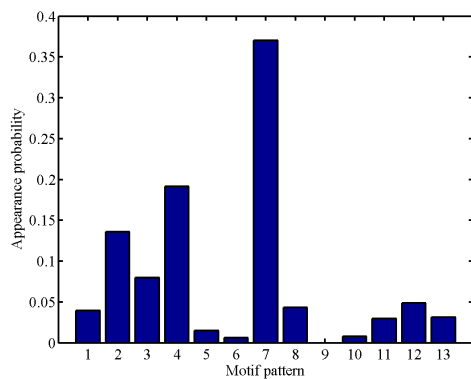
Cluster-8 は MP4,MP1,MP2 の片方向エッジ MP から構成される。この点で Cluster-6 と同様に購入の順序が明確に決まっている事が示されている。加えて極端に MP4 が多いことから 2 つの商品で購入が完結する傾向にあるといえる。これは小説の上下巻などで観察されるパターンである。

4.5 主成分分析

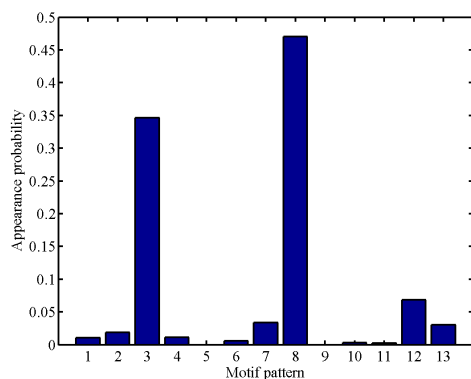
4.4 で構築した 141 モチーフベクトルに対し、主成分分析を行った。また、表 2 に示す様に、累積寄与率が第四主成分で 85% に達していることから、分析の対象を第四主成分までとする。まず、第一主成分から第四主成分の固有ベクトルの各要素 (MP) について、正負それぞれ絶対値の大きいものを 2 つずつ抽出し、各主成分が示す特徴について分析する。図 10 に固有ベクトルを示す。



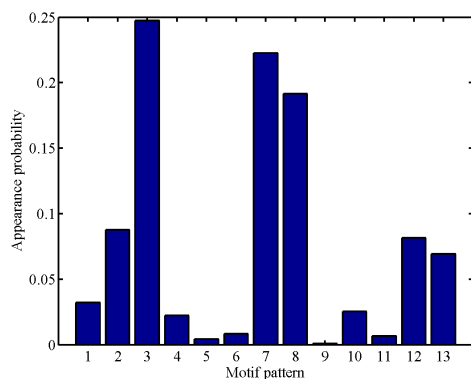
(a) Cluster-1 モチーフ分布



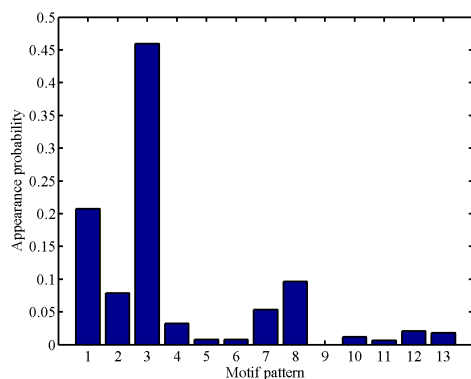
(b) Cluster-2 モチーフ分布



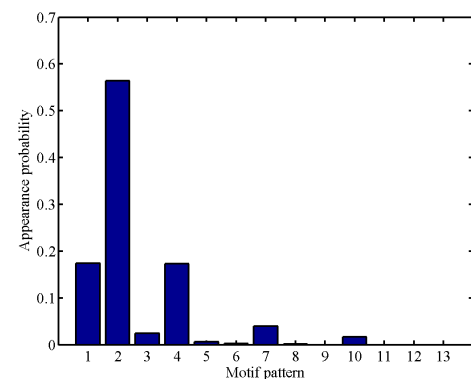
(c) Cluster-3 モチーフ分布



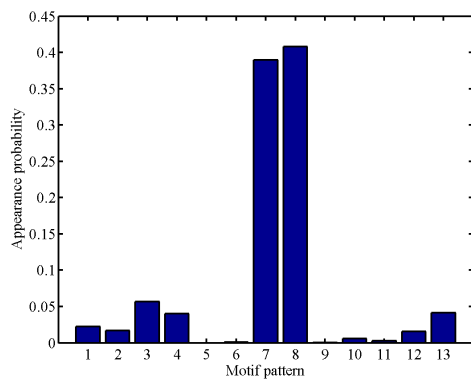
(d) Cluster-4 モチーフ分布



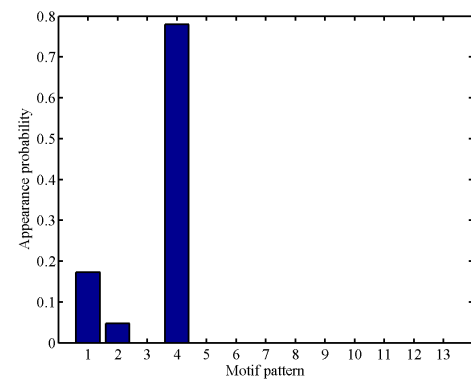
(e) Cluster-5 モチーフ分布



(f) Cluster-6 モチーフ分布



(g) Cluster-7 モチーフ分布



(h) Cluster-8 モチーフ分布

図9 クラスタ別モチーフ分布

表 2 寄与率

	PC1	PC2	PC3	PC4
寄与率	0.3324	0.2655	0.1511	0.1022
累積寄与率	0.3324	0.5979	0.7490	0.8512

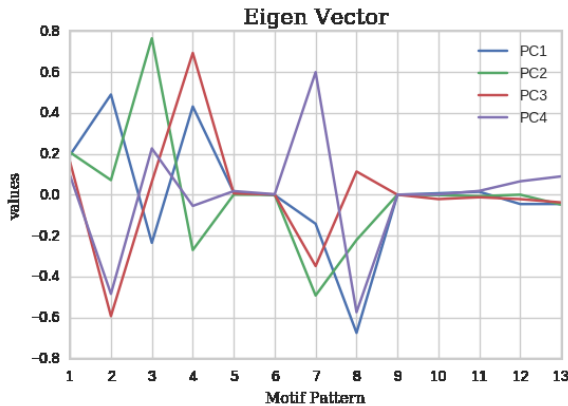


図 10 固有ベクトル

- 第一主成分 (PC1)

大きな正の値を示したのは MP2 次いで MP4, 負の値では MP8 次いで MP3 である。両者の違いは双方向エッジの有無である。MP2,4 は単方向エッジのみから構成される MP, MP8,3 は双方向エッジを含む MP である。さらに MP2,4 のみに着目すると, MP2 は完全な順序性を示すのに対し, MP8 は順序性を一切示さない。以上のことから, 第一主成分は順序性を表していると考えられる。

- 第二主成分 (PC2)

大きな正の値を示したのは MP3 次いで MP1, 負の値では MP7 次いで MP4 である。エッジ方向について, MP3,4 は 1 ノードから周辺 2 ノードへ流れ出る傾向にあり, MP7,4 は周辺 2 ノードから 1 ノードへ集まる傾向にある。以上のことから, 第二主成分はコアノードと周辺ノードの間のエッジ方向を表していると考えられる。

- 第三主成分 (PC3)

大きな正の値を示したのは MP4 次いで MP1, 負の値では MP2 次いで MP7 である。両者の違いは構造的対称性であると考えられる。MP4,1 は線対称な MP であるが, MP2,7 はいずれも非線対称な MP である。また, MP1 に僅差で続く MP8 も対称性を示す。以上のことから, 第三主成分は MP の構造的対称性を表していると考えられる。

- 第四主成分 (PC4)

大きな正の値を示したのは MP7 次いで MP3, 負の値では MP8 次いで MP2 である。これらの違いはコアノードから見た入エッジと出エッジの比であると考えられる。MP7,3 において入出比はそれぞれ 2:1, 1:2 となっている。一方 MP8,2 の入出比は 1:1 である。以上のことから, 第四主成分はコアノードから見たエッジの入出比を表していると考えられる。

表 3 各主成分の特徴

	正	負
PC1	順序性が高い	順序性が低い
PC2	放出型	集中型
PC3	対称性が高い	対称性が低い
PC4	入出比が大きい	入出比が小さい

5. 考察

5.1 購買行動パターン間の関係

クラスタリング結果及び主成分分析の結果から, クラスタ (購買行動パターン) 間の関係について考察する。図 11 に第一主成分から第四主成分及びクラスタリングの散布図行列を示す。まず, 第一主成分において高い値を示すのは Cluster-6,8 であり, セントロイドのモチーフベクトルからも順序性があることを把握することができる。加えて, 第三主成分を観察すると Cluster-8 は高い値を, Cluster-6 は低い値を示す。これは, Cluster-8 を構成する主な MP4 が線対称な構造であること, 対して Cluster-6 の主な MP が非線対称な MP2 であることが要因である。具体的には, 購買行動が短いシーケンスで完結するか, さらなる発展につながるかという違いである。以上のことから, 第一主成分 (順序性) と第三主成分 (構造的対称性) に着目すると, 順序性の高い購買行動だが短いシーケンスで完結するパターン (Cluster-8) と順序性の高い購買行動でさらなる購買に発展するパターン (Cluster-6) の 2 つの購買行動パターンを抽出することができる。

次に, 第二主成分において, Cluster-5 が大きい値, Cluster-2,7 が低い値を示している。これは, コアノードと周辺ノードについて, 前者は放出型であり, 後者は集中型であることを示している。加えて, 第一主成分について, Cluster-2,5 は高い値, Cluster-3,7 は低い値を示している。これらのことから, Cluster-2 は順序性が高い集中型, Cluster-5 は順序性が高い放出型, Cluster-3,7 はそれぞれ順序性が低い放出型及び集中型であるといえる。以上のことから, コアノードと周辺ノードから成る構造に対して, 第一主成分 (順序性) と第二主成分 (放出・集中) に着目することで幾つかのパターンを抽出することが可能であるといえる。

以上のように, クラスタリング結果と主成分分析結果を照らし合わせることで, 購買行動パターン間の関係を把握することが可能である。

5.2 コアノード・周辺ノード

5.1 から, コアノードと周辺ノード及びその関係性が, 購買行動パターンを抽出するにあたって重要な局所構造であることが示された。本節では, これらの構造についてより詳細な分析を行う。図 3 におけるノードロールについて表 4 のように分類する。さらにコアロールは, コアノードからの出エッジが多い「放出」, コアノードへの入エッジが多い「流入」, 入出比が同じ「両方」の 3 種類に分けることができる。ここで, ノード i に対して, コアロールとして出現する回数 $y_{ir} (r = \{1, 3, 6, 9, 16, 19\})$ の和を $\sum_{r=1}^{30} y_{ir}$ で除したものを「ノード i のコア比率」と定

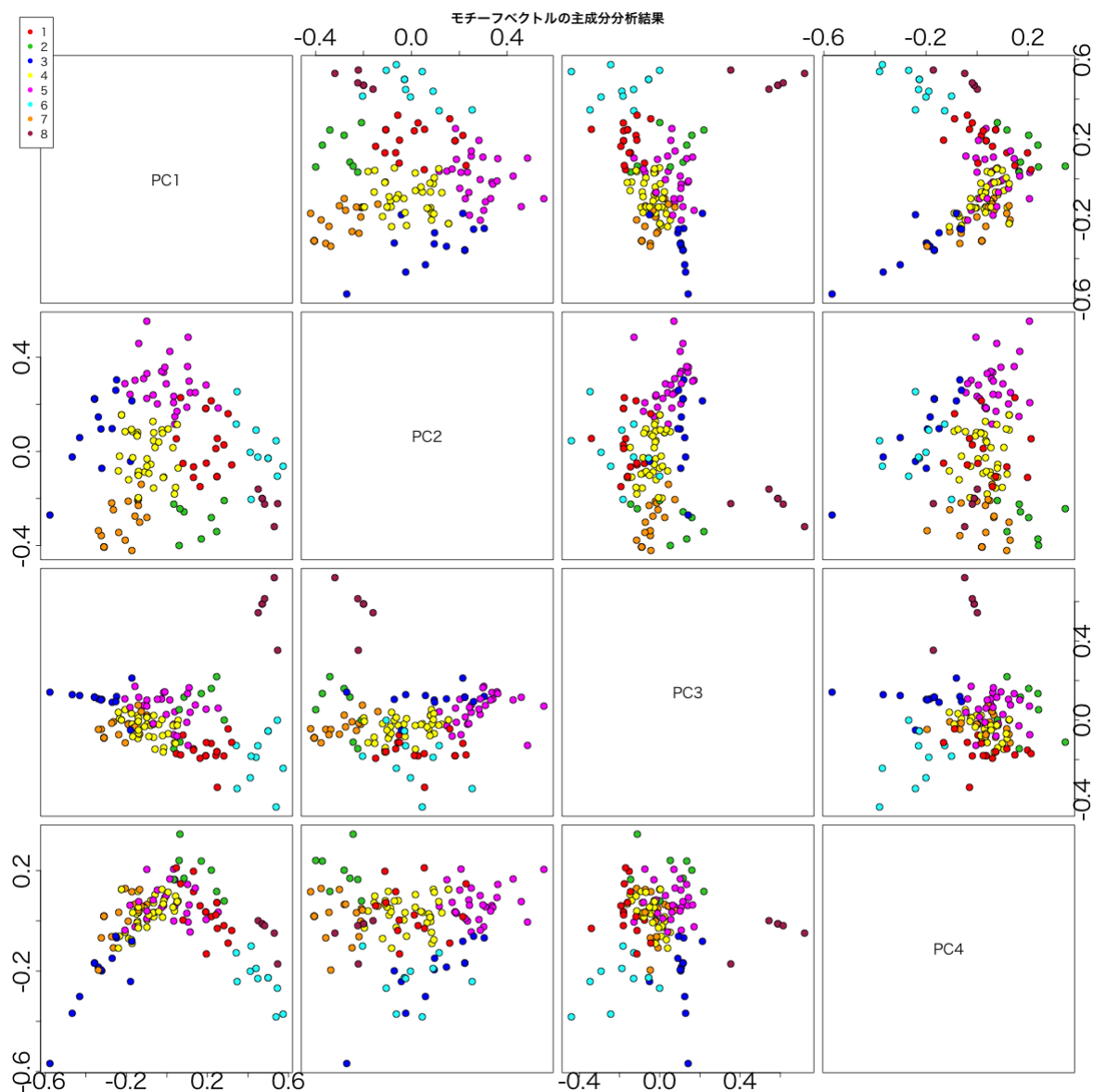


図 11 モチーフベクトルに対する主成分分析及び K-means クラスタリング結果

義する。周辺ロール ($r = \{2, 4, 5, 7, 8, 10, 17, 18, 20\}$) についても同様に「周辺比率」を定義する。縦軸に「コア比率」及び「周辺比率」、横軸に比率のランキングを取り、図 12 にプロットした。まず、コア比率が高いノードは全体のうちごくわずかしかな存在しないことがわかる。一方、周辺比率については比率が 1 のノードが全体の半分近く存在する。すなわち、コアノードとの間以外にエッジが存在しない商品が多数あるということであり、コアノードの重要性を示している。ここで、コア比率が 1 のノードをすべて抽出し、「放出」「流入」「両方」の内訳を図 13 にプロットした。縦軸に「放出」「流入」「両方」の比率をとり、横軸は各ノードに対応している。3 種類の比率によって、コアノードの性質を把握することができる。例えば、ノード 17 は「流入」のノードロールとしてのみ出現していることがわかる。実際のデータを調べたところ、ノード 17 の商品はスーツケースベルトであり、周辺ノードは複数種のスーツケースであった。このことから、ユーザは各々のニーズに合わせた

スーツケースを購入後、共通規格のオプション商品として販売されているスーツケースベルトを購入すると考えられる。また、「放出」の比率が高いノード 18 は万年筆、周辺ノードはカートリッジ関連商品であった。これは万年筆を購入後、サプライ品を購入したためと考えられる。このようにノードロールを用いることでコア比率の高いノードを抽出し、さらにその性質について分析が可能である。

表 4 ノードロール

コアロール			周辺ロール
放出ロール	両方ロール	流入ロール	-
1,6	3,19	9,16	2,4,5,7,8,10,17,18,20

6. ま と め

本研究では、レビューデータから構築したレビュー順序グラフのモチーフベクトルに対し、 k -means クラスタリング及び

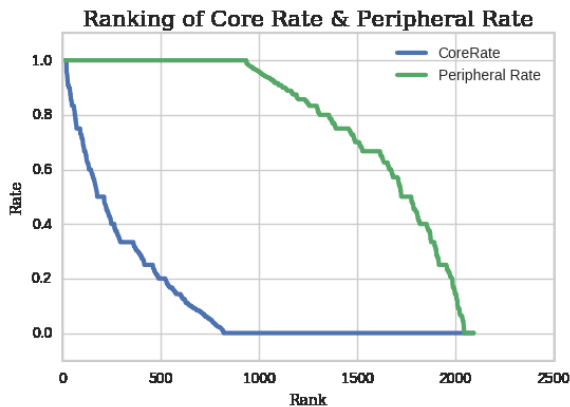


図 12 コア比率及び周辺比率ランキング

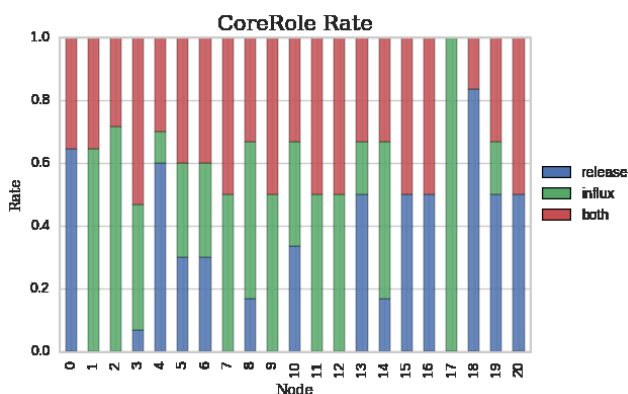


図 13 コアノードにおけるノードロール種別内訳

主成分分析を用いて購買行動パターンの抽出・分析を行った。楽天市場のレビュー履歴から有向グラフを構築し、WCC 毎にモチーフ分析、取得したモチーフベクトルに k -means クラスタリングを行う事で 8 つの購買行動パターンを抽出した。さらに主成分分析によって購買行動パターン間の関係が「順序性」「(コアノードと周辺ノードから成る構造における) 流出・集中」「構造的対象性」「入出比」の 4 つの局所構造によって位置付けられることを示した。同時に、購買行動パターンにおけるコアノードの重要性を示し、その詳細について分析した。

7. 謝 辞

本研究は JSPS 科研費 JP16H02907 の助成を受けたものです。また、楽天株式会社が国立情報学研究所の協力により研究目的で提供している「楽天公開データ」を利用しました。ここに記して謝意を示します。

文 献

- [1] A. Hayashi, M. Kohjima, T. Matsubayashi, and H. Sawada. Regularity measure and influence weight for analysis and visualization of consumer's attitude. In *2015 19th International Conference on Information Visualisation*, pages 290–299, July 2015.
- [2] F. Isinkaye, Y. Folaajimi, and B. Ojokoh. Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3):261 – 273, 2015.
- [3] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda. Topic tracking model for analyzing consumer purchase behavior. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI'09*, pages 1427–1432, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.
- [4] J.-G. Liu, Q. Guo, and Y.-C. Zhang. Information filtering via weighted heat conduction algorithm. *Physica A: Statistical Mechanics and its Applications*, 390(12):2414–2420, 2011.
- [5] J.-G. Liu, T. Zhou, and Q. Guo. Information filtering via biased heat conduction. *Physical Review E*, 84:037101, Sep 2011.
- [6] S. Salihoglu. Networks as vectors of their motif frequencies and 2-norm distance as a measure of similarity. Technical report, Stanford University, 2006.
- [7] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web, WWW '01*, pages 285–295, New York, NY, USA, 2001. ACM.
- [8] A. Srivastava. Motif analysis in the amazon product co-purchasing network. *CoRR*, abs/1012.4050, 2010.
- [9] P. Symeonidis, E. Tiakas, and Y. Manolopoulos. Product recommendation and rating prediction based on multi-modal social networks. In *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11*, pages 61–68, New York, NY, USA, 2011. ACM.
- [10] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.
- [11] F. Yu, A. Zeng, S. Gillard, and M. Medo. Network-based recommendation algorithms: A review. *Physica A: Statistical Mechanics and its Applications*, 452:192–208, 2016.
- [12] T. Zhou, Z. Kuscsik, J. Liu, M. Medo, J. Wakeling, and Y. Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515, 2010.
- [13] T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang. Bipartite network projection and personal recommendation. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 76(4):046115+, 2007.

- [1] A. Hayashi, M. Kohjima, T. Matsubayashi, and H. Sawada. Regularity measure and influence weight for analysis and visualization of consumer's attitude. In *2015 19th International Conference on Information Visualisation*, pages 290–299, July 2015.
- [2] F. Isinkaye, Y. Folaajimi, and B. Ojokoh. Recommendation systems: Principles, methods and evaluation. *Egyptian In-*