

推薦論文

トライアド推移に基づく購買行動の成長分析

稲福 和史^{1,a)} 伏見 卓恭^{2,b)} 佐藤 哲司^{3,c)}

受付日 2018年5月31日, 採録日 2019年1月15日

概要: 近年, オンラインショッピングの拡大にともな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. 我々はこれまで, 有向グラフの最小単位であるトライアドの構造に着目し, ユーザの購買履歴から構築したグラフから典型的なアイテム関係を抽出してきた. 本論文では, トライアド構造の推移に着目したグラフの成長傾向を分析する手法を提案する. 実運用されているオンラインショッピングサイトのレビューの投稿履歴を購買行動データと見なし, 購買行動の成長分析を行った. その結果, 同じ商品カテゴリでも異なるトライアド構造推移パターンが複数存在することや, 推移パターンの種別によって成長時期が異なることが明らかとなり, 提案手法が購買行動の成長分析に有用であることを確認した.

キーワード: トライアドパターン, トライアド推移, 成長分析, 購買行動, ネットワーク分析

Growth Analysis of Purchasing Behaviors Based on Triad Transition

KAZUFUMI INAFUKU^{1,a)} TAKAYASU FUSHIMI^{2,b)} TETSUJI SATOH^{3,c)}

Received: May 31, 2018, Accepted: January 15, 2019

Abstract: Recent years, a huge data of customer purchase history has become available with the expansion of online shopping. We tried to extract typical relationship of items with the focus on the structure of triads. In this paper, we propose a method to analyze the growth of the graph, focusing on the transition of the triad structure and maintenance of that structure. We applied to proposal method to datasets of the actual online shopping site. As a result, even in same item category, it found that there are differences in transition pattern of the triad structure and the time period maintained.

Keywords: triad pattern, transition of triad, growth analysis, purchase action, network analysis

1. はじめに

近年, 携帯型インターネット接続端末の普及や計算機の性能向上, ストレージの低廉化などを背景として, 様々なサービスにおいて, 大量のユーザ行動履歴データの取得・蓄積が可能となった. たとえば, オンラインショッピングサイトにおける購買履歴, インターネットテレビにおける

視聴履歴, 観光地を巡る旅行者の移動履歴などがあげられる. ユーザ行動履歴を分析することは, 各種サービスにおけるマーケティングや施策決定のために重要であり, 実際にこれらのデータを用いたユーザ行動の推定 [1], [5] やアイテム推薦 [4], [13], [16] の研究がさかに行われている.

オンラインショッピングサイトの購買履歴データを対象とした分析では, 多くのユーザの購買履歴データを用いることで, アイテムがどのような順序で購入されるのか, というアイテム間の関係を抽出することができる. 図 1 にアイテム間の関係の例を示す. 図 1 (a) のスマートフォンとアクセサリのように, 1つの中心的なアイテムを起点として複数の関連アイテムへと購買が発展する関係や,

¹ 筑波大学大学院図書館情報メディア研究科
Graduate School of Library, Information and Media Studies,
University of Tsukuba, Tsukuba, Ibaraki 305-8550, Japan

² 東京工科大学コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technol-
ogy, Hachioji, Tokyo 192-0982, Japan

³ 筑波大学図書館情報メディア系
Faculty of Library, Information and Media Science, Univer-
sity of Tsukuba, Tsukuba, Ibaraki 305-8550, Japan

a) inafuku@ce.slis.tsukuba.ac.jp

b) takayasu.fushimi@gmail.com

c) satoh@ce.slis.tsukuba.ac.jp

本論文の内容は 2017 年 6 月のマルチメディア, 分散, 協調とモバイル (DICOMO2017) シンポジウムで報告され, グループウェアとネットワークサービス研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である.

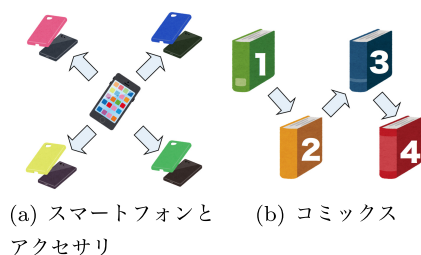


図 1 購買行動に基づくアイテム間の関係例

Fig. 1 Relationship of items based on purchasing behaviors.

図 1(b) のコミックスのように、連続して購入される関係など、アイテムのジャンルや特徴によって異なる関係が考えられる。そのため、アイテム推薦の場面などでは、これらの関係、すなわち「アイテムの買われる順序」に応じて推薦を行うことが有効であると考えられる。

筆者らは、このような関係を抽出するために、購買履歴からグラフを構築し、各連結成分におけるトライアドパターンの分布に基づいて分析する手法を提案している [2]。その結果、アイテムの購買順序が明確に決まっているタイプや、2 商品どうしの関係が長く連鎖するタイプ、中心的な商品と複数の関連商品からなるタイプが特に典型的な関係として抽出できることを明らかにしている。さらに、トライアドパターンの推移（以下、推移パターン）に着目し、アイテム間の関係の時系列変化をとらえる手法を提案している [6]。

本論文では、トライアドの推移パターンに加え、各トライアドパターンの構造をどの程度維持したか、を示す滞留度を用いた分析手法を提案する。具体的には、ユーザによって連続購買されたアイテム間に有向エッジを張ることで重み付きグラフを構築する。このとき、購買時刻順にエッジを張ることで、動的な成長グラフとする。この成長グラフに対して、14 種類のトライアドパターンの推移傾向および各構造における滞留度を算出する。これにより、商品カテゴリや推移パターンによる推移傾向や滞留度の違いを明らかにできると考えられる。成長傾向の差異を明らかにすることで、発売初期のタイミングや定番商品として定着したタイミングなど、各成長段階に合わせた商品の推薦などに応用できると期待している。

本論文の構成は次のとおりである。2 章でグラフ構造を用いた分析手法や推薦手法に関する関連研究について説明する。3 章でトライアド推移を用いた本研究の提案手法について詳述する。4 章で実データを用いた実験評価を行い、5 章でそれらの結果について考察する。最後に 6 章で本論文の成果について述べる。

2. 関連研究

本研究において核となるのは、アイテムからなるグラフのモチーフ（トライアド）分析である。本章では、EC サ

イトでのアイテムの購買パターンに関する研究、グラフのベクトル表現に関する研究、動的グラフの成長に関する研究、トライアドの構造変化に関する研究に焦点を当て説明する。

2.1 EC サイトでのアイテム購買パターン抽出

Srivastava 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 におけるアイテムの共購買関係を無向グラフとして分析している [15]。Srivastava は、グラフ全体に対するモチーフ分析を行ったのち、グラフ全体をコミュニティに分割、各コミュニティをノードとしたグラフを構築し、同様にモチーフ分析を行った。これにより、マクロレベルとミクロレベルの両方の視点から分析が可能になることを示した。評価実験の結果、アイテムの需要予測、アイテムカテゴリにおける購買傾向、アイテム間の関係を明らかにできることを確認した。本研究では、購買順序を有向グラフで表現している点、また、購買が行われることによるグラフの成長における傾向を抽出しようとしている点で異なる。

2.2 グラフのベクトル表現

グラフ構造を低次元のベクトルで表現する研究は多く存在する。大きく分けると、グラフ内のサブグラフ構造の出現頻度により表現する手法 [3], [12], [14]、グラフスペクトルなどの固有ベクトルによる手法 [7]、グラフ間の距離からベクトルを学習する手法 [11] に大別される。本研究の提案手法は、グラフ内のサブグラフ構造の出現頻度による手法に分類される。中でも最小単位であるトライアドに着目した研究として、Salihoglu は無向グラフを 3 ノードおよび 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 [12]。

2.3 トライアドの構造変化分析

実世界の動的グラフに対して分析した研究として、伏見らの研究がある [18]。文献 [18] では、Cookpad におけるユーザ間のフォロー数関係、Enron 社における社員間の Email 送受信関係、Twitter における Reply と Retweet 関係の有向グラフに対して、トライアドの推移パターンを抽出し、コミュニケーションの性質上、「片方向リンクから相互リンク」、「片方向リンクのまま」というパターン推移が有意に見られた点で、ソーシャルメディアによって構造が異なっていることを確認した。本研究と同様、重み付き有向グラフを対象としているが、推移パターンの滞留度に着目する点で大きく異なる。また、コミュニケーションの構造変化を分析する研究として、中田らの文献 [17] がある。トライアドリンク構造の状態遷移図を用いて、友人関

係ネットワークの成長過程を分析しているが、重みなし有向グラフを対象としている点で異なる。Tuomo は、トライアドの推移に着目し、対象とするパターンが多く出現するような人工グラフの生成モデルを提案している [8]。

3. 提案手法

本論文の仮説を検証するとともに、典型的な購買パターンを検出するための提案手法について説明する。提案手法では、全ユーザの購買履歴の集合が与えられたとき、アイテムをノードとし、購買順序に基づきアイテム間に有向エッジを付与することで購買履歴グラフ (PHG: Purchase History Graph) を構築する。全エッジを付与した PHG に対して、確率的ブロックモデルによりいくつかのコミュニティに分割することで、購買関係が密なアイテム群を抽出する。各コミュニティから、連結する 3 ノードのリンクパターンを表すトライアドパターンを抽出する。購買順序に基づきエッジが付与される PHG を動的グラフとしてとらえると、抽出したトライアドパターンは他のパターンからエッジが追加されて変化したものと考えることができる。すなわち、各トライアドパターンを状態とした際の状態推移と見なすことができるため、各トライアドの推移系列と推移系列における各状態の滞留度を算出する。滞留度を要素としたベクトルをクラスタリングし、類似の傾向を持つ系列に分類する。各コミュニティの滞留度ベクトルが、クラスタに偏在しているかを調べることで、典型的なパターンを抽出する。提案手法の流れをまとめると以下のとおりである。

- (1) 購買履歴グラフ構築；
 - (2) コミュニティ分割；
 - (3) 各コミュニティに対して、トライアド推移系列の数え上げ；
 - (4) 滞留度ベクトルトライアド推移系列のクラスタリング；
 - (5) クラスタ偏在度の定量化；
- 各ステップの詳細を、次節以降で説明する。

3.1 購買履歴グラフ構築

ユーザ集合を U 、アイテム集合を I と定義する。ユーザ $u \in U$ が時刻 t にアイテム i を購入した場合、その購買行動を $r = (u, i, t)$ と表す。ここで、ユーザ u が時刻 t までに購入したアイテムの時刻順のリストを $I(u)^{(t)} = [i_1, i_2, \dots]$ とし、 $I(u)^{(t)}$ において連続するアイテムをペアにしたリストを

$$SI(u)^{(t)} = [(i_1, i_2), (i_2, i_3), \dots] \subset I(u)^{(t)} \times I(u)^{(t)}$$

とする。

いま、すべてのユーザの連続購買アイテムペアリスト $SI(u)^{(t)}$ から、PHG $G^{(t)} = (V, E^{(t)})$ を構築することを考える。すなわち、PHG のノードはアイテムであるからノー

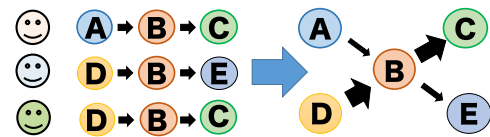


図 2 PHG の構築手法

Fig. 2 Construction of PHG.

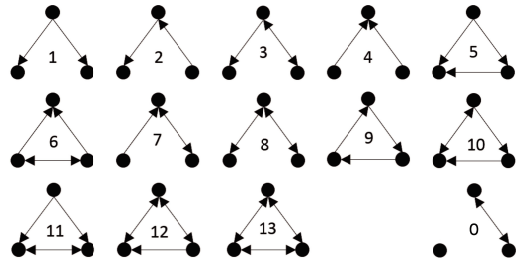


図 3 トライアドパターン

Fig. 3 13 triad patterns for three nodes.

ド集合は $V = I$ であり、エッジ集合は以下のように定義する：

$$E^{(t)} = \left\{ (i, j) \in \bigcup_{u \in U} SI(u)^{(t)} \right\}.$$

一般的に、各アイテムは異なる時刻に複数のユーザにより購買されることがあるため、アイテムペア (i, j) の出現回数をエッジ重み $w(i, j)$ として定義する。図 2 に、PHG 構築の模式図を示す。図 2 では、3 人のユーザが A, B, C, D, E の 5 種のアイテムを購入した場合に構築される PHG の例を示している。アイテム (D, B) のペアは、複数のユーザに $D \rightarrow B$ の順序で購買されたため、PHG におけるエッジの重み (太さ) が大きくなっている。この $G^{(t)} = (V, E^{(t)})$ で表される有向グラフを時刻 t までの動的 PHG と呼ぶことにする。当然、時刻 t がすすむにつれユーザの購買履歴は蓄積され、新たなアイテムペア間にエッジが追加され、または、すでにあるエッジの重みが大きくなり、PHG は成長していく。

次に、購買履歴データにおける最終時刻を T としたとき、最終時刻 T における PHG $G^{(T)}$ を確率的ブロックモデル [10] により複数のサブグラフに分割する。なお、分割したサブグラフをコミュニティと呼ぶ。

3.2 トライアド推移と滞留度

有向グラフには図 3 に示すように、連結 3 ノード (トライアド) 間のリンクパターンが 13 種ある [9]。これにパターン 0 を加えた 14 種に関してみていく (以下、各パターンを TP0, TP1, ..., TP13 というように表記する)。上述したように、時刻 t がすすむにつれユーザによるアイテム購買履歴が蓄積され、PHG の構造も動的に変化する。すなわち、PHG に含まれる各トライアドパターンの出現頻度も変化する。たとえば、TP1 の構造にエッジが 1 本追加

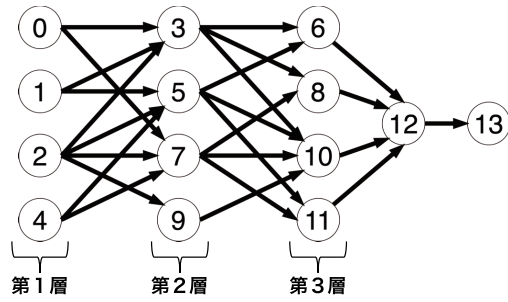


図4 トライアド推移パターン

Fig. 4 Triad transition pattern.

された場合、そのトライアドの状態は TP3 か TP5 に推移する。あるいは、重複エッジが追加される場合は、TP1 の状態にとどまる。本研究では、PHG が成長していく過程における 14 種のトライアドパターン間の推移を観察することで、ユーザによる商品購買の傾向を分析する。

図 4 に、トライアドパターン間の可能な推移パターンを示す。図 4 において、第 1 層の TP0, 1, 2, 4 は 2 種のエッジからなるパターンである。第 2 層の TP3, 5, 7, 9 は、第 1 層のトライアドに 1 本エッジが追加された場合に推移しうるパターンである。提案手法では、確率的ブロックモデルにより分割した各コミュニティに対して、時刻 t がすすむにつれて変化するトライアドの推移パターンを数え上げる。たとえば、ある時刻 t において m 番目のコミュニティ $G_m^{(t)}$ で観測された TP1 が、時刻 $t+1$ における $G_m^{(t+1)}$ では TP3 に推移したことを数え上げる。著者らの既存研究 [2] から、TP8 が高頻度で出現し、重要な役割を果たすといえる。したがって、TP8 を含む第 3 層の TP6, 8, 10, 11 への全 28 推移パターンを対象とする。

上述したように、重複エッジが追加される場合はトライアドパターン間の推移が起こらず、そのトライアドパターンに状態がとどまる。本論文では、トライアドパターン間の状態推移が起きずに同一の状態に滞留する期間（追加エッジ数でカウント）をトライアドパターンの滞留数とする。たとえば、コミュニティ m において TP a が観測され、エッジ追加によりグラフが成長する過程で他の TP に推移せず TP a の状態にある程度滞留したとする。その後 TP b に推移し、またある程度滞留した後 TP c に推移したとする。このとき、コミュニティ m 内で観測された推移パターン $a \rightarrow b \rightarrow c$ における第 l 層の滞留数を $x_m^{(l)}(a \rightarrow b \rightarrow c)$ とし、全 3 層の滞留数の和を $x_m(a \rightarrow b \rightarrow c) = \sum_{l=1}^3 x_m^{(l)}(a \rightarrow b \rightarrow c)$ と表す。そして、全層の和に対する割合を $y_m^{(l)}(a \rightarrow b \rightarrow c) = x_m^{(l)}(a \rightarrow b \rightarrow c) / x_m(a \rightarrow b \rightarrow c)$ とし、第 l 層の滞留度と呼ぶ。各層の滞留度を要素とする 3 次元ベクトル $\mathbf{y}_m(a \rightarrow b \rightarrow c)$ をコミュニティ m の推移パターン $a \rightarrow b \rightarrow c$ の滞留度ベクトルとして定義する。1 つのコミュニティにおいて、最大で 28 推移パターン（図 4 参照）が起こりうるため、コミュニ

ティごとに 28 種の 3 次元滞留度ベクトルが得られる。

3.3 滞留度ベクトルのクラスタリングとクラスタ偏在度

全 M コミュニティの 28 種の滞留度ベクトルそれぞれに対して、 k -means 法により K 個のクラスタに分類する。ベクトル間のメトリックには、ユークリッド距離を用いる。滞留度を要素とするベクトルをクラスタリングすることにより、滞留度とコミュニティの関係や滞留度と 3 層目のトライアドパターンの関係、あるいは、第 1 層のトライアドパターンとの関係に傾向があるかを分析する。

次に、各クラスタにおける滞留度ベクトルの偏在度を定量化する。 K 個のクラスタに分類した際、あるクラスタには推移パターン $a \rightarrow b \rightarrow c$ が、別のクラスタには d から始まる推移パターン $d \rightarrow X \rightarrow X$ が、また別のクラスタには e で終わる推移パターン $X \rightarrow X \rightarrow e$ が偏って分類されているかなどを定量化する。いま、クラスタ C_k に M 個中 n 個のコミュニティの推移パターン $h: a \rightarrow b \rightarrow c$ が割り振られた場合、パターン h のクラスタ C_k での出現頻度を $c_{h,k} = n$ とする。推移パターン h の出現確率を $p_h = \sum_{k=1}^K c_{h,k} / L$ 、クラスタ C_k の割り当て確率を $q_k = \sum_{h=1}^{28} c_{h,k} / L$ とする。ここで、 $L = \sum_{k=1}^K \sum_{h=1}^{28} c_{h,k}$ は全推移パターンの総出現回数を意味する。クラスタと推移パターンの間に偏りがなく、各推移パターンが一様ランダムにクラスタに割り振られると仮定すると、周辺確率 p_h と q_k を用いて期待値 $L \cdot e_{h,k} = L \cdot p_h q_k$ を計算できる。実際の出現頻度と期待値との差、および、標準偏差により以下の Z スコアを計算する：

$$z_{h,k} = \frac{c_{h,k} - L \cdot e}{\sqrt{L \cdot e_{h,k}(1 - e_{h,k})}}. \quad (1)$$

Z スコア $z_{h,k}$ はランダムな場合と比較した際の偏り具合を表し、値が正で大きいほどクラスタ k に推移パターン h が統計的に多く存在するといえる。Z スコアの値が大きい推移パターンに着目し、典型的な推移パターンを抽出する。また、クラスタをコミュニティに置き換えることで、各コミュニティにおける推移パターンの偏在度を算出する。

4. 評価実験

4.1 データセット

本研究では、レビューの投稿順序を商品の購買順序と見なし、楽天市場のレビューデータ^{*1}を使用した。利用者の購買順序を求めるために、6,500 万レビューから、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。

^{*1} <http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html>

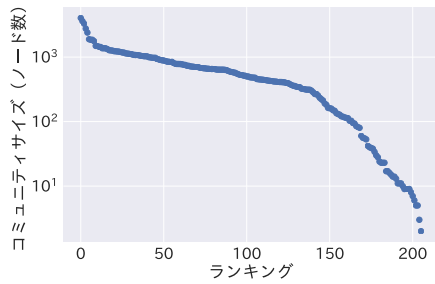


図 5 コミュニティサイズ分布

Fig. 5 Distribution of community size.

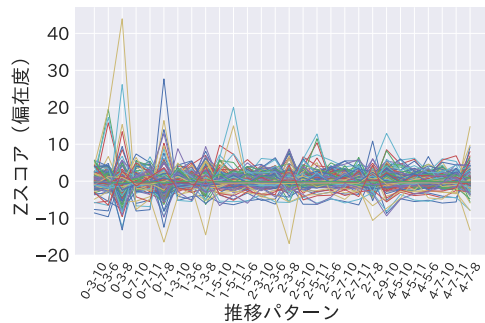


図 6 コミュニティにおける推移パターン偏在度

Fig. 6 Z-score of transition pattern by community.

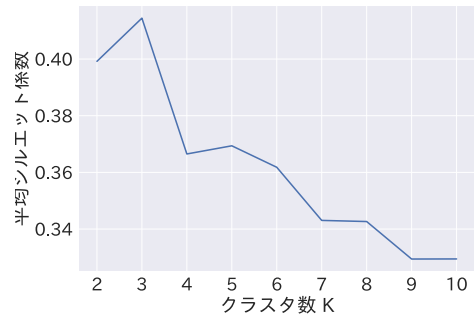
4.2 購買履歴グラフの構築

4.1 節で抽出した全データの、商品をノード、複数のユーザから連続して購買された商品ペア間にエッジを付与し、PHG $G^{(T)}$ を構築した。続いて、確率的ブロックモデルにより複数のコミュニティ（サブグラフ）に分割した。以上の処理を行い、273,994 ノード、459,110 エッジ、207 コミュニティからなる PHG を実験対象とした。

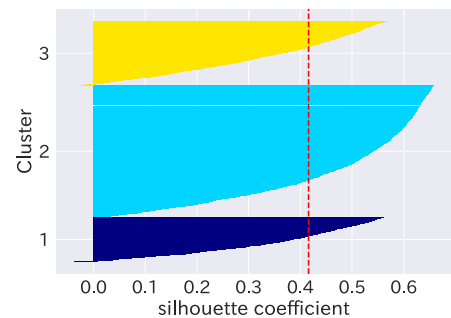
コミュニティのサイズ（ノード数）分布を図 5 に示す。縦軸はコミュニティサイズ（ノード数）、プロット点はコミュニティに対応する。これを見ると、大半のコミュニティがおおよそ数百ノードから一千ノード程度で構成されていることが分かる。

4.3 コミュニティにおける推移パターン偏在度

207 コミュニティそれぞれについて、3.2 節で説明した 28 種の推移パターンがどのような分布で出現するか、Z スコアを用いて算出した。図 6 にコミュニティ別の推移パターンの偏在度を示す。横軸が各推移パターン、縦軸が偏在度を表す Z スコア、各プロット線がコミュニティに対応している。一般に Z スコアの絶対値が 2 を超えると、全体の分布に対して出現数が偏っているといえる。実験では、207 コミュニティ中 194 コミュニティにおいて、いずれかの推移パターンの Z スコア絶対値が 2 を超えていることが確認された。このことから、多くのコミュニティにおいて推移パターンが偏在している、すなわち、成長傾向に違いがあるといえる。



(a) 平均シルエット係数



(b) シルエット図

図 7 シルエット分析

Fig. 7 Silhouette analysis.

4.4 滞留度ベクトルのクラスタリング

207 コミュニティそれぞれに対して、3.2 節で定義した 28 種の滞留度ベクトルを取得した。207 コミュニティ $\times$ 28 種 = 5,796 個のベクトルが取得できる。実際には、コミュニティによって出現しない推移パターンがあったため、実験に用いる滞留度ベクトルは 5,052 個である。得られた 3 次元の滞留度ベクトルについて、 k -means クラスタリングを行った。クラスタ数 K は、 $K=2$ から $K=10$ までのシルエット分析を行い決定した。図 7(a) は、横軸をクラスタ数 K 、縦軸を平均シルエット係数とした、平均シルエット係数の推移である。この図で、 $K=3$ のときに最も高い値 0.414 を示したことから、 $K=3$ 時のシルエット図を図 7(b) に示す。クラスタの大きさに偏りはあるものの、クラスタ間でのシルエット係数の差は小さい。これらの結果からクラスタ数は $K=3$ とした。

クラスタ別の滞留度ベクトルを図 8 に示す。横軸は各層、縦軸は各層における滞留度、プロット線は各滞留度ベクトルを表す。また、プロット線の色は推移パターンの種別に対応しているが、誌面の都合上、凡例は割愛する。これらを観察すると、いずれのクラスタも 3 層目の滞留度が最も大きい*2。クラスタ間の違いは 1, 2 層目にあることが読み取れる。

*2 第 3 層の滞留度と全層の滞留数の和の関係を明らかにするために、両者の相関係数を求めたところ -0.008 であり、相関は見られなかった。

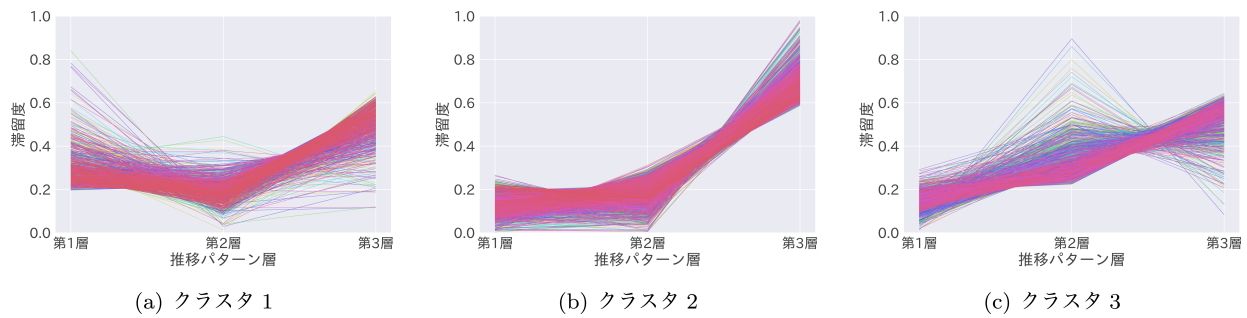


図 8 滞留度ベクトル
Fig. 8 Stagnate vectors.

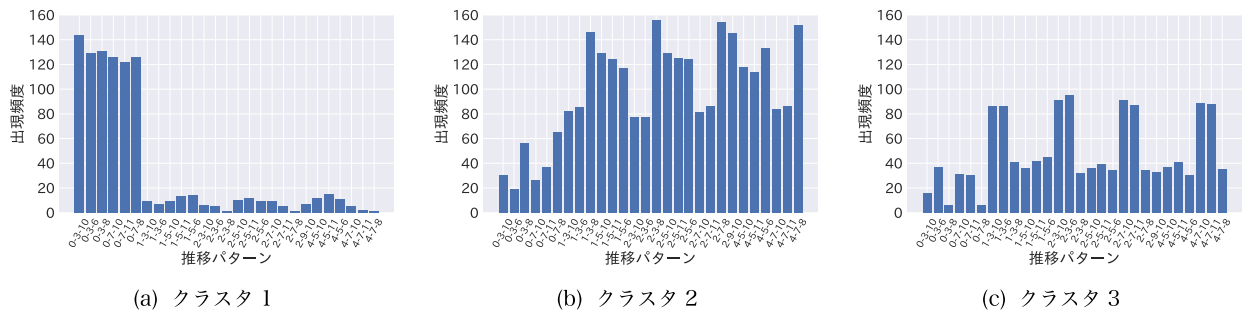


図 9 推移パターンの出現頻度
Fig. 9 Appearance frequency of transition patterns.

クラスタ 1 は比較的 1 層目の滞留度が大きく、初期構造に長く滞留する傾向にあることが分かる。一方、クラスタ 2 は 1 層目、2 層目にはほとんど滞留せず、すぐに複雑な構造へと推移することが分かる。また、クラスタ 3 は一定のペースで各層の滞留時間が長くなっている。

滞留度ベクトルに加えて、各クラスタにおける推移パターンの出現頻度分布を図 9 に示す。横軸は各推移パターン、縦軸は推移パターンの出現数を示す。また、図 10 にクラスタ別に、起点となる TP および到達する TP ごとの推移系列数を示す。分布の違いを明確にするため、合計が 1 になるように正規化した。横軸は各クラスタ、縦軸は推移系列比を示す。図 10(a) から、クラスタ 1 は、ほぼすべて TP0 から始まる推移パターンであることが読み取れる。一方、クラスタ 2, 3 に TP0 を起点とする推移パターンはごく僅かしか含まれていない。このことから、TP0 を起点とする推移パターンは、第 1 層 (TP0) の滞留度が大きい推移パターンであるといえる。また図 9(b), (c)からは、クラスタ 2, 3 の推移パターン出現頻度の高低がちょうど逆であることが読み取れ、これは補完的な関係であるといえる。そのなかでも、図 10(b) のクラスタ 2 に着目すると、TP8 に到達する推移パターンの出現頻度が小さいことが分かる。このことから、TP8 に到達する推移パターンは、第 1 層、第 2 層の滞留時間が非常に短い傾向にあるといえる。

4.5 クラスタにおける推移パターンの偏在度

4.4 節で得られた 3 つのクラスタそれぞれにおける推移

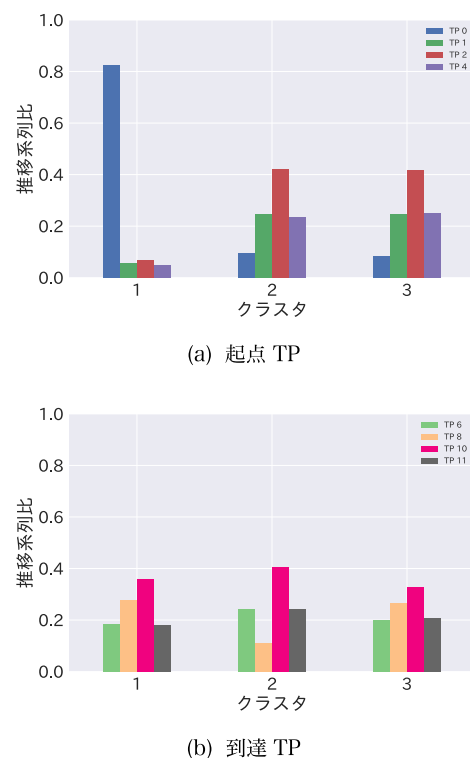


図 10 クラスタ別 推移系列比

Fig. 10 Appearance probability of transition sequence by cluster.

パターンの偏在度を図 11 に示す。横軸が推移パターン、縦軸が偏在度を表す Z スコア、各プロット線がクラスタに対応する。Z スコアが高ければ高いほど、全体の分布に対し、そのクラスタに多く出現 (偏在) することを意味する。

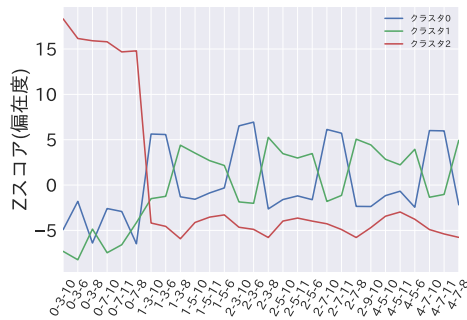


図 11 クラスタにおける推移パターン偏在度

Fig. 11 Z-score of transition pattern by cluster.

一般に Z スコアが 2 を超えると有意に多く出現しているといえる。これを見ると、6 つの推移パターンがクラスタ 1 に多く出現していることが分かる。これは図 10(a) で示した TP0 を起点とする推移パターンである。また、クラスタ 3 に偏って出現している 8 つの推移パターンは図 10(b) で示した TP8 に到達しない推移パターンであると考えられる。このように、異なる傾向の滞留度ベクトルからなるクラスタにおいて、頻出する推移パターンが有意に異なることから、推移パターンによって滞留度には一定の傾向があることが定量的に示された。

5. 考察

5.1 実験結果の考察

4.3 節では、出現する推移パターンはコミュニティごとに偏っていることが明らかになった。また、4.4 節では、滞留度の違いは推移パターンに依存していることが明らかになった。以上の結果から、購買行動を構造的観点からとらえる推移パターンには、コミュニティに依存した偏りがあることが確認された。一方、時間的観点からとらえる滞留度には、推移パターンに依存した偏りがあることが確認された。これらの結果から、PHG の成長過程は、構造的観点および時間的観点のいずれにおいても、コミュニティによって大きな違いがあるといえる。

図 12 にコミュニティにおける推移パターンの偏在度に顕著な傾向が見られたものを示す。横軸が各推移パターン、縦軸が偏在度を表す Z スコア、各プロット線がコミュニティに対応している。また、コミュニティに属するノード（アイテム）の属性からカテゴリを調査したところ、コミュニティ 23 および 110 は主に食品カテゴリからなるコミュニティであった。これらのコミュニティに出現する推移パターンは、TP6, 10, 11 を第 3 層とするが、途中経路となる第 1, 2 層には大きな差異があった。たとえば、コミュニティ 23 で最も高い偏在度を示す推移パターン 2-9-10 は、コミュニティ 110 においてマイナスの偏在度を示している。同じ食品カテゴリを主な構成要素とする 2 つのコミュニティが異なる偏在度を示した要因として、食品は価格帯が類似していること、また食品間の組合せも多様であるこ

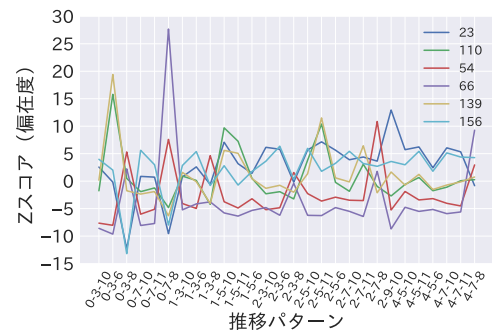


図 12 コミュニティにおける推移パターン偏在度（一部抜粋）

Fig. 12 Z-score of transition pattern by community (partially excerpted).

とが、多様な購買行動の遷移を引き起こしたことが考えられる。

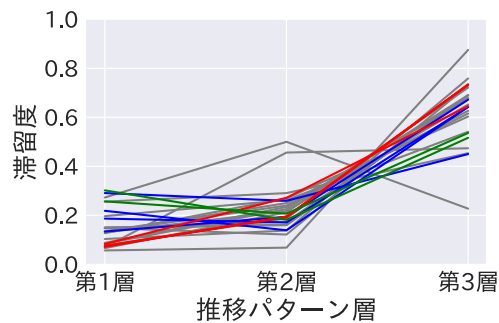
コミュニティ 54 と 66 はそれぞれ女性ファッションカテゴリと男性ファッションカテゴリからなるコミュニティである。図 12 で両コミュニティの推移パターンの偏在度を観察すると、多くの推移パターンで類似した偏在度が観察されるが、コミュニティ 66（男性ファッション）においては、推移パターン 0-7-8 が極端に大きな値を示している。一方で、コミュニティ 54（女性ファッション）では、TP8 に至る 5 種の推移パターンが偏在度 5 近くを示す傾向にある。類似するファッションカテゴリでも、女性ファッションカテゴリに比べ、男性ファッションカテゴリのコミュニティには特徴的な成長傾向（0-7-8）があるといえる。

コミュニティ 139 および 156 はインテリアカテゴリからなるコミュニティである。図 12 で推移パターンの偏在度を見ると、コミュニティ 139 は推移パターン 0-3-6 の偏在度が高いことが分かる。一方、コミュニティ 156 では極端に高い偏在度を示す推移パターンはなく、コミュニティ 139 と 156 は大きく異なる推移パターンからなるといえる。

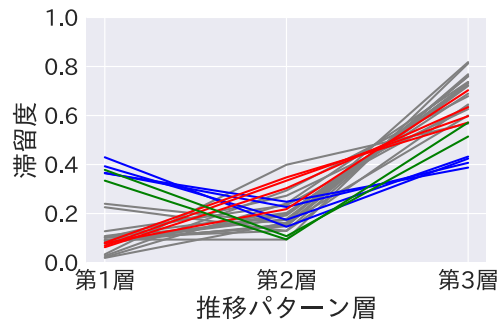
ここで、コミュニティの滞留度ベクトルに着目する。4.4 節で説明したように、滞留度は推移パターンの種別に大きく左右される。図 13 にコミュニティ 139 および 156 それぞれの滞留度ベクトルを示す。横軸が各層、縦軸が滞留度、プロット線の色は、TP0 起点かつ TP8 終点の推移パターンが緑、TP0 起点の推移パターンが青、TP8 終点の推移パターンが赤、その他の推移パターンが灰色に対応している。これを見ると、コミュニティ 139 に比べ、コミュニティ 156 の第 1 層への滞留度が大きい推移系列が多いことが分かる。これら 3 つのケースから、同じカテゴリのアイテムからなるコミュニティであっても、その成長傾向が異なる場合があることが示された。

5.2 グラフ構造の可視化

コミュニティ 139 および 159 について、グラフの可視化を行った。一部抜粋した結果を図 14 に示す。ノードの大きさは回数に対応している。コミュニティ 139 の可視化結



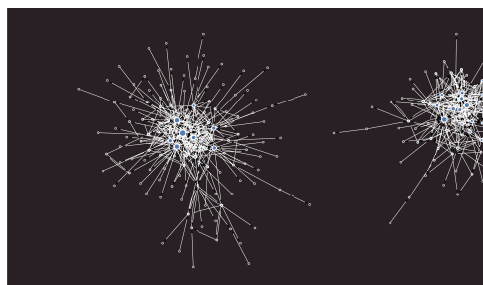
(a) コミュニティ 139



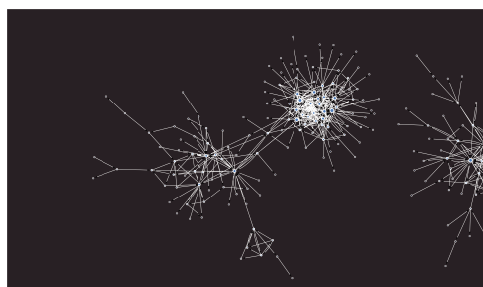
(b) コミュニティ 156

図 13 滞留度ベクトル

Fig. 13 Stagnate Vectors.



(a) コミュニティ 139



(b) コミュニティ 156

図 14 グラフ構造の可視化結果 (一部抜粋)

Fig. 14 Visualization of network (partially excerpted).

果からは、1章で述べた図1(a)のように、1つのアイテムを中心として複数のアイテムがリンクする構造が観察される。これは1つのメインアイテムに対し、ユーザの嗜好に合わせた複数のオプションアイテムが存在する構造である。このため、複数のユーザがメインアイテムを購入する

際に、個々人の嗜好に合ったオプションアイテムを同時購入する機会が多いと考えられる。そのため早い段階で複数のノードとリンクする、すなわちTPが推移するので、第1層への滞留度が小さいと考えられる。

一方、コミュニティ156の可視化結果からは、図1(b)のように、長いパスが複数観察される。これは、アイテムどうしの購買関係が明確に決まっている場合に見られる構造であり、多くのユーザが特定の組合せや特定の順序でアイテムを購入する。この場合には、アイテム間のリンクが限定的となるため、新たなノードとリンクすることによるTP推移が発生しにくく、結果として第1層の滞留度が大きくなったと考えられる。

これを推薦に応用すると、前者のようなアイテムは、成長初期の段階から推薦するアイテムをなるべく多様にすることで、購買実績を重ね、さらなる購買を誘引できると考えられる。逆に後者のようなアイテムは、推薦するアイテムを限定的にし、すでに購買実績のあるアイテムを中心とした推薦が効果的だと考えられる。特に、長いチェーンを構成するようなアイテム群については、チェーンの開始アイテムを購入したユーザに対する「まとめ買い」推薦なども有効であろう。チェーンを構成するアイテムは、ユーザの嗜好に依らないアイテムであり、購買時の満足度も高いと期待できる。「まとめ買い」により購買実績が伸びることで、前者同様さらなる購買を誘引できると考えられる。

以上のことから、アイテム推薦の場面において、アイテムの属性や共購買情報に加え、グラフの成長傾向を考慮することで、推薦するアイテムのより適切な選択が可能になるといえる。

5.3 今後の課題

今後の課題としては以下の2点があげられる。まず、滞留度ベクトルの算出手法について、本論文では、トライアドの構造を変化させるエッジが1本でも追加された場合、トライアドが次層に推移したと定義して、各層の滞留度を算出している。3つのノード a, b, c からなるトライアド $triad_{abc}$ を考える。今、 $triad_{abc}$ には $a \rightarrow b$ と $a \rightarrow c$ の2種のエッジが付与されており、トライアド構造は第1層のTP1である。ここで、新規に $b \rightarrow c$ のエッジが1本付与されたとすると、構造は第2層のTP5へと推移する。さらに、 $a \rightarrow c$ のエッジが続けて5回付与された場合、この5本のエッジは第2層TP5の滞留度として計上される。しかし、このケースではTP1に5回滞留したと計上する方がより実際に即した結果になると考えられる。このことから、滞留度ベクトルの算出において、トライアドが推移を判定する閾値を動的に設定する必要があるといえる。

次に、グラフの成長評価の指標について、本論文では、推移パターンや滞留度に着目してその成長度合いを分析している。これに加え、コミュニティごとのクラスタ係数

やノード数など既存の評価指標を動的に算出することで、PHGの成長傾向をより明確にとらえられると考えられる。

6. おわりに

本論文では、トライアドの推移パターンおよび滞留度に着目した購買行動の成長分析手法を提案している。提案手法では、動的に購買履歴グラフを生成し、推移パターンの偏在度の算出や滞留度ベクトルのクラスタリングを行っている。動的に成長する購買履歴データに適用し、類似するアイテムカテゴリでも異なる成長傾向を示すことを確認した。

謝辞 本研究はJSPS科研費JP16H02904の助成を受けたものです。本研究の実装・評価に際し、楽天株式会社と国立情報学研究所が提供する「楽天データ」を利用しました。

参考文献

- [1] Hayashi, A., Kohjima, M., Matsubayashi, T. and Sawada, H.: Regularity measure and influence weight for analysis and visualization of consumer's attitude, *2015 19th International Conference on Information Visualization*, pp.290–299 (July 2015).
- [2] Inafuku, K., Fushimi, T. and Satoh, T.: Extraction method of typical purchase patterns based on motif analysis of directed graphs, *Proc. 18th International Conference on Information Integration and Web-based Applications and Services, iiWAS '16*, pp.86–95, ACM (2016).
- [3] Inokuchi, A., Washio, T. and Motoda, H.: An apriori-based algorithm for mining frequent substructures from graph data, *Proc. 4th European Conference on Principles of Data Mining and Knowledge Discovery, PKDD '00*, pp.13–23, Springer-Verlag (2000).
- [4] Isinkaye, F., Folajimi, Y. and Ojokoh, B.: Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, Vol.16, No.3, pp.261–273 (2015).
- [5] Iwata, T., Watanabe, S., Yamada, T. and Ueda, N.: Topic tracking model for analyzing consumer purchase behavior, *Proc. 21st International Joint Conference on Artificial Intelligence, IJCAI'09*, pp.1427–1432, Morgan Kaufmann Publishers Inc. (2009).
- [6] Inafuku, K., Fushimi, T. and Satoh, T.: Growth analysis of purchasing behavior in online shopping, *Multimedia, Distributed, Cooperative, and Mobile Symposium*, 5F-5 (2017) (in Japanese).
- [7] Luo, B., Wilson, R.C. and Hancock, E.R.: Spectral embedding of graphs, *Pattern Recognition*, Vol.36, No.10, pp.2213–2230 (2003).
- [8] Mäki-Marttunen, T.: An algorithm for motif-based network design, *IEEE/ACM Trans. Computational Biology and Bioinformatics*, Vol.14, No.5, pp.1181–1186 (2017).
- [9] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network motifs: Simple building blocks of complex networks, *Science*, Vol.298, No.5594, pp.824–827 (Oct. 2002).
- [10] Nowicki, K. and Snijders, T.A.B.: Estimation and prediction for stochastic blockstructures, *Journal of the American Statistical Association*, Vol.96, No.455, pp.1077–1087 (2001).
- [11] Riesen, K. and Bunke, H.: *Graph Classification and Clustering Based on Vector Space Embedding*, World Scientific Publishing Co., Inc. (2010).
- [12] Salihoglu, S.: Networks as vectors of their motif frequencies and 2-norm distance as a measure of similarity, Technical Report, Stanford University (2006).
- [13] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J.: Item-based collaborative filtering recommendation algorithms, *Proc. 10th International Conference on World Wide Web, WWW '01*, pp.285–295, ACM (2001).
- [14] Sidere, N., Heroux, P. and Ramel, J.-Y.: A vectorial representation for the indexation of structural informations, da Vitoria Lobo, N., Kasparis, T., Roli, F., Kwok, J.T., Georgiopoulos, M., Anagnostopoulos, G.C. and Loog, M. (Eds.), *Structural, Syntactic, and Statistical Pattern Recognition*, pp.45–54, Springer Berlin Heidelberg (2008).
- [15] Srivastava, A.: Motif analysis in the amazon product co-purchasing network, *Computing Research Repository*, abs/1012.4050 (2010).
- [16] Symeonidis, P., Tiakas, E. and Manolopoulos, Y.: Product recommendation and rating prediction based on multi-modal social networks, *Proc. 5th ACM Conference on Recommender Systems, RecSys '11*, pp.61–68, ACM (2011).
- [17] 中田豊久, 加藤義彦, 國藤 進: 友人ネットワークの状態遷移図による分析, 情報処理学会論文誌数理モデル化と応用 (TOM), Vol.2, No.1, pp.87–97 (2009).
- [18] 伏見卓恭, 佐藤哲司: レシピ投稿サイトにおけるユーザ間コミュニケーションのネットワーク分析, 電子情報通信学会技術研究報告, Vol.115, No.230, pp.71–76 (Sep. 2015).

推薦文

DICOMO2017の発表論文の中で特に評価が高い論文であった。購買行動の成長を見るというアプローチと、レビューデータを購買行動と見なす手法は、ともに興味深く新規性が高いため推薦する。

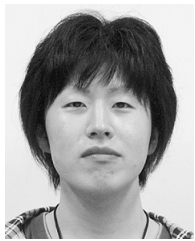
(グループウェアとネットワークサービス研究会主査

齊藤典明)



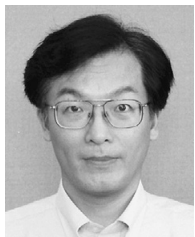
稲福 和史

2018年筑波大学知識情報・図書館学類卒業。現在、同大学大学院図書館情報メディア研究科博士前期課程在学中。複雑ネットワーク、購買行動分析に関する研究に従事。



伏見 卓恭 (正会員)

2011 年静岡県立大学大学院経営情報学研究科修士課程修了。2014 年静岡県立大学大学院経営情報イノベーション研究科博士後期課程修了。同年静岡県立大学大学院経営情報学部客員研究員。2015 年筑波大学図書館情報メディア系特別研究員 (PD)。2017 年より東京工科大学コンピュータサイエンス学部助教。複雑ネットワーク、可視化の研究に従事。博士 (学術)。人工知能学会、日本データベース学会各会員。



佐藤 哲司 (正会員)

1980 年山梨大学工学部電子工学科卒業。同年日本電信電話公社 (現, NTT) 武蔵野電気通信研究所に入所。1994 年工学博士 (大阪大学)。2007 年より筑波大学図書館情報メディア系教授。情報検索・知識発見, 社会ネットワーク分析, 社会インタラクションに興味を持つ。電子情報通信学会, 日本データベース学会各会員。