

## ACOに基づく階層型時系列パターンマイニング法の提案

The proposal of the layered sequential pattern mining method based on ACO algorithm

坪井 一晃 篠田 孝祐 諏訪 博彦 栗原 聡  
 Kazuaki Tsuboi Kosuke Shinoda Hirohiko Suwa Satoshi Kurihara

電気通信大学大学院情報システム学研究科

Graduate School of Information System, The University of Electro-Communications

Consumers are various. When considering marketing, we need to get to know consumers' action. Then we can consider the strategy. In order to understand consumers, it is necessary to get to know the pattern which occurs frequently from action at consumers' retail store. In old pattern mining, correspondence was not completed for the needs of the changing consumers. And computation time is very huge. Then, we propose the layered sequential pattern mining method based on ACO algorithm. We take in the adaptability and dynamic stability which are the features of ACO algorithm to pattern mining.

## 1. はじめに

我々は、大量の消費者行動データ（店頭アクセスデータ）から消費者インサイトを抽出・可視化する技術を開発することを目指している。マーケティングを考える上で、消費者の行動や心理といった消費者インサイトを理解することは重要である。多様な消費者ニーズを正しく理解し、的を絞った商品やサービスが求められている [1].

従来、コンビニやスーパーマーケットなどにおけるマーケティングは、POS データのような消費者が買った商品情報をデータ化し、解析することによってマーケティング戦略を考えている。しかし、情報技術が発展するに伴い、動画情報もデータとして利用できるようになりつつある。現実の店舗において、動画情報を用いることで購買結果のみではなく、購買に至るまでの消費者の一連の行動をデータとして活用できるようになっている。これらのデータを用いて、消費者の行動の背景にある心理や意識を抽出することは重要であり、実際に研究も行われている [2].

しかし、現実において、消費者というものは流行に左右されやすい傾向にあり、絶えず変化するという特徴をもつ。具体的なマーケットを考えてみても、季節や気候の変化によって消費者のニーズが変化することは容易に想像がつく。また、店舗でのマーケティングを考えたときに、商品を消費者に対して販売する小売店におけるマーケティング戦略を考えるのか、小売店に商品を卸すメーカーにおけるマーケティングを考えるのかといった、誰にとってもマーケティング戦略を考えるのかといった対象によっても変化が生じる。

そこで、本研究では大量の消費者行動データから時間変化する頻出するパターンを抽出することを考える。想定する消費者行動データは、実際の店舗において購入商品を選択している場面の動画データをもとに、どのような順序で商品を手に取りながら、どの商品を購入しているかというデータである。この消費者行動データを用いて、時間変化に対応した購買パターンを抽出することを課題とする。この課題に対して、本稿では、シミュレーションデータから、連続して複数の商品を手にするパターンを抽出するアルゴリズムを提案する。提案するアルゴ

リズムの性能を評価するために、シミュレーションデータでの解析を行う。

本論文の構成を次に示す。2章では、マーケティングにおけるパターンマイニングの研究例や、本研究での提案手法となるACO アルゴリズムの関連研究について述べる。3章では、提案手法となるACO に基づいたパターン抽出のアルゴリズムについて述べる。4章では、消費者が連続して手に入る商品のパターン抽出を、シミュレーションデータを用いた検証実験について述べる。5章では本研究の現在の進捗を踏まえたうえで、今後の研究に関する展望について述べる。

## 2. 関連研究

頻出するパターンを発見するためのアルゴリズムとしてアプリアリアルゴリズム [3] が有名である。これは、ある基準を満たすアイテムの組み合わせやアイテム間の関係を表すルールをすべて抽出するための手法である。実際に、マーケティングにアプリアリアルゴリズムを適用した研究がある [2]。他にもデータベース上から頻出パターンを発見するためのアルゴリズムとして時系列を考慮したパターンを発見する Apriori All [4] や、時系列パターンを発見するにあたって時間制約などの条件をつけることで計算処理の高速化を図った GSP アルゴリズム [5]、*lattice* という概念を用いて頻出パターンの候補をグループ分割することで計算コストの高速化を図った SPADE アルゴリズム [6] の提案が行われている。しかし、これらのアルゴリズムでは、全ての入力データを等しく参照しているため、消費者の変化するニーズに対応した頻出パターンの抽出は達成できない。

そこで、我々は、ACO アルゴリズムに着目する。アントコロニー最適化 (Ant Colony Optimization, ACO) アルゴリズム [8] とは自然界におけるアリの採餌行動をモデル化したもので、最適化問題である巡回セールスマン問題の解法として紹介された。ACO アルゴリズムでは、アリは通過した経路にフェロモンを残すことと、アリはフェロモンの濃度が濃い経路を好んで経路選択をするというアリの行動を前提として考えられる。この前提によって、アリが行動するごとに最短経路を通過するアリが多くなり、結果として最短経路に残るフェロモンの濃度が濃くなる。一方で、フェロモンは気体であり、時間経過に伴い蒸発する。結果として、アリが経過する頻度が少な

連絡先: 坪井一晃, 電気通信大学大学院情報システム学研究科, 〒182-8585 東京都調布市調布ヶ丘 1-5-1, 042-443-5664, tsuboi@uec.ac.jp

い、最短経路以外の経路に関してのフェロモン濃度が薄くなる。残ったフェロモンの濃度が最短経路問題の解となる。ACO アルゴリズムの特徴として、高い適応性および頑健性を有することが挙げられる。

玉置らはセンサが人の行動を読み取ることに着目し、センサの隣接関係を ACO アルゴリズムを用いて推定する研究を行っている [7]。ACO アルゴリズムを用いたことで、センサの故障や移動といった環境の変化や、空間内における複数人の同時行動、センサの誤認識によるノイズに対応できるセンサの隣接関係の推定を達成した。この研究ではセンサの隣接関係のみを推定しているが、この隣接するセンサが反応する一連の流れを追うことができると、人間の行動パターンとして抽出できると考えられる。

このように、もともと最適化技術として知られていた ACO アルゴリズムであるが、単純な行動ルールに基づくエージェントの移動と環境に対するフェロモンの付加・蒸発をアルゴリズムに应用することで最適化以外の分野でも活躍できる。そこで、本研究においても、ACO アルゴリズムが有する動的安定性や頑健性に着目し、ACO アルゴリズムの応用を採用する。

### 3. ACO によるパターン抽出手法

本研究では小売店における消費者の行動をパターンとして抽出することを考える。消費者が小売店において、購入する商品を検討する際、消費者が触った商品、商品に触った時間、次の商品を手に取るまでの時間といった一連のデータが、入力となる時系列データである。この時系列データを用いて、ACO アルゴリズムを応用した ACO 型パターン抽出アルゴリズムによって消費者の行動パターンを抽出する。

#### 3.1 時系列データ

提案アルゴリズムでは、小売店において、それぞれの消費者がハブラシやハミガキを手にとりながら購入する商品を考えている様子をデータ化したものを入力として扱う。本研究で使用するデータセットは消費者が手に取った商品と商品を手にとってから次の商品を手に取るまでの時間、および商品を棚に戻してから次の商品を手に取る時間間隔から成り立つ。各商品はアイテムセットとして番号で管理する。このデータセットの二列目が消費者が手に取った商品  $i$  を表し、三列目が商品を手にとってから次の商品を手に取るまでの時間幅  $t$  を表す。なお、現在の進捗では消費者が商品を手にとってから次の商品を手に取るまでの時間幅の値は活用していない。ただし、今後の研究では、ごく短時間しか見なかった商品については無視するなどの活用方法を考えている。

提案アルゴリズムを適用するにあたって、入力データセットは各消費者単位で切り分けられたデータを使用する。本研究では、次の商品を手に取るまでの時間幅に着目し、次の商品を手に取るまでの時間幅  $t$  がある閾値  $\phi$  を超えたとき、商品を選んでいる消費者が変わったものとして考える。つまり  $t > \phi$  となるごとに商品を選ぶ消費者が変化したものとして扱う。そうすることによって、図 1 のような各消費者が順に手に取った商品を切り出すことができる。なお、 $c$  は顧客を表す。

#### 3.2 ACO 型パターン抽出アルゴリズム

ACO アルゴリズムは環境へのフェロモンの付加とフェロモンの蒸発から成り立つ。そこで、本研究においては、入力データセットを用いて環境にフェロモンを付加し、入力データセットにおける顧客の変化を時間経過とみなしてフェロモンの蒸発を考える。初め、フェロモンの付加は、二つのノード間で行われる。二つのノードの間にフェロモンが付加されることによって、二つのノードが連結した新たなノードを生成する。よって、

$$\begin{aligned} c_1 &= \{i_1, i_2, i_3, i_4\} \\ c_2 &= \{i_5, i_6, i_7\} \\ c_3 &= \{i_8, i_9, i_{10}, i_{11}\} \\ c_4 &= \{i_{12}\} \\ &\vdots \\ &\vdots \end{aligned}$$

図 1: 入力データセット

上位層では、要素が多数となるパターンが抽出できる。全ての入力データセットについて、フェロモンの付加と蒸発を繰り返すことでパターンの抽出を試みる。

##### 3.2.1 仮想グラフの用意

まず、フェロモンを付加するための環境を用意する。環境は仮想グラフ  $G$  で表す。 $G = (V, E)$  は仮想空間上の有向グラフである。グラフ  $G$  における各ノード  $v_i \in V$  は実環境における各商品  $s_i \in S$  に対応し、各エッジ  $e_{i,j} \in E$  は商品  $v_i$  を触った後に商品  $v_j$  を触る経路を表す。また、グラフ  $G$  上に付加されるフェロモンの分布は  $\tau$  で表すことにする。なお、前提条件として頻出パターンに対する予備知識がないとして、 $\tau$  の初期値は  $\tau(0) = 0$  とした。

##### 3.2.2 フェロモンの付加及び蒸発

顧客  $c_t$  の入力データごとにフェロモン  $\tau$  の付加と蒸発を行う。まず、入力データ  $c_t$  から、消費者が商品間で移動した度数分布  $d_{i,j}(t)$  を作成する。度数分布  $d_{i,j}(t)$  は消費者  $i$  が商品  $s_i$  を触った後に商品  $s_j$  を触った回数を表す。得られた分布度数  $d_{i,j}(t)$  を用いて式 1 に従って、フェロモン  $\tau$  の付加を行う。

$$\tau_{i,j}(t) = \tau'_{i,j}(t-1) + d_{i,j}(t) \quad (1)$$

次にフェロモンの付加をした後に、フェロモンの蒸発を行う。フェロモンの蒸発は蒸発率  $\rho$  に従う。蒸発の計算式を式 2 に示す。

$$\tau'_{i,j}(t) = \tau_{i,j}(t) (1 - \rho) \quad (2)$$

このフェロモンの蒸発により、古い探索情報は少しずつ破棄され、解析結果には常に新しい探索情報が一定の割合で反映されることになる。蒸発率  $\rho$  は更新の速さを表す。 $\rho$  が大きいほど、フェロモンの蒸発は速くなり、新しい情報に重みをおく。一方で、 $\rho$  が小さいほど、フェロモンの蒸発は遅くなり、過去の情報を蓄積しやすい。

##### 3.2.3 抽出パターンの階層化

連続した二つのノード間でのフェロモンを付加することで、上位層のノードを生成する。二つのノード間にフェロモンが付加されることで、一つ上位の層に新たなノードを生成する。一人の消費者が商品 1,2,3,4,5 と連続して商品に手を触れた場合 (図 2) を例にして階層化を説明する。

消費者が商品 1,2 と連続した触れた場合にノード 1 と 2 の間にフェロモンが付加される。それと同時に一つ上位の層である二階層目にノード 1 → 2 が生成される。その直後に消費者が商品 2,3 と連続して触れた場合、ノード 2 と 3 の間にフェロモンが付加される。先と同様に二階層目には、ノード 2 → 3 が生成される。このとき、二階層目に着目すると、ノード 1 → 2 が生成された直後にノード 2 → 3 が生成されている。同一階層内で連続して起きた事象として、ノード 1 → 2 とノード 2 → 3 の間にフェロモンの付加を行う。このフェロモンの付加と同時に三階層目となるノード 1 → 2 → 3 が生成される。このよう

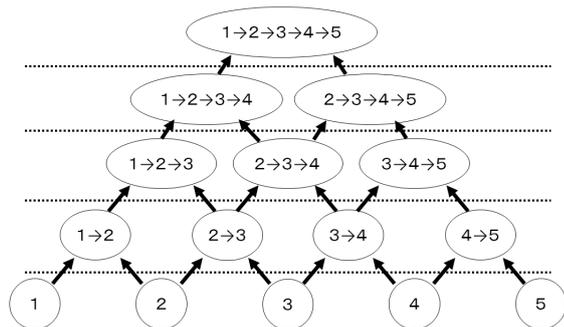


図 2: 階層化イメージ

に、同一階層内で同一消費者の行動によって、ノード間にフェロモンが付加された場合、連続したノードが一つのノードとして一つ上位の階層に生成される。これによって、上位の階層では、連続したより多数の商品に触れるパターンを抽出する。

## 4. 実験

提案したアルゴリズムが想定した頻出パターンを抽出できることを検証するために、シミュレーションデータを用いた実験を行った。

### 4.1 シミュレーションデータの生成

想定する解となる頻出パターンが抽出できることを確認するために、解となるパターンを埋め込んだデータセットを生成する。商品であるアイテムセットの種類は 26 種類とし、 $i \in \{1, 2, \dots, 26\}$  とした。このとき、アイテムを手にとって時間幅  $a$  がとりうる値は  $a \in \{1, 2, 3, 4, 5\}$  であり、次の商品を手にとるまでの時間幅  $t$  がとりうる値は  $t \in \{1, 2, \dots, 100\}$  とした。

シミュレーションデータセットは基本的には上記の数値の中から一様乱数で生成する。このとき、ある一定の確率  $p$  で、解として抽出したいパターンアイテム群の埋め込みを行う。今回の実験では商品 1, 2, 3 を連続で触るパターン  $P_1$  と商品 11, 12, 13, 14 を連続で触るパターン  $P_2$  を想定する。なお、パターン  $P_1$  の発生確率を  $p_1 = 1\%$  とし、パターン  $p_2$  の発生確率を  $p_2 = 5\%$  とした。想定したパターンを生成している際の次の商品を手にとるまでの時間幅は小さく設定した。本研究では、シミュレーションデータとして、10000 アイテム分のデータを生成した。

### 4.2 パターン抽出の検証実験

生成したシミュレーションデータを用いて、提案アルゴリズムの検証を行う。シミュレーションデータで想定した頻出パターンを抽出できるかを確かめる。今回の実験では、フェロモンの蒸発率  $\rho$  の値を  $\rho = \{0, 0.005, 0.01\}$  の 3 種類用意した。なお、消費者の単位に分けるために閾値  $\phi$  は  $\phi = 50$  に設定した。

#### 4.2.1 フェロモン蒸発率 $\rho = 0$

このとき、フェロモンの蒸発が起らない。つまり、過去の情報も新しい情報も等しく扱い、情報を蓄積していく。単調に出現するパターンを網羅的に記録するため、データ構造に変化が生じた場合には対応できないが、静的環境下では優秀な性能を有する。3 つの商品が連続して手にとられる階層における

フェロモンが濃く残る 5 ノードを表 1 に示す。表 1 をみると、フェロモンが濃い上位 3 つが  $12 \rightarrow 13 \rightarrow 14$ ,  $11 \rightarrow 12 \rightarrow 13$ ,  $1 \rightarrow 2 \rightarrow 3$  となり想定した頻出パターンが抽出できていることがわかる。4 つの商品が連続して出現する階層においては  $11 \rightarrow 12 \rightarrow 13 \rightarrow 14$  に最も濃いフェロモンが付加されることが確認できた。

表 1:  $\rho = 0$  のときの残留フェロモンが強い上位 5 ノード

ノード	フェロモン
$12 \rightarrow 13 \rightarrow 14$	427
$11 \rightarrow 12 \rightarrow 13$	427
$1 \rightarrow 2 \rightarrow 3$	89
$13 \rightarrow 14 \rightarrow 11$	34
$14 \rightarrow 11 \rightarrow 12$	29

#### 4.2.2 フェロモン蒸発率 $\rho = 0.005$

実験結果より蒸発率  $\rho = 0$  のときと同様に  $11 \rightarrow 12 \rightarrow 13$ ,  $12 \rightarrow 13 \rightarrow 14$ ,  $1 \rightarrow 2 \rightarrow 3$  及び、 $11 \rightarrow 12 \rightarrow 13 \rightarrow 14$  のパターンが抽出できていることがわかった。

#### 4.2.3 フェロモン蒸発率 $\rho = 0.01$

フェロモンが濃く残る 5 ノードを表 2 に示す。表 2 をみると  $11 \rightarrow 12 \rightarrow 13$  と  $12 \rightarrow 13 \rightarrow 14$  のパターンは抽出できているのに対して、 $1 \rightarrow 2 \rightarrow 3$  のパターンはフェロモンの濃度が小さい。一度の出現でフェロモンが 1 だけ付加されることを考慮に入れるとパターンが抽出できたとはいきれない。

表 2:  $\rho = 0.01$  のときの残留フェロモンが強い上位 5 ノード

ノード	フェロモン
$12 \rightarrow 13 \rightarrow 14$	7.17
$11 \rightarrow 12 \rightarrow 13$	7.17
$1 \rightarrow 2 \rightarrow 3$	1.55
$24 \rightarrow 11 \rightarrow 12$	1.40
$13 \rightarrow 14 \rightarrow 24$	1.15

### 4.3 アルゴリズムの適応性の検証

想定する頻出パターンが変化したときに、抽出されるパターンが変化することを確認する。

#### 4.3.1 入力データセットの生成

想定する頻出パターンが変化したときの抽出されるパターンを確認するため、入力用のデータセットは前半と後半に分ける。前半は 4.1 で使用したデータセットと同一のものを使用した。後半は改めて、前半のパターンと同様に、10000 アイテム分のデータを、5, 6, 7 と連続で触るパターン  $P_3$  と商品 15, 16, 17, 18 と連続で触るパターン  $P_4$  を抽出したいパターンとして埋め込んだデータセットを生成した。前半のデータセットと後半のデータセットを順に並べ、適応性検証用のデータセットとした。

#### 4.3.2 適応性の検証実験

適応性検証用のデータセットを用いて、検証実験を行う。ここでも、4.2 同様に、フェロモンの蒸発率  $\rho$  の値を  $\rho = \{0, 0.005, 0.01\}$  の 3 種類で検証する。アルゴリズムを適用した結果、3 つの商品が連続して手にとられる階層において、最終的な残留フェロモンが強く残る上位 10 ノードを表 3 に示す。

表 3: 適応性検証実験における残留フェロモンが強い上位 10 ノード

$\rho = 0$		$\rho = 0.005$		$\rho = 0.01$	
ノード	フェロモン	ノード	フェロモン	ノード	フェロモン
15 → 16 → 17	436	15 → 16 → 17	20.33	15 → 16 → 17	10.07
16 → 17 → 18	436	16 → 17 → 18	20.33	16 → 17 → 18	10.07
12 → 13 → 14	427	5 → 6 → 7	5.97	5 → 6 → 7	3.86
11 → 12 → 13	427	17 → 18 → 18	3.18	17 → 18 → 18	2.42
5 → 6 → 7	93	17-18 → 15	2.06	17 → 18 → 15	1.04
1 → 2 → 3	89	18 → 15 → 16	1.72	18 → 15 → 16	1.01
13 → 14 → 11	34	17 → 18 → 16	1.58	7 → 15 → 16	1.01
17 → 18 → 15	31	17 → 18 → 10	1.35	6 → 7 → 12	1.00
17 → 18 → 11	30	7 → 15 → 16	1.28	7 → 12 → 26	1.00
14 → 11 → 12	29	6 → 7 → 15	1.17	6 → 7 → 15	0.99

#### 4.4 考察

上記の実験より、求める頻出パターンの出現頻度が小さいとき、フェロモンの蒸発率  $\rho$  を小さく設定する必要があることがわかる。一方で、より環境変化にすばやく対応するためにはフェロモンの蒸発率  $\rho$  は大きい方が好ましい。よって、実際のマーケティングで活用させるためには求めたいパターンの出現頻度に応じた蒸発率を的確に設定する必要がある。

また、実際のマーケティングについて考えたとき、変化にも複数種類があることが伺える。例えば、季節の変化であったり、曜日の変化である。季節の変化について考えたとき、気温が高いときは冷たいものが食べなくなるが、気温が低いときには暖かいものが食べなくなる。一方で、曜日の変化というものは、例えば平日と休日に分けると、人の行動パターンが違うものであることは容易に想像がつく。このように、時系列の変化を追うときに、想定する変化の周期を考える必要がある。この問題に対応するには蒸発率の設定が重要になると考えられる。求めたいパターンの周期が大きくなると、蒸発率は小さく設定する必要がある。

#### 5. おわりに

本稿では、消費者の行動パターン抽出のために、ACO アルゴリズムを用いたパターンマイニングを提案した。結果からわかるように、フェロモンの蒸発率は検討を重ねる必要があることがわかる。また、階層化するにあたって、より上位の階層に属するパターンは出現頻度がより小さくなることが想定できる。したがって階層ごとに異なる蒸発率を設定していく必要がある。

今後、実際のマーケティングで使用できるような GUI の開発を進める。マーケティングにおいて、システム利用者は小売店販売店やメーカなど多様にわたることが予想され、抽出したいパターンの商品単位やカテゴリ単位といった解析粒度の項目や季節性や曜日効果といった時間軸の項目の自由度を考える必要があると考える (図 3)。

#### 参考文献

- [1] 経済産業省: 経済産業ジャーナル 2013 年 12・1 月号, 4-13, 2013
- [2] 梶井昌邦: アプリオリアルゴリズムを活用した店舗のブランド選択の意思決定に関する考察, 福岡大学経済学論叢, 福岡大学, pp. 37-54, 2010



図 3: GUI

- [3] A g rawal, R. and Srikant, R : Fast algorithms for mining association rules in large database, Proc. Int. Conf. on Very Large Data Bases, pp. 487-499, 1994
- [4] A g rawal, R. and Srikant, R : Mining Sequential Patterns, Proc. of The 11th Int'l Conf. on Data Engineering, pp. 3-14, 1995
- [5] A g rawal, R. and Srikant, R : Mining Sequential Patterns: Generalizations and performance improvements, In Proc. of The 5th Int'l Conference on Extending Database Technology, 1996
- [6] Hohammed j.Zaki : Spade: an efficient algorithm for mining frequent sequences, In Machine Learning Journal, special issue on Unsupervised Learning, pp. 31-60, 2001
- [7] 玉置洋, 福井健一, 沼尾正行, 栗原聡: フェロモンを介したエージェント協調モデルによるセンサー隣接関係構築法の提案, 情報科学技術レターズ, Vol.6, pp.153-156, 2007
- [8] Dorigo, M and Gambardella, LM : Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem, IEEE Transaction on Evolutionary Computation, 1(1), 53-66, 1997