

# 群知能メカニズムを用いた時系列階層型パタンの 抽出手法の提案

坪井 一晃<sup>1,a)</sup> 須賀 聖<sup>2</sup> 金光 博明<sup>3</sup> 坂本 圭司<sup>4</sup> 栗原 聡<sup>2</sup>

**概要:** 自宅やオフィス、通勤などの日常における行動様式や、店舗における曜日や季節による商品の売れ方などには、様々なパターンが含まれる。既存の系列的なパターン抽出技術では、主に複数の系列におけるデータ間でのパタンの頻出の是非を問題としており、一つの長大な系列データ内からパターンを抽出することを苦手としている。さらに、マーケティング等においては、消費者行動は常に変化するために、時間経過に伴うパタンの変動を考慮できる動的なパターン抽出が課題となっている。そこで、本研究では群知能型の探索手法を改良し、一つの時系列データから動的に変化するパタンの自動抽出を可能とする手法を提案する。また、実データとなるマーケティングデータについて分析し、消費者の動的な購買パターンについて考察する。

## A Proposal of Hierarchical Sequential Pattern Mining Based on Swarm Intelligence

KAZUAKI TSUBOI<sup>1,a)</sup> SATOSHI SUGA<sup>2</sup> HIROAKI KANEMITSU<sup>3</sup> KEIJI SAKAMOTO<sup>4</sup> SATOSHI KURIHARA<sup>2</sup>

### 1. はじめに

創発というキーワードは大変興味深い。創発とは、部品や要素が集合することによって、単独では見られなかった様相が発現することである。生物は創発という現象を上手く活用している。たとえば、アリやハチといった昆虫たちである。それぞれのアリやハチの単体では優れた知能を持つとは言いがたいが、一つの集団として観察すると極めて優れた知能を発揮している。アリが群れとして行動する様子から着想を得て、巡回セールスマン問題などの組み合わせ最適化問題の解法としてアントコロニー最適化アルゴリズム (ACO:Ant Colony Optimization) [1][2] が提案され

ている。これは、自然界のアリの採餌行動における、アリ単体では餌をみつけ巣に戻るときにはフェロモンという揮発性物質を移動経路に付加することと、餌を探索するにあたりフェロモンが残る経路を好んで探索しやすいという行動指針をアリエージェントに組み込んでいる。個々のエージェントは先の単純な行動指針のみだが、これを集団として行うことによって、餌場と巣におけるほぼ最短の経路を獲得できる。自然界のアリがモデルということもあり、柔軟な適応性や高い頑健性が報告されている。

インターネットの急激な進化やコンピュータの高性能化および低価格化を背景に、社会のいたるところに IT を活用したシステムが導入されるようになってきている。導入されたシステムを利活用するたび、システムからはデータが生成され続けている。我々が意思を向けずとも稼働しているシステムも存在し、日々多種多様にわたるデータが大量に生成・蓄積されている。近年、この大量に蓄積されるデータを用いることで有用な情報を得られるのではないかという期待が高まっている。このように、蓄積されたデータから有用な知見を得るための技術としてデータマイニングに

<sup>1</sup> 電気通信大学  
The University of Electro-Communications, Chofu, Tokyo  
182-8585, Japan  
<sup>2</sup> 慶應大学  
Keio University  
<sup>3</sup> 株式会社トライアルカンパニー  
Trial Company, Inc.  
<sup>4</sup> 株式会社トライアルホールディングス  
Trial Holdings, Inc.  
a) tsuboi@uec.ac.jp

注目が集まっている。データマイニングにおける重要な技術の一つに、データベースの中から頻出するアイテムの組み合わせを抽出するパターンマイニングがある。パターンマイニングを用いることで、例えばコンビニエンスストアやスーパーマーケットなどの小売店におけるよく同時に購入される商品などを効率的に発見できるようになる。さらに、このパターンマイニングを現実世界において度々重視される時系列に着目し発展させた、時系列パターンマイニング技術が提案されている。時系列性を考慮したパターンを抽出することによって、この商品が購入された次にはこの商品が購入されるはずであるという予測することができるようになる。代表的な時系列パターンマイニング技術では、複数の系列データを入力し、それぞれのアイテムの組み合わせがいくつの系列データ内に出現しているのかが頻出の是非であり、ひとつの系列データ内における頻出の是非は問題にしていない。また、時系列パターンマイニングでは、パターン内のアイテムの前後関係には注目しているが、その時間間隔にはふれていない。

現実世界を考えると、日常における自宅やオフィスなどをつなぐ通勤といった行動様式や小売店における曜日や季節などによる商品の売れ方などには、様々なパターンを内包している。行動様式においては一人の行動ログから頻出する行動パターンなどの抽出や、マーケティングにおいては一人の顧客の購買パターンなど、一つの系列データ内において頻出するパターンを抽出する必要がある。また、マーケティングを考える際には、1日おきに起こるパターンなのか、1ヶ月おきにおきるパターンなのかは区別したいものである。

そこで、本研究では群知能アルゴリズムを代表する ACO アルゴリズムを用いて、時系列パターンマイニング手法を提案する。一つの長大な系列データ内に頻出するアイテムの組み合わせを、アイテム間の間隔も含め抽出する。また、入力する時系列データにおいて、パターンは変化するという前提のもと、ACO アルゴリズムを応用することで、動的なパターン抽出アルゴリズムを提案する。

本稿の構成を次に示す。2章では、パターンマイニングや ACO アルゴリズムに関する関連研究について述べる。3章では、提案手法となる群知能メカニズムによる時系列階層構造を有するパターンの抽出手法について述べる。4章では、提案手法に対してマーケティングデータを用いた実験を行い消費者の購買パターンについて考察する。5章では、まとめと今後の課題について述べる。

## 2. 関連研究

大量のデータから頻出するアイテムセットの組み合わせをパターンとして抽出する手法がパターンマイニングである。データベース上から頻出するアイテムセットのパターンを抽出するためのアルゴリズムとして Apriori アルゴリ

ズム [3] が有名である。Apriori アルゴリズムは、あるアイテムセットの出現頻度が高いためにその部分集合のアイテムセットの出現頻度が高い必要があるという考えから、小さいアイテム集合から順にデータベースにおける出現回数を数える幅優先探索を実行する。あるアイテムセットが頻出とならない場合、その上位となるアイテムセットである、そのアイテムセットは包括するより大きいアイテムセットについて探索する必要がないとみなせる。このことから解の探索範囲空間を減らしている。しかし、解の候補となるアイテム集合が生成されるたびに出現回数を数えるために全ての入力データベースを調べなければならなくなる。そこで、入力データベースが膨大なものとなったときに、計算時間や計算機のメモリなどに問題が生じる。また、Apriori アルゴリズムでは、アイテムの時系列性を考慮していないといった課題が指摘されている。

そこで、データベース上の時系列性を考慮したアイテムのパターン抽出を行うために Apriori アルゴリズムを拡張した AprioriAll アルゴリズム [4] やパターン抽出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] が提案されている。さらに、ID - List という見方をしたデータベースを活用した SPADE アルゴリズム [6] や ID-List の管理をビットマップで表現することで高速化を達成した SPAM アルゴリズム [7] が提案されている。また、深さ優先型の探索法を採用したシーケンシャルパターンマイニング手法として PrefixSpan アルゴリズム [8] が提案されている。

上記の、GSP アルゴリズムや、SPADE, SAPM, PrefixSpan アルゴリズムなどの既存の代表的な時系列パターンマイニング技術では、静的なアルゴリズムであり、すべてのデータがそろったときに、その入力に対して最小支持度という基準を満たすパターンを抽出している。つまり、データが増えるごとに解析はすべてやり直しであり、時間経過にともなう変化の様子を抽出することは処理が大変になる。

そこで、インクリメンタルに入力データを挿入しつつ、パターンを抽出できるアルゴリズムとして IncSPAM [9] が提案されている。しかし、これらのアルゴリズムは、複数の系列データ間においてそのパターンが出現する系列数のより頻出の是非を判断している。つまり、一つの系列データの中からアイテムの組み合わせが頻出かどうかを判断することができない。

一方、[10] では、一つの長大な系列データを入力として時系列パターンを抽出することを目的としたアルゴリズムを提案している。しかし、このアルゴリズムでは、パターンにおけるアイテム間の間隔を考慮されていない。

さらに、時系列の前後関係は考慮していても、その前後する時間の間隔は考慮されていない。解析結果をマーケティングなどで活用することを考えると、二つの商品の購

買間隔が1日であるものと1週間であるものは区別して扱いたい。

そこで、我々は群知能分におけるACOアルゴリズムに着目している。ACOアルゴリズムの特徴として、どのような環境に対しても解を生成することができる適応性や、環境の変化に対しても柔軟に解を再探索できる頑健性が挙げられる。このような特徴は現実問題に対して有効であると考えられる。

そもそも、ACOアルゴリズム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あり、最適化問題の巡回セールスマン問題に対する解法として提案された。ACOアルゴリズムにおいて、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だ経路選択を行うという二つをアリの行動の前提とする。この前提によって、アリが集団として行動するにつれて、最短経路を探索するアリの数が増加しフェロモンの濃度がますます濃くなる。一方、フェロモンは揮発性物質であるために時間経過とともに蒸発する。結果的に、アリが通過する頻度が低い経路のフェロモンの濃度は薄くなる。環境に残留するフェロモンが最短経路問題の解となる。結果として、ACOアルゴリズムの適応性や頑健性の高さが特徴となっている。

ACOアルゴリズムをデータマイニングのクラスタリング手法として応用した研究としてAnt Miner+アルゴリズム [11] が提案されている。このアルゴリズムは分類結果がわかりやすい分類器でありながら精度が高く、データが分散した環境にも適するとされる。また、ACOアルゴリズムをパターンマイニングに応用した研究としてTamakiら [12] の研究がある。Tamakiらは、センサが人の行動を読み取り反応することから、連続した人の行動から隣接するセンサは連続して反応するという前提によって、センサの隣接関係の推定を行っている。ACOアルゴリズムを応用したことで、複数の人による連続してセンサが反応する場合への対応や、センサ自体の故障や移動といった出来事にも柔軟に対応している。

このように、もともとは最適化技術として提案された群知能分野のアルゴリズムであるが、単純な行動ルールに基づいたエージェントの移動と環境に対するフェロモンの付加および蒸発を応用することで、柔軟なシステムの構築が達成できる。本研究においても、ACOアルゴリズムが有する適応性や頑健性に着目し、群知能メカニズムによる時系列階層構造を有するパターン抽出を試みる。

### 3. 群知能メカニズムによる時系列階層構造を有するパターン抽出手法

提案アルゴリズムでは、アイテム集合  $I = \{i_1, i_2, \dots, i_m\}$  で構成させる一つの時系列データ  $D_n = d_1, d_2, \dots, d_n$  (ただし、 $d \subset I$ ) から、時系列にそって頻出するアイテムの組

み合わせをパターンとして順序関係および出現間隔を考慮しながら抽出する。

提案するアルゴリズムでは時系列データから時刻順に入力し、各時刻の入力に対して嗜好有したアリエージェントたちが探索しマップに対する宇フェロモンの付加を行う。また、時間経過に伴うフェロモンの蒸発を行う。提案するアルゴリズムを擬似コード (アルゴリズム 1) に示す。

---

#### Algorithm 1

---

**Input:** 時系列データ  $D_n$

**Output:** フェロモンが付加された仮想マップ  $M_t$

$t := 0$

**while**  $t < n$  **do**

アリエージェントの生成・追加

アリエージェントによる探索および  $M_t$  に対するフェロモンの付加

フェロモンの蒸発

探索期間が終了したアリの削除

$t := t + 1$

**end while**

---

#### 3.1 仮想フェロモンマップ $M_t$ の用意

まず、フェロモンを付加するための環境として仮想マップ  $M_t$  を用意する。時刻  $t$  における仮想マップ  $M_t = (P, v(P))$  は、パターン集合  $P \ni p$  およびそのパターン  $p$  に対するフェロモン量  $v(p)$  を記録した仮想マップである。パターン  $p = (F, S, l)$  は、順序関係を考慮した複数のアイテムまたは既出のパターン ( $F$ ) および、その複数のアイテムまたはパターンの時間間隔 ( $S$ )、パターンとしての階階数 ( $l$ ) からなる。

#### 3.2 事前情報の取得

アリエージェントの初期生成のために、事前知識としてアイテム単体の出現頻度  $c_i \in C$  を計測する。

#### 3.3 アリエージェントの生成

各時刻において、アリエージェントは  $ANTS_{max}$  だけ追加生成する。それぞれのアリエージェントはパターンを探索するにあたって、順序性を有する複数の嗜好  $F \subset \{I, P\}$  とそれぞれの嗜好に対しての好む間隔  $S$  を有する。また、一匹のアリエージェントが探索できる探索期間  $r$  が与えられる。

アリエージェントには2種類あり、1つは最小のアイテムから嗜好を選択するタイプであり、もう1つは生成されたフェロモンマップを基にアイテムの組み合わせから嗜好を選択するタイプである。

それぞれのアリエージェントにおける嗜好は、事前知識またはフェロモンマップ  $M_{t-1}$  に基づいて、アイテムまたはパターンから決定される。アリが生成された時刻にお

るアイテム集合からランダムに第一の嗜好として設定される。アルゴリズム実行初期段階として時刻  $T_1$  まで、多数のパターンを構築するために生成されるすべてのアリエージェントのそれぞれの嗜好  $f$  は、出現頻度の割合を確率としてアイテムから選択される。つまり、アイテム  $i$  が選択される確率  $q(i)$  は式 1 に従う。

$$q(i) = \frac{c_i}{\sum_{j \in I} c_j} \quad (1)$$

パターンからパターンを階層的に構築するために、フェロモンマップから嗜好を選択するアリエージェントのそれぞれの嗜好  $f$  は、フェロモンマップ  $M$  におけるフェロモン量  $v(P)$  の割合を確率としてパターンから選択される。つまり、パターン  $p$  が選択される確率  $q(p)$  は式 2 に従う。

$$q(p) = \frac{v_p}{\sum v(P)} \quad (2)$$

また、アリエージェントに設定する嗜好に対する間隔  $S$  は一様乱数によって定める。アルゴリズムを??に示す。

---

#### Algorithm 2 アリエージェントの生成・追加

---

```

ANT := 0
while ANT < ANTSmax do
  if t < T1 | ANT < ANTθ then
    s := random(1, smax)
    dt から f1 の選択
    dt+s から f2 の選択
  else
    Mt-1 から f1 と f2 の選択
    if f1 を構成する嗜好の 2 番目のものと f2 を構成する嗜好の
    1 番目のものが同一 then
      s := random(0, smax)
    else
      s := random(1, smax)
    end if
  end if
  f1 探索モードに設定
  ANT := ANT + 1
end while

```

---

### 3.4 アリエージェントによる探索及びフェロモンの付加

それぞれのアリエージェントは設定された嗜好および嗜好に対する間隔に基づいて入力された時刻におけるデータに対して探索を行う。アリエージェントが嗜好および嗜好に対する間隔からなるアイテムまたはパターンの組み合わせを発見することができた場合、そのアイテムまたはパターンの組み合わせを一つのパターン  $p$  とみなし、そのパターンに対するフェロモン量を 1 だけ増加させる。なお、発見したパターンが、新規登場であった場合にはパターン番号として順次固有のナンバリングを割り当てる。また、アリエージェントが嗜好およびその間隔から発見したパターンにおいて、アリエージェントが有する嗜好のうち最大の階層数に 1 加えたものをそのパターンの階層数とする。なお、初期のアイテム集合はそれぞれ階層数を 1 とする。

---

#### Algorithm 3 アリエージェントによる探索

---

```

ANT := 0
while ANT < ANTSall do
  if f1 探索モードか then
    if Dt に f1 が出現 then
      f2 探索モード (嗜好までのカウント s1) に設定
    end if
  else
    if 嗜好までのカウントが 1 then
      if Dt に探索 f が含まれる then
        if fnext があるか then
          fnext 探索モード (嗜好まで snext) に設定
        else
          発見パターンに対してフェロモンの付加
        end if
      else
        f1 探索モードに戻る
      end if
    else
      嗜好まで (のカウント):= -- 1
    end if
  end if
  ANT := ANT + 1
end while

```

---

### 3.5 蒸発にともなうフェロモン値の減算

時間経過に伴い、残留するフェロモン量を減少させる。フェロモン量の減少に際して、フェロモン量自体は指数関数的に減少することおよび、パターンのパターンとなっているような深い階層のパターンに関しては出現率が低いと考えられるため、減少するフェロモン量は少ないことを考慮した。

$$v = (\exp(\log(v+1) - 0.00001) - 1) \times (1 - 10^{-l}) \quad (3)$$

## 4. テストデータを用いた評価実験

提案するアルゴリズムを用いて時系列に応じて変化するパターンを抽出できることを検証するために、ランダムな時系列データの中に、抽出されるべきパターンが埋め込まれたテストデータを用いた評価実験を行う。

### 4.1 テストデータ

テストデータとして、アイテム集合  $I$  は「a」から「z」の 26 種類のアルファベットからなり、30000 の時系列データ  $D_{30000}$  において各時刻ひとつのアルファベットがランダムに埋め込まれている。さらに、時系列データは、前半、中盤、後半として 10000 ずつに分割する。抽出されるべきパターンとして前半には「agent」の連続したアルファベットの系列と「a\*\*g\*\*e\*\*n\*\*t」のように「\*」で表された任意の文字が含まれた系列、「a\*\*\*\*\*g\*\*\*\*\*e\*\*\*\*\*n\*\*\*\*\*t」のように 5 つの任意の文字を間に含まれた系列を 50 個ずつ埋め込む。中盤には、さきほどの「agent」と「a\*\*g\*\*e\*\*n\*\*t」、  
「a\*\*\*\*\*g\*\*\*\*\*e\*\*\*\*\*n\*\*\*\*\*t」を 25 個ずつ埋め込むと

ともに、新しく「komei」の連続したアルファベットの系列と「k\*\*o\*\*m\*\*e\*\*i」のように2つの任意の文字列を間に含めた系列、「k\*\*\*\*\*o\*\*\*\*\*m\*\*\*\*\*e\*\*\*\*\*i」のように5つの任意の文字を間に含めた系列を25個ずつ埋め込む。後半には、「komei」、「k\*\*o\*\*m\*\*e\*\*i」、「k\*\*\*\*\*o\*\*\*\*\*m\*\*\*\*\*e\*\*\*\*\*i」を50個ずつ埋め込む。このテストデータを用いた実験を行い、「a」、「g」、「e」、「n」、「t」に関連した文字列が抽出されつつ、後編に向けて「k」、「o」、「m」、「e」、「i」に関連した文字列が抽出されるように変化の様子を獲得する。

#### 4.2 テストデータを用いた評価実験

生成したテストデータを用いて、提案アルゴリズムの検証を行う。それぞれのアリエージェントが有する嗜好は2つとし、探索するパターンは  $p = (f_1, s, f_2, l)$  からなる。各アリエージェントの探索範囲  $r$  は1000とする。実験に際して1時刻を1ステップとし、各ステップにおいて10000のアリエージェントを生成する。1000ステップまでに生成されるすべてのアリエージェントは、アイテム集合であるアルファベットから事前知識を用いて嗜好を選択する。1000ステップ移行は100のアリエージェントはアイテム集合であるアルファベットから嗜好を  $s$  円卓し、残りの9900のアリエージェントはフェロモンマップ  $M_{t-1}$  から嗜好を選択する。また、2つの嗜好に対する間隔は最大6最小0の一樣乱数から選択される。なお、2つの嗜好に対する間隔において0が選択されるのは、アリエージェントの有する2つの嗜好をみたときに一つ下の階層において重複が許容される場合のみである。

今回の実験において、入力は一時刻一文字の入力データであり、単純な文字列が入力とみなせる。そこで、一つの長大な文字列から頻出するパターンとして、変数として任意の文字を許容しつつ頻出するパターンを抽出するアルゴリズムとしてMaxMotif??が提案されており、この実験において競合している。ただし、我々の提案するアルゴリズムはインクリメンタルに各時刻のステップが入力されるたびにその時刻における頻出だとされるパターンが出力される。一方で、MaxMotifはバッチ処理的に、入力された文字列に対して最小支持度を満たすパターンを網羅的に列挙する。すべての時刻における頻出パターンを抽出するために、MaxMotifにおいては  $n$  文字までの入力を  $1 < n < 130000$  で繰り返し実行していく。

#### 4.3 実験結果

テストデータを入力として、提案アルゴリズムを実行した。前半、中盤、後半、それぞれの最終ステップである10000、20000、30000ステップ時において、パターンの内容が5つのアルファベットからなるパターンのうち、フェロモン量が多く残っていた上位10のパターンとテストデー

タに埋め込んだパターンに関する主なパターンをそれぞれ表1、2、3に示す。それぞれのパターンは階層数が4や5であり、パターンのパターンからなっているため、そのパターンが示す具体的なパターンも表中に示した。例として、 $p = (1531, 3, 3695, 4)$  であれば、本来  $1531 = (a, 3, g, 2)$  というパターンと  $p = (653, 0, 1566, 3)$  からなり、さらに  $653 = (e, 3, n, 2)$  というパターンと  $1566 = (n, 3, t, 2)$  というパターンが構成要素となっている。それぞれの結果からわかるようにパターンが表す内容が同じでも、パターンからパターンを構成している関係でパターンの構成のバリエーションは複数存在している。

表1をみると、前半で抽出されるべきパターンとして埋め込んだa, g, e, n, tに関連した3つのパターンが、複数のパターン構成のバリエーションにて抽出されていることが確認できる。また、20000ステップ終了時には、パターンが変化していく様子として、a, g, e, n, tに関連した3つのパターンが引き続き抽出されている一方で、k, o, m, e, iに関連したパターンが3つのうち2つが抽出されだしていることが確認できる。最終的な30000ステップ終了時には、k, o, m, e, iに關係した3つのパターンが抽出されていることがわかる。埋め込んだパターンの変化に応じつつ、パターンの抽出ができていく。

表1 10000ステップ終了時におけるフェロモン量

パターン	フェロモン	パターンが示す内容
1531,3,3695,4	28.64	a**g**e**n**t
32688,0,32789,5	15.60	a**g**e**n**t
33259,0,34794,4	8.37	agent
6393,3,1566,4	8.23	a**g**e**n**t
140,6,3592,4	7.91	a*****g*****e*****n*****t
6937,0,3695,4	5.24	a**g**e**n**t
6393,0,3697,4	5.18	a**g**e**n**t
33259,1,10,4	4.20	agent
3561,6,1368,4	3.49	g*****e*****n*****t*****g

表2 20000ステップ終了時におけるフェロモン量

パターン	フェロモン	パターンが示す内容
32688,0,32789,5	12.68	a**g**e**n**t
1531,3,3695,4	9.85	a**g**e**n**t
140,6,3592,4	3.52	a*****g*****e*****n*****t
6393,3,1566,4	3.06	a**g**e**n**t
33259,0,34794,4	2.72	agent
33259,1,10,4	2.32	agent
103696,0,133494,4	2.29	k*****o*****m*****e*****i
6393,4,794,4	1.99	a**g**e**d*****e
3656,2,2705,4	1.99	a*****e**n**m***s
2040,4,3656,4	1.99	g*****h***a*****e**n
974,1,98861,4	1.49	komei
101696,0,98859,4	1.49	komei
109812,0,98861,4	1.49	komei

表 3 30000 ステップ終了時におけるフェロモン量

パターン	フェロモン	パターンが示す内容
974,1,98861,4	18.77	komei
96928,3,29227,4	16.31	k**o**m**e**i
101696,0,98859,4	13.30	komei
32688,0,32789,5	10.29	a**g**e**n**t
101696,1,28829,4	8.77	komei
410,3,101998,4	5.78	k**o**m**e**i
974,5,101998,4	4.87	k**o**m**e**i
101696,4,2208,4	4.07	kom***m*i
1478,1,109260,4	3.22	e**kom*i
1531,3,3695,4	3.21	a**g**e**n**t
103696,6,938,4	2.42	k****o****m****e****i
141,6,137134,4	2.18	k****o****m****e****i

#### 4.4 考察

実験により、ランダムな時系列データに対して抽出されるべきパターンとして埋め込んだパターンが抽出されることがわかった。また、前半、中盤、後半と埋め込むパターンに変化をつけたが、時系列で変化するパターンに対して適応的にパターンを抽出できることを示した。パターンのパターンを繰り返し階層を挙げることで大きな階層は残りやすくした結果、階層数が5になった(32688,0,32789)のパターンは実験終了後も大きくフェロモン量が残ってしまった。また、一方で20000ステップ終了時にはまだ「k\*\*o\*\*m\*\*e\*\*i」に関するパターンが抽出できていない。これは、図??をみてわかるように、全体的にk, o, m, e, iに関連したパターンの抽出が遅れていることがうかがえる。大きなパターンはパターンのパターンとして積み重ねていくことと、すべてにフェロモン量が多く残るa, g, e, n, tに関係したパターンが嗜好として選択されやすいことが影響していると考えられる。探索するときのリエージェントに与えるパラメータおよび、探索させるリエージェント数が重要な要素となると考えられる。

#### 5. おわりに

本研究では、群知能アルゴリズムに代表されるACOアルゴリズムを基にしたパターン抽出手法を提案するとともに、実際のマーケティングデータを用いて消費者の購買パターンについて抽出する実験を行った。実験の結果、30代女性の隔日で缶飲料を買っていたパターンが、連日の小型ペットボトルを購入するパターンに変化する様子が確認できた。

今後の課題として、リエージェントに対する適切なパラメータに対する知見の獲得が考えられる。また、現在のアルゴリズムでは、一度リエージェントにパラメータを設定したらそのパラメータに従い探索し続けるが、各リエージェントが入力データに対して柔軟に嗜好や嗜好間の間隔などを変化されることによってより適応性を向上させられることも考えられる。

また、実際のマーケティングデータなどの時系列データに対して提案するアルゴリズムを適用させることで、消費者の購買パターンなどの変化する様子を抽出することを目指す。

#### 参考文献

- [1] Dorigo, M., Maniezzo, V., & Colomi, A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1), 29-41.
- [2] Stulzle, T., & Hoos, H. H. (2000). MAX UTF2013MIN ant system. *Future generation computer systems*, 16(8), 889-914.
- [3] Agrawal, Rakesh, & Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215. 1994.*
- [4] Agrawal, Rakesh, & Ramakrishnan Srikant. "Mining sequential patterns." *Data Engineering, 1995. Proceedings of the Eleventh International Conference on. IEEE, 1995.*
- [5] Srikant, Ramakrishnan, & Rakesh Agrawal. "Mining sequential patterns: Generalizations and performance improvements." *Springer Berlin Heidelberg, 1996.*
- [6] Zaki, Mohammed J. "SPADE: An efficient algorithm for mining frequent sequences." *Machine learning* 42.1-2 (2001): 31-60.
- [7] Ayres, J., Flannick, J., Gehrke, J., & Yiu, T. (2002, July). Sequential pattern mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 429-435). ACM.
- [8] Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M. (2001, April). Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *proceedings of the 17th international conference on data engineering* (pp. 215-224).
- [9] Ho, C. C., Li, H. F., Kuo, F. F., & Lee, S. Y. (2006, December). Incremental mining of sequential patterns over a stream sliding window. In *Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on* (pp. 677-681). IEEE.
- [10] Iwanuma, K., Ishihara, R., Takano, Y., & Nabeshima, H. (2005, November). Extracting frequent subsequences from a single long data sequence a novel anti-monotonic measure and a simple on-line algorithm. In *Data Mining, Fifth IEEE International Conference on* (pp. 8-pp). IEEE.
- [11] Martens David, et al. "Classification with ant colony optimization" *IEEE Transactions on Evolutionary Computation* 11.5, pp.651-665, 2007.
- [12] Tamaki, Hiroshi, et al. "Pheromone Approach to the Adaptive Discovery of Sensor-Network Topology." *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 02. IEEE Computer Society, 2008.*
- [13] Arimura, H., & Uno, T. (2007). An efficient polynomial space and polynomial delay algorithm for enumeration of maximal motifs in a sequence. *Journal of combinatorial optimization*, 13(3), 243-262.
- [14] Ryang, Heungmo, Unil Yun, and Keun Ho Ryu. "Fast algorithm for high utility pattern mining with the sum of item quantities." *Intelligent Data Analysis* 20.2 (2016): 395-415.