

# 非負値多重行列因子分解と決定木学習による 行動パターンと属性情報の分析

小島 世大<sup>1,a)</sup> 石樽 隼人<sup>1</sup> 坂田 美和<sup>1</sup> 武藤 敦子<sup>1</sup> 森山 甲一<sup>1</sup> 犬塚 信博<sup>1</sup>

概要：近年 IC カードの普及やセンサデバイスの発達により、多岐にわたるデータの収集・蓄積が可能となり、それらのデータを用いて人の行動分析に関する研究がされている。それらのデータの多くは、非負の値を要素にもつ行列として表現することができる。これらを解析する手法として行列因子分解に基づく手法が注目を集めており、パターン抽出や推薦システムなどに利用されている。本研究では、データから非負値多重行列因子分解 (NMMF) と決定木学習を用いて人の頻出行動パターンと属性情報（性別、年齢など）の関係を分析する手法を提案する。最後に、われわれは提案手法を用いて入退室データの分析を行い、提案手法の有効性を確認する。

## An Analysis of the Relationship between Users' Behavior Patterns and Their Attributes Information Using Non-negative Multiple Matrix Factorization and Decision Tree Learning

### 1. はじめに

近年 IC カードの普及やセンサデバイスの発達などにより、多岐にわたるデータの収集・蓄積が可能となった [1]。これらのデータを用いて人の行動分析の研究が増えている [2][3][4][5][8][9][10]。企業などでは電子錠と IC カードにより、人の入退を制御する入退室管理システムの導入が増えている。現在では、企業は単純な入退室制御だけではなく、社員や部門ごとの入退室傾向の分析や社員がどの部屋にいるのかを把握したいなどの要望があり [6]、入退室管理システムから得られる入退室データを用いて社員属性ごとに行動履歴の定量的分析の研究もされている [7]。また、マーケティング環境において POS システムの普及が進んでおり、POS システムと会員カードを組み合わせることによって顧客別の購買履歴を捕捉している [11]。POS データを用いて、企業のマーケティング担当者は、市場の消費者をセグメントに分割し、各セグメントの特性を調べることによって、新商品のターゲットとする消費者群の決定やプロモ-

ーション計画の策定などの戦略立案を実施している [12]。以上のように、人の行動分析において属性情報は着目されている。また、これらの収集・蓄積されたデータの多くは、非負の値を要素にもつ行列として表現することができる。これらを解析する手法として行列因子分解に基づく手法が注目を集めており、パターン抽出や推薦システムなどに利用されている [13][14][15]。行列因子分解に基づく手法は人の行動分析にも利用されているが、その考察法は定まっていない [1][16][17][18][19]。

本研究では、データから行列因子分解に基づく手法のひとつである非負値多重行列因子分解 (Non-negative Multiple Matrix Factorization, NMMF) [1][20] を用いて人の頻出行動パターンを抽出し、決定木学習を用いて抽出された頻出行動パターンと属性情報（性別、年齢など）の関係を分析する手法を提案する。NMMF とは画像、文書、音声などさまざまなデータ分析において有効性が示されている非負値行列因子分解 (Non-negative Matrix Factorization, NMF) [13][21] を基礎とし拡張させ、複数の入力行列から頻出パターンを同時に抽出できる手法である [1]。NMMF を用いることで、たとえば、購買ログデータからは人がどこで何を買ったかに関する行動パターンを抽出でき、入退室データからはどのくらいの時間でどこに移動したかに関する行

<sup>1</sup> 名古屋工業大学 大学院 工学研究科 情報工学専攻  
Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology  
Showa, Aichi, 466-8555, Japan

<sup>a)</sup> s.kojima.571@nitech.jp

動パターンを抽出することができる。NMMF と決定木学習を組み合わせることで、抽出された各頻出行動パターンに当てはまる人はどのような属性情報で分割されていくのかを分析することができる。と考える。

最後に提案手法を用いて多くの組織で普及が進んでいる入退室管理システムから得られる入退室データの分析を行い、その有効性を確認する。

## 2. 関連研究

### 2.1 IC カード利用履歴を用いた行動分析に関する研究

IC カードの利用者数の増加により、IC カード利用履歴データが人の行動分析手段として注目を集めている。鈴木ら [3] は 交通 IC カードの利用履歴を用いて人が駅の改札口を出て、次の改札口に入るまでの間にその人の滞在目的があるとし、生活パターンを定量的に表す生活行動属性を提案した。カード利用者の生活圏(駅)によらず似た生活パターンをもつ人を抽出しマーケティングなどへ活用すると示した。また、嶋本ら [2] は英国・ロンドンで導入されている Oyster Card の 4 週間分の利用履歴データを用いて公共交通の変動を把握した。料金支払い形態に応じて利用者の利用属性を分類することで、1 人あたりの利用回数の変動の 4 割以上を説明できることを示した。坂田ら [7] は、企業の働き方改革などへの利用のため、入退室データを用いて会議室の入室回数と滞在時間を考慮した新たな指標(会議参加数)を活躍の指標として、社員属性ごとに分析する手法を提案した。提案手法を用いて入退室データを分析したところ職級、年代、部署で有意差のある結果が得られ、提案手法の活用の可能性を示した。

### 2.2 NMMF

NMMF[1][20] とは NMF[13][21] を拡張した手法である。NMF とは非負値で表された  $I \times J$  サイズの入力行列  $X$  を因子分解し、下記のように  $I \times R$ ,  $J \times R$  サイズの因子行列  $A, B$  の積の形へ分解する手法である。 $R$  は因子数である。

$$X \approx AB^T \quad (1)$$

因子分解では因子行列  $A, B$  の積と入力行列  $X$  の距離を最小化することで分解結果を得ることができ、得られた分解結果がデータ中のパターンに対応する。たとえば、購買データへ適用し、入力行列  $X$  (ユーザ  $\times$  商品) から潜在的な購買のパターンを因子行列  $B$  の因子  $r$  ( $1 \leq r \leq R$ ) の列を見ることで把握することができ、各ユーザがどの購買パターンにどの程度基づいて購買するのかを因子行列  $A$  の各ユーザ  $i$  ( $1 \leq i \leq I$ ) の行を見ることで把握することができる [1]。

NMMF[1][20] は NMF を複数の入力行列を対象として因子行列を共有することや補助行列と因子行列を共有して因

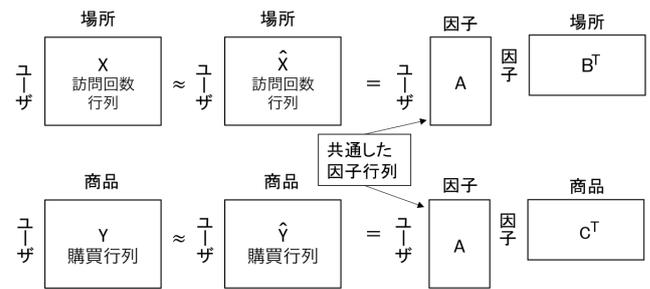


図 1 NMMF の例

Fig. 1 NMMF example

子分解するよう拡張した手法である。入力行列と因子行列の積の距離計算方法は入力行列によって複数存在するが、今回は本研究でも用いている行の指し示す事柄(ユーザなど)が等しい  $I \times J$  サイズ,  $I \times K$  サイズの入力行列  $X, Y$  の因子分解手法を説明する。この二つの入力行列  $X, Y$  は下記のように  $I \times R, J \times R, K \times R$  サイズの因子行列  $A, B, C$  に分解できる。

$$X \approx AB^T, Y \approx AC^T \quad (2)$$

ユーザがもつパターンを表す因子行列  $A$  を、二つの入力行列の因子分解に共通して用いることで同時にパターン抽出を可能とする [1]。たとえば、図 1 に示す訪問・購買データに適用することで、因子  $r$  ( $1 \leq r \leq R$ ) のパターンに強く基づくユーザは、因子行列  $B$  の因子  $r$  の訪問パターンに基づいてある場所によく訪問し、因子行列  $C$  の因子  $r$  の購買パターンに基づいて、ある商品を購入するといった結果が得られる [1]。NMMF に関する研究はほかにもされており、幸島ら [19] は消費者行動パターン抽出のために、従来の NMMF にユーザや商品などの属性情報に関する入力行列を加えた、属性情報を考慮した非負値多重行列因子分解法 (Non-negative Micro Macro Mixed Matrix Factorization, NM4F) の提案をした。提案手法と従来の NMMF の定量評価や因子行列をクラスタに分けて提案手法の定性評価を行っていた。

### 2.3 クラスタリングと決定木学習を用いた研究

クラスタリングとは、分類対象のデータ集合を類似するデータが同じクラスタになるように自動的に分類する教師なしデータ解析手法の一種である [22]。

決定木学習は学習データから自動的に決定木を作成する機械学習における一つの手法である。ここで決定木はデータ集合の値で入力データを分類する探索ノードからなるデータ分類の記述方式である。決定木学習は結果が木構造になっており、枝をたどることによってどのような属性の集まりでどのような結果が得られるのかを手で読み解きやすいという特徴をもっている [23]。

この二つの手法を用いて分析をしている研究がある。古川ら [26] はインターネットの YouTube にアップロードさ

れた国内外のチェリスト 46 名に対して、著者が目視によりチェリストのさまざまな属性についてのデータを獲得し、特徴を表すベクトルを用いてチェリストをクラスタリングし、説明変数にユーザ属性情報、目的変数をクラスタとして決定木分析を用いてクラスタの特徴付けを行った。

### 3. 提案手法

ユーザは性別、年齢などさまざまな属性情報を持っており、行動する際にも購買ログデータの分析では購買商品に対する移動経路の違い、移動履歴データの分析では移動時間や移動経路などさまざまな要素を持って行動しているため、ユーザの行動も属性情報もさまざまな要素から特徴を抽出することができる。これまでの方法では、単属性対単行動の相関分析 [24]、多属性対単行動の決定木学習を用いた分析 [26] が行われている。一方で、多属性対多行動の分析ができれば、さまざまな属性を持った人の駅の移動経路や平日休日の生活パターンの違いなど社会のセンシングへの応用が期待できる。

そこで本研究では多属性対多行動の分析のための、NMMF の因子分解結果を用いたユーザクラスタリングと決定木学習によるユーザの頻出行動パターンと属性情報の関係を分析する手法を提案する。本提案の想定する入力行列は購買ログデータや入退室データなどから得られるユーザの購買行動パターンや移動パターンなどを表現する複数の特徴行列である。購買ログデータであれば入力行列はユーザ×商品の購買回数の行列とユーザ×場所の訪問回数の行列などである。ここで、提案手法の流れを以下に示す (図 2)。

- (1) NMMF を用いて頻出行動パターンを抽出する。
- (2) 抽出された因子行列のうち、ユーザの各因子への寄与度を表した因子行列の各行を特徴量ベクトルとし、クラスタリングする。
- (3) クラスタリング結果を元に、説明変数を属性情報、目的変数をクラスタに属するユーザを 1、そのほかのユーザを 0 とし、各クラスタに対して決定木学習を用いて分類木を作成する。
- (4) 各クラスタ中心をクラスタに含まれるユーザの特徴とし、因子行列と分類木を照らし合わせることで、頻出パターンと属性情報の関係を分析する。

本提案手法の特徴は多属性対多行動の分析のために、NMMF と決定木学習を組み合わせる点である。決定木学習によりユーザがどのような属性情報で分かれていくのかを考察することができ、さらに NMMF により複合データから抽出した頻出行動パターンとの関係まで分析することができる。本提案の NMMF により得られた因子行列をクラスタに分けて属性情報の分析をする点で [19] と似ているが、クラスタリング結果を決定木学習に用いる点で異なる。ユーザは性別、年齢などさまざまな属性情報を持っており、そのユーザの行動パターンも単一ではなく、経路や経過時間

など複数のパターンを持っている。幸島ら [19] で提案した手法でも多属性対多行動の分析は可能であり、クラスタごとに各因子行列を見て、因子の中の値が高い順に並べることで、どの属性情報、どの行動がそのパターンに基づくのかといった分析ができるが、属性情報の集まりがどういった行動をするのかという分析はできない。しかし、本提案では行動と属性の関係の分析に決定木学習を用いるため、各クラスタに含まれるユーザの属性情報の集まりを分類木の枝をたどることで把握することができる。したがって、本提案手法を用いることで、ユーザの行動をさまざまな要素から特徴を抽出することができ、単属性の分析に限らず属性情報の組み合わせと行動パターンの関係を詳細に分析する際に有効である。

### 4. 提案手法を用いた入退室データの分析

#### 4.1 入退室データ

本研究では (打刻日、打刻時刻、場所、操作、社員 ID) の 5 つの属性で構成された入退室データを扱う。操作には、部屋を退室もしくは入室したかが記録されている。移動時間に着目するために、退室と入室がセットになっているものを結合し、(移動開始日、移動開始時刻、移動時間、移動元、移動先、社員 ID) の 6 つの属性で構成されたデータ構造 (以後移動データと呼ぶ) に変換する。ここで移動時間は退室して入室するまでの時間差 (入室時刻 - 退室時刻) とする。分析には協力企業の 2016 年 6 月の 332 名の移動データのうち、移動時間が 90 分以内の移動のみを用いた。

#### 4.2 提案手法に用いる入力行列の算出

提案手法の NMMF に入力する非負値の特徴行列を算出する。非負値で表した社員ごとの移動時間に関する特徴行列  $X$  (社員×移動時間) と場所に着目した部屋間の組に関する特徴行列  $Y$  (社員×部屋間の組) の二つの特徴行列の説明をする。

##### 4.2.1 移動時間特徴行列 $X$

$I \times J = 332 \times 9$  サイズの移動時間に関する特徴行列  $X$  (社員 × 時間区分) を算出する。移動時間特徴行列の要素  $x_{ij}$  は、社員  $i$  ( $1 \leq i \leq I$ ) が時間区分  $j$  ( $1 \leq j \leq J$ ) で移動した回数の割合とする [24]。分析に用いる時間区分  $j$  を時間帯 (午前・昼休憩・午後) × 移動時間 (0-5 分・5-20 分・20-90 分の 3 次元) の 9 種類で表した (表 1 左)。

##### 4.2.2 部屋間の組特徴行列 $Y$

$I \times K = 332 \times 6$  サイズの部屋間の組に関する特徴行列  $Y$  (社員 × 部屋間の組) を算出する。部屋間の組特徴行列の要素  $y_{ik}$  は、社員  $i$  が部屋間の組  $k$  ( $1 \leq k \leq K$ ) で移動した回数の割合とする。たとえば、ある社員が A 室から B 室の移動を 10 回しており、ある期間の総移動回数が 10 回の場合、部屋間の組特徴行列の要素の値は 1.0 となる。入力行列  $Y$  の部屋間の組は a: 移動前と移動後同じ部屋、b:

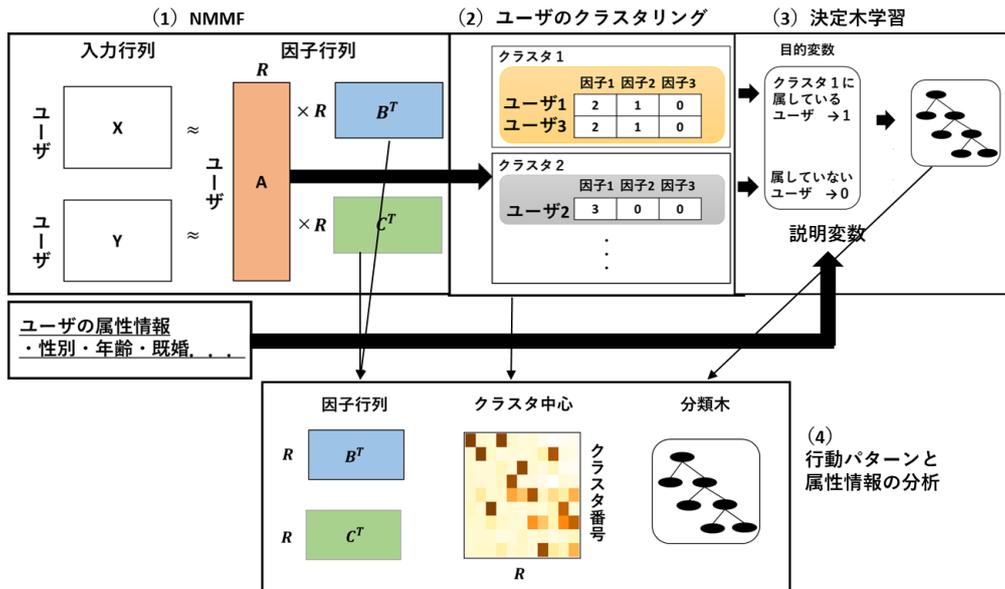


図 2 提案手法の流れ

Fig. 2 Flow of proposed method

表 1 時間区分と部屋間の組  
Table 1 Time divisions and pairs between rooms

時間区分			部屋間の組	
記号	時間帯	移動時間	記号	
11	午前	0-5 分	a	移動前と移動後同じ部屋
12	午前	5-20 分	b	移動前と移動後同じビル, 同じ階で違う部屋
13	午前	20-90 分	c	移動前と移動後同じビルで 違う階
21	昼休憩	0-5 分	d	第 1 ビルと第 2 ビルの間の移動
22	昼休憩	5-20 分	e	第 1 ビルと第 3 ビルの間の移動
23	昼休憩	20-90 分	f	第 2 ビルと第 3 ビルの間の移動
31	午後	0-5 分		
32	午後	5-20 分		
33	午後	20-90 分		

移動前と移動後同じビル, 同じ階で違う部屋, c: 移動前と移動後同じビルで違う階, d: 第 1 ビルと第 2 ビルの間の移動, e: 第 1 ビルと第 3 ビルの間の移動, f: 第 2 ビルと第 3 ビルの間の移動の 6 種類で表した (表 1 右).

### 4.3 NMMF によるパターン抽出

分析に用いる NMMF は [1] を参考にした. 本節では入退室データを NMMF で因子分解した行列分解形, 分析に用いる NMMF, 因子行列の初期値の設定, 因子数の決定方法と最後に抽出した因子行列について説明をする.

#### 4.3.1 行列分解形

4.2 節で算出した特徴行列  $X, Y$  と因子数  $R$  を入力として NMMF を用いると図 3 のように因子分解できる. NMMF での分解結果は, それぞれ  $I \times J, J \times R, K \times R$  サイズの非負値の因子行列  $A, B, C$  である. 因子行列  $A, B$  の積が入力行列  $X$ , 因子行列  $A$  と  $C$  の積が入力行列  $Y$  の近似  $\hat{X}, \hat{Y}$  である. 因子行列  $B^T, C^T$  は各因子  $r$  ( $1 \leq r \leq R$ )

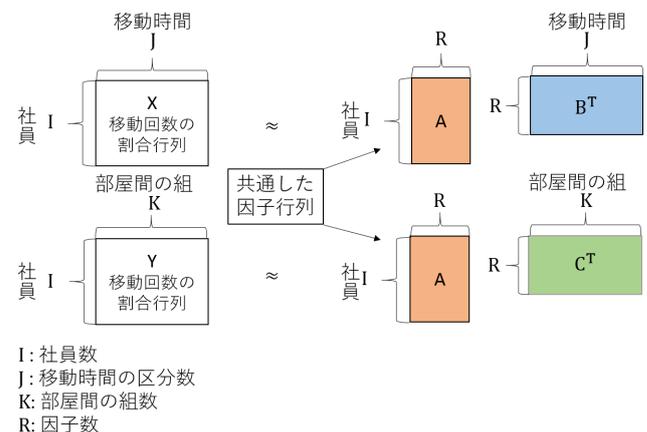


図 3 行列分解形

Fig. 3 Form of matrix decomposition

が移動の特徴を表わしており, 因子行列  $A$  は各社員がどの因子にどの程度基づくのかを表す.

#### 4.3.2 定式化

図 3 の近似  $\approx$  の尺度には, 式 (3) に示すユークリッド距離  $d_{EU}$  を利用した.  $x_{ij} \in X$  が入力行列の要素,  $\hat{x}_{ij} \in \hat{X}$  が近似行列の要素である.

$$d_{EU}(x_{ij}||\hat{x}_{ij}) = \frac{1}{2}(x_{ij} - \hat{x}_{ij})^2 \quad (3)$$

$d$  は行列の要素同士の距離であり, 行列同士の距離  $\mathcal{D}_{EU}$  は式 (4) のように表せる.

$$\mathcal{D}_{EU}(X||\hat{X}) = \sum_{i,j=1}^{I,J} d_{EU}(x_{ij}||\hat{x}_{ij}) \quad (4)$$

ここで, 入力行列  $X$  と因子行列の積  $\hat{X}$  の距離を  $\mathcal{D}_{EU}(X||\hat{X})$  (入力行列  $Y$  も同様  $\mathcal{D}_{EU}(Y||\hat{Y})$ ) とし, NMMF は式 (5) に

示す最適化問題を解くことで因子行列を出力する。因子行列を更新していく上で入力行列と因子行列の積の誤差（距離）を損失関数の値とする。

$$\arg \min_{A,B,C} \{D_{EU}(X||\hat{X}) + D_{EU}(Y||\hat{Y})\}$$

$$s.t. A, B, C \geq 0 \quad (5)$$

この最適化問題を解くアルゴリズムは複数存在するが、ここでは実装上の簡易さから利用されることの多い式(6)-(8)の乗法更新則に基づくアルゴリズム [1] を利用した。

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J x_{ij} b_{jr} + \sum_{k=1}^K y_{ik} c_{kr}}{\sum_{j=1}^J \hat{x}_{ij} b_{jr} + \sum_{k=1}^K \hat{y}_{ik} c_{kr}} \quad (6)$$

$$b_{jr} \leftarrow b_{jr} \frac{\sum_{i=1}^I x_{ij} a_{ir}}{\sum_{i=1}^I \hat{x}_{ij} a_{ir}} \quad (7)$$

$$c_{kr} \leftarrow c_{kr} \frac{\sum_{i=1}^I y_{ik} a_{ir}}{\sum_{i=1}^I \hat{y}_{ik} a_{ir}} \quad (8)$$

#### 4.3.3 因子行列の初期値の設定

入力行列  $X, Y$  は要素が割合で表してあることから各行の合計が1になることが分かる。そこで、本論文では更新をスムーズに進めるために、因子行列の初期値を式(9)-(11)に示すように設定した。

$$A \sim Uniform \left\{ 0, \frac{1}{\sqrt{JR}} + \frac{1}{\sqrt{KR}} \right\} \quad (9)$$

$$B \sim Uniform \left\{ 0, \frac{2}{\sqrt{JR}} \right\} \quad (10)$$

$$C \sim Uniform \left\{ 0, \frac{2}{\sqrt{KR}} \right\} \quad (11)$$

この初期値は入力行列の各行の合計が1になることを考慮し、因子行列同士を掛け合わせた行列の各行の合計が1に近くなるように設定している。これによって、入力行列に近い初期値を与えることが可能となる。

#### 4.3.4 因子数の決定方法

因子行列は因子数を増やすほど損失関数の値は減少していくため、損失関数の値の減少幅が明らかに小さくなった時、最適な因子数とする [27]。入力行列  $X, Y$  を因子分解した結果、因子数は損失関数の値の減少幅が明らかに下がった因子数9とする。因子分解により得られた3つの因子行列  $A, B, C$  のヒートマップを図4に示す。色の濃いマスほど値が大きく、色の薄いマスほど値が小さくなるように表示している。

#### 4.4 因子行列によるクラスタリング

社員がどの因子にどの程度基づくのかを示す因子行列  $A$  (図4)の各行を社員の特徴量ベクトルとし、社員のクラスタリングを行う。クラスタリングには k-means 法を用い、クラスタ数は elbow 法により 11 とした。クラスタ中心のヒートマップを図5に、クラスタに含まれる社員数を表2に示す。

表2 クラスタ内に含まれる社員数

Table 2 Numbers of employees included in clusters

クラスタ番号	1	2	3	4	5	6	7	8	9	10	11
社員数	21	108	11	29	5	13	46	8	14	26	51

#### 4.5 決定木学習を用いた分類木の作成

次に、クラスタリング結果を用いて決定木学習を行う。決定木学習に用いる説明変数を社員属性（性別、部署、職種、採用種別、年代、社歴、会議参加数 (NP)[7]）をとした。NP は1日あたりの平均会議参加数を表した指標であり、本研究で用いた入退室データと同期間のデータを用いた。決定木学習の最大の階層を3とした。全クラスタの分類木のうち、クラスタ人数の最も多いクラスタ2の結果を図6に示す。

#### 4.6 考察

##### 4.6.1 クラスタ2の分析

図4, 図5, 図6を用いたクラスタ2に対する頻出行動パターンと社員属性の関係を分析する。まず頻出行動パターンと社員属性を別々に解釈し、最後にそれぞれの関係を照らし合わせて分析する。

##### 頻出行動パターンの解釈

図5より、クラスタ2の108名(表2)の社員は因子7によく当てはまる。次に、図4の因子行列  $B$  と  $C$  の  $R7$  (因子7)を見て頻出行動パターンを分析する。移動時間の特徴を表した因子行列  $B$  より、すべての時間帯において20-90分かかる移動が多く、昼休憩には5-20分かかる移動も多いことが分かる。同様に部屋間の組の特徴を表した因子行列  $C$  より、同じ部屋に戻る移動が多いことが分かる。

##### 属性情報の解釈

図6の分類木のカバー率の高い点に着目して、クラスタ2の社員属性の特徴を分析する。分類木はある6つの部署で社歴が0-9年もしくは15-19年の職種が Manager, Sr.Specialist, 出向、役員、14名が93%の確率でクラスタ2に分類されていた。

##### 頻出行動パターンと属性情報の関係の分析

この6つの部署の社歴が短いもしくは長い、Sr.Specialist, 出向や職級の高い社員 (Manager, 役員) はすべての時間帯で長時間の移動が多い。協力企業への聞き取り調査によると20-90分の午前と午後の移動は社外で仕事、昼休憩では社外にランチに行っていると推測できる。

このように頻出行動パターンと属性情報を別々で解釈し、その後二つを照らし合わせることで、容易に頻出行動パターンと属性情報の関係を分析することができる。

##### 4.6.2 そのほかのクラスタの分析

そのほかの分析のうち特徴的なものを以下に紹介する。

- ある6つの部署の比較的職級が高い社員 (Manager, Sr.Manager, 役員, Specialist, 出向) の会議参加数が

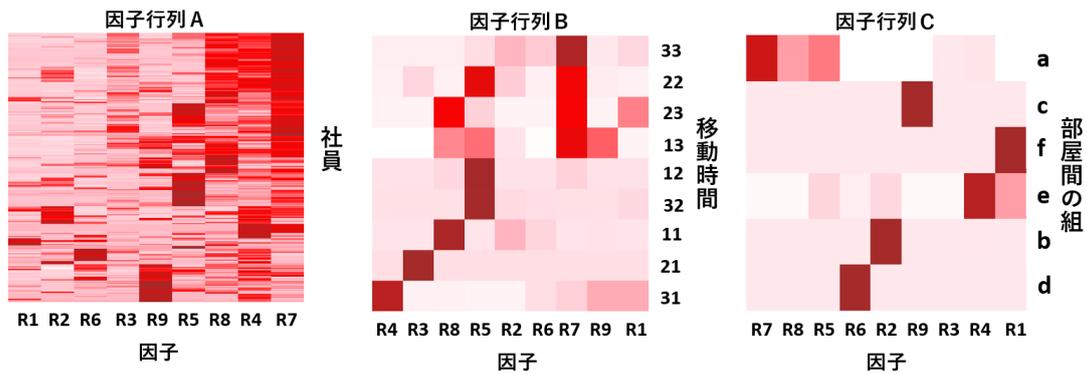


図 4 NMMF による因子分解の結果

Fig. 4 Results of factorization using NMMF

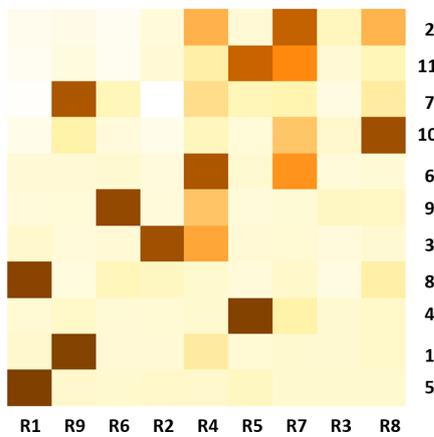


図 5 因子行列 A のクラスタリングによるクラスタ中心のヒートマップ, 行がクラスタ番号, 列が因子

Fig. 5 Heat map of cluster centers using clustering of the factor matrix A, the row is cluster numbers, the column is factor

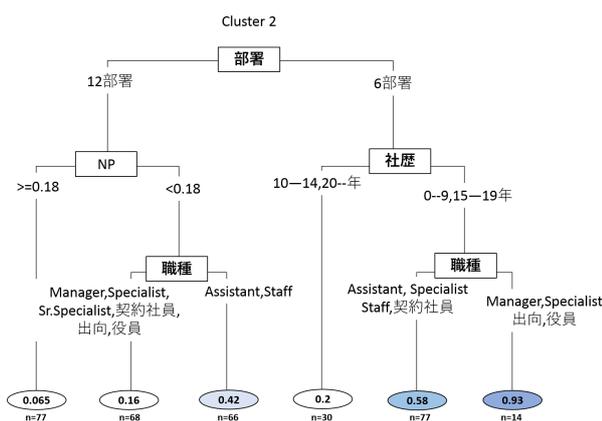


図 6 クラスタ 2 の分類木

Fig. 6 Classification tree of cluster 2

平均的な社員は午前社外で仕事することが多く、同じビルで違う階に移動することから、さまざまな場所に移動する傾向がある。

- ある 4 つの部署で職級が高い社員 (Sr.Manager, 役員)

や Staff の会議が少ない社員は小休憩を挟む傾向があるが、中には午前と午後に社外に仕事に行き、昼休憩には社外でランチをとる特徴ももつ。

これらの考察を比較すると、似た社員属性 (高い職級 (Sr.Manager, 役員)) をもつクラス同士でも部署 (それぞれの部署はすべて異なる) により違う行動をすることが分かった。このように決定木学習を用いることで、頻出行動パターンと属性情報の関係を分析する際にクラス同士の比較を多くの要素から比較できることも利点であると確認できる。

## 5. まとめ

本研究では NMMF により抽出された人の頻出行動パターンと属性情報の関係を決定木学習を用いて分析する手法を提案した。提案手法を用いてオフィスの入退室データから社員の移動時間と部屋間の組に関する頻出行動パターンと社員属性の分析を行ったところ、入退室データの分析により、結果の解釈の容易さや、クラス同士の比較を行うことで似た役割をもつクラスでも部署で行動パターンが異なるなど、提案手法を用いて頻出行動パターンと属性情報の関係を分析することの有効性を確認した。しかし、クラス数の決定方法に elbow 法を用いたが、分析結果の中で同じ特徴を持ったクラスがあったため、最適なクラス数とは言い切れないと考える。

今後の方針として、クラス数の決定方法の見直しや決定木学習の最大の階層の指定方法などが挙げられる。

**謝辞** 本研究を進めるにあたり、入退室データを提供して頂いた協力企業に感謝の意を表す。本研究は JSPS 科研費 JP18K18160 の助成を受けたものです。

## 参考文献

- [1] 幸島匡宏, 松林達史, 澤田宏: 複合データ分析技術と NTF [1] ——複合データ分析技術とその発展——, 電子情報通信学会誌, Vol. 99, No. 6, 2016.
- [2] 嶋本寛, 北脇徹, 宇野伸宏, 中村俊之: IC カード利用履

- 歴データを用いた公共交通需要変動分析, 土木学会論文  
集 D3(土木計画学), Vol.70, NO.5(土木計画学研究・論文  
集 第 31 卷), pp.605-610, 2014.
- [3] 鈴木敬, 相蘭敏子: 交通 IC カード利用履歴を用いた生活行  
動属性指標の提案, 信学技報, IEICE, Technical Report,  
LOIS2011-84, pp. 67-72, 2012.
- [4] 西内裕晶, 轟朝幸: 交通マーケティング手法検討のための  
IC カードデータを活用した利用者行動特性の把握, 土木  
学会論文集 F3(土木情報学), Vol. 68, No. 2, II.8-II.17,  
2012.
- [5] 矢部直人, 倉田陽平: 東京大都市圏における IC 乗車券を用  
いた訪日外国人の観光行動分析, Theory and Applications  
of GIS, Vol. 21, No. 1, pp. 35-46, 2013.
- [6] 佐藤雅之, 及川和彦, 永嶋規充: 入退室管理システムに  
おける通行履歴の応用, 情報処理学会第 77 回全国大会,  
2G-05, pp. 481-482, 2015.
- [7] 坂田美和, 小島世大, 森木田一真, 武藤敦子: 行動履歴の  
定量的分析によるオフィスワーカーの活躍評価, 第二回  
計算社会科学ワークショップ, 2018.
- [8] 岡田将吾, 神谷祐樹, 佐藤祐作, 藤田義弘, 山田敬嗣, 新  
田克己: センサ環境を利用したオフィスワーカーの行動  
パターン分析, 第 27 回人工知能全国大会, 1C4-1in, pp.  
1-4, 2013.
- [9] 桑原教彰, 野間春生, 鉄谷信二, 萩田紀博, 小暮潔, 伊関  
洋: ウェアラブルセンサによる看護業務の自動行動計測,  
情報処理学会論文誌, Vol. 44, No. 11, pp. 2638-2648,  
2003.
- [10] 大内一成, 土井美和子: 携帯電話搭載センサによるリアル  
タイム生活行動認識システム, 情報処理学会論文誌, Vol.  
53, No. 7, pp. 1675-1686, 2012.
- [11] 守口剛: 潜在クラス分析を用いたマーケット・セグメン  
テーション, 商学研究科紀要 (66), pp.1-13, 2008.
- [12] 阿部誠, 近藤文代: マーケティングの科学— POS データ  
の解析—, 朝倉書店, 2005.
- [13] D.D. Lee and H.S. Seung: *Learning the parts of objects  
by non-negative matrix factorization*, Nature, vol. 401,  
No. 6755, pp. 788-791, 1999.
- [14] Y. Koren, R. Bell, and C. Volinsky: *Matrix factoriza-  
tion techniques for recommender systems*, Computer,  
42.8, pp. 30-37, 2009.
- [15] S. Purushothm, L. Yan, and C-C.J. Kuo: *Collabora-  
tive Topic Regression with Social Matrix Factorization  
for Recommendation Systems*, In Proc. 29th International  
Conference on Machine Learning (ICML-12), pp. 759-  
766, 2012.
- [16] 平山明彦, 原直, 阿部匡伸: 非負値行列因子分解による  
PC 操作ログからの勤務パターン抽出, 信学技報, 114(32),  
pp. 33-38, 2014.
- [17] 李龍, 若宮翔子, 角谷和俊: Tweet 分析による群衆行動を  
用いた地域特徴抽出, 情報処理学会論文誌, データベ  
ース, Vol.5, No.2, pp. 36-52, 2012.
- [18] 桑野将司, 木下礼央, 福山敬, 谷本圭志, 菅原一孔, 川村  
尚生: 非負値テンソル因子分解を用いた検索ログデータ  
からのバス利用特性の分析, 土木学会論文集 D3 (土木計  
画学), Vol.73, No.5 (土木計画学研究・論文集第 34 卷),  
pp. L569-L578, 2017.
- [19] 幸島匡宏, 松林達志, 澤田宏: 属性情報を考慮した消費者  
行動パターン抽出のための非負値多重行列因子分解法, 人  
工知能学会論文誌, Vol. 30, No. 6, SPI-G, pp. 745-754,  
2015.
- [20] K. Takeuchi, K. Ishiguro, A. Kimura, and H. Sawada:  
*Non-negative Multiple Matrix Factorization*, Proceed-  
ings of the 23rd international conference on Artificial In-  
telligence, pp.1713-1720, 2013.
- [21] Lee, D.D. and Seung, H.S.: *Algorithms for Non-negative  
Matrix Factorization*, Proc. Neural Information Process-  
ing Systems (NIPS), pp. 556-562, 2001.
- [22] 吉田哲也, 岡山一宏: 制約を反映するグラフ構造に基づ  
く射影による半教師ありクラスタリング, 情報処理学会  
論文誌, 数理モデル化と応用, Vol. 4, No. 1, pp. 62-71,  
2011.
- [23] 古宮嘉那子, 但馬康宏, 小谷善行: 決定木を用いた敬語  
の説明ルールの獲得, 情報処理学会論文誌, Vol. 49, No.  
7, pp. 2679-2691, 2008.
- [24] Seidai Kojima, Hayato Ishigure, Miwa Sakata, Atsuko  
Mutoh, Koichi Moriyama and Nobuhiro Inuzuka: *An  
Analysis Method of Traveling-Time Patterns Between  
Rooms from Entry and Exit Data of Office Workers*,  
2018 IEEE 7th Global Conference on Consumer Elec-  
tronics(GCCE2018), pp.267-270, 2018.
- [25] 小島世大, 石樽隼人, 坂田美和, 武藤敦子, 森山甲一, 犬塚信  
博: オフィスワーカーの入退室データを用いた移動時間パ  
ターンの分析, 第 32 回人工知能全国大会, 302-OS-1b-01,  
2018.
- [26] 古川康一, 升田俊樹, 西山武繁: チェロ演奏動画の目視に  
よるデータ獲得と演奏スタイルの分類, 第 30 回人工知能  
学会, 1M4-OS-14a-3, 2016.
- [27] 安川武彦: 非負値行列因子分解を用いたテキストデータ  
解析, 計算機統計学, Vol.28, No.1, pp.41-55, 2015.