

トピックモデルに基づく宅内行動分析技術

八田俊之^{†1} 三輪祥太郎^{†1} 中川愛梨^{†2} 守谷一希^{†2} 諏訪博彦^{†2}
藤本まなと^{†2} 荒川豊^{†2} 安本慶一^{†2}

概要：従来の行動認識技術は教師付き学習に基づいて研究されてきたが、教師付きデータベースの作成に手間がかかる上、認識の対象行動を事前に定義しなければならないという問題があった。この問題に対して、本研究では教師なし学習であるトピックモデルに基づく行動モデリングを提案する。本稿では、家の中での実生活における加速度データベースを被験者実験により作成し、これに提案手法を適用した。その結果、提案手法によって行動のジャンルに相当する「移動の動作」「上半身を用いた動作」等の行動トピックを推定できた。また推定された行動トピックを用いて、実際の生活における行動を分類できることを確認した。

キーワード：トピックモデル，行動分析，ウェアラブルセンシング

1. はじめに

近年、スマートフォンやスマートウォッチ等のセンサデバイスが普及し、日常生活における人の行動センシングが活発になっている。また IoT (Internet of Things) や M2M (Machine to Machine) による家電をはじめとする機器のネットワーク化が進んでおり、これらの機器について人の行動を予測した制御のための研究が盛んになっている。日常生活における人の行動予測のためには、センサデータからの人の行動認識が重要となる。

従来の研究において、行動認識は教師付き学習の問題として扱われてきた。このため、事前に認識の対象となる行動を定義し、対象行動のラベルを付加した教師付き行動データベースを作成する必要があった。しかしながら、教師付き行動データベースは作成に多大な手間がかかるだけでなく、日常生活においては人の行動に制限がないため、事前に対象行動を定義すること自体が困難である。よって日常生活における人の行動認識のため、対象行動の事前定義が不要な行動モデリング手法が必要である。

上記の課題に対して、著者らは教師なし学習であるトピックモデルに基づく行動モデリングを提案した[1]。トピックモデルは自然言語処理分野で多く利用されており、文書にトピックモデルを仮定することで、単語の頻度分布で表された文書から、特定の単語の頻度分布で表されるトピックを推定できる。推定されたトピックは、文書の背後にある「料理」「政治」「スポーツ」といったジャンルに相当しており、文書は複数のトピックの混合率として表現できる。

提案手法では、行動のセンサ値を単語、時間窓で区切られた行動を文書と対応づけて、人の行動にトピックモデルを仮定する。これにより対象行動を事前に定義することなく、行動の背後にあるジャンルに相当する行動トピックを推定し、人の行動を複数の行動トピックの混合率として表

現できる。

文献[1]では基礎検討として、腰に装着したスマートフォンに搭載されている加速度センサを用いて、宅外での移動に関わる行動データベースを作成し、行動トピックを推定した。結果、推定された行動トピックは「静止している」「前後左右に動揺している」「足踏みしている」「着地している」「浮いている」といった行動のジャンルに相当することを確認した。これにより、宅外での移動に関わる行動は上記の行動トピックの混合率で表現できることが分かった。

本稿では、宅内における行動データベースにトピックモデルを適用することで、宅内での行動の背後にあるジャンルに相当する行動トピックを推定する。さらに推定された行動トピックの混合率による表現について、従来の行動認識研究[2]で使用されてきた行動ラベルを真値として用いることで分析し、その妥当性を考察する。

行動データベースは、奈良先端科学技術大学院大学内に設置された実験用住宅設備において、被験者に実際に生活してもらうことで作成した。また、センサデバイスとしてスマートフォンとスマートウォッチに搭載された加速度センサを用いた。

以降、2章では関連研究について述べる。3章では提案手法であるトピックモデルに基づく行動モデリングの詳細について述べる。4章では宅内における行動データベースの作成について述べる。5章では行動トピックの推定実験について述べる。6章では行動トピックによる行動の表現の分析について述べる。最後に、7章で結論を述べる。

2. 関連研究

トピックモデルを用いた人の行動モデルの関連研究について述べる。文献[3][4]では、日々の行動パターンの分析を目的として、本稿と同様に手首と腰に装着した加速度センサによる行動データベースに対してトピックモデルを適用している。加速度の統計的特徴量からクラスタリングによって単語を生成し、数十分の時間窓で区切られた行動を文書と対応付けることで、「通勤」や「夕食」といった抽象

^{†1} 三菱電機株式会社
Mitsubishi Electric Corporation

^{†2} 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

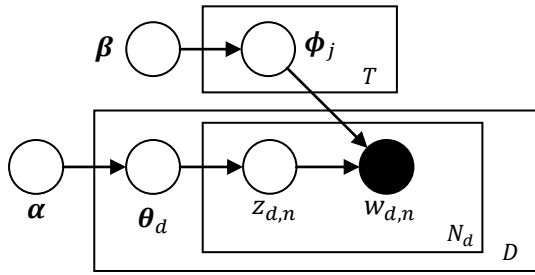


図 1 LDA のグラフィカルモデル

度の高い行動のジャンルに相当する行動トピックを推定している。

これに対して、本稿では宅内における行動パターンの分析を目的としている。加速度センサ値を単語、1分未満の短い時間窓で区切られた行動を文書と対応づけることで、より抽象度の低い行動のジャンルに相当する行動トピックを推定し、宅内における行動を表現する。

3. トピックモデルに基づく行動モデリング

提案手法では、トピックモデルにおける代表的な手法である潜在的ディリクレ配分法 (LDA; Latent Dirichlet Allocation) [5]を用いる。LDA のグラフィカルモデルを図 1 に示す。

提案手法において、行動は時間窓により D 個の区間に分割されており、各区間における行動がそれぞれひとつの文書に相当する。また、行動のセンサ値集合 (行動データベース) を $\mathbf{w} = \{\{w_{d,n}\}_{n=1}^{N_d}\}_{d=1}^D$ とする。ここで、 $w_{d,n}$ は区間 d における n 番目のセンサ値であり、1から V までの自然数をとる。 N_d は区間 d におけるセンサ値のデータ点数である。

区間 d における行動は、 T 個の行動トピックの混合率である行動トピック分布 $\theta_d = (\theta_{d,1}, \dots, \theta_{d,T})$ で表される。ここで、 $\theta_{d,j}$ は区間 d における行動トピック j の出現確率 (混合率) である。また、行動トピック j はセンサ値分布 $\phi_j = (\phi_{j,1}, \dots, \phi_{j,V})$ として表される。ここで、 $\phi_{j,v}$ は行動トピック j におけるセンサ値 v の出現確率である。

LDA によって仮定されるセンサ値 $w_{d,n}$ の生成過程を以下に示す。

- (1) ディリクレ事前分布 $\text{Dir}(\alpha)$ から、区間 d の行動を表す行動トピック分布 θ_d を生成する ($d = 1, \dots, D$)。
- (2) ディリクレ事前分布 $\text{Dir}(\beta)$ から、行動トピック j を表すセンサ値分布 ϕ_j を生成する ($j = 1, \dots, T$)。
- (3) 区間 d において、多項分布 $\text{Mult}(\theta_d)$ から、 n 番目の行動トピック $z_{d,n}$ を生成する ($n = 1, \dots, N_d$)。
- (4) 区間 d において、多項分布 $\text{Mult}(\phi_{z_{d,n}})$ から、 n 番目のセンサ値 $w_{d,n}$ を生成する ($n = 1, \dots, N_d$)。

ここで、 $\alpha = (\alpha_1, \dots, \alpha_T)$ と $\beta = (\beta_1, \dots, \beta_V)$ はそれぞれ θ_d と ϕ_j が従うディリクレ事前分布のパラメータ (ハイパーパラメータ) である。また、 $z_{d,n}$ はセンサ値 $w_{d,n}$ を生成する行動ト

ピックであり、1から T までの自然数をとる。

上記過程によって生成されたセンサ値 $w_{d,n}$ の頻度分布は、 T 個の行動トピックを表すセンサ値分布 $\{\phi_j\}_{j=1}^T$ の、行動トピック分布 θ_d を混合率とした混合分布となる (図 2)。

以上のように、提案手法を用いることで、行動を複数の行動トピックの混合率として表現できる。

3.1 行動トピックの推定

行動トピック j を表すセンサ値分布 ϕ_j は、センサ値集合 \mathbf{w} および行動トピック集合 $\mathbf{z} = \{\{z_{d,n}\}_{n=1}^{N_d}\}_{d=1}^D$ 、ハイパーパラメータ α, β によって推定できる。本稿ではギブスサンプリング [6] によって行動トピック集合 \mathbf{z} を推定する。ここで、行動トピック $z_{d,n}$ のサンプリング確率を以下に示す。

$$P(z_{d,n} = j | w_{d,n} = v, \mathbf{w}_{\setminus(d,n)}, \mathbf{z}_{\setminus(d,n)}, \alpha, \beta) \propto \frac{C_{j,\setminus(d,n)}^d + \alpha_j}{C_{\cdot,\setminus(d,n)}^d + \alpha} \cdot \frac{C_{v,\setminus(d,n)}^j + \beta_v}{C_{\cdot,\setminus(d,n)}^j + \beta} \quad (1)$$

ここで、 $C_{j,\setminus(d,n)}^d$ は行動 d における行動トピック j の生成回数、 $C_{v,\setminus(d,n)}^j$ は行動トピック j におけるセンサ値 v の生成回数であり、 \mathbf{w}, \mathbf{z} により得られる。また、 $C_{\cdot,\setminus(d,n)}^d = \sum_{j=1}^T C_{j,\setminus(d,n)}^d$ 、 $C_{\cdot,\setminus(d,n)}^j = \sum_{v=1}^V C_{v,\setminus(d,n)}^j$ 、 $\alpha = \sum_{j=1}^T \alpha_j$ 、 $\beta = \sum_{v=1}^V \beta_v$ であり、 $\setminus(d,n)$ は行動 d における n 番目の要素を除いた回数もしくは集合を表す。

また、ハイパーパラメータ α, β は、不動点反復法 [7] によって行動データベースから推定する。反復式を以下に示す。

$$\alpha_j \leftarrow \alpha_j \frac{\sum_{d=1}^D \Psi(C_{j,\setminus(d,n)}^d + \alpha_j) - D\Psi(\alpha_j)}{\sum_{d=1}^D \Psi(C_{\cdot,\setminus(d,n)}^d + \alpha) - D\Psi(\alpha)} \quad (2)$$

$$\beta_v \leftarrow \beta_v \frac{\sum_{j=1}^T \Psi(C_{v,\setminus(d,n)}^j + \beta_v) - T\Psi(\beta_v)}{\sum_{j=1}^T \Psi(C_{\cdot,\setminus(d,n)}^j + \beta) - T\Psi(\beta)} \quad (3)$$

ここで、 $\Psi(\cdot)$ はディガンマ関数である。

行動トピック j を表すセンサ値分布 ϕ_j は、行動トピック集合 \mathbf{z} の推定とハイパーパラメータ α, β の推定を規定の回数だけ交互に繰り返した後、以下の式により推定できる。

$$\hat{\phi}_{j,v} = \frac{C_{v,\setminus(d,n)}^j + \beta_v}{C_{\cdot,\setminus(d,n)}^j + \beta} \quad (4)$$

また、行動トピック分布 θ_d は以下の式により推定できる。

$$\hat{\theta}_{d,j} = \frac{C_{j,\setminus(d,n)}^d + \alpha_j}{C_{\cdot,\setminus(d,n)}^d + \alpha} \quad (5)$$

行動トピック j の相当する行動のジャンルは、混合率 $\hat{\theta}_{d,j}$ がとくに大きくなる区間 d における被験者の実際の行動との対応から確認できる。

3.2 新しい行動における行動トピックの混合率推定

推定された行動トピックを表すセンサ値分布を用いて、未知の行動における行動トピックの混合率 (行動トピック分布) を推定できる。以下に EM アルゴリズムを用いた手順を示す。

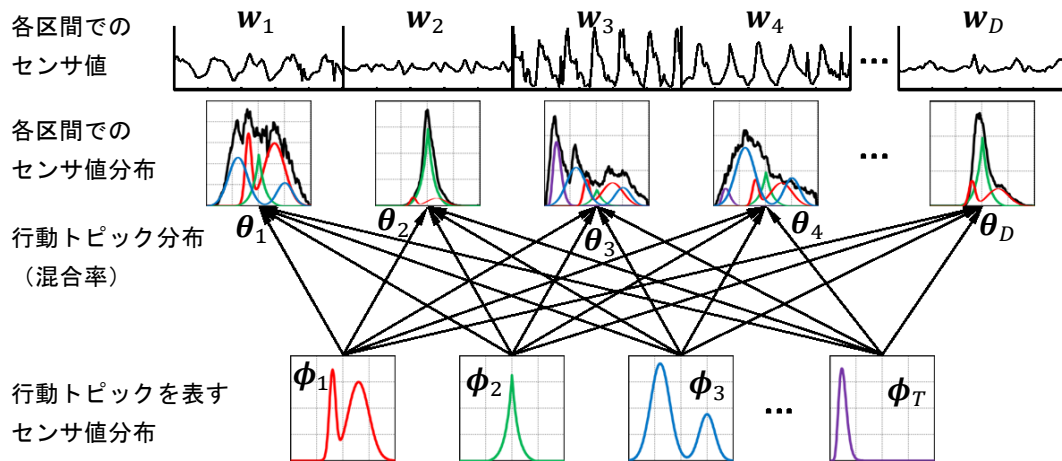


図 2 トピックモデルに基づく行動モデリング

(1) 未知の区間 d' における行動の行動トピック分布 $\theta_{d'} = (\theta_{d',1}, \dots, \theta_{d',T})$ の初期値を設定する。

(2) E-Step: 未知の区間 d' において、行動トピック j が n 番目のセンサ値 $w_{d',n}$ を生成した確率 $\psi_{d',n,j}$ を推定する ($j = 1, \dots, T$).

$$\psi_{d',n,j} = \frac{\theta_{d',j} \phi_{j,w_{d',n}}}{\sum_{k=1}^T \theta_{d',k} \phi_{k,w_{d',n}}} \quad (6)$$

(3) M-Step: 未知の区間 d' における行動の行動トピック j の混合率 $\theta_{d',j}$ を推定する ($j = 1, \dots, T$).

$$\theta_{d',j} = \frac{1}{N} \sum_{n=1}^{N_{d'}} \psi_{d',n,j} \quad (7)$$

(4) 手順 (2), (3) を規定の回数繰り返す。

以上の手順により、未知の行動における行動トピックの混合率を推定できる。

4. 宅内における行動データベースの作成

本章では、本研究において作成した宅内における行動データベースについて述べる。

図 3 に示すスマートホーム設備 (奈良先端科学技術大学院大学内に設置された実験用住宅設備) において、被験者 8 名 (男性 6 名, 女性 2 名, いずれも 20 代) にそれぞれ 2 日間実際に生活してもらい、計 16 日間の宅内における行動データベースを作成した。被験者には右腰にスマートフォン (Samsung GALAXY S5), 利き手とは逆の手首にスマートウォッチ (LG Watch Urbane) を装着してもらい、両デバイスに搭載されている 3 軸加速度センサによって、センサ値をサンプリング周波数 20 Hz で収集した。

実験期間中、被験者には最低でも 1 日 3 時間は活動し、10 時間以上 (睡眠時間含む) はスマートホーム設備で自由に生活するよう依頼した。ただし、あらかじめ定義された 15 種類の行動 (料理, 洗面, 食器洗い, 掃除, PC 使用, 食事, ゲーム, スマートフォン使用, 読書, テレビ視聴, 入浴, 風呂掃除, 洗濯, 外出, 睡眠) をできるだけ 1 回以

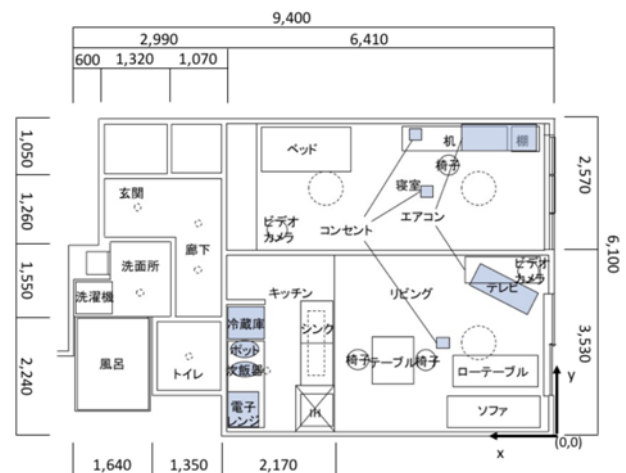


図 3 スマートホーム設備の間取り

上行うよう依頼した。また、外出や入浴, 風呂掃除, 睡眠の際はセンサデバイスを装着しないよう依頼した。

4.1 センサデータ処理

スマートフォンおよびスマートウォッチから収集された 3 軸加速度に対して、それぞれノルムを算出し、0.69 g から 1.31 g まで 0.02 g 刻みで量子化した。これにより両デバイスから 31 種類ずつ、合わせて 62 種類のセンサ値 (単語) を得た。また、センサ値を 1 分毎の区間に分割し、1 分間 (1200 点) の行動がそれぞれひとつの文書に相当するとした。

4.2 行動の記録および行動ラベルの付加

実験期間中の生活の様子を、寝室とリビングに設置した 2 台のビデオカメラとスマートフォンの音声記録アプリによって記録した。これらの記録は被験者本人の管理の下、本人のみが視聴できる。またこれらの記録を用いて、被験者本人によって各時刻における 11 種類の行動ラベル (料理, 洗面, 食器洗い, 掃除, PC 使用, ゲーム, スマートフォン使用, 読書, テレビ視聴, 行動ラベルなし) を付加した。これらの行動ラベルは、従来の行動認識研究 [2] において使用されてきたものであり、本稿では提案手法による行動の表現の分析に用いる。

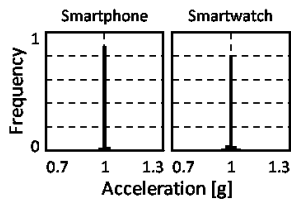


図 4 1日の行動の
センサ値分布

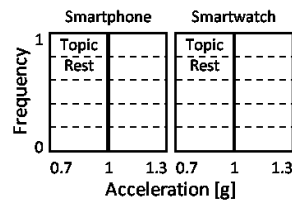


図 5 静止行動トピック
を表すセンサ値分布

5. 行動トピックの推定

本章では、前章で述べた宅内での行動データベースに対して、行動トピックの推定を行う。また、推定された行動トピックが相当する行動のジャンルについて確認する。

5.1 実験条件

行動トピックの推定は、計 16 日間ある行動データベースについて 1 日間ずつ行った。ギブスサンプリングの回数 1 回、不動点反復法の回数 50 回として、これを交互に 200 回繰り返した。行動トピックの相当する行動のジャンルは、特定の行動トピックの混合率が特に高くなる区間について、実験期間中に撮影した映像から被験者本人によって確認した。

また図 4 に、1 日の行動から得られた 62 種類のセンサ値の頻度分布の例を示す。図より、重力加速度(1 g)に対応するセンサ値の頻度が極めて高いことが分かる。これは宅外での移動に関わる行動[1]にはなかった特徴である。宅内では「PC 使用」「食事」「ゲーム」「スマートフォン使用」「読書」「テレビ視聴」といった動きの少ない静的な行動があるため、1 g 付近のセンサ値が頻出していると推測できる。上記のように特定のセンサ値が行動データベース内で頻出する場合、推定される行動トピックに悪影響を与える可能性がある。自然言語処理においては特定の高频出語をストップワードとして、トピックを推定する処理の対象から除く場合が多い。

本稿では、スマートフォンおよびスマートウォッチの両デバイスから 0.09 g 以上 1.01 g 未満のセンサ値が 0.2 秒 (4 点) 以上連続して得られたとき、それらを「静止状態」に相当する静止行動トピック (図 5) から生成されたと仮定し、行動トピックを推定する処理の対象から除いた。またこのとき、行動データベースの各区間で処理の対象となるセンサ値のデータ点数が減少するため、データ点数が 10% (120 点) 未満となる区間を行動データベースから除いた。

5.2 実験結果

行動トピックを表すセンサ値分布の推定結果の一例を図 6 に示す。まず、静止状態を事前に除かず、トピック数 $T = 5$ として推定した結果を図 6(a) に示す。0.09 g 以上 1.01 g 未満のセンサ値が高频度である行動トピックが複数推定されており (トピック 1, 2, 5), これらはすべて「静止状態」に相当すると考えられる。

表 1 行動トピックが相当する行動のジャンル

行動トピック	被験者の状態
静止行動トピック	静止
トピック a	移動の動作 例: リビングを歩く 例: 掃除機をかけながら移動する
トピック b	上半身を用いた動作 例: ドライヤーで髪を乾かす 例: 洗濯物をたたむ
トピック c	手を用いた動作 例: 片栗粉を袋から出す 例: 食器を洗う
トピック d	手をついた動作 例: ソファで姿勢を変える 例: 机で文字を書く

これに対して、処理対象から静止状態を事前に除き、トピック数 $T = 4$ として推定した結果を図 6(b) に示す。1 行目には事前に仮定した静止行動トピックを示している。推定された 4 つの行動トピックを表すセンサ値分布の形状はそれぞれ異なり、「静止状態」以外のジャンルに相当すると考えられる。また、推定された行動トピックのセンサ値分布は、異なる被験者間で類似した形状となった。

つぎに、それぞれの区間における行動トピック分布の推定結果の一例を図 7 に示す。横軸は時刻であり、縦軸は行動トピックの混合率である。ただし、ここでの混合率とは、各区間から「静止状態」に相当するセンサ値が除かれた残りのデータ点数における混合率である。

特定の行動トピックの混合率がとくに高くなる区間について、実際の行動を確認した結果をグラフの上部に示した。この図より、トピック a は「歩く」ときに混合率が高いほか、トピック c は「片栗粉を袋から出す」「豚肉を豆腐に巻く」など手を用いているときに混合率が高いことが分かる。このようにして確認された行動トピックが相当する行動のジャンルを表 1 にまとめる。

以上より、トピック数 $T = 4$ として推定した宅内での行動トピックは、それぞれ「移動の動作」「上半身を用いた動作」「手を用いた動作」「手をついた動作」といった行動のジャンルに相当することが分かった。

6. 行動トピックによる行動の表現

本章では、前章で推定された行動トピックの混合率による行動の表現について、従来の行動認識研究[2]で使用されてきた 11 種類の行動ラベルを真値として用いることで分析し、その妥当性を考察する。

6.1 実験条件

各被験者の実験期間 2 日間のうち、一方の 1 日から推定された行動トピックを用いて、他方の 1 日における行動の行動トピックの混合率を算出した。また、前章で推定した行動トピックに加えて、事前に仮定した静止行動トピックを合わせて用いた。EM アルゴリズムの繰返し回数は 20 回とした。

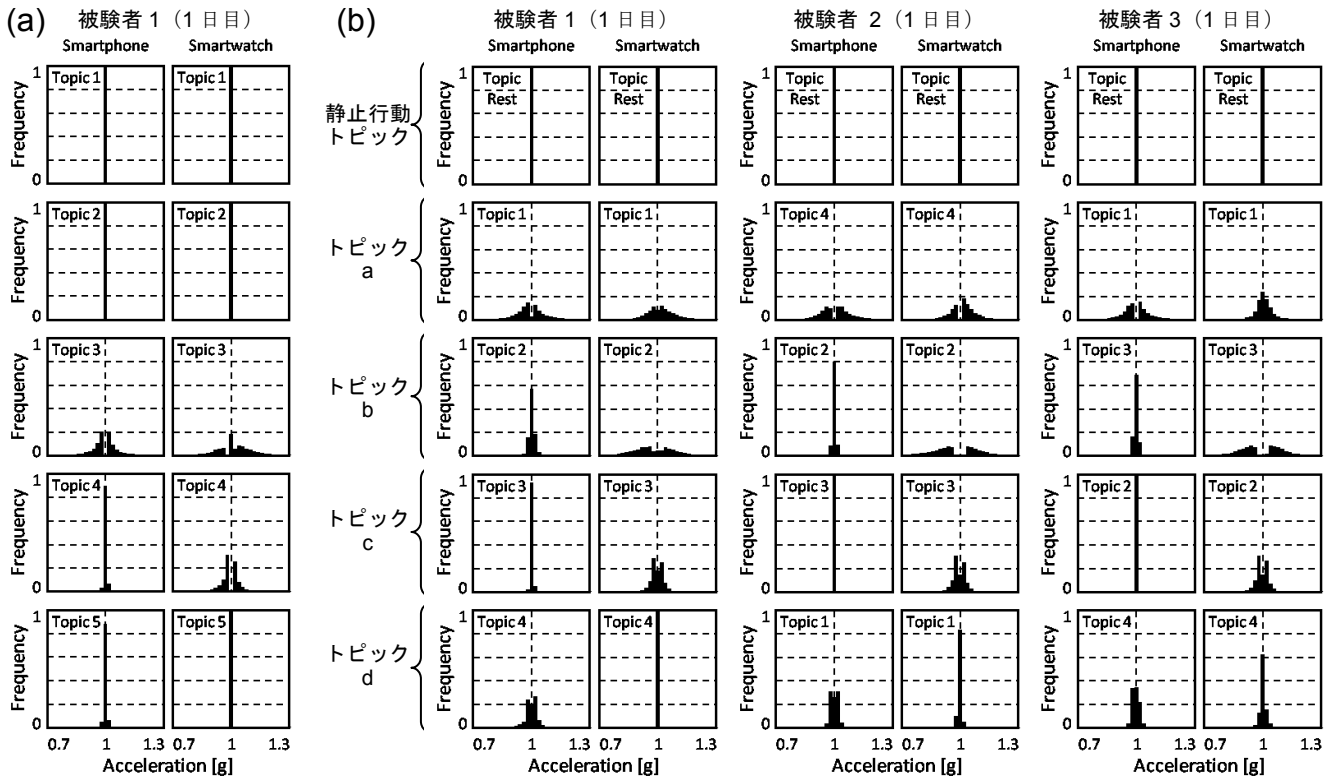


図 6 行動トピックを表すセンサ値分布 (a) 静止状態を除かない場合, (b) 静止を除いた場合

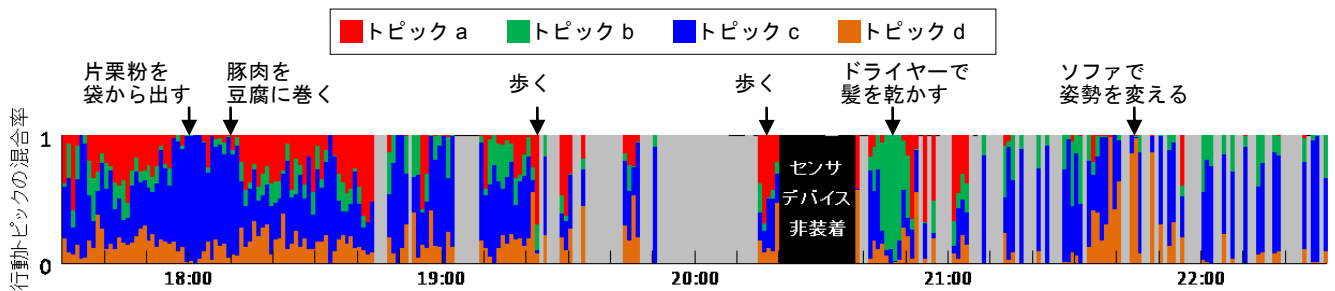


図 7 行動トピックと対応する実際の行動 (被験者 1 (1 日目))

6.2 実験結果

結果を図 8 に示す. 縦軸は行動トピックの混合率, 横軸は時刻である. グラフの上部に示した行動ラベルは, 見やすさのため 5 分以上継続した行動ラベルのみを示した. 図より, 同一の行動ラベルが付加された多くの区間で, 類似した混合率となっている.

本稿では, 1 日のうち同一の行動ラベルが付加された区間における行動トピックの平均混合率を算出し, これを行動ラベルに対する代表的な混合率とした. 例として被験者 3 名の「料理」「食器洗い」「掃除」「PC 使用」の平均混合率を図 9 に示す. 図より, 同一の行動において行動トピックの平均混合率が類似していることが分かる. また, 「食器洗い」では「手を用いた動作」(トピック c), 「掃除」では「移動の動作」(トピック a), 「PC 使用」では「静止」(静止行動トピック) がいずれの被験者においても最も平均混合率が高く, 直感と一致した結果となっている. なお, 「料

理」ではいずれの行動トピックの平均混合率も同程度である. これは「料理」が多様な動作の混ざった行動ラベルであり, 平均混合率が平滑化されたためと考えられる.

また, 11 種類の行動ラベル間における行動トピックの平均混合率について, 類似度をユークリッド距離によって評価し, 類似度行列を作成した (図 10). 被験者 8 名の 2 日間について行動ラベルが 11 種類あるため, 類似度行列は 176×176 の対称行列となった.

類似度行列の対角成分に注目すると, 同一の行動ラベルにおける平均混合率が類似していることが分かる. また, 6 種類の静的な行動ラベル (「PC 使用」「食事」「ゲーム」「スマートフォン使用」「読書」「テレビ視聴」) はすべて高い類似度となった. その他の活動的な行動ラベルでは, 移動を中心とした「掃除」と, 手を中心とした「食器洗い」は他に強く類似する行動がないが, 「行動ラベルなし」「料理」「洗面」(歯磨きや洗顔の動作を含む) が高い類似度を示し

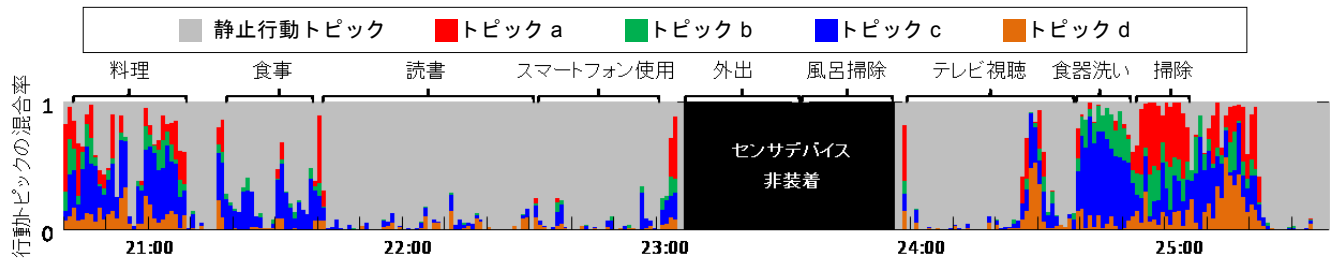


図 8 行動トピックの混合率と行動ラベル (被験者 2 (2 日目))

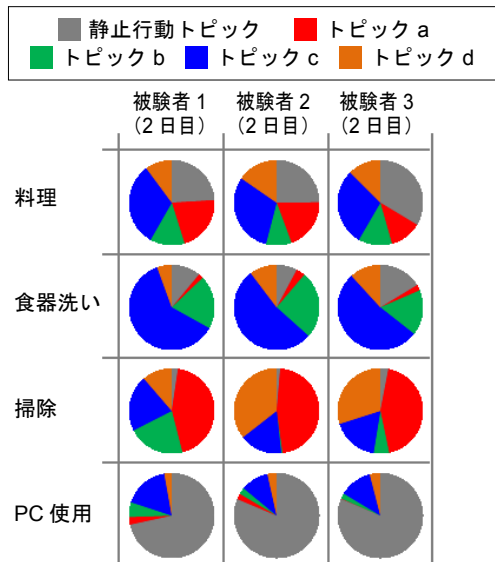


図 9 各行動ラベルにおける平均混合率

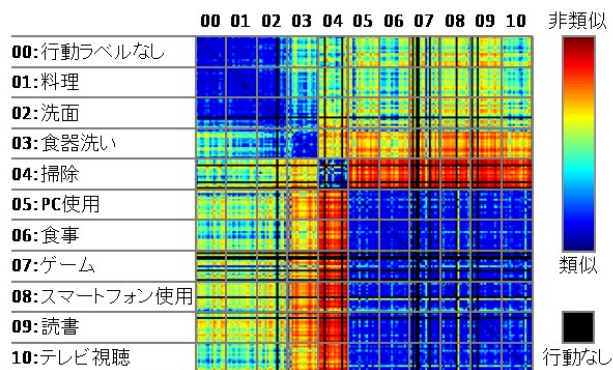


図 10 平均混合率の類似度行列

た. これらは多様な動作の混ざった行動ラベルであるため平滑化された平均混合率となり, 類似度が高くなったと考えられる.

以上より, 本稿で推定した行動トピックを用いて, 異なる被験者において共通的な行動の表現が可能であることが分かった. また, 従来の行動認識研究で使用されてきた行動ラベルのうち, 活動的な行動ラベルについて3つに分類(移動を中心とした行動ラベル, 手を中心とした行動ラベル, 多様な動作の混ざった行動ラベル)できることが分かった. また, 動きの少ない静的な行動ラベルは分類できないことを確認した.

7. おわりに

本稿では, 日常生活における室内での人の行動モデリング手法として, 対象行動の事前定義が不要なトピックモデルに基づく行動モデリングを提案した.

スマートホーム設備を用いた被験者実験によって加速度センサによる行動データベースを作成し, 提案手法を適用した. その結果, すべての被験者から共通的に「静止」「移動の動作」「上半身を用いた動作」「手を用いた動作」「手をついた動作」に相当する行動トピックを得た. これらの行動トピックを用いて, 異なる被験者間で共通的な行動の表現が可能である. また事前に定義された行動ラベルのうち, 活動的な行動ラベルについて, 3つに分類できることを確認した. 以上より, 提案手法は室内での活動的な行動に対して有用であることが分かった.

今後, 室内における静的な行動を分類するため, 装着した加速度センサだけでなく, 環境に固定されたセンサ(位置センサ, 家電機器の電力センサなど)の利用が有用と考えられる. また, 体温や心拍といった生体センシングによって, 人の感情や気分の要素を含んだ行動トピックの推定が可能になると考えられる.

参考文献

- [1] 八田俊之, 三輪祥太郎. トピックモデルに基づく人行動分析技術. 第14回情報科学技術フォーラム (FIT2015) 講演論文集, 第4分冊, pp. 367-370 (2015).
- [2] Ueda, K., Suwa, H., Arakawa, Y. and Yasumoto, K.. Exploring Accuracy-Cost Tradeoff in In-Home Living Activity Recognition based on Power Consumptions and User Positions. Proc. of the 14th IEEE International Conference on Ubiquitous Computing and Communications (IUCC 2015), pp. 1130-1137 (2015).
- [3] Huynh, T. and Fritz, M.. Discovery of Activity Patterns using Topic Models, Proc. of the 10th ACM International Conference on Ubiquitous Computing (UbiComp 2008), pp. 10-19 (2008).
- [4] Sun, F. T., Yeh, Y. T., Cheng, H. T., Kuo, C. and Griss, M.. Non-parametric Discovery of Human Routines from Sensor Data. Proc. of the 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom 2014), pp. 11-19 (2014).
- [5] Blei, D. M., Ng, A. Y. and Jordan, M. I.. Latent Dirichlet Allocation. Journal of Machine Learning Research, Vol. 3, pp. 993-1022 (2003).
- [6] Griffiths, T. L. and Steyvers, M.. Finding Scientific Topics. Proc. of the National Academy of Sciences, Vol. 101, pp. 5228-5235 (2004).
- [7] Minka, T. P.. Estimating a Dirichlet Distribution. Technical Report (Massachusetts Institute of Technology), pp. 1-8 (2000).