

# 運動状態のパターンを用いた 異なる期間の人の生活特性比較に関する一考察

磯村 洋<sup>†</sup> 新谷 隆彦<sup>†</sup> 大森 匡<sup>†</sup>

<sup>†</sup> 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

E-mail: [fisomura.hiroshi@hol.is.uec.ac.jp](mailto:fisomura.hiroshi@hol.is.uec.ac.jp), [†{shintani,omori}@is.uec.ac.jp](mailto:†{shintani,omori}@is.uec.ac.jp)

**あらまし** 我々はリストバンド型センサで人の活動に伴う動きを常時測定し、いつからいつまでどのような運動状態であったかを意味する運動データをライフログとして収集している。本研究では、これらのデータから異なる2つの期間の生活を比較する手法を提案する。提案手法は日常的に行われる行動を示す運動データのパターンをエピソードマイニングによって抽出し、抽出した運動データのパターンの中で各期間において行われたパターンの違いによって生活を比較する。報告者が収集した602日間に渡る運動データを用いた実験を行い、提案手法において、抽出したパターンと行動の関係における問題点を解決することで生活を比較することができることを示した。

**キーワード** ライフログ, エピソードマイニング, データマイニング, 実データ解析

Hiroshi ISOMURA<sup>†</sup>, Takahiko SHINTANI<sup>†</sup>, and Tadashi OHMORI<sup>†</sup>

<sup>†</sup> Graduate School of Information Systems, The University of Electro-Communications

1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585 Japan

E-mail: [fisomura.hiroshi@hol.is.uec.ac.jp](mailto:fisomura.hiroshi@hol.is.uec.ac.jp), [†{shintani,omori}@is.uec.ac.jp](mailto:†{shintani,omori}@is.uec.ac.jp)

## 1. はじめに

近年、パーソナルコンピュータや小型端末、安価なセンサの普及により人の生活に関する長期間のデータ（ライフログ）が蓄積されるようになってきている。さらに、蓄積されたライフログを利用して、人の生活の振り返りを促すサービスが行われている。例えば、自分が読んだ本を記録していくことで、どのような本を読んでいたのか振り返る「読書メーター」[1]、測った体重と体脂肪率を自動的に記録し、変化を確認する「からだログ」[2]等が存在する。このようなライフログによるサービスの1つとして、加速度センサを携帯することで利用者の運動状態に関するデータを蓄積し、1日の歩行数、運動量、睡眠時間等を図示する製品・サービスが増えてきており、例えば「Fibit」[3]、「Jawbone Up」[4]、「NIKE+FuelBand」[5]等がある。これらのサービスを使うことで利用者は自身の過去の生活を振り返ることができる。

これまでのライフログサービスは、過去の一時点のライフログデータを提示することで短期間の生活の振り返りを支援を行っていた。今後期待されるライフログサービスの1つとして、長期間の生活の振り返りの支援が挙げられる。

本研究では、数週間以上の生活の振り返りの支援のために、異なる2つの期間の生活を比較する手法を提案する。生活の比

較とは、どの程度同じような生活をしてきたかを評価することである。ある期間の生活を振り返る際に、同じ生活をしてきた別の期間、違う生活をしてきた別の期間が分かれば、当時の生活を振り返ることに役立つ。本研究ではそれぞれの期間で同じ行動が多く行われていた場合は同じ生活、違う行動が多く行われていた場合は違う生活をしてきたとする。提案手法では、それぞれの期間に行った行動の違いによって生活を比較するアプローチを採る。実際に行った行動の違いによって生活を比較する場合、具体的な行動内容を長期間に渡って記録する必要がある。しかし、具体的な行動内容を長期間に渡って記録することは困難である。このため本研究では、リストバンド型センサで得られた運動状態のデータを行動内容の代わりに用いる。運動状態のデータは行動を表していると考えているためである。

異なる2つの期間の生活を比較する手法として、本研究では日常的に行われる行動を示す運動データのパターンをエピソードマイニングで抽出し、それぞれの期間で行われたパターンの集合の類似度によって生活を比較するアプローチを提案する。さらに、抽出した運動データのパターンが持つ問題点を解決するため、頻度の尺度の変更、サブパターンの除外、運動データ間の時間差の制限、運動データの運動時間による区別を行う。そして、報告者が蓄積した602日間のライフログデータを利用して、提案する手法によって生活を比較することができるかを

検証する。

## 2. 関連研究

加速度等から人の行動や状態を推定する研究はこれまで数多く行われてきている。これらの研究が進み、人の行動のデータを正確に得ることができるになれば、人の生活の評価に利用できるようになる。このような人の行動認識の研究の1つとして、利用者の移動状態を推定するシステムの研究に小林らの積迦 [7] がある。積迦は、携帯電話上の GPS や加速度センサ、マイクを複合的に用いてユーザの移動状況を推定する。また、人間行動のセンシングを通じて行動認識のための大規模なデータベースの構築を目指す HASC [6] のようなプロジェクトもある。本研究で用いるライフログデータはこれらの研究が目指すような具体的な行動内容では無い。

人の生活を評価する研究は昔から行われてきた。人の生活を評価する研究の1つに、Monk らによる SRM (Social Rhythm Metric) [8] がある。SRM は生活の規則正しさをアンケートにより集計したある期間内の日常的な 17 種類の生活行動の時刻 (例: 起床時刻, 働き始めた時刻, 夕食の時刻等) の分散によって評価する指標である。SRM は我々が一般的に想像する「規則正しさ」の考え方に合致していて感覚的に受け入れやすいこともあり、広く利用されている。実際に双極性障害の予兆発見等で一定の有用性が確認されている [9]。他に、睡眠と覚醒の状態の記録から老人の生活リズムについての調査 [10] 等がある。このような研究では、評価に用いるデータを主にアンケート等の手作業で収集していた。

センサが安価に普及するようになったことで、センサを用いて自動的に収集した情報から人の生活の状態を評価する研究が行われてきた。青木ら [13] は室内センサのデータを 10 分単位で 11 種類の行動にラベリングして作ったデータから過去 30 日分のデータを用い、ある 1 時間の行動が日常化非日常か判断する研究を行った。老人の見守りの研究には他に、木川ら [14] による、室内センサのデータから行動パターンを作り、1 日の全ての行動パターンに対する各行動パターンが占める割合を用いて生活リズムの変化を示すシステムの研究等がある。本研究では部屋にいる間だけでなく、1 日中の行動を表すライフログデータを用いる。

室内センサ以外には、河本ら [11] による、本研究と同じライフログデータを利用して人の生活の典型的な日の遷移列によって生活を表すモデルの提案を行った研究がある。

## 3. 本研究で用いるライフログデータ

### 3.1 運動データの定義

本研究で扱うライフログデータについて述べる。本研究において報告者は日立製作所製の日立リストバンド型生活モニタ装置 HT-PB3 [12] を常に装着し、1 日中の体の動きに関するデータを収集している。HT-PB3 は 3 軸加速度や温度を常時測定すると共に、この加速度データからいつからいつまでの程度の運動状態を続けていたか示すデータを算出する。この運動状態のデータを運動データと呼ぶ。運動データはある

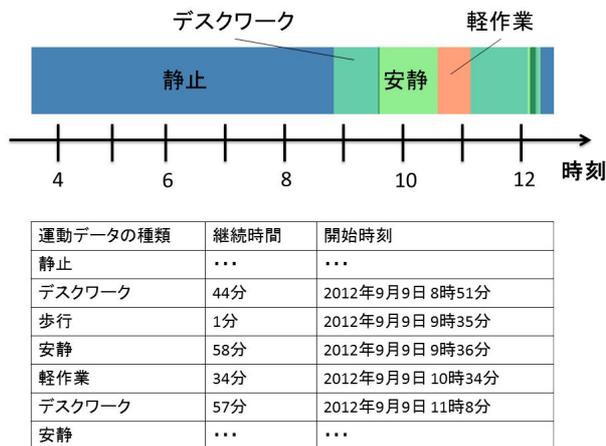


図 1 運動データ

時刻  $ts$  からある時刻  $te$  まで ( $ts < te$ ) の運動の激しさの程度を 9 種類に分けて表すデータで、途切れることなく連続して得ることができる。1 つの運動データは運動状態  $m$ , 運動時間  $d = te - ts$ , 開始時刻  $ts$  を持つ。時刻と継続時間の単位は“分”である。

運動状態の種類は 1. 静止, 2. 安静, 3. デスクワーク (座位活動), 4. 軽作業, 5. 作業, 6. 運動, 7. 歩行, 8. ジョギング, 9. 非装着の 9 種類である。静止から運動までの運動時間状態は、順に激しい運動をしていたことを示している。歩行は周期的な運動をしていたこと、ジョギング周期的で強い運動をしていたことを示している。HT-PB3 が算出した運動データの誤差は本研究では考慮しない。

図 1 の表は報告者の実際の運動データの一部である。図 1 の上の図は下の表の動作データを図示したものである。青が静止状態、水色がデスクワーク、黄緑が安静、橙色が軽作業、緑が歩行を示している。

### 3.2 期間内の運動データ

本研究では 2 つの異なる期間を指定し、その期間内に行われた運動データを用いて生活を比較する。期間  $T$  は開始日と終了日で指定される。期間  $T$  内に行われた運動データは、開始日の 0 時 0 分 ( $start$ ) 以後に開始し ( $start \leq ts$ )、終了日の次の日の 0 時 0 分 ( $end$ ) までに終了した ( $te < end$ ) 全ての運動データである。

## 4. 生活の比較

### 4.1 運動データを用いた生活の比較

本研究ではそれぞれの期間で同じ行動が多く行われていた場合は同じ生活、違う行動が多く行われていれば違う生活をしたと考える。このため提案手法では、実際に行った行動の違いによって生活を比較する。ここで行動とは生活において目的を持った一連の活動を指す。例えば、大学の授業期間内の 2 つの期間について考えた場合、どちらも通学や授業、勉強等の主な行動が同じであるため、2 つの期間は同じ生活をしたとする。逆に、毎日大学に通学している期間と、夏休みの旅行中の期間について考えた場合、後者は普段と異なる移動や活動

(レジャーや観光)等の行動を行っているため、前者とは行動が異なる。このように2つの期間において行われる行動が異なる場合、2つの期間は違う生活をしていたとする。

運動データは実際に行われた行動に伴う動きを表しているため、具体的な行動内容の代わりに運動データを用いて生活を比較することは実際に行った行動を考慮して生活を比較することができる。

#### 4.2 単純手法

単純な方法として、2つの期間における各運動データがどの程度の時間を占めていたか、その分布の違いで生活を比較する方法が考えられる。そこで、各運動データの種類の運動時間の和の分布に差があるか検定を行った。同じ生活をしていた2つの期間の運動データと、違う生活をしていた2つの期間の運動データに対し、「分布に差が無い」という仮説に対し分布の差の検定を行った。6.6節で詳細を述べるが、結果としてすべての場合で有意確率が非常に小さい値となり、このよう単純な手法では比較ができなかった。

### 5. 提案手法

#### 5.1 運動データと行動

単純手法で生活を比較することができなかった理由は、個別の運動状態にのみ着目して生活を比較したことが問題だったと考えられる。1つの運動データが1つの行動を表さない場合がある。多くの場合、1つの行動は複数の運動データによって構成されている。例えば、コンビニへの買い物という行動は、1. コンビニまで行く(歩行)、2. 商品を物色する(軽作業)、3. 歩いて家に変える(歩行)のように、3つの運動データで構成されている。同じ歩行という運動データでも、家からコンビニへ行く場合と駅から大学へ行く場合とは違う行動の中の歩行である。このように、行動は単一の運動データではなく、複数の運動データで表される。このため、行動を比較することは、運動データの順序パターンを比較することに相当すると考えられる。

#### 5.2 提案するアプローチ

普段日常的に行われた行動(日常行動と呼ぶ)に着目し、それぞれの期間においてどの日常行動がどのくらい行われていたかを取得する。そして両方の期間に共通する日常行動がどのくらい存在するかによって生活を比較する。

まず初めに日常行動を表す運動データのパターンを利用者の全てのデータからエピソードマイニング[16]によって抽出する。ここで抽出された運動データのパターンを日常エピソードと呼ぶ。次に、それぞれの期間で一定の頻度以上行われた日常エピソードを取得する。この日常エピソードを局所頻出であるという。日常エピソードの頻度はエピソードマイニングと同じ基準で計算する。そして、それぞれの期間で局所頻出となった日常エピソードの集合から、Jaccard係数による類似度を計算する。ここで、集合X, YのJaccard係数による類似度は $sim = \frac{|X \cap Y|}{|X \cup Y|}$ である。この提案手法では、2つの期間で同じような生活をしていた場合、共通する日常エピソードが多くなり、類似度が大きくなる。2つの期間で違う生活をしていた場合、共通する日常エピソードが少なくなり、類似度が小さくなる。

以上を踏まえて、提案手法のアプローチをまとめる。

(1) 全ての運動データから日常エピソードをエピソードマイニングによって抽出する。

(2) 抽出した日常エピソードが各期間で局所頻出となるか調べ、それぞれの期間で局所頻出である日常エピソードを列挙する。

(3) それぞれの期間の日常エピソードの集合の類似度を計算する。

#### 5.3 運動データのパターンの取得

##### 5.3.1 エピソードマイニング

本研究では日常行動をある程度回数行われた行動とするため、日常行動を表す運動データのパターンはエピソードマイニングによって抽出する。

通常のエピソードマイニングでは、対象となるデータはデータの種類とタイムスタンプのみで構成されている。しかし、本研究で扱う運動データは開始時刻と終了時刻が有り、継続時間を持つデータである。以下で運動時間を持つデータに対するエピソードマイニングの定義を行う。

エピソードマイニングの対象であるデータをシーケンスと呼ぶ。シーケンス $S = \langle s_1, s_2, \dots, s_n \rangle$ は、運動状態の種類 $m_i \in M$ と運動時間 $d_i$ 、そして開始時刻 $t_i$ を持つ運動データ $s_i = (m_i, d_i, t_i)$ を時刻順に記録したものである。 $M$ は運動状態の種類の全体集合である。全ての開始時刻 $t_i (1 \leq i \leq n-1)$ に対し、 $t_i < t_{i+1}$ が成り立つ。 $S$ の長さ $|S|$ は $S$ の要素数 $n$ とする。図1に示したデータは次のシーケンスで表現できる。((静止, 300分, 2011年6月10日0時0分)(安静, 95分, 2011年6月10日5時0分)(軽作業, 100分, 2011年6月10日6時35分)(デスクワーク, 165分, 2011年6月10日8時35分))。

エピソードマイニングで得られるエピソード $\alpha$ (serial episode)を次のように記述する。 $\alpha = \langle a_1 a_2 \dots a_k \rangle (a_j \in M)$ 。 $\alpha$ の長さ $|\alpha|$ を $\alpha$ の要素数 $k$ とする。エピソードは各要素が先頭から順に発生するパターンを意味している。シーケンス $S = \langle s_1, s_2, \dots, s_n \rangle$ において、エピソード $\alpha = \langle a_1 a_2 \dots a_k \rangle$ の各要素 $a_j$ がそれぞれ $\alpha$ における順序を保持しながら $S$ の運動データ $s_i$ に含まれる場合( $a_j \in s_i$ )、 $\alpha$ は $S$ に出現するという。このとき $\alpha$ の各要素を含む $S$ の運動データの発生時刻を時刻順に取り出した数値列 $o = \langle o_1, o_2, \dots, o_k \rangle$ を $\alpha$ のインスタンスと呼ぶ。エピソード $\beta = \langle b_1 b_2 \dots b_n \rangle$ に、 $\alpha$ のすべての要素が順序を保持したまま存在するとき、 $\alpha$ を $\beta$ のサブエピソードとよび、その関係を $\alpha \sqsubseteq \beta$ と記述する。 $\beta$ が $\alpha$ を含む、あるいは包含するとよぶ。

エピソード $\alpha$ の評価値として頻度を使う。 $S$ における $\alpha$ の頻度を $freq(S, \alpha)$ で示す。 $\alpha$ の頻度には複数の決定方法がある。これを頻度の尺度と呼ぶ。与えられたシーケンス $S$ と最小頻度 $min\_freq$ に対し、 $freq(S, \alpha) \geq min\_freq$ を満たすエピソード $\alpha$ を頻出エピソードとよぶ。エピソードマイニングは全ての頻出エピソードを列挙する問題である。

##### 5.3.2 エピソードの頻度の尺度

エピソードの頻度にはいくつかの基本的な尺度がある[15]。ス

ライド窓が含むエピソードを数え上げる fixed-win-freq [16] が最も基本的な頻度の尺度である。他に、重なり合わないエピソードのインスタンスのみを数え上げる non-overlapped-freq [17] 等がある。

fixed-win-freq によるエピソード  $\alpha$  の頻度は、時間幅  $w$  を持つライド窓をシーケンス  $S$  上で単位時間毎に移動させた際に  $\alpha$  が出現するライド窓の数である。この定義により、fixed-win-freq は  $\alpha$  の同じインスタンスを重複して数え上げる。また、継続時間が短いエピソード程頻度が大きくなる性質を持っている。このような性質から、fixed-win-freq によるエピソードの頻度は、行動の回数を示していない。

シーケンス  $S$  中の  $\alpha$  の 2 つのインスタンス  $o, o'$  が  $o_1 > o'_k$  または  $o'_1 > o_k$  を満たすとき  $o$  と  $o'$  は non-overlapped であるという non-overlapped-freq による  $S$  における  $\alpha$  の頻度は non-overlapped であるインスタンスの数である。この定義から、non-overlapped-freq は同じインスタンスを重複して数え上げず、時間的に重なるインスタンスも数え上げない頻度の尺度である。

#### 5.4 抽出されたエピソードの問題点

本節では単純に抽出した日常エピソードをそのまま生活の比較に用いることの問題点を議論する。

##### 5.4.1 エピソードの包含関係

エピソードマイニングでは、頻出であるエピソード  $A$  と共に  $A$  の全てのサブエピソードも抽出される。これら全てが日常エピソードとなる。ここで、ある 2 つの日常エピソード  $A$  と  $C$  について日常エピソード  $B$  は  $A$  と  $C$  両方のサブエピソードとする。2 つの期間のうち片方の期間で片方の期間で  $A$  が行われて  $C$  が行われず、もう片方の期間で  $A$  が行われず  $C$  が行われた場合、2 つの期間で共通する行動は無いが、 $A$  と  $C$  の一部を表す  $B$  が共通であるため類似度が高くなる可能性がある。

例えば、期間 1 で局所頻出である日常エピソードの集合が  $X_1 = \{ \langle \text{歩行, 軽作業} \rangle, \langle \text{歩行, 安静} \rangle \}$ 、期間 2 で局所頻出である日常エピソード集合が  $X_2 = \{ \langle \text{歩行, 軽作業} \rangle, \langle \text{静止, 安静} \rangle \}$  であるとする。共通する日常エピソードは 1 個、全てのエピソードは 3 個であるため、類似度が  $1/3$  となる。しかしこの場合、期間 1 はサブエピソードとして歩行と軽作業と安静を、期間 2 は歩行と軽作業と静止と安静も同時に抽出される。このサブエピソードも含めて類似度を計算すると  $4/7$  となる。共通する行動は 1 つだが、類似度が  $4/7$  になる。

この問題を解決するため、抽出された日常エピソードから、他のエピソードのサブエピソードとなるエピソードを全て除外する。残った日常エピソードは他のどの日常エピソードのサブエピソードにもならないエピソードである。つまり極大エピソードと等しい。

##### 5.4.2 運動データ間の時間差

単純にエピソードマイニングを行った場合、連続して行われた複数の異なる行動が 1 つの日常エピソードとして抽出されてしまう場合がある。例えば、日常エピソード  $\langle \text{歩行, 軽作業, 安静, 歩行, 作業} \rangle$  が抽出された場合、 $\langle \text{歩行, 軽作業} \rangle$ 、 $\langle \text{歩行, 作業} \rangle$  が示す行動が 1 つの行動となる。しかし、上記の

データで安静の運動時間が 180 分だった場合、作業は軽作業の 180 分後に行われたことになる。このような行われた時刻が離れた運動データは別の行動を表していると考えられるが、エピソードマイニングにおいては 1 つの日常エピソードとして抽出されてしまう。

この問題を解決するため、運動データと運動データの間の時間（ギャップ）に制限を設ける。この制限によって、指定した時間以上のギャップが 1 つでもあるインスタンスを除外できる。

##### 5.4.3 運動時間を考慮した運動データの区別

運動データを加工せずに日常エピソードを抽出した場合、複数の行動が 1 つの日常エピソードとして抽出される場合がある。例えば  $\langle \text{安静, 歩行} \rangle$  という日常エピソードは報告者の生活において、休憩した場合（飲み物を買に行き、戻った後ゆっくりする）場合と、通学して講義に出席するという行動の場合（大学まで歩き、講義を聞く）がある。この 2 つは異なる行動だが、同じ日常エピソードとして抽出されてしまう。

そこで、異なる行動が別の日常エピソードとして抽出されるようにするために、運動データを運動時間によって区別することにする。本研究では単純に、運動データの種類毎に、分けた後の運動データを当頻度分割する。これによって、休憩した場合の日常エピソードは  $\langle \text{歩行 (短), 安静 (短)} \rangle$  として、通学して講義に出た場合の日常エピソードは  $\langle \text{歩行 (長), デスクワーク (長)} \rangle$  のように、1 つの行動が 1 つの日常エピソードとして抽出されるようになる。

## 6. 評価実験

### 6.1 検証に利用する期間の組み合わせ

報告者の 602 日間の運動データを用いて、提案手法が生活を比較することができるか検証する。自分が同じ生活をしていたと判断する期間の組み合わせ 2 つと、違う生活をしていたと判断する期間の組み合わせ 3 つの計 5 つを用意した。これを以下に示す。

- 組 1（同期間）前期授業期間 5 月 / 6 月
- 組 2（同期間）後期授業後期 年明け前 / 年明け
- 組 3（異期間）2011 年秋 / 2012 年秋
- 組 4（異期間）授業期間 / インターンシップ
- 組 5（異期間）就職活動後 / 就職活動中

組 1 と組 2 は同じ生活をしていたと報告者が判断している期間の組み合わせ（同期間）で、組 3 から組 5 は違う生活をしていたと判断している期間の組み合わせ（異期間）である。

同じ生活をしていた期間の組 1 は 2011 年の前期授業期間のゴールデンウィークが終わった後の 8 週間を前後 4 週間ずつに分けた組み合わせである。この期間は合同輪講も無く、テスト期間でも無かったため、同じ生活をしていた。組 2 は 2011 年の後期授業期間の冬休み前と冬休み後の組み合わせである。年や季節が異なるが、後期授業を履修し大学に通う行動も同じであるため同じ生活をしていた。

違う生活をしていた期間の組 3 は 2011 年の秋と 2012 年の秋の期間の組み合わせである。この 2 つの期間は季節が同じだが、前者は授業を履修していて、後者は授業を履修していない

という違いがある。組4は2011年前期授業前期全体と夏季インターンシップ中の3週間である。この2つの期間には、前者の移動は電車、バス、自転車等があるが、出社は徒歩のみという違いがある。組5は就職活動後の2012年の秋の期間と、夏休みの就職活動のため研究室に来ていない期間である。この2つの期間には、前者では大学への通学や研究活動等が行われているが、後者においてはあまり行われなかったという違いがある。

## 6.2 頻度の尺度の評価

提案するアプローチにおいて、fixed-win-freqを用いた場合と、non-overlapped-freqを用いた場合の各組の類似度を比べた。結果を表1に示す。2列目は期間の組み合わせの名前である。3列目は提案するアプローチにおいてfixed-win-freqを用いた場合の各組み合わせの類似度、4列目はnon-overlapped-freqを用いた場合の類似度である。

non-overlapped-freqの方がfixed-win-freqよりも同じ生活をしてきた期間と違う生活をしてきた期間類似度の差が大きい。このためnon-overlapped-freqの方が生活の比較に適していると考えられる。

表1 頻度の尺度を変えた場合の2つの期間の類似度

期間の組		fixed-win	non-overlapped
同期間	1	0.878	0.837
	2	0.867	0.842
異期間	3	0.728	0.660
	4	0.734	0.637
	5	0.663	0.607
類似度の差の範囲		0.133 - 0.215	0.177 - 0.235

## 6.3 エピソードの包含関係の評価

提案手法において、全てのエピソードを用いた場合と極大エピソードを用いた場合の各組の類似度を比べた。結果を表2に示す。2列目は期間の組み合わせの名前である。3列目は全てのエピソードを生活の比較に利用した場合の類似度、4列目は極大エピソードのみを生活の比較に用いた場合の類似度である。どの組みでも、全ての期間の組み合わせで全てのエピソードを用いる場合より極大エピソードを用いた場合の方が類似度が低い。包含関係にあるエピソードが減ることで共通する日常エピソードが減り、類似度が下がったと考えられる。特に、違う生活をしてきた期間の組み合わせである組3-組5においては類似度が大きく低下している。同じ生活をしてきた期間と違う生活をしてきた期間では類似度の差が増えている。このことから極大エピソードの方が生活を比較する際に用いるエピソードに適していると考えられる。

## 6.4 ギャップの制限の評価

ギャップに対する強い制限(0分)を入れた場合、ギャップに対する弱い制限(30分)を入れた場合、ギャップに対する制限をほぼ入れなかった場合(360分)の類似度について調べた。結果を表3に示す。2列目は期間の組み合わせの名前、3列目は強い制限(0分)の場合の類似度、4列目は弱い制限(30分)の場合の類似度、5列目は制限をほぼ入れなかった(360分)場

表2 比較に利用するエピソードを変えた場合の2つの期間の類似度

期間の組		全てのエピソード	極大エピソード
同期間	1	0.837	0.712
	2	0.842	0.718
異期間	3	0.660	0.388
	4	0.637	0.442
	5	0.607	0.286
類似度の差の範囲		0.177 - 0.235	0.270 - 0.432

合の類所度である。

弱い制限(30分)入れた場合と制限をほぼ入れなかった(360分)場合は類似度の差が大きい。制限をほぼ入れなかった場合は期間3と期間4で類似度が大きくなっている。このため、弱い制限を入れた場合がもっとも生活の比較に適していると考えられる。

表3 ギャップを制限した場合の2つの期間の類似度

期間の組		0分	30分	360分
同期間	1	0.652	0.712	0.890
	2	0.744	0.718	0.820
異期間	3	0.564	0.388	0.605
	4	0.429	0.442	0.707
	5	0.600	0.286	0.407
類似度の差の範囲		0.052 - 0.315	0.270 - 0.432	0.215 - 0.483

## 6.5 運動時間による運動データの区別の評価

運動データを分割しない場合、2分割する場合、3分割する場合の類似度について調べた。結果を表4に示す。2列目は期間の組み合わせの名前、3列目は運動データを分割しない場合、4列目は2分割する場合、5列目は3分割する場合の類似度である。

運動データを3分割した場合が最も類似度の差が大きい。しかし、運動データを2分割した場合は類似度の差が小さかった。これは、運動データを適切に区別できなくなったためと考えられる。今回は単純に運動データを等頻度分割したため、適切に区別できた場合とできなかった場合が出た。今回最も良い結果が出た3つに当頻度分割した場合よりもさらに適切に運動データを分割する方法がある可能性がある。

表4 運動時間により運動データを区別した場合の類似度

期間の組		なし	2分割	3分割
同期間	1	0.712	0.687	0.840
	2	0.718	0.692	0.750
異期間	3	0.388	0.553	0.231
	4	0.442	0.579	0.261
	5	0.286	0.406	0.235
類似度の差の範囲		0.270 - 0.432	0.108 - 0.286	0.489 - 0.600

6.6 単純手法との比較

単純な方法による生活の比較と提案手法による生活の比較の結果を比べる。単純な手法として、各期間の運動データの種類の運動時間の分布に差があるか調べた。「各期間の分布に差が無い」という仮説に対し分布の差の検定を行った。結果の有意確率を表5に示す。提案手法は頻度の尺度に fixedwin-freq を用い、単純なエピソードを比較に用いた場合の結果と、頻度の尺度に non-overlapped-freq を用い、単純なエピソードの問題を解決した場合(極大エピソードの利用, ギャップの弱い制限(30分)の利用, 運動時間による運動データの区別(3分割))を行った場合の結果を示す。2列目は期間の組み合わせの名前である。3列目は分布の差の検定による有意確率, 4列目は単純なエピソードを用いた場合の提案手法による類似度, 5列目は単純なエピソードの問題を解決した場合の提案手法である。

単純な手法ではすべての場合で有意確率  $10 \times 10^{-18}$  以下となってしまう, 非常に小さい値となった。このため, このよう単純な手法では生活を比較できなかった。提案手法は単純な運動データのパターンを利用した場合にも比較することができているが, エピソードの問題を解決した場合はさらに明確な類似度の差がある。

表5 単純手法との比較

期間の組	単純手法の有意確率	単純な	問題を解決した	
		エピソードを用いた場合の類似度	エピソードを用いた場合の類似度	
同期間	1	$1.0 \times 10^{-18}$ 以下	0.837	0.840
	2	$1.0 \times 10^{-18}$ 以下	0.842	0.750
異期間	3	$1.0 \times 10^{-18}$ 以下	0.660	0.231
	4	$1.0 \times 10^{-18}$ 以下	0.637	0.261
	5	$1.0 \times 10^{-18}$ 以下	0.607	0.235
類似度の差の範囲		0.177 - 0.235	0.489 - 0.609	

7. おわりに

本研究では, ある期間の生活の振り返りを支援する機能を実現する技術の1つとして, それぞれの期間に行われた行動の違いによって異なる2つの期間の生活を比較する方法を提案した。比較に用いるライフログデータとして, リストバンド型センサ装置から得られる運動データを用いた。日常的な行動を表す運動データの順序パターンの取得にはエピソードマイニングを用いた。また, 単純なエピソードマイニングによって得られるエピソードと行動の関係を考察し, 問題点と解決方法を提案した。

報告者が蓄積した602日間のライフログデータを利用した評価実験により, 提案するアプローチで異なる2つの期間の生活を比較できるかを検証した。提案手法において, エピソードマイニングでは頻度の尺度は non-overlapped-freq を用いること, 比較に利用する日常エピソードを極大エピソードのみに限定すること, 日常エピソードの運動データ間のギャップを制限すること, 運動時間によって運動データを区別することで生活を比較できることを示した。

[1] 読書メーター, <http://book.akahoshitakuya.com/>, (2013/1/21 アクセス)

[2] からだログ, <http://karada.goo.ne.jp/>, (2013/1/21 アクセス)

[3] Fitbit, <http://www.fitbit.com/>, (2013/1/21 アクセス)

[4] Jawbone Up, <https://jawbone.com/up>, (2013/1/21 アクセス)

[5] Nike+, <http://nikeplus.nike.com/plus/>, (2013/1/21 アクセス)

[6] HASC Challenge 2012, <http://hasc.jp/hc2012/index.html>, (2013/1/21 アクセス)

[7] 小林亜令, 岩本健嗣, 西山智, “釈迦: 携帯電話を用いたユーザ移動状態推定・共有方式,” 情報処理学会研究報告. UBI, vol.2006 no.54, pp.15-22, 2006.

[8] T H Monk, J F Flaherty, Frank E Hoskinson, D J Kupfer, D. J. “The social rhythm metric: an instrument to quantify the daily rhythms of life,” Journal of Nervous and Mental Disease vol.178, pp.120-126, 1990

[9] Szuba, “Disruption of social circadian rhythms in major depression: A preliminary report,” Psychiatry Research, vol.42, pp.221-230, 1992

[10] 湯浅美千代, 正木治恵, 佐藤弘美, 酒井郁子, 野口美和子, 施設・病院に入っている老人の生活リズムの乱れとその看護, 老年看護学, 1, pp.79-89, 1996

[11] 河本健, 栗山裕之, 田中毅, 新谷隆彦, 鈴木敬, “典型日遷移列に基づく生活モデルの提案,” 電気学会研究会資料. CMN, 通信研究会 2010, 19-24, 2010

[12] 鈴木敬, 栗山裕之, 山下春造, 矢野和男, “ライフ顕微鏡: 20人のライフタベストーリーが語る人とセンサとITの未来,” 日立評論, 2007年, 12月号.

[13] 青木茂樹, 岩井嘉男, 大西正輝, 小島篤博, 福永邦雄, “独居高齢者の行動パターンに注目した非日常状態の検出,” 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J87-D-II, 1083-1093, 2004-05

[14] 木川 泰, 小栗宏次, “高齢者在宅データ遷移確率に基づく生活リズム変動解析,” 電子情報通信学会 技術研究報告. MBE,ME とバイオサイバネティクス 102, 41-44, 2003

[15] M.Gan, H.Dai, “A study on the accuracy of frequency measures and its impact on knowledge discovery in single sequences,” Proceedings of 3st IEEE Internatioanl Workshop on Reliability Issues in Knowledge Discovery, pp.859-866, 2010

[16] H. Mannila, H. Toivonen, and A. I. Verkamo, “Discovery of Frequent Episodes in Event Sequences,” Data Mining and Knowledge Discovery, vol.1, no.3, pp.259-289, 1997.

[17] S. Laxman, P. Sastry, and K. Unnikrishnan, “A Fast Algorithm for Finding Frequent Episodes in Event Streams,” Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining Pages 410-419, 2007