

## 頻出長時間エピソードを用いた生活比較に関する一考察

## Consideration of Lifestyle Comparison with Frequent and Long Duration Episodes

中山 恭明<sup>†</sup>新谷 隆彦<sup>‡</sup>大森 匡<sup>‡</sup>藤田 秀之<sup>‡</sup>

Yasuaki Nakayama

Takahiko Shintani

Tadashi Ohmori

Hideyuki Fujita

## 1. はじめに

近年安価なセンサが普及し、人の生活に関する長期間データであるライフログデータを大量に収集することができるようになった。ライフログ活用のための、長期間の生活の振り返りを支援する技術として二つの期間の生活比較の研究をしてきた [1]。リストバンド型センサで取得した運動状態データにエピソードマイニング [2] を適用して頻出エピソードを抽出し、それぞれの期間で行われた行動の違いによって生活を比較した。生活比較により、同様の生活をしていた期間、異なる生活をしていた期間を提示することが可能となり、当時の生活を振り返ることに役立つ。しかし、生活比較においては行われた回数が多い行動に相当する頻出エピソードのみで生活を特徴づけることは不十分であり、長い時間を占めた行動も重要である。また、これまでは頻出エピソードによって何を行ったかを示す運動状態のパターンで考慮していた。人の生活ではどの行動をいつからどのくらいの時間行っていたかも重要であるが、これまでは考慮されてこなかった。

本研究では、継続時間を考慮した長時間エピソード [3] を生活比較に導入すること、および、頻出長時間エピソードとして抽出された行動を開始時刻と継続時間を考慮して分類することを検討する。

## 2. 本研究で用いるライフログデータ

本研究では、加速度センサを搭載したリストバンド型センサで取得した腕の動きに関するデータをライフログデータとして利用する。リストバンド型センサとして、(株) 日立システムズのリストバンド型ライフログレコーダ UW-301BT を使用した。UW-301BT を手首に装着することで、ユーザの腕の動きを常時観測し、ユーザがいつからいつまでどの程度の運動を継続したかを示す運動状態データを得ることができる。運動状態データは、運動状態  $m$  とその開始日時  $t_s$ 、終了日時  $t_e$  ( $t_s < t_e$ ) の組からなるイベント  $(m, t_s, t_e)$  が、開始日時の順に並べられたシーケンスデータである。運動状態は、静止、安静、軽作業、作業、デスクワーク、歩行、ジョギング、非装着、データなしの 9 種類で表現される。

## 3. 生活比較

本研究における生活比較は、二つの期間に同じ生活をしてきたか、異なる生活をしてきたかを判断することである。人の生活は日々の様々な行動の積み重ねであり、行動は目的を持って行った一連の活動である。人の生活はユーザが行った行動によって特徴づけることができるため、それぞれの期間に行った行動がどのく

らい同じであるか、または違うかを調べることによって生活比較を行うことができる。同じ行動が多く行われていた場合に同じ生活、違う行動が多く行われていた場合に異なる生活を送っていたと判断する。これを実現するには、実際に行った行動内容で比較する必要がある。具体的な行動内容を人が手作業で長期間にわたって記録し続けることが必要となるため、実現することは困難である。そこで、ライフログデータから抽出した運動状態のパターンによって生活比較を行ってきた [1]。この手法では、普段日常的に行われていた行動である日常行動に着目した。二つの期間について、それぞれの期間でどの日常行動が行われていたかを取得し、両方の期間に共通する日常行動がどのくらいあるかで生活を比較する。運動状態データは実際の動きを反映したデータであり、ユーザの行動を表しているため、具体的な行動内容の代わりに運動状態データを利用することができる。運動状態データはシーケンスデータであるため、エピソードマイニングによって頻出エピソードを抽出することが可能である。抽出された頻出エピソードは行われた回数の多い行動に相当するため、これを日常行動とし、それぞれの期間で出現した頻出エピソードの集合の類似度によって生活を比較する。

## 4. 提案手法

## 4.1. 長時間エピソードの導入

これまでは、日常行動として運動状態データにエピソードマイニングを適用することによって抽出された頻出エピソードを用いた。行動は複数の運動状態の連なりによって構成されるため、運動状態データに多く現れる運動状態のパターンである頻出エピソードは一つの行動を示している。しかし、人の生活においては行われた回数が多い行動だけでなく、長い時間を占めた行動も重要な行動である。長い時間を占めた行動は 1 回行うために長い時間を要するため、行うことができる回数に限度があり、頻度が低い傾向がある。そのため、頻度を閾値とする従来のエピソードマイニングでは長い時間を占める行動に相当するエピソードを抽出することができない。生活比較においては生活を特徴づける日常行動を知る必要があるため、頻出エピソードのみでは不十分である。

本研究では、頻出エピソードだけでなく、長い時間を占めた行動に相当する長時間エピソードを用いて生活比較を行う。頻出エピソードによって行った回数が多い行動を知るだけでなく、長時間エピソードによって行った回数は頻出ではない場合でも行った時間の合計が長い時間を占める行動を知ることができる。頻出エピソードと長時間エピソードを合わせて頻出長時間エピソードと呼ぶ。頻出長時間エピソードを用いることで生活比較の精度向上が期待できる。

<sup>†</sup>電気通信大学情報理工学部情報通信工学科<sup>‡</sup>電気通信大学大学院情報理工学研究科

長時間エピソードは、頻度ではなく、総継続時間を閾値として運動状態データから抽出する [3]。エピソードは運動状態のリストであるため、総継続時間はエピソードが運動状態データに現れた区間であるインスタンスの継続時間の合計で求めることができる。インスタンスはイベントのリストであるため、インスタンスの継続時間は先頭のイベントの開始日時から末尾のイベントの終了日時までの時間差となる。つまり、総継続時間はエピソードのすべてのインスタンスの継続時間の和となる。また、長すぎる継続時間となるインスタンスを回避するための制約条件として、個々のインスタンスの継続時間の最大値である最大インスタンス継続時間が設定される。

ユーザが指定した二つの期間に対する、頻出エピソードと長時間エピソードを用いた生活比較の手順を以下に示す。

1. 頻度が最小頻度以上となる頻出エピソードと、総継続時間が最小総継続時間以上となる長時間エピソードを抽出する。ここで、頻度が  $\frac{\text{総継続時間}}{\text{最大インスタンス継続時間}}$  以上となるエピソードのうち最小総継続時間を満たすものを長時間エピソードとする。
2. 極大な頻出エピソードと極大な長時間エピソードをそれぞれ選出し、日常行動を示すエピソードである日常頻出エピソード、日常長時間エピソードとする。
3. 比較するそれぞれの期間について、局所的に頻出となる日常頻出エピソードと局所的に長時間となる日常長時間エピソードを取得する。それぞれの期間で取得された日常頻出長時間エピソードの集合から、Jaccard 係数を算出し、その値を生活比較の結果として出力する。

#### 4.2. エピソードの分類

単純な頻出長時間エピソードは運動状態のリストで表現される。エピソードを構成する運動状態が等しい場合には一つの行動を示すことになる。しかし、構成する運動状態が等しいが、実際の行動内容は異なる場合がある。例えば、エピソード”歩行-デスクワーク”となる行動でも、学校に行き研究室で研究を行ったとき、学校に行き授業を受けたとき、家に帰り PC を操作したときなどがあてはまる。この場合、実際の行動内容が異なる行動が一つのエピソードとなるため、生活比較の精度が低下する原因となる。このような行動はエピソードを構成する運動状態が等しくとも、いつからどのくらいの時間行ったかで区別することができる。頻出長時間エピソードが現れた区間はインスタンスであるため、同一のエピソードのインスタンスの中で開始時刻と継続時間が近いインスタンスは同様の時刻に開始し、同程度の時間行った行動を示しており、また、開始時刻と継続時間に大きな差があるインスタンスは異なる行動を示していると考えられる。各エピソードのインスタンスを開始時刻と継続時間からなるベクトルで表現し、クラスタリングを行う。同一のクラスタに属するインスタンスを一つのエピソードで表現し、これらを用いて生活比較を行う。

分類したエピソードによる生活比較の手順は、4.1 節で示した 2 と 3 の手順の間に次の手順を加える。

- それぞれの日常頻出長時間エピソードに対してインスタンスを取り出し、開始時刻と継続時間でクラスタリングを行う。ここで、各クラスタに属するインスタンスが最小頻度または最小総継続時間を満たすようにする。本稿では、インスタンスの距離関数として標準化ユークリッド距離を用い、k-means 法によってクラスタリングを行う。同一のクラスタ内のインスタンスを一つのエピソードで表現する。

### 5. 実験結果

実験参加者が収集した実際の運動状態データを用いて実験を行った。提案手法による生活比較を行った結果を表 1 に示す。ここで、Clustered はエピソードの分類を行なった場合、L と F はそれぞれ頻出エピソードと長時間エピソードを用いたことを示す。

表 1: 生活比較の結果

	Clustered	L or F	F	L
同じ生活	0.485	0.465	0.780	0.420
異なる生活	0.0664	0.126	0.732	0.014
類似度の差	0.419	0.339	0.049	0.407

この結果から、分類したエピソードを用いることによって同じ生活と異なる生活の間の類似度の差が大きくなっているため、提案手法が生活比較に適していることがわかる。

### 6. おわりに

生活比較において、継続時間を考慮した長時間エピソードを導入した手法、および、頻出長時間エピソードを開始時刻と継続時間を考慮して分類する手法を提案した。実データを用いた評価実験により、提案手法により生活比較の精度を改善できることを確認した。

#### 謝辞

本研究は、JST、CREST、JPMJCR1503 の支援を受けたものである。

#### 参考文献

- [1] 磯村洋, 新谷隆彦, 大森匡, ” 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”, DEIM Forum, E5-4, 2013.
- [2] H. Zhu, P. Wang, X. He, Y. Li, W. Wang, and B. Shi, *Efficient Episode Mining with Minimal and Non-overlapping Occurrences*, IEEE ICDM, pp.1211- 1216, 2010.
- [3] 櫻田滋大, 新谷隆彦, 大森匡, 藤田秀之, ” 継続時間を閾値としたエピソードマイニングの提案”, 電子情報通信学会総合大会, D-4-9, 2015.