

オカレンスの非重複を考慮した探索候補削減による  
 複数の時間帯に注目した長時間エピソード抽出手法  
 Consideration on candidate pruning method by excluding overlapped intervals of  
 occurrences for mining long-duration episodes in multiple time-ranges

松崎 太一<sup>1)</sup> 新谷 隆彦<sup>1)</sup> 大森 匡<sup>1)</sup> 藤田 秀之<sup>1)</sup>  
 Taichi Matsuzaki Takahiko Shintani Tadashi Ohmori Hideyuki Fujita

## 1 はじめに

近年では時計型ウェアラブル端末の普及によりライフログの活用に関心が高まっている [1]. 我々はいつどの程度の動きを行ったかを表す運動状態データを収集し、ユーザの生活特性を示す長い時間を占めた行動を長時間エピソードとして抽出する研究を進めてきた. 長時間エピソードマイニングはユーティリティエピソードマイニング [2] において、継続時間を評価値とした手法である. さらに、行動した時間帯に着目した長時間エピソードの抽出の研究を進めてきた [3]. これまでは 1 つの行動時間帯に注目した長時間エピソードの抽出に限定されていた. しかし、人間の生活特性を示す行動は複数の時間帯に行われることがある. そこで本研究では複数の行動時間帯に注目した手法について検討する. 行動した時間帯に着目した長時間エピソードの抽出では、オカレンスと呼ばれるエピソードが行われた区間のうち、時間帯内にある継続時間のみを評価するため、閾値を低く考慮する必要があり、探索候補数が増大してしまう. そこで、未探索候補が取り得る時間帯内の継続時間の最大値を計算する探索候補削減手法が提案されたが [3], 複数の時間帯に着目した場合は探索候補を削減する効果が小さくなった. 本研究ではオカレンスの非重複を考慮して計算し、冗長な見積りの排除を試みた.

## 2 複数の時間帯に注目した長時間エピソード

### 2.1 長時間エピソード

処理対象となるイベントシーケンスデータ  $D$  は発生したイベントが記録されたデータであり、イベントタイプ  $m_j \in M$ , 開始日時  $t_{s_j}$ , 終了日時  $t_{e_j}$  からなるイベント  $d_i = (m_j, t_{s_j}, t_{e_j})$  を開始日時の順に並べたデータである. イベントが重複する期間はない.  $M$  はイベントタイプの集合である.

エピソード  $m_1-m_2-\dots-m_k$  とは、イベントタイプ  $m_i \in M (1 \leq i \leq k)$  のリストである.  $m_i$  と  $m_{i+1}$  では  $m_i$  が先に発生することを表す. イベントシーケンスデータ  $D$  にエピソードのイベントタイプと同じ順でイベントが出現する区間  $[t_s, t_e]$  をオカレンス  $occ$  という. ここで、区間の開始日時が  $t_s$ , 終了日時が  $t_e$  である.  $(t_e - t_s)$  をオカレンスの継続時間という. オカレンスの継続時間が最大継続時間の制約  $maxspan$  を超えるとき、またはオカレンスを構成するイベントの間隔が制約  $maxgap$  を超えるとき、オカレンスとは認めない. これら 2 つの制約により、冗長なオカレンスを除外する. 極小なオカレンスとは、オカレンス内に同一エピソードのオカレンスが存在しないことをいい、エピソード  $\alpha$  の極小なオカ

レンスの集合を  $MO(\alpha)$  という. 非重複なオカレンスの集合とは、どの 2 つのオカレンスも互いに重複区間を持たないオカレンスの集合である. 非重複オカレンスの集合を計算するとき、重複する 2 つのオカレンスのうち先に発生したものをオカレンスとして認める. エピソード  $\alpha$  の極小かつ非重複なオカレンス集合を  $MANO(\alpha)$  という. エピソード  $\alpha$  の総継続時間  $tdur$  とは、 $MANO(\alpha)$  のオカレンスの継続時間の総和である.

時間帯を開始時刻を  $r_s$ , 終了時刻を  $r_e$  として  $t_r = [r_s, r_e]$  と表す. 時間帯集合  $TR$  を、 $t_r$  の集合とする.  $TR$  の要素のうち、他の時間帯と連続する、もしくは重複区間を持つ時間帯を認めない. また、時間帯  $t_r$  の指定長の合計が 24 時間以上となるような時間帯集合を認めない. オカレンスの時間帯継続時間  $r_d(occ)$  とは、オカレンス  $occ = [t_{s_d}, t_{e_d}]$  と、時間帯  $[r_s, r_e]$  が重なる区間の長さである. オカレンスの  $TR$  との時間帯継続時間の総和を複数時間帯継続時間  $r_{d_s}(occ)$  とし、 $r_{d_s}(occ) = \sum_{t_r \in TR} r_{d_r}(occ)$  となる. エピソードの総時間帯継続時間  $trdur$  とは、 $MANO$  のオカレンスの複数時間帯継続時間の総和である. エピソードの時間帯偏り度は  $tb = \frac{trdur}{tdur}$  である.

### 2.2 複数の時間帯に注目した長時間エピソード抽出問題

複数の時間帯に注目した長時間エピソード抽出問題は、イベントシーケンスデータ  $D$ , 時間帯集合  $TR$ , 閾値である最小総時間帯継続時間  $mintrdur$ , 最小時間帯偏り度  $mintb$  が与えられ、 $(trdur \geq mintrdur) \wedge (tb \geq mintb)$  となる行動時間帯の偏りを考慮した長時間エピソードをすべて抽出することである.

## 3 複数の時間帯に注目した長時間エピソード抽出手法

### 3.1 抽出手法

複数の時間帯に注目した長時間エピソードの抽出は、行動した時間帯に着目した長時間エピソードの抽出 [3] と同様で以下の手順で行う.

- 手順 1 イベントシーケンスデータをスキャンする. 各イベントタイプについて  $MANO$  を計算し、 $tdur, trdur$  の計算を行う.  $trdur, tb$  が閾値を満たすものを出力する.  $UpperLimit \geq mintrdur$  を満たすイベントタイプ集合を  $FI$  とする.  $UpperLimit$  の計算は後述する.
- 手順 2 イベントタイプ  $e \in FI$  の直後にイベントタイプ  $e' \in M$  を付加してできるエピソード  $\alpha$  について、 $MO(e)$  と  $MO(e')$  から  $\alpha$  の  $MO(\alpha), MANO(\alpha)$  を計算し、 $tdur, trdur$  の計算を行う.  $trdur, tb$  が閾値を満たすならば  $\alpha$  を出力する.
- 手順 3 エピソード  $\alpha$  が  $UpperLimit \geq mintrdur$  を満たすならば、エピソード  $\alpha$  の末尾にイベントタイプ

1) 電気通信大学大学院情報理工学研究科 Graduate School of Informatics and Engineering, The University of Electro-Communications

$e'' \in M$ を付加してできるエピソードについて再帰的に手順 2,3 を行う。

エピソード末尾にイベントタイプを付加し, 新しいエピソードを探索する行程を, "エピソードを成長させる" という. エピソード  $\alpha$  を成長させる前に, 成長後のエピソードが取り得る  $trdur$  の最大値  $UpperLimit(\alpha)$  を計算し,  $mintrdur$  を満たし得ない場合は成長後エピソードの探索を省略する [3].  $UpperLimit$  は,  $MO(\alpha)$  のオカレンス終了時刻を継続時間が  $maxspan$  となるように変更し, 総時間帯継続時間を計算することで得られる.

図 1 の  $occ_1, occ_2$  について  $UpperLimit$  の計算例を示す.  $occ_1$  は終了時刻が  $t_{e1}'$  に,  $occ_2$  は終了時刻が  $t_{e2}'$  に変更されることで指定時間帯  $t_{r1}, t_{r2}$  と重なる区間  $[t_{s1}, r_{e1}], [r_{s2}, t_{e1}'], [r_{s2}, t_{e2}')$  の長さの総和が  $UpperLimit$  となる.  $occ_1$  の区間  $[r_{s2}, t_{e1}')$  に注目する. 複数の時間帯を指定する場合は, オカレンスの末端が時間帯と短い重複区間を持つ場合が考えられる. しかし, オカレンスの継続時間が  $maxspan$  程となる成長は稀であり, 多くの場合, このような区間の加算は過剰となる.

### 3.2 NOEWU を利用した探索候補削減の提案

複数の時間帯に注目した長時間エピソードの  $UpperLimit$  計算について過剰な加算区間を排除するため,  $NOEWU$ [4] の処理を拡張した手法を提案する.  $NOEWU$  は, 成長後エピソードのオカレンスが非重複になるように成長を仮定し, 取り得る最大の総継続時間を見積もる.  $NOEWU$  の処理を利用した  $UpperLimit$  の計算は以下のようにして行う.

$MO$  のオカレンスについて, 継続時間が  $maxspan$  となるように終了日時を変更する. ただし, 次のオカレンスの開始日時を超える場合, 終了日時は次のオカレンスの開始日時に設定する. これにより, 非重複で最大の継続時間を持つようなエピソードの成長を仮定する. 終了日時が変更されたオカレンスを時刻によるオカレンスに変換したあと, 総時間帯継続時間を計算し,  $UpperLimit$  とする.

図 1 に  $NOEWU$  処理後の区間を示した.  $UpperLimit$  計算の前に  $NOEWU$  処理を行い,  $occ_1$  の終了日時は  $t_{s2}$  に変更され, 重複区間であった  $occ_1$  の区間  $[r_{s2}, t_{e1}')$  は,  $UpperLimit$  の計算から除外される.

図 1 について指定時間帯の間隔  $r_{s2} - r_{e1}$  が  $maxspan$  未満であるような場合について,  $t_{r1}$  に開始時刻をもつすべてのオカレンスの末端が  $t_{r2}$  と重なることが考えられ,  $Upperlimit$  が過剰に加算されてしまうが, その一部を非重複を考慮することによって排除することができる. 本手法における探索候補の多くは, 時間帯付近に開始時刻があるオカレンスをもつエピソードとなる. よって図 1 の  $occ_1, occ_2$  ように 1 日の各時間帯内にオカレンスを持つケースも多く, 重複区間の発生による  $Upperlimit$  の見積もりの精度向上が期待できる.

## 4 評価実験

3.2 節の提案手法による探索候補数の減少を評価する.

3.1 節の  $UpperLimit$  計算手法を基本手法とし, 提案手法と基本手法それぞれの探索候補数を比較する. 本実験では, 報告者の運動状態をイベントとした運動状態データを利用する. イベントタイプ集合は{静止, 安静, 軽作業, ジョギング, 歩行, 作業, デスクワーク}である. 1

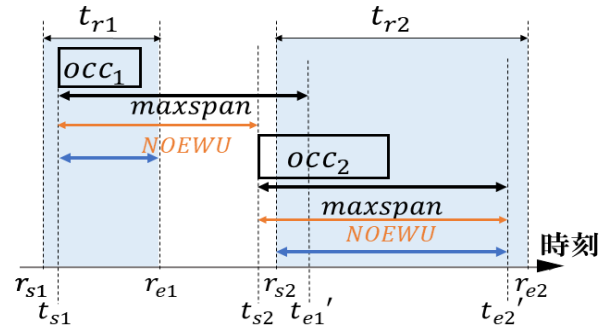


図 1 NOEWU を利用した  $UpperLimit$  計算

つの時間帯の指定幅は 4 時間, 時間帯の間隔を最小 1 時間とする.  $|TR| = 1, 2, 3$  として, 時間帯指定数ごとにすべての時間帯指定の場合で 1 回ずつ, 基本手法と提案手法で行動時間帯に偏りのあるエピソードマイニングを行う. 探索候補数の合計を時間帯指定数ごと, 手法別に集計する. パラメータは  $maxspan = 360$ (分),  $maxgap = 30$ (分),  $mintrdur = 1117 * |TR|$ (分) とした.  $mintb$  は探索候補削減手法がなく, 探索候補数に影響を与えない.

表 1 探索候補数の比較

$ TR $	提案手法	基本手法	減少率 (%)
1	17370	18140	4.2
2	112302	120858	7.1
3	300186	324864	7.6

いずれの  $|TR|$  についても提案手法による候補の減少が確認できた. 時間帯が複数になると減少率が大きくなり, 複数の時間帯に拡張した場合の問題点が解消されていることが考えられる. また, 各  $|TR|$  のエピソードマイニングの平均ユーザ実行時間についても基本手法から減少しており,  $|TR| = 3$  では 3.9% 減少した. 新たに追加された  $UpperLimit$  計算行程の総処理量が, 提案手法によって探索されなくなったエピソードのオカレンス作成の総処理量を下回ったためであると考えられる.

## 5 おわりに

本研究では時間帯の偏りを考慮した長時間エピソードマイニングについて, 複数の時間帯に注目した手法を提案した. 提案手法は時間帯における継続時間の見積りにおいて重複を考慮して過剰な加算を減らすことで探索候補の削減を実現した. 評価実験を行い, 時間帯の指定数ごとの探索候補の減少率, 候補減少に伴う実行時間の減少を確認した. 今後は時間帯偏り度による探索候補削減手法について検討をおこなう.

### 参考文献

- [1] A. Ksibi, A.S. Alluhaidan, A. Salhi, S.A. El-Rahman, Overview of Lifelogging: Current Challenges and Advances, IEEE Access, 2021.
- [2] C. Wu, Y. Lin, P. Yu, V. Tseng, Mining High Utility Episodes in Complex Event Sequences, ACM SIGKDD, 2013.
- [3] 安井, 新谷, 大森, 藤田, 行動時間帯に偏りのある長時間エピソード抽出における発生区間の範囲による探索候補の枝刈りの提案, FIT2021, 2021
- [4] 橋本, 新谷, 大森, 藤田: 長時間エピソードマイニングにおける冗長な重複の回避によるオカレンス作成処理の削減手法, FIT, 2022.