

時系列イベントデータからの行動パターン発見 ～ 行動歩数記録データの前処理方法の検討～

Preprocessing Algorithm for Discovering Behavioral Patterns in Time Series Event Data

植野 研^{*1}
Ken UENO

櫻井 茂明^{*1}
Shigeaki SAKURAI

^{*1} (株)東芝 研究開発センター 知識メディアラボラトリー
Knowledge Media Laboratory, Corporate Research & Development Center,
TOSHIBA CORPORATION

In these days ubiquitous sensors are used in various personal situations, such as trekking mountains with mobile GPS tools, joining exhibitions with RFID tags and so on. These ubiquitous computing devices provide situation-dependent services for users in a real time way. However, with the ubiquitous equipments, tremendous amounts of user's behavioral trace data are being accumulated into time series databases, which often contain complex data structures. We propose preprocessing algorithm for discovering method in order to find behavioral patterns in the time series event data logs with human behavioral characteristics.

1. 問題の背景

近年、健康改善を目的としたユビキタスセンシング技術に注目が集まっている。[野村総研 02]では、ユビキタス健康安心システムを取り上げ、その場の状況に合わせて運動時の健康管理を支援する、きめ細かなサービスを提供するコンシェルジェ型健康支援システム構想を提案している。ウェアラブルセンサやユビキタスセンサは、常にユーザの状況をモニタリングできるので、このようなきめ細かなサービスを実現するのに必須の技術であると考えられる。

川村らは、ユビキタス環境において、個人の行動履歴をセンサデータとして獲得し、この情報をもとに群衆の行動を如何に最適化するかの問題をテーマパーク問題として定義し、マルチエージェントによって群衆行動の調整を解決する方法について検討している[川村 03]。また、RFID タグから群衆のマクロ的な移動履歴を収集し、グラフを用いてロケーションモデルを作成するシステムがある[村上 02]。これらは群衆の通過地点の空間的位置関係を把握するのが目的である。

しかしながら、これらの方法では、群衆の位置情報を分析することが目的であり、健康支援のような、個人の行動履歴をきめ細かく分析することは困難である。そこで、個人の位置検出装置を使って、個人のコミュニケーションをサポートする試み[Nakanishi 02]や、ユーザモデルを生成する試み[Samulowitz 00]などがなされている。しかしながら、位置取得センサを用いるのでは、センサの基地局近辺にユーザが近づいたときのみその地点にいると判定されるため、ユーザの状態を常にモニタリングすることは困難である。また、センサのみでは、行動の意味までは推し量ることができない。

本研究は、これらの個人行動履歴収集に関する問題を、携帯型センシングと行動記録により解決する。携帯型センサとして歩数計を、行動記録手段として PDA を用いた。これにより、歩数計から運動状況を、行動記録から活動状況を抽出する。これらの関係性を明らかにすることで健康行動を支援することが最終目的であるが、本稿では、健康・不健康パターン抽出のための前処理方法を検討する。本方法を用いると、個人行動コーパスを作成することができる。このコーパスを利用することで、パターン発見器は、データクレンジング後のデータを用いてパターンを抽出することが可能であるので、応用領域依存の部分の前処理にて吸収できる。したがって、領域非依存な発見器を用いることができる[Fayyad 93]。

また、本処理手法を用いると、小規模なデータにおいては、時系列的なイベントデータからの行動パターン発見が容易になる。モバイルマーケティングの分野では、ひとりひとりの顧客行動を認識する技術を用いて販促戦略を立てる方法が考えられている[神場 02]。しかしながら、この技術は現在発展段階にあり、認識システムを高度化するには、人間行動を認識するいわば行動辞書が必要であると述べられている。

ここでは、階層行動モデルを用いることで、行動記録から個人行動グラフを生成する方法を紹介する。これらの定量・定性データが混在した個人行動グラフを分析することで、日常の運動状況と活動状況との関連性を分析することができる。この関係性が明らかになれば、生活習慣病防止のための健康改善支援[健康日本21 01]などに応用が可能であると考えられる。また、マーケティング分野での顧客行動分析技術にも応用可能であると考えられる。本稿では、歩数計測と行動記録により行動データを時系列イベントデータとして蓄積する方法と、行動記録モデルを用いた行動パターン発見のためのデータ前処理手法を提案する。また、小規模な日常行動データを用いた実験結果を最終章で述べる。

連絡先: 植野 研, E-Mail: ken.ueno@toshiba.co.jp,
〒212-8582 川崎市幸区小向東芝町1
(株)東芝 研究開発センター 知識メディアラボラトリー,
(044)-549-2398

2. 行動データ収集実験

2.1 収集方法

2ヶ月間に渡り被験者3名に日常行動を歩数計により計測させ行動を記録させた。行動記録にあたっては「主に出社してから帰宅するまでの行動を記録していただきたい」「記録できる範囲内で行動記録を取っていただければよい」「休日のプライベートな時間については記録しなくてもよい」との教示を与えた。

2.2 行動記録方法

「行動記録データ」を「行動イベント」、「行動」、「上位行動」の3種類に分類した。「行動イベント」とは、ある地点でのその時刻における行き先を示したベクトルとその行動ラベルである。例えば、「会社から出張先へ出発する」は行動イベントの1例である。行動とは、「行動イベント」の対である。例えば、「会社から出張先へ出発する」と「出張先から会社へ帰還する」の対は、行動「出張」を示している。「上位行動」とは、特殊な形の「行動」であり、「行動」対または一つの「行動イベント」と一つの「行動」からなる対である。例えば、行動「出勤」と行動「退勤」は行動「通勤」を表現できるし、行動「退勤」と、行動イベント「会社からAストアに立ち寄り」とで行動「寄道帰宅」を表現できる。ここでは上位行動自体が行動であり、再帰的な定義になっている点に注意が必要である。この上位行動を導入することで階層的な行動を表現できる。実験ではこれらの3つの定義を以下のように記録させた。

まず、行動記録簿に出発地点(FROM)、目的地点(TO)の2つの基本要素とそれに付随する時刻、歩数値、メモを記入させた(図1)。地点には現在地指定(#)を付けるようにした(図2)。例えば家から会社へ出発するのを示すのに FROM に#home を TO に office を記入する。この時点での時刻、歩数値、メモも同時に記入する。行動記録簿では各行が行動記録の最小単位となる。

行動記録簿では、時刻、歩数、地点、メモを書き込んでいく。しかしながら、これだけでは各行動記録データがどんな内容を表しているのかは不明である。そこで、これらのデータがどんな内容を示しているのかを行動イベント定義集合(図3)に書き込んでおく。行動イベント定義集合とは行動記録中の FROM と TO から行動イベントを生成するルールベースである。例えば、先ほどの例では FROM に#home が、TO に office が指定された場合、「家から会社へ出発する」という行動イベントの定義を書き込んでおく。この定義を参照することで行動記録から行動イベントを生成することができる。行動イベント定義集合は基本的なものは実験者側で用意したものを使用するが、それ以外のユーザによる定義は、その都度、被験者に書き込んでもらうようにした。

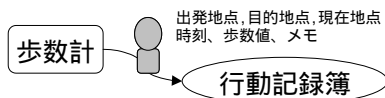


図1 行動記録方法

「行動」と「上位行動」は行動辞書集合に生成ルールを書き込ませた。「行動」と「上位行動」は、ここでは別個にせず、行動辞書に混在させた。「行動」と「上位行動」は、相互に関係性がある。関係性を矛盾なく間違いなく記録させるために、これを一括して記録させた(図4)。この行動辞書も、実験者が用意したデフォルトのもの以外は、その都度、被験者に作成してもらった。

		2003	2	13	thu	
時	分	歩数	FROM	TO	メモ	
9	31	0	#home	office		
10	32	3038	home	#office	via-薬局	
12	23	3038	#office	lun		
13	4	3310	office	#lun		
13	29	3689	#office	hon-mtg	図書委員	
14	10	4375	hon-mtg	#office		
17	5	4375	#office	e-mtg		
18	38	4535	e-mtg	#office	勉強会	
19	24	4535	#office	din		
20	3	5224	din	#office		
21	7	5256	#office	home		
21	29	7543	office	#vsta		
22	1	8339	vsta	#home		

図2 行動記録簿の一例

FROM	TO	EVENT
#home	office	家から会社へ出発
home	#office	家から会社へ到着
#office	home	会社から家へ出発
office	#home	会社から家へ到着
#office	lun	昼食へ出発
lun	#office	昼食から帰還
#office	din	夕食へ出発
din	#office	夕食から帰還

図3 行動イベント定義集合の例

ACT1	ACT2	BHV
家から会社へ出発	家から会社へ到着	出勤
会社から家へ出発	会社から家へ到着	退勤
昼食へ出発	昼食から帰還	昼食
夕食へ出発	夕食から帰還	夕食

図4 行動辞書集合の例

3. 行動記録モデル

筆者等の6ヶ月に渡る予備実験と、2ヶ月に渡る3名の被験者による本実験から、各被験者の行動記録データは、独特の構造をもっていることが分かった。ユーザは、詳細に一つ一つのデータを記録する場合もあれば、状況によって行動記録を柔軟に変更することもある。この予備実験結果を考察すると、行動記録の取り方は、以下の5つの型に分類できることが分かった。これらの型を、行動記録モデルと呼ぶことにする。このモデルは以下に示す5つの型に分類できる。

(1) 基本型 (BHV: A1→A2)

FROM	TO	Label
#A1	A2	→ E1
A1	#A2	→ E2
---辞書---		
EV1	EV2	BHV
E1	E2	<u>B</u>

(2) 回帰型 (BHV: A1→A2→A1)

FROM	TO	Label
#A1	A2	→ E1
A2	#A1	→ E2
---辞書---		
EV1	EV2	BHV
E1	E2	<u>B</u>

(3) デイジーチェーン型 (BHV: A1→A2→A3)

FROM	TO	Label
#A1	A2	→ E1
A2	#A3	→ E2
---辞書---		
EV1	EV2	BHV
E1	E2	<u>B</u>

(4) 寄道型 (BHV: A1→ [A3→] A2)

FROM	TO	Label
#A1	A2	→ E1
A1	#A3	→ E2
A1	#A2	→ E3

---辞書---

EV1(BHV)	EV2	BHV
E1	E2	B1
B1	E3	B2

(5) 梯子型 (BHV: A1→A5< A2A3>→A4)

FROM	TO	Label
#A1	A2	→ E1
A3	#A4	→ E2

---辞書---

EV1(ER1)	EV2(ER2)	BHV(ER3)
A2	A3	A5
E1	E2	B

詳細な記録が可能な場合、被験者の多くは(1)の基本型を用いる傾向がある。これは「行動イベント」のベクトル部分を記録していることになる。この基本型では、出発する地点 = 出発した地点であり、かつ目的予定の地点 = 到着地点となっている。「行動イベント」のイベント自体は、行動イベント定義に書き込んである。(2)の回帰型は、出発地点から出発し、ある地点を経て出発地点へ戻る行動イベント対を記録することで「行動」を表現している。これは、通勤などの行動を表現する場合に用いられる。行動を生成するための行動イベントまたは行動対は、行動辞書に書き込んである。同様に、行動辞書を使った行動記録の仕方が(3)-(5)に見られた。

4. 行動記録データの前処理手法

3章で考察した行動記録モデルより、「行動記録」から「行動」を生成するための前処理方法を本章にて提案する(図5)。まず、行動イベント定義を用いて、行動記録簿から行動イベントを抽出する。そして、これらの行動イベントから、行動辞書を用いて、個人行動パターンを生成する。個人行動パターンは、個人行動グラフとして図示される。これらより、個人行動パターン抽出のための前処理アルゴリズム `bp_pre`(図6)を定義できる。

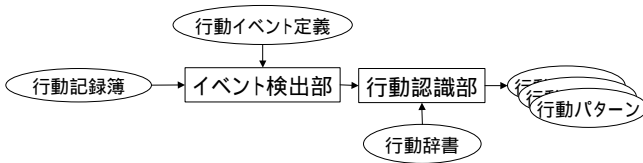


図5 処理手順

以下に、具体的な処理手順を示す。

- (1) 地点集 L : $L = \{ l_1, l_2, \dots, l_n \}$ n 個の地点集合を定義する。
- (2) 行動記録 $er(EID, LF, LT, T, CP, M)$ を定義する。
ただし、 EID は各記録のIDで整数、 LF は出発地点で $LF \#L$ 、 L 、 LT は到着地点で $LT \#L$ であり、 $\#$ は現在地を表す記号である。ここで、 $\#L$ は、集合 L の各要素の頭に $\#$ 記号をつけた集合であると定義する。 T は時刻で表記は「時:分」、 CP は累積歩数で整数値を、 M はその時点での覚書きを格納する文字列変数である。
- (3) 行動イベント集合 $LogDataSet$ を記録する。

行動イベント集合 $LogDataSet$ は m 個の行動記録 er からなる集合であり、以下のように表現する。
 $\{ er(1, LF_1, LT_1, T_1, CP_1, M_1), er(2, LF_2, LT_2, T_2, CP_2, M_2), \dots, er(m, LF_m, LT_m, T_m, CP_m, M_m) \}$

- (4) 行動イベント定義集合 $DefDataSet$ と、行動辞書集合 $DicDataSet$ を定義する。

行動イベント定義集合 $DicDataSet$ は、行動イベント定義 $def(VID, VL, LF, LT)$ を q 個含み、 $\{ def(1, VL_1, LF_1, LT_1), def(2, VL_2, LF_2, LT_2), \dots, def(q, VL_q, LF_q, LT_q) \}$ で表現する。 VL_q はイベントラベルである。また、 VL をイベントラベル集合とする。また、行動辞書集合 $DicDataSet$ は、行動辞書項目 $dic(DID, DL, A_1, A_2)$ を r 個含み、 $\{ dic(1, DL_1, A_{11}, A_{21}), dic(2, DL_2, A_{11}, A_{22}), \dots, dic(r, DL_r, A_{1r}, A_{2r}) \}$ で表現する。ただし r は行動辞書集合の集合サイズ、 DL_r は行動ラベルである。また、 DL を行動ラベル集合とする。 A_{1r} A_{2r} は、ともに行動イベント、または行動ラベルであり、 A_{1r} VL DL 、 A_{2r} VL DL である。

- (5) 行動集合 $BhvSet$ を定義する。
(4)の結果と(3)から、行動集合 $BhvSet$ を生成する。行動 $bhv(BID, BL, A_1, A_2, BT, TT, CP, PV, M)$ を p 個含み、 $\{ bhv(1, BL_1, A_{11}, A_{12}, BT_1, TT_1, CP_1, PV_1, M_1), bhv(2, L_2, A_{21}, A_{22}, BT_2, TT_2, CP_2, PV_2, M_2), \dots, bhv(p, BL_p, A_{p1}, A_{p2}, BT_p, TT_p, CP_p, PV_p, M_p) \}$ で表現する。ただし、 p は集合サイズ、 BL_p は、行動イベント定義から得られる行動イベントラベル、もしくは、行動辞書から得られる行動ラベルで BL_p VL DL 、 TT_p は所要時間、 CP_p は開始時累積歩数、 PV_p は所要歩数である。
- (6) アルゴリズム `bp_pre`を実行する。

ただし、数値変数中の n/a は該当数値がないことを、文字列変数中の " $"$ " は NULL を示すこととする。また、 $BhvSetA$ は行動集合であり、検査対象の行動集合として定義し、 $bhv(AID, BL, LF, LT, T, TT, CP, PV, M)$ と表記する。また、 $\#(X)$ は L に $\#$ が指定されているときは TRUE、ついていないときは FALSE を返すブール関数とする。

Algorithm: `bp_pre(LogDataSet, DefDataSet, DicDataSet)`

```

{
  BhvSetA :=
  BhvSetA := event_dtct(LogDataSet, DefDataSet)
  BhvSet := BhvSetA
  BhvSet := bhv_idntfy(BhvSetA, DicDataSet, BhvSet)

  return BhvSet
}
    
```

Procedure: `event_dtct(LogDataSet, DefDataSet)`

```

{
  BhvSetA := , EID:=1, VID:=1
  for each er(EID, LF, LT, T, CP, M) LogDataSet
    for each def(VID, VL, VF, VT) DefDataSet
      if LF=VF LT=VT then
        BhvSetA := BhvSetA bhv(EID, BL, LF, LT, T, n/a, CP,
                                n/a, M)

  next VID
  next EID

  return BhvSetA
}
    
```

Procedure: `bhv_idntfy(BhvSetA, DicDataSet, BhvSet)`

```

{
  if (BhvSetA )
    NbhvSet :=
    for each bhv(BID, BL, LF, LT, BT, TT, CP, VP, M) BhvSet
    
```

```

if TT n/a then
  NbhvSet
  := NbhvSet bhv(BID,BL,LF,LT,BT,TT,CP,VP,M)
next BID

return NBhvSet
else
for each bhv(AID,BL,LF,LT,BT,TT,CP,VP,M) BhvSetA
BhvSetA := BhvSetA \ bhv(BL,LF,LT,BT,TT,CP,VP,M)
for each dic(DID, DL,A1,A2), DicDataSet
if BL=A1 then
for each bhv(BID,BL2,LF2,LT2,BT2,
              TT2,CP2,VP2,M2) BhvSet
if BL2=A2 then MN:= M + M2
if #(BL) #(BL2) then TN:= T1+T2, AN:=A1+A2
else TN:= T2-T1, AN:= A2 - A1, AID++,BID++
BhvSet := BhvSet bhv(AID,DL, BL, BL2,BT,
                    TN,CP,AN,MN)
BhvSetA:=BhvSetA bhv(BID,DL, BL, BL2,BT,
                    TN,CP,AN,MN)

Next BID
Next DID
Next AID
return bhv_idntfy(BhvSetA,DicDataSet)
}

```

図6 アルゴリズム bp_pre

5. 処理結果

これらのアルゴリズムより、2日分の実験データを処理した結果を図7と図8に示す。図7と図8を比較すると、図7の場合は、帰りにタクシーを利用したが、利用しない日とほとんど所要時間が変わらないにもかかわらず、一日の歩数総数が約 4,000 歩もすくなくなってしまうことが分かる。これらから、図8の例は図7の例よりも効率よく、しかも多く歩行運動をした行動を示しているものと考えられる。

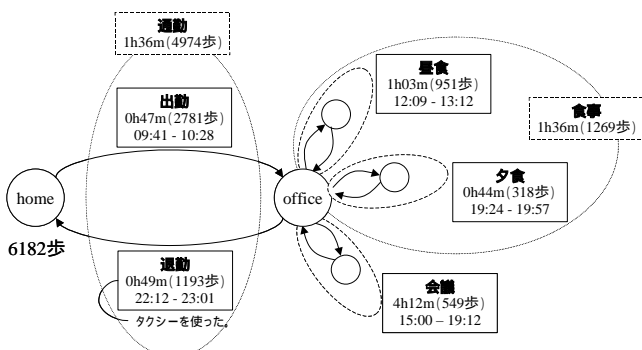


図7 処理結果の一例

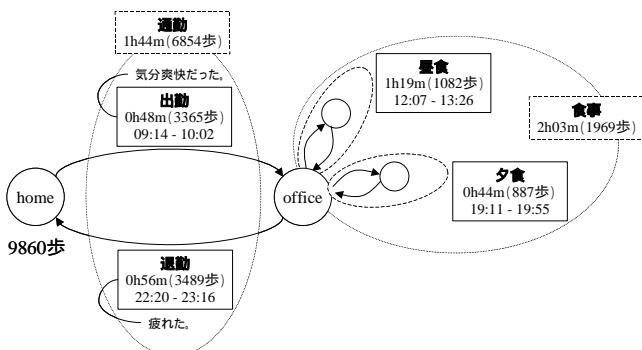


図8 処理結果の一例

6. 考察

「行動記録データ」を「行動イベント」、「行動」、「上位行動」の3種類に分類し、行動記録モデルを用いることで、簡潔な前処理アルゴリズムを作成することができた。行動記録モデル(1)(2)(3)(5)は、行動イベント定義と行動辞書を用いることで、問題なく行動を生成することができた。

また、行動記録モデル(4)により、階層的な個人行動グラフを生成する方法を明らかにすることができた。特に、寄道型の行動記録の場合、起点Aから終点Bに向かう途中でサブゴールがn個存在する(C_n)。この場合、再帰処理を用いることで、これらの記録を一つの行動にまとめることができる。すなわち、 $((\dots((A \ C_1) \ C_2) \dots \ C_i) \ B)$ のように定義できる。これを擬似コードに変形すると、

```

do A:=act(A,Ci) while i<n.
return act(A,B).

```

となり、行動辞書を用いた再帰処理により、これらの行動記録から行動グラフが生成できることが分かった。

7. 今後の課題とまとめ

歩数計測と行動記録により、日常行動の行動履歴収集実験を行った。行動履歴から行動記録モデルを考案し、この記録モデルに基づいて、行動定義・行動辞書を用いて個人行動グラフを生成する前処理アルゴリズムを実装した実装プログラムを用いて、実験データから個人行動グラフを生成することができた。本方式は、時系列パターン発見のための前処理として十分機能することを確認した。今後は、健康分野だけでなく、定量・定性混在型の個人行動履歴データからの行動パターン生成がキラーアプリケーションとなる分野を見定めることが課題である。

参考文献

[川村 03] 川村 秀憲, 車谷 浩一, 大内 東: テーマパーク問題のマルチエージェントによる定式化と調整アルゴリズムに関する検討. 情処研報 Vol.2003, No.8, p.25-30, 2003-ICS-131, 2003.

[神場 02] 神場 知成: モバイルマーケティングと AI. 人工知能学会誌 17 巻 5 号 p.532-537, 2002.

[健康日本21 01] (財)健康・体力づくり事業財団: 「健康日本21」ホームページ, <http://www.kenkounippon21.gr.jp/>

[Samulowitz 00] M.Samulowitz: Towards Context-aware User Modeling , Trends in Distributed Systems: Towards a Universal Service Market, Proc of the 3rd International IFIP/GI Working Conference, pp.272-277, 2000 .

[Nakanishi 02] Y.Nakanishi, K.Takahashi, T.Tsuji, K.Hakozaki: A Mobile Communication Tool Using Location and Schedule Information , Proc of the 1st Pervasive Computing, pp.239-252, 2002 .

[野村総研 02] 野村総合研究所, コビキタス・ネットワークと新社会システム, p.145-169, 2002.

[Fayyad 93] U. Fayyad, G. Pietetsky-Shapiro, P. Smyth, R. Uthurusamy: Advances in Knowledge Discovery and Data Mining. MIT Press, 1996.

[村上 02] 村上 朝一, 中西 健一, 桐原 幸彦, 徳田 英幸: ロケーションモデル作成支援機構の構築 日本ソフトウェア学会ソフトウェアシステム研究会 SPA サマワークショップ, 2002.