

# 生活習慣と健康状態の時系列データ解析における重み付けの検討 (I)

## 一日毎の任意係数による重みづけ

黛 勇氣<sup>†</sup> 竹内 裕之<sup>‡</sup> 児玉 直樹<sup>‡</sup>

<sup>†</sup>高崎健康福祉大学大学院 健康福祉学研究科 〒370-0033 群馬県高崎市中大類町 37-1

<sup>‡</sup>高崎健康福祉大学 健康福祉学部 医療情報学科 〒370-0033 群馬県高崎市中大類町 37-1

E-mail: <sup>†</sup>0910404@takasaki-u.ac.jp, <sup>‡</sup>htakeuchi@takasaki-u.ac.jp, <sup>‡</sup>kodama@takasaki-u.ac.jp

あらまし 我々は、個人の生活習慣と健康状態に関する日常のデータをインターネット上に蓄積し、生活習慣と健康状態に関する個人毎の相関ルールを抽出する（健康データマイニング）システムを試行中である。健康データマイニングでは、相関ルールを抽出する際に、ある期間の生活習慣データの加算値と健康データの増減の時系列相関を基にデータセットを生成する。これまで、生活習慣データの加算は単純加算で処理している。本研究では、時系列データ解析の精度をより高めることを目的とし、生活習慣データの加算処理へ重み付けの概念を導入した。ここで発見できる適切な重み係数は、ユーザの生活習慣の改善にとっての有用な情報となることが考えられる。

キーワード 健康データマイニング, 時系列データ解析, 重み付き加算, 生活習慣, 健康状態, 相関ルール

## Study of Weighting in Time-Series Data Analysis of Lifestyles and Health Conditions (I) - Weighting with daily arbitrary coefficients -

Yuuki MAYUZUMI<sup>†</sup> Hiroshi TAKEUCHI<sup>‡</sup> and Naoki KODAMA<sup>‡</sup>

<sup>†</sup>Post-Graduate Course of Health and Welfare, Graduate School of Takasaki University of Health and Welfare, 37-1 Nakaorui-machi, Takasaki-shi, Gunma, 370-0033 Japan

<sup>‡</sup>Department of Healthcare Informatics, Faculty of Health and Welfare, Takasaki University of Health and Welfare  
37-1 Nakaorui-machi, Takasaki-shi, Gunma, 370-0033 Japan

E-mail: <sup>†</sup>0910404@takasaki-u.ac.jp, <sup>‡</sup>{htakeuchi, kodama}@takasaki-u.ac.jp

**Abstract** A personal dynamic healthcare system which provides a healthcare-data-mining service and notifies users, through their mobile phones, of important rules relevant to their health and lifestyle on the basis of users' stored data is now on a test run. In the healthcare-data-mining, time-series correlation between a summation of lifestyle data and a variation of health data is examined. In this study, a concept of weighting is introduced in the summation of daily lifestyle data. The appropriate weighting coefficients obtained in this analysis could be useful information for users to improve their lifestyles.

**Keyword** Healthcare-data-mining, Time-series data analysis, Weighting

### 1. はじめに

インターネットや家庭内 LAN などコンピュータ・ネットワークの急速な発展に伴い、アプリケーション・ソフトウェアはクライアントサイドからサーバサイドへと移行しつつある。従来の ASP (Application Service Provider) は SaaS, クラウドコンピューティングと発展しながら、少しずつではあるが着実に我々の生活に入り込んでいる。この潮流は健康医療分野も例外ではなく、近年では民生用の血圧計や体重計、歩数計などを PC (Personal Computer) や BT (Bluetooth) などのワイヤレス・ネットワークに接続し、データを DB (Data Base) サーバに蓄積することが可能となりつ

つある。我々はいち早くこれらのデータソースの活用に着目し、自動健康データマイニングをインターネット上でサービスする個人健康管理システム (Personal Dynamic Healthcare System : PDHS) を開発してきた [1, 2]。このシステムは、携帯電話 (Cell Phone) を情報端末として、個人が自分自身の生活習慣と健康状態のデータをインターネット上の DB サーバに蓄積する。一定数のデータが蓄積されると、自動健康データマイニングのサービスがユーザ毎にデータを解析し、その個人にユニークな相関ルールを抽出し、結果をユーザに通知するというシステムである。現在はこのシステムをさらにブラッシュアップし [3, 4], PC からのデー

タ入力・閲覧・修正や、BT 通信機能付き血圧計や体重計などからの計測時点でのデータ入力，つまり計測時点情報管理に対応し，さらに保健指導者がこれらのデータを参照しながら WEB を介して各ユーザに指導コメントを付加する機能も備えるまでに至っている [5]．本研究では，このシステムの核である健康データマイニングのアルゴリズムにおいて重要な役割を担っている，生活習慣の蓄積と健康状態の変化に関する時系列データ解析の精度をより高めるため，時系列データ解析における生活習慣データの加算処理へ重み付けの概念を導入した．

## 2. 研究方法

### 2.1. 健康データマイニング

我々が開発中の健康データマイニングでは，生活習慣と健康状態の関係を「生活習慣の蓄積が遅延して健康状態の変化に影響を及ぼす」というシンプルな概念でモデル化している (図 1)．つまり，ある健康状態の変化 (体重や体脂肪率，体温の増減など) を目的変数とし，その前日から数日前までに起こった生活習慣 (摂取，消費カロリーや睡眠時間，運動量など) の蓄積を説明変数として，個人毎にユニークな相関ルールを抽出する．すなわち，個人における健康状態の変化は，その個人の生活習慣と何らかの関連があり，その生活習慣の蓄積 (加算) が個人の健康状態の変化に反映されているというモデルである．

健康データマイニングのアルゴリズムでは，データのスクリーニング，相関ルールの生成と評価というスキームで相関ルールを抽出する．まずデータの事前処理として，注目する健康データ項目とある程度関係がありそうな生活習慣データ項目を事前に評価するため，式 (1) で表されるピアソンの積率相関係数を評価して

$$r(\Delta h_{nm}, e_{ij}^t) = \frac{\text{Cov.}(\Delta h_{nm}, e_{ij}^t)}{S.D.(h_{nm})S.D.(e_{ij}^t)} \quad (1)$$

時系列相関を求め，目的変数毎に入力変数をスクリーニングする．

ここで，

$$\Delta h_{nm} = h_n - h_m \quad (2)$$

は目的変数である健康状態  $h$  の差分値であらわされる変化であり，

$$e_{ij}^t = e_i + e_{i-1} + \dots + e_j \quad (3)$$

は説明変数である生活習慣の加算値である (図 2)．式 (1) における  $S.D.(\Delta h_{nm})$  は，ある健康データ  $h$  のあ

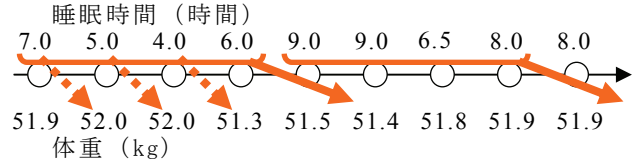


図 1 時系列データ処理の模式的説明図

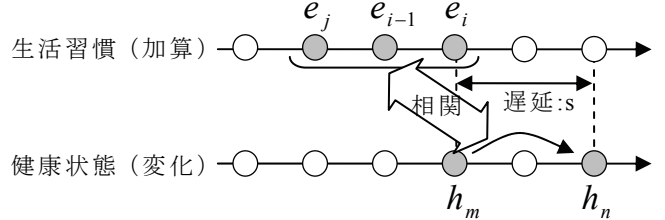


図 2 時系列データ処理の概念図

る区間の差分値の標準偏差， $S.D.(e_{ij}^t)$  は遅延を考慮したある生活習慣データ  $e$  のある区間の積算値の標準偏差，また  $\text{Cov.}(h_{nm}, e_{ij}^t)$  はそれらの共分散である．この  $r(h_{nm}, e_{ij}^t)$  がある閾値を超えたものを，対象とする目的変数 (健康状態) についての健康データマイニングの入力変数として採用する．

次に，時系列相関の評価によるスクリーニングで採用された生活習慣の蓄積を入力変数  $Y$  とし，対象とする健康状態をその時系列データが「高い」「中間」「低い」の 3 つのシンボル値を持つ出力変数  $X$  として，ITRULE アルゴリズム [1] を用いた相関ルールマイニングを行う [2]．ITRULE アルゴリズムは，

$$\text{If } Y=y, \text{ then } X=x \text{ with probability } p \quad (4)$$

という相関ルールを生成する．このアルゴリズムでは，多くのデータセットから有効な相関ルールを抽出するため，(5) 式で表される  $J$  測度を用いてルールを評価する．

$$J(x|y) = p(y) \left( p(x|y) \log \frac{p(x|y)}{p(x)} + (1-p(x|y)) \log \frac{(1-p(x|y))}{(1-p(x))} \right) \quad (5)$$

$J$  測度は  $Y=y$  という事象が起きた場合に， $X$  の値に関して得られる情報量の大きさ，つまり  $Y=y$  という前提が有る場合と無い場合で， $X$  の値に関する確率分布がいかに異なるかという尺度に，実際に  $Y=y$  という事象が起きる確率  $p(y)$  を掛けたものであり，この値が大きいほどよい相関ルールということになる [1]．

### 2.2. 重み付き時系列相関

健康データマイニングは 2.1. で述べたように，健康状態の変化と生活習慣について時系列に解析する際，生活習慣の積み重ね (加算) や遅れ (遅延) を考慮すると，1 日単位で比較するより，高い相関係数が得られる場合があるという発見に基づいている．このとき，

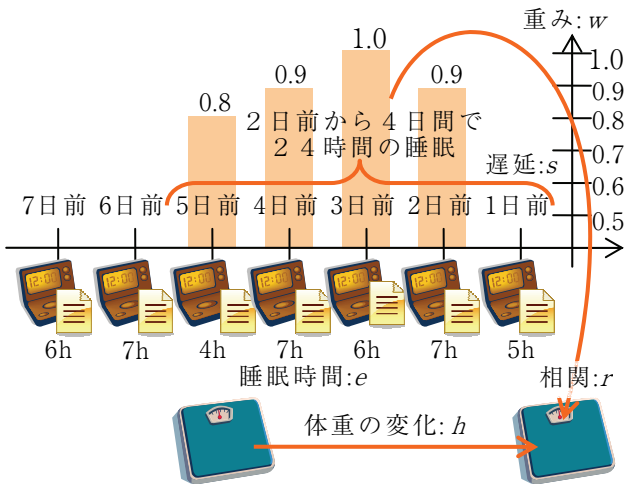


図3 生活習慣データに考慮する重みの概念図

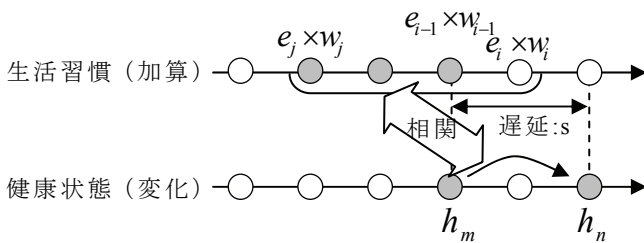


図4 重みつき加算の概念図

最大の相関を示す生活習慣の加算範囲や遅延日数には個人差があり、さらにデータの種類によっても異なるため、時系列データ処理ではこの加算範囲と遅れを動的に算出する。ここで、単純加算によって算出された加算値は、加算範囲における全ての値が一樣な重み ( $w$ ) として単純化されているといえる。つまり、本来であれば加算範囲における個々の値が、それぞれ独立した重みをもつことが考えられるが、これを一樣の重み ( $w=1$ ) として加算している。また、図2における遅延日数 ( $s$ ) は、図3における1日前の重みがゼロ ( $w=0$ ) になる場合と解釈することができる。重み付き時系列相関解析では図3の例に示すように、まず動的だった加算の範囲を固定、それぞれの値に独立した重み ( $w$ ) を考慮し、 $w$  と  $e$  の積の加算値を算出する。このとき、 $w$  の取りうる値の範囲はゼロ～1.0 とする。つまり、図4に示すように生活習慣データ ( $e$ ) を  $i$  から  $j$  まで加算するとき、それぞれの値に対応する重み ( $w$ ) との積によって値を漸減する。このとき、先に述べたように遅延日数 ( $s$ ) は、遅延に相当する日の値の重み ( $w$ ) がゼロであるとみなすことができる。図4において生活習慣 ( $e$ ) における加算値を求める場合、これまで動的に扱っていた  $i$  と  $j$  を、 $i=n-1$  として静的に扱う代わりに、それぞれの日の生活習慣 ( $e$ ) の値が独立でもつ重み ( $w$ ) を変化させることによって加算値を算出する。すなわち、重みを考慮した加算値は、

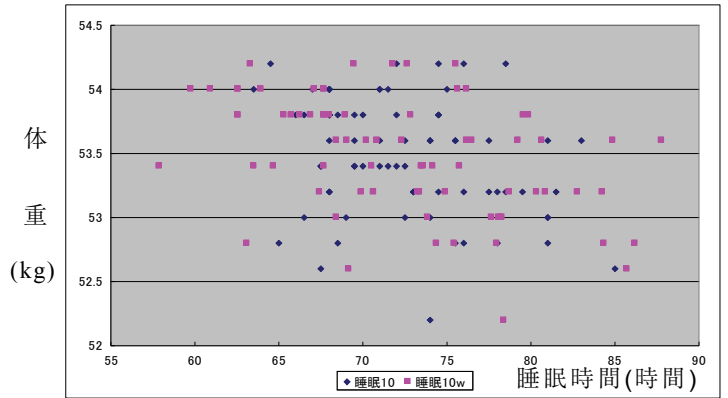


図5 睡眠時間の単純加算および重み付け加算と体重の散布図

$$e'_j w'_j \frac{i-j}{i-j-(1-w'_j)} = e'_j w'_j \times \left( \frac{i-j}{i-j-(1-w'_j)+(1-w'_{j-1})+\dots+(1-w'_j)} \right) \quad (6)$$

という式によって導かれる。式(6)では、 $(i-j)/(i-j)$  の分母に軽さ  $(1-w)$  の和を考慮することで、重み付けをした加算値を補正する。これは、生活習慣 ( $e$ ) と  $w$  との単純な積を加算した場合、それぞれの  $e$  を重み  $w$  によって漸減した分、期間の加算値がずれるのを防ぐためである。ここで得られる重み付き加算値を、式(1)における  $e$  の加算値と置き換え、時系列相関を算出する。重みを考慮した時系列相関係数は、

$$r \left( \Delta h_{nm}, e'_j w'_j \frac{i-j}{i-j-(1-w'_j)} \right) = \frac{\text{Cov} \left( \Delta h_{nm}, e'_j w'_j \frac{i-j}{i-j-(1-w'_j)} \right)}{S.D.(h_{nm}) S.D. \left( e'_j w'_j \frac{i-j}{i-j-(1-w'_j)} \right)} \quad (7)$$

という式によって求められる。

例として、2日前から10日間の睡眠時間の単純加算を睡眠10、(7)式を用いた加算を睡眠10wとして計算した散布図を図5に示す。重み付けを行った睡眠10wでは、 $w$  によって値が減少した分だけ移動している。この例において、それぞれの相関係数を算出すると、睡眠時間の単純加算と体重では  $r=-0.318$ 、重み付き加算と体重では  $r=-0.439$  が最大値となる。このように、重み付き加算では考慮した重みによって異なる相関係数が得られる。実際の重み付けでは、この相関係数の絶対値が最大となる重みのパターンを探索することにより、時系列相関における精度の向上を図る。

### 3. 解析結果

#### 3.1. ユーザAにおける重み付き加算

ユーザAの睡眠時間と体重変化の時系列データを用いて、遅延を2とした睡眠時間の1~10日間の単純加算と重み付き加算について、体重との相関係数を算出し

表1 ユーザ A の睡眠時間と体重の相関係数

生活習慣の加算範囲	ピアソンの積率相関係数
睡眠時間単純加算 1 日	-0.186
睡眠時間単純加算 2 日	-0.317
睡眠時間単純加算 3 日	-0.298
睡眠時間単純加算 4 日	-0.354
睡眠時間単純加算 5 日	-0.261
睡眠時間単純加算 6 日	-0.222
睡眠時間単純加算 7 日	-0.199
睡眠時間単純加算 8 日	-0.256
睡眠時間単純加算 9 日	-0.281
睡眠時間単純加算 10 日	-0.318
睡眠時間重み付き加算 11 日	-0.439

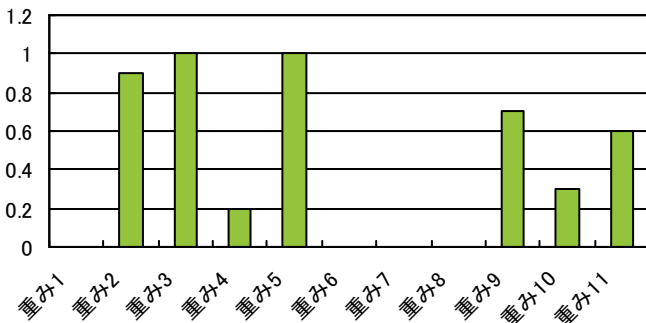


図7 ユーザ A の睡眠時間の加算における重み係数

た結果を表1に示す。睡眠時間の単純加算では、2日前から4日間の加算で負の相関  $r = -0.354$  で最大となったのに対し、重み付き時系列相関では11日間について図7の重みで相関係数  $r = -0.439$  となり、睡眠と体重の間では、この相関係数の絶対値が最大となった。このとき、重み1が前日、重み2が一昨日に考慮する重みに相当し、重み11までがゼロ~1.0の間で0.1刻みの独立した値をとる。なお、この例では重み1がゼロとなっており、この場合は2日前からの生活習慣の加算値を示す。これが従来の遅延日数 ( $s$ )に相当し、 $s = n - 2$  を意味する。前述のとおり表1に示した睡眠1~10には遅延2日を考慮していることから、単純加算10日間（遅延2日）が、重み付き加算11日で重み1がゼロとして表現されている。

### 3.2. ユーザ B における重み付き加算

ユーザ B の睡眠時間と体重変化の時系列データを用いて、睡眠時間 1~11 日までの間の単純加算と重み付き加算について、11日前からの体重の変化との相関係数を算出した結果を表2に示す。体重の変化について、単純加算では11日間の加算が正の相関  $r = 0.436$  で最大となったのに対し、重み付き時系列相関では図8の重みで相関係数  $r = 0.510$  となり、睡眠と体重変化の間では、この相関係数の絶対値が最大となった。なお、表2に示した睡眠1~11には遅延を考慮していない。しかし、遅延を考慮すると若干相関係数は変動するも

表2 ユーザ B の睡眠時間と体重変化の相関係数

生活習慣の加算範囲	ピアソンの積率相関係数
睡眠時間単純加算 1 日	0.153
睡眠時間単純加算 2 日	0.141
睡眠時間単純加算 3 日	0.143
睡眠時間単純加算 4 日	0.268
睡眠時間単純加算 5 日	0.367
睡眠時間単純加算 6 日	0.332
睡眠時間単純加算 7 日	0.360
睡眠時間単純加算 8 日	0.355
睡眠時間単純加算 9 日	0.324
睡眠時間単純加算 10 日	0.384
睡眠時間単純加算 11 日	0.436
睡眠時間重み付き加算 11 日	0.510

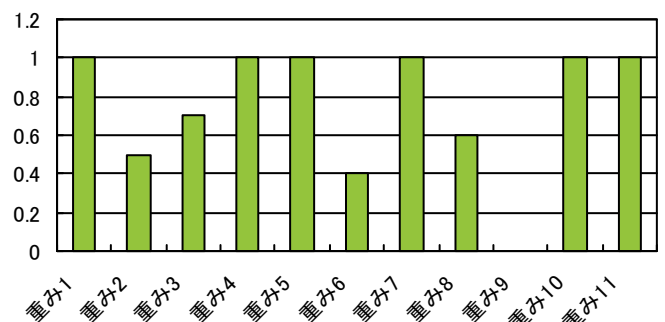


図8 ユーザ B の睡眠時間の加算における重み係数

の、重み付き相関を上回る値は算出されなかった。

## 4. 考察

### 4.1. 単純加算と重み付き加算

これまでの健康データマイニングでは生活習慣データに関して、いずれの日の生活習慣も健康状態の変化に一律に寄与すると仮定し、単純加算で処理していた。本研究で生活習慣の加算処理に重みの概念を導入することで、「特に何日前の生活習慣との関連があるのか」という情報が得られたと考えている。この情報は、ユーザが相関ルールを参考にして実際に自身の健康管理を行う際に、より具体的に個人の傾向を表した情報として役立てることができる。今回の試みにおいて、ユーザ A と B でそれぞれ重み付けを行ったところ、全く異なる重み係数となった。これは、そもそもユーザ A は負の相関であり、ユーザ B は正の相関を示していることから、ユーザの睡眠時間の大小と体重との関係に影響する別の生活習慣因子が絡んでいる可能性があることが考えられる。

### 4.2. 重みの表現方法

これまでの健康データマイニングでは、「 $s$  日前から  $i-j$  日間の生活習慣  $e$  が・・・なら、健康状態  $h$  が・・・の傾向がある」といった相関ルールを抽出しているが、特定保健指導に健康データマイニングを活用する研究

[5,6]のフィールドテストにおいて、このルールの表現における加算日数や遅延が、ユーザや保健指導者にとって理解しにくいという意見が出ている[3,4]。これらの課題については、関連ルールの文章表現による解決を試みているが、重みという要素が加わった場合に、ユーザにとって関連ルールがさらに難解になることを避ける必要がある。本研究をルール生成に適用すると、関連ルールは「生活習慣  $e$  が・・・なら、健康状態  $h$  が・・・の傾向がある」と簡略化され、一方で「重み (0.0 0.9 1.0・・・)」というような情報が新たに付加されることになる。しかし、重みを情報として関連ルールに付加するにあたってユーザが理解しやすい情報とはいえない。また、3.で示したような表もしかりである。重み係数を図 7,8 のようなグラフで表す方法も考えられるが、時系列に並べる場合、図 3 のように左から若い日付となるが、図 7,8 では重みの番号が 1 日前、2 日前と対応するので、この番号が右から若い順に並ぶことに注意が必要である。あるいは、ルールそのものを表で現す方法が考えられる。図 9 は、ユーザ A について抽出された場合の関連ルールを、ガントチャートを活用して表した例である。この図では重みを各日の濃度によって表現しているが、これは線グラフや柱グラフに置き換えてもよい。このように多少工夫することで、重みを考慮した関連ルールは視覚的な表現が可能であると考えている。

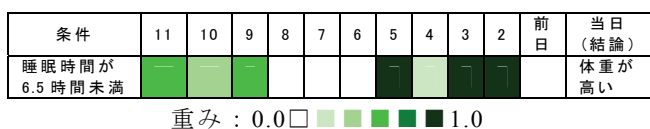


図 9 関連ルールのガントチャート

### 4.3. 新たな課題

本研究のように、個々の重み係数の値をゼロ～1.0 の 11通りとし、全ての重みの組合せから最大の相関係数を探索する手法では、相関を求めるデータセットを飛躍的に増大させる。実際の重み係数は、漸減すると相関係数が下がる ( $w=1$ ) 場合と、漸減によって相関が上がり続ける ( $w=0$ ) 場合、漸減によって相関が上がり、ある程度を超えると下がる ( $w=0.9\sim 0.1$ ) 場合がある。これらを考慮し、より効率的に重みを探索する手法を考える必要がある。

### 5. まとめ

重み付き時系列相関では、入力変数が数値であり時系列加算が行われれば、そこに重み付けを考慮することができる。ここで発見できる個人毎の重み係数は、ユーザの生活習慣と健康状態に関する傾向をより詳細に表現した情報として、ユーザの生活習慣の改善にお

ける有用な情報となることが期待できる。しかし、いたずらな重み付けはデータの組合せを極端に増加させる原因ともなるので、新たに明らかとなった課題について解決を図り、さらに効率化された重み付けを追究する。

### 謝辞

本研究は文部科学省科研費 (課題番号: 20300222) の助成を受けている。併せて日立 HiRDB アカデミック支援プログラムの適用を受けている。

### 文 献

- “personal dynamic healthcare system,” Proc. 2<sup>nd</sup> Int. Conf. on Computational Intelligence in Medicine and Healthcare, pp.37-43, Lisbon, Portugal, Jun. 2005.
- [2] 竹内裕之, 児玉直樹, 橋口猛志, 林 同文, “個人健康管理を目的とした健康データマイニングシステム,” DEWS2006 1B-i11, Mar. 2006.
  - [3] 上村勝美, 黛 勇気, 児玉直樹, 竹内裕之, “特定健診・保健指導を支援するシステムの開発,” DEWS2008, C5-4. Mar. 2008.
  - [4] 黛 勇気, 児玉直樹, 竹内裕之, 矢部貴之 “健康データマイニングを活用した特定保健指導支援システム,” DEIM2009, E2-2. Mar. 2009.
  - [5] 黛 勇気, 竹内裕之, 児玉直樹, 佐藤恵一 “データマイニング機能を備えた健康管理 ASP システムの開発,” DEIM2010, E7-2. Mar. 2010.
  - [6] 竹内裕之, 児玉直樹, “生活習慣と健康状態に関する時系列データ解析手法の開発,” DEWS2008 E1-5, Mar. 2008.
  - [7] P. Smyth and R. M. Goodman, An information theoretical approach to rule induction from databases, IEEE Trans. Knowledge and Data Engineering, vol.4, no.4, pp.301-316, Aug. 1992.
  - [8] 竹内裕之, 児玉直樹, 橋口猛志, 林 同文, “個人健康管理システムのための自動関連ルール抽出アルゴリズム”, DBSJ Letters, vol.5, no.1, pp.25-28, Jun. 2006.