

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 ()

- 少数パラメータで特徴付けられる重み付け -

竹内 裕之[†] 黛 勇気[‡] 児玉 直樹[†]

[†]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

[‡]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

E-mail: [†]{htakeuchi, kodama}@takasaki-u.ac.jp, [‡]0910404@takasaki-u.ac.jp

あらまし 「生活習慣の蓄積が健康状態に変化をもたらす、その影響は時間遅れをもって現れることがある」という極めてシンプルなモデルをベースとして「健康データマイニング」手法の開発を行っている。時間遅れは代謝など生体内の複雑な反応を反映していると考えられる。本研究では、生活習慣データの蓄積（加算）に重み付けの概念を導入し、健康状態の変化（差分）との間の時系列相関解析を試みた。その結果、少数のパラメータで表現できる重み付けパターンにより、健康に関する個人差を特徴付けられる可能性が示唆された。

キーワード 健康データマイニング, 時系列データ解析, 重み付け

Study of Weighting in Time-Series Data Analysis of Lifestyles and Health Conditions ()

- Weighting characterized by a small number of parameters -

Hiroshi TAKEUCHI[†] Yuuki MAYUZUMI[‡] and Naoki KODAMA[†]

[†]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

[‡]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

E-mail: [†]{htakeuchi, kodama}@takasaki-u.ac.jp, [‡]0910404@takasaki-u.ac.jp

Abstract The time-series data analysis described here is based on the simple idea that the accumulation of the effects of lifestyle events such as ingestion, alcohol intake, and exercise could affect personal health condition with some delay. The delay may reflect complex bio-reactions such as those of metabolism in a human body. In the analysis, the accumulation of the effects of lifestyle events is represented by a summation of daily lifestyle data whose time-series correlation to variations of health data is examined (healthcare-data-mining). In this study, a concept of weighting is introduced in the summation of daily lifestyle data. As a result, it is suggested that personal health nature could be represented by a weighting pattern characterized by a small number of parameters.

Keyword Healthcare-Data-Mining, Time-Series Data Analysis, Weighting

1. はじめに

健康医療分野の情報化は、従来の医療機関を中心とした情報化の波に加え、e-healthというキーワードに象徴されるように健康分野で大きな進展をみせている。特にインターネットを存分に活用するクラウドコンピューティングの一応用分野として、健康情報サービスが新たな局面を迎えている。最近の医療情報に関する国際学会の潮流としても、m(mobile)-health や p(personalized)-health といった概念が浸透してきた[1]。我々はいち早く、インターネット上で動く自動健康データマイニングシステムを核とした個人健康管理シ

テム (Personal Dynamic Healthcare System: PDHS) を開発してきた[2]。このシステムは、携帯電話 (mobile phone) を情報端末として、個人の日常の生活習慣や健康に関するデータをインターネット上のサーバに蓄積し、サーバ上で生活習慣と健康状態の相関ルール抽出 (健康データマイニング) を行って個人に通知する、正にクラウドコンピューティングをいち早く具現化したものである[3]。現在、我々は、このシステムを健康データマイニング機能を備えた ASP (Application Service Provider) システムとしてさらにブラッシュアップを図っている[4]。

本研究では、我々が開発しているシステムの核となっている「健康データマイニング」のアルゴリズムの中心部を占める、生活習慣と健康状態に関わる時系列データ解析の精度をより高めるために、時系列データ解析の加算処理に重み付けの概念を導入した。

2. 研究方法

2.1. 時系列データ解析手法

我々が開発している健康データマイニングでは、「生活習慣の蓄積が健康状態に変化をもたらす、その影響は時間遅れをもって現れることがある」という極めてシンプルなモデルをベースとしている[5]。すなわち、ある健康状態の変化を目的変数とし、時間遅れを考慮したある期間の生活習慣の蓄積を入力変数として相関ルールを抽出する。相関ルールの抽出には、あらかじめ、ある期間の時系列データを基にして、目的変数である健康状態の変化に影響を及ぼす生活習慣の蓄積をスクリーニングする。スクリーニングには式(1)で表される時系列データ間のピアソンの積率相関係数を用いる。

$$r(\Delta h_{nm}, e^{t_{ij}}) = \frac{\text{Cov}(\Delta h_{nm}, e^{t_{ij}})}{SD(\Delta h_{nm})SD(e^{t_{ij}})} \quad (1)$$

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 の変化を表す差分値であり、

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

は生活習慣 e の蓄積を表す、ある期間に亘る加算値である。時間遅れは遅延期間 $s = n - i \geq 1$ で表現する(図1参照)。式(1)において、 r は相関係数、 $SD(\Delta h_{nm})$ は Δh_{nm} の時系列区間における標準偏差、 $SD(e^{t_{ij}})$ は $e^{t_{ij}}$ の時系列区間における標準偏差、 $\text{Cov}(\Delta h_{nm}, e^{t_{ij}})$ は Δh_{nm} と $e^{t_{ij}}$ の共分散である。

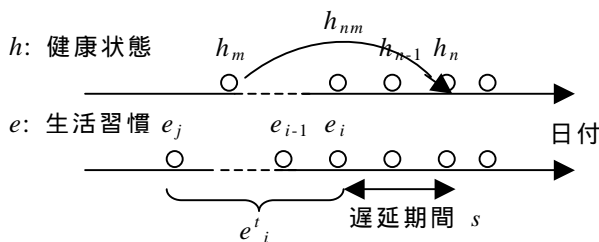


図1 健康状態の変化と生活習慣蓄積の時系列相関の見方

実際のスクリーニングは、対象とする健康状態 h と各種生活習慣 e のある期間の時系列データについて、 $n-m, i-j, s$ をパラメータとして式(1)のピアソンの積率相関係数を評価する。すなわち、各 $(n-m, i-j, s)$ のセットにつき式(1)を評価し、相関係数の絶対値が最大となるセットを探す。そして最大値がある閾値 r_s より大きければ、そのセットに対応する e の蓄積を h に対する入力変数として採用し[6]、ITRULEアルゴリズム[7]によるルールマイニングを実行する[8]。

2.2. 本研究の着眼点

式(3)では、何日間かの生活習慣データを単純加算し、加算日数と遅延期間をパラメータとして時系列相関を評価している。しかし、生活習慣データを過去に遡って毎日に単純に加算するというのは、健康状態との関わりにおいて複雑なはずの生体反応をあまりにも単純化しすぎているかもしれない。過去の生活習慣が現在の健康状態に変化を与えるという観点からは、即効性のあるもの遅効性のあるもの様々なパターンの影響が考えられる。実際に、これまでの研究から遅延期間を考慮して加算すると相関係数の絶対値が著しく増大するケースが多々あった。

そこで、過去の生活習慣データを毎日に適当な重みをつけて加算することで、より精度の高い時系列相関解析ができるのではないかと考えた。重み付け加算の手法として、(1) 毎日に任意の重み付け係数を定義して相関係数の絶対値が最大になるようにそれぞれの係数を調整する方法、(2) 少数のパラメータでパターンを特徴付けられる重み付け関数を定義して相関係数の絶対値が最大になるようパラメータを調整する方法、等が考えられるが、本研究では(2)を検討した。

具体的には、これまでの時系列データ解析結果を参考に、式(4)で表される正規分布関数に着目し、 σ と μ の2つのパラメータで重み付けを特徴付ける方法を考案した。従来の単純加算方法と比較すると、加算日数が大きいことは σ が大きいことに対応し、遅延日数が大きいことは μ が大きいことに対応する。ここで、式(4)を重み付け関数として使うときは、 x の値はゼロもしくは正の整数値とし、得られた w の値を過去の日付に降順に割り当てる。

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (\sigma, \mu > 0) \quad (4)$$

いくつかの σ と μ の値に対応する重み付け関数のパターンを図2に示す。ここでは、健康データを取得した日の前日から過去11日間の生活習慣データを重みを付けて加算することを想定している。 $x = 0, 1, 2, \dots, 10$ に対応する重み w の値を w_0, w_1, w_2, \dots ,

w_{10} とすると、重み付き生活習慣データの加算は、

$$w_0 e_{n-1} + w_1 e_{n-2} + w_2 e_{n-3} + \dots + w_{10} e_{n-11} \quad (5)$$

となる。ここで、健康データを取得した日付のカーソルが n 日にあると仮定している。従って、この重み付き加算と健康状態の変化 h_{nm} との時系列相関を各種 μ と σ の組み合わせについて評価することになる。この場合、従来の遅延期間という概念は重み付けパターンの中に吸収されてしまっている。

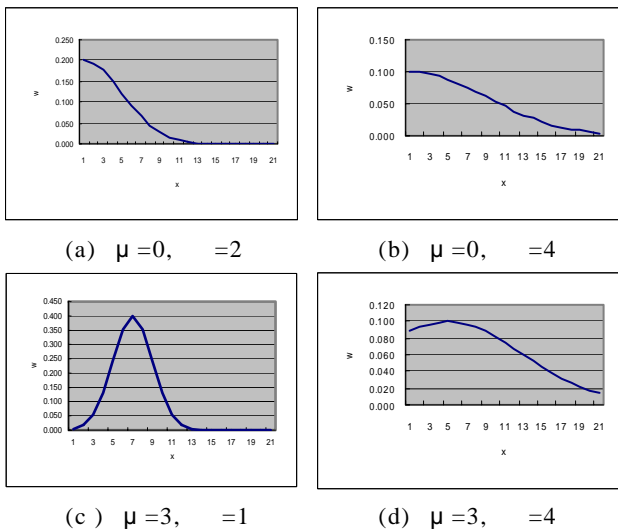


図2. 各種 μ と σ の値に対応する重み付け関数のパターン

時系列相関評価の結果、図2(a)のようなパターンの重み付けで相関係数の絶対値が最大になれば、比較的直近の生活習慣が遅延無く健康状態に影響を与えることになり、(b)や(d)のようなパターンの重み付けで相関係数の絶対値が最大になれば、比較的長期の生活習慣が尾を引いていることになる。また、(c)のようなパターンの重み付けで相関係数の絶対値が最大になれば、比較的短期間の生活習慣が遅れをもって健康状態に影響を与えていることになる。

3. 解析結果

本研究では、59歳男性が約半年間に亘りほぼ毎日取得した体脂肪率、総消費カロリー、総摂取カロリーのデータについて、重み付けを導入した時系列相関解析を行い従来の解析結果と比較検討した。ここで、体脂肪率は毎朝タニタの体組成計を用いて取得され、総消費カロリーはオムロンの歩数計で計測された日毎の歩数消費カロリーとその他運動による消費カロリーの和である。また総摂取カロリーとは、朝、昼、夜、それぞれの食事で摂取したエネルギーの日毎の総和である。

3.1. 体脂肪率と総消費カロリーの相関

解析に用いた重み付け関数は、(4)式において、 $\mu=0, 1, 2, 3, \dots =0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0$ の各組み合わせの32種類のパターンである。ここで、 μ の値を整数値としたのは、従来法における遅延日数とほぼ等価なパラメータであるがためである。一方、 σ は従来法における加算日数そのものでなく、その大小の尺度パラメータに相当するので小刻みに設定した μ の範囲は $x=0\sim 10$ で、体脂肪率を取得した日の前日から11日間の総消費カロリーデータの重み付け加算を行った。実際の各重み w_i は

$$\sum_{i=0}^{10} w_i = 1 \quad (6)$$

となるように規格化した値を用いた。

解析対象のデータは2005年3月1日～5月31日の3ヶ月間および2005年3月1日～8月31日の6ヶ月間の2種類とした。

解析の結果、どちらの対象期間においても、 $\mu=2, \sigma=1.0$ の重み付けパターンで相関係数の絶対値が最大となり、相関係数はそれぞれ、 $-0.482, -0.367$ であった。それぞれ対応する散布図を図3、4に示す。相関

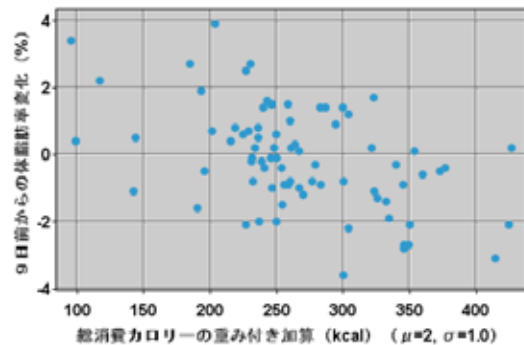


図3. 体脂肪率変化と総消費カロリー重み付き加算の散布図(対象データ期間3ヶ月)($n=79, r=-0.482$)

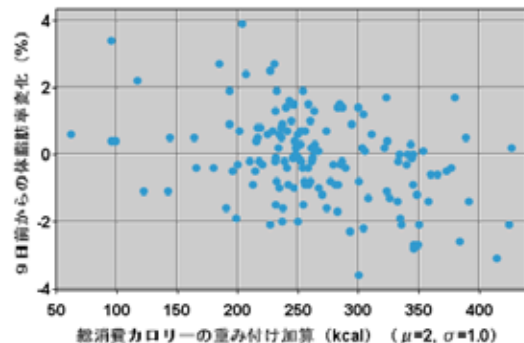


図4. 体脂肪率変化と総消費カロリー重み付き加算の散布図(対象データ期間6ヶ月)($n=148, r=-0.367$)

検定の結果、1%水準で負の相関有りという結論とな

った．比較のために従来の遅延期間を考慮した単純加算で解析した結果を図5, 6に示す．2日前から4日間の総消費カロリーに加算と最大の相関を示し, 相関係数はそれぞれの対象期間において, -0.452 , -0.323 であった．いずれの値も重み付け加算解析での最大値を下回っているが, ともに1%水準で相関があるという結論に影響を与えるような差異ではなかった．

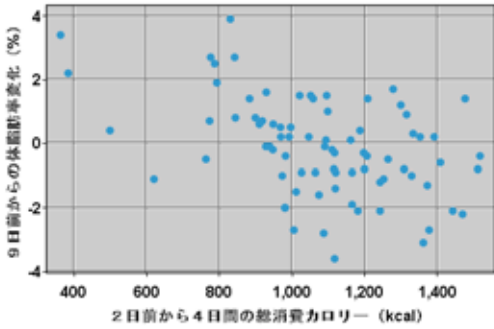


図5．体脂肪率変化と総消費カロリー4日間加算の散布図(対象データ期間3ヶ月)($n=79, r=-0.452$)

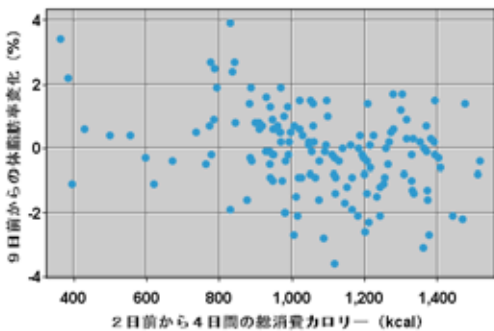


図6．体脂肪率変化と総消費カロリー4日間加算の散布図(対象データ期間6ヶ月)($n=148, r=-0.323$)

3.2. 体脂肪率と総摂取カロリーの相関

前節と同様に, 体脂肪率と総摂取カロリーとの時系列相関解析を行い, 重み付けした場合と従来法を比較した．図7, 8に重み付けした場合の解析結果を示す．対象期間が3ヶ月間の場合には, $\mu=3, \sigma=2.0$ の重み付けパターンで相関係数が最大となり, 0.400 であった．また対象期間が6ヶ月間の場合には, $\mu=3, \sigma=1.0$ の重み付けパターンで相関係数が最大となり, 0.342 であった．いずれも, 1%水準で有意な正の相関があった．従来法での解析結果を図9, 10に示す．どちらの対象期間においても, 3日前から4日間の総摂取カロリーの加算と最大の相関を示し, 相関係数はそれぞれ, 0.446 , 0.352 であった．今度は, どちらも単純な従来法での解析の方がかえって高い相関係数を示した．即ち, 重み付けをして加算することにより, より高い相関を示す条件を見出すには至らなかった．

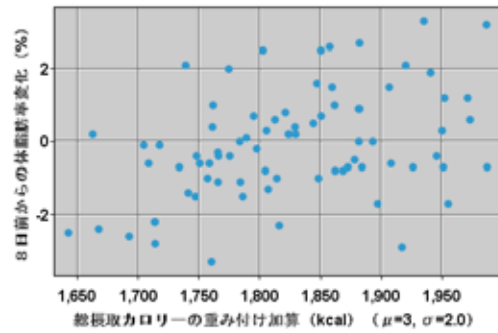


図7．体脂肪率変化と総摂取カロリー重み付け加算の散布図(対象データ期間3ヶ月)($n=79, r=0.400$)

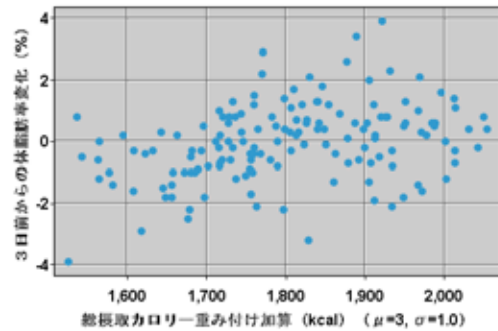


図8．体脂肪率変化と総摂取カロリー重み付け加算の散布図(対象データ期間6ヶ月)($n=148, r=0.342$)

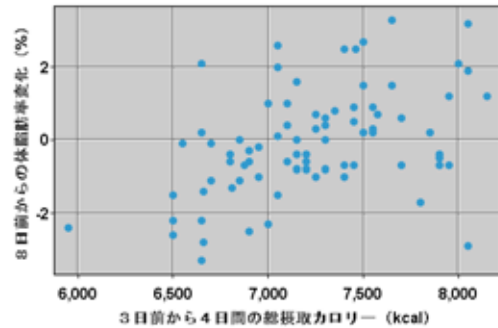


図9．体脂肪率変化と総摂取カロリー4日間加算の散布図(対象データ期間3ヶ月)($n=78, r=0.446$)

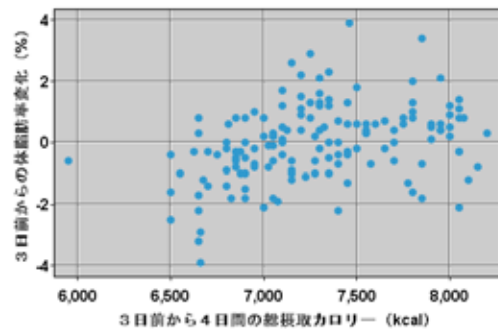


図10．体脂肪率変化と総摂取カロリー4日間加算の散布図(対象データ期間6ヶ月)($n=147, r=0.352$)

4. 結果の検討

4.1. 従来法との比較

体脂肪率と総消費カロリーとの関係において、絶対値が最大の相関係数を示した重み付けパターン($\mu = 2, \sigma = 1.0$)を図11に示す。図には、従来方法で最大の相関が得られた「2日前から4日間の加算」と等価な重み付けパターンも同時に示してある。従来法での解析結果と照らし合わせて、32種類の重み付けパターンの中から($\mu = 2, \sigma = 1.0$)の組み合わせが選ばれたことが妥当であることが判る。また、滑らかな形で重み付けされたことにより、相関係数の絶対値が大きくなっており、散布図もより自然に散らばっている(図3, 4と図5, 6を比較)。

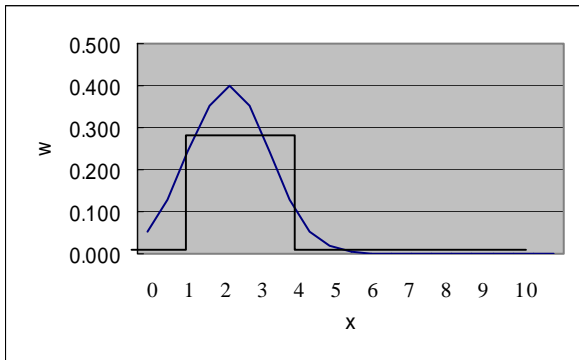


図11. 体脂肪率と総消費カロリーの関係において最大の相関が得られた重み付けパターン($\mu = 2, \sigma = 1.0$) (矩形のパターンは従来法で最大の相関が得られた条件を重み付けとして表したもの)

体脂肪率と総摂取カロリーとの関係において、最大の相関係数を示した重み付けパターン($\mu = 3, \sigma = 1.0, 2.0$)を図12に示す。図には、従来方法で最大の相関が得られた「3日前から4日間の加算」と等価な重み付けパターンも同時に示してある。従来方法による解析結果と照らし合わせて、やはり32種類の重み付けパターンから($\mu = 3, \sigma = 1.0, 2.0$)の組み合わせが選ばれたことが妥当であることが判る。しかし、滑らかな形で重み付けをしたにも関わらず、相関係数は従来法で得られた最大値より小さかった。ただし、散布図はより自然に散らばっている(図7, 8と図9, 10を比較)。

これらの結果から、生活習慣データの加算に重み付けの概念を導入することによって、より精度の高い時系列相関解析を狙ったが、その後のルールマイニング過程に影響を及ぼすほど大きな効果は得られなかった。その理由として、従来法における遅延期間という概念の導入がかなり本質的で、その他は単純加算でも大きな傾向は掴めること、注目している生活習慣デー

タ蓄積(加算)の精度よりもその他の生活習慣変動の影響が大きいこと、等が考えられる。

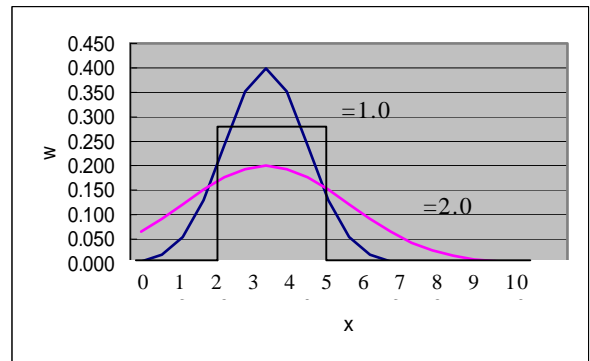


図12. 体脂肪率と総摂取カロリーの関係において最大の相関が得られた重み付けパターン($\mu = 3, \sigma = 1.0, 2.0$) (矩形のパターンは従来法で最大の相関が得られた条件を重み付けとして表したもの)

4.2. パラメータによる個人の特徴付け

本研究において、 μ と σ の2つのパラメータで特徴付けられる重み付けパターンを用いて時系列データ解析を行うことにより、従来と同等あるいはそれ以上の精度で相関関係を抽出できることが判った。

遅延期間と加算日数という従来のパラメータと比較して、 μ と σ の組み合わせはデータ項目毎に個人の傾向を特徴づけるパターンとして見えやすい。例えば、自身の健康状態への生活習慣の影響の仕方が、図2で例示したいくつかのパターンのなかでどれなのか知っていることは有意義と考えられる。また、大勢の人の傾向をクラスタリングする場合や、加齢に伴う変化を追跡するパラメータとしても使いやすいと考えられる。

5. まとめ

健康データマイニングの前段部の処理である生活習慣と健康状態の時系列相関解析の精度を高める目的で、生活習慣の蓄積を表す日毎の加算処理に重み付けの概念を導入した。重み付け関数のパターンとして、正規分布関数の活用を試みた。

その結果、健康データマイニングの後段部の処理であるルールマイニングの入力変数選択に影響を与えるほどの効果はみられなかった。しかし、従来方法の遅延期間と加算日数に代わり、正規分布関数の μ と σ の2つのパラメータで適当な重み付けパターンを表現できることが判った。これにより、生活習慣と健康状態の関係において、個人の特徴が「見える化」され、今後のクラスタリング解析や加齢による変化の解析に有用であることが示唆された。

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。また，日本データベース学会と日立製作所による日立 HiRDB アカデミック制度の適用を受けている。

参 考 文 献

- [1] H. Kumpusch, D. Hayn, K. Kreiner, M. Falgenhauer, J. Mor, and G. Schreier, "A Mobile Phone Based Telemonitoring Concept for the Simultaneous Acquisition of Biosignals and Physiological Parameters", Proc. 13th World Congress on Medical and Health Informatics (Medinfo2010), pp. 1344-1348, 2010.
- [2] H. Takeuchi, T. Hashiguchi, and T. Shintani, "Personal Dynamic Healthcare System Utilizing Mobile Phone and Web Technologies", Proc. 2nd Int'l Conf. Advances in Biomedical Signal and Information Processing, pp. 304-307, 2004.
- [3] 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, 2006.
- [4] H. Takeuchi, Y. Mayuzumi, N. Kodama, and K. Sato, "Application Service Provider System for Healthcare with Data Mining Function", Proc. 13th World Congress on Medical and Health Informatics (Medinfo2010), 2010.
- [5] 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.
- [6] H. Takeuchi, N. Kodama, T. Hashiguchi, and N. Mitsui, "Healthcare Data Mining Based on a Personal Dynamic Healthcare System", Proc. 2nd Int'l Conf. on Computational Intelligence in Medicine and Healthcare, pp. 37-43, 2005.
- [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, 1992.
- [8] H. Takeuchi, N. Kodama, T. Hashiguchi, and D. Hayashi, "Automated Healthcare Data Mining Based on a Personal Dynamic Healthcare System", Proc. 28th IEEE EMBS Annual Int'l Conf., pp.3604-3607, 2006.