

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

竹内 裕之[†] 児玉直樹[‡]

高崎健康福祉大学 健康福祉学部 医療福祉情報学科 〒370-0033 群馬県高崎市中大類町 37-1

E-mail: † htakeuchi@takasaki-u.ac.jp, ‡ kodama@takasaki-u.ac.jp

あらまし 2008年度から実施される、メタボリックシンドロームの予防を目的とした特定健診・保健指導においては、個人それぞれの生活習慣や日々の健康状態を把握してきめ細かなフォローアップを行うことが肝要である。本研究では、生活習慣と健康状態に関する時系列データを解析し、両者の間に潜む相関ルールを抽出することを目的としている。生活習慣データに関してはその蓄積に、健康データに関してはその変化に着目して、時系列データ間の相関を遅延パラメータの概念を導入してチェックする。遅延パラメータは、生活習慣の蓄積がある遅れをもって健康状態の変化に反映されることを想定しており、実際に両者の時系列データ間の相関に大きな影響を及ぼすことが判った。

キーワード メタボリックシンドローム, 時系列データ解析, 生活習慣データ, 健康データ, 相関ルール

Time-Series Data Analyses for Lifestyle and Daily Health Condition

Hiroshi TAKEUCHI[†] 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

E-mail: † htakeuchi@takasaki-u.ac.jp, ‡ kodama@takasaki-u.ac.jp

Abstract A health-check system for protection against metabolic-syndrome will start in 2008 in our country. In this new system, checking one's health based on medical data that specifies the relationship between personal lifestyle and health data is very important. The purpose of this study is to extract association rules between personal lifestyle and health conditions by time-series data analyses. These analyses were based on the simple idea that accumulation of the effects of a person's lifestyle could influence daily health with some delay. In these analyses, correlations of time-series data are checked focusing on the accumulation of effects of daily lifestyle, and on variation of daily health data. A retardation parameter, which is introduced assuming that lifestyle data affect health data with some delay, was found to change correlation coefficients dramatically.

Keyword Metabolic-syndrome, Time-series data analysis, Lifestyle data, Health data, Association rule

1. はじめに

メタボリックシンドロームに象徴される生活習慣病の予防は少子高齢化社会に突入した我が国にとって最重要課題の一つである。政府はその対策として、2008年度から医療保険者にその40歳以上のすべての加入者を対象にメタボリックシンドロームの予防のための特定健診および特定保健指導を義務付ける。この特定保健指導においては、加入者個人それぞれの生活習慣や日々の健康状態を把握して、きめ細かいフォローアップを実施することが肝要である。

我々はこれまで携帯電話と Web テクノロジーを活用した個人健康管理システムの開発を行ってきた[1]-[3]。このシステムは、携帯電話を端末として日常の生活習慣と健康に関するデータをインターネット経

由で時系列的にサーバコンピュータに蓄積する仕組みである。蓄積されたデータはその統計を判りやすいグラフ表示で見ることができ、個人の生活習慣と健康状態の間に何らかの規則性が見出せれば相関ルールとしてユーザの携帯電話に通知する[4]-[6]。そして、これらの情報を参考にユーザが自分で自分の健康管理を行うことを期待している。

本研究では、個人健康管理システムにおいて開発した、相関ルールを抽出するための前段処理として行っている時系列データ解析手法を実データに適用することによりその有効性を実証した。さらに、この時系列データ解析手法が特定保健指導を支援する強力なツールになりうることを示唆した。

2. 解析方法

2.1. 時系列データ解析の概念

個人の日常生活においては様々な生活習慣データが変動しており、これらがその人の健康データに複雑に関連していると考えられる。従って、ある1つの生活習慣が関心のある健康データに影響しているかどうかを検証するためには、何らかの統計的手法に頼らざるを得ない。本研究の手法は、ある期間に亘る個人の生活習慣と健康状態の時系列データ間の相関をみることによって、両者の関係を抽出することに特徴がある。図1に本手法と従来の疫学的手法の比較を示した。

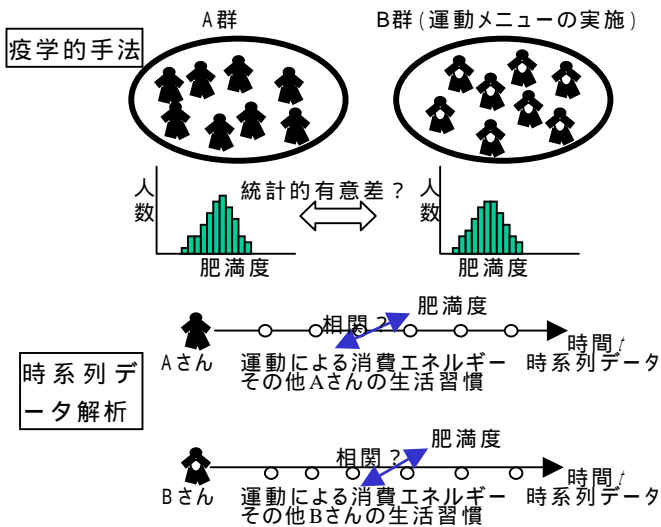


図1 時系列データ解析の概念

Fig.1 Concept of Time-Series Data Analysis

疫学的手法では、ある規模の集団を無作為に2つの群に分け、ある生活習慣について介入(図1の例では運動メニューの実施)を行った群と行わない群で関心ある健康データ(図1の例では肥満度)に統計的な有意差がどうかで個人差に埋もれた関係をあぶりだす。しかしこの手法では、統計的有意差を判定するのに多数の n (人数)が必要であり、また集団としての差異をみることはできても、必ずしも個人にとって有用な情報が得られるわけではない。本研究では、あくまでも個人の時系列データを統計処理の対象にし、他の生活習慣データも時間的に変動するなかで、ある生活習慣データと関心ある健康データとの相関を抽出する。具体的には、ある生活習慣データの蓄積と関心ある健康データの変化に着目し、散布図を確認しながらピアソンの積率相関係数を算出することで時系列データ間の相関の程度を押し量る。そして、ある程度相関が認められる生活習慣データ項目を後段の相関ルールマイニングの入力変数として採用する[6]。ただし、採用された入力変数がすべて相関ルールの生成に寄与

するとは限らない。これは、相関ルールは因果関係であり、相関関係とは異なることに起因する[7]。

2.2. 時系列データの処理

時系列的に蓄積されたある生活習慣データと関心ある健康データを図2に示す手法でチェックする。

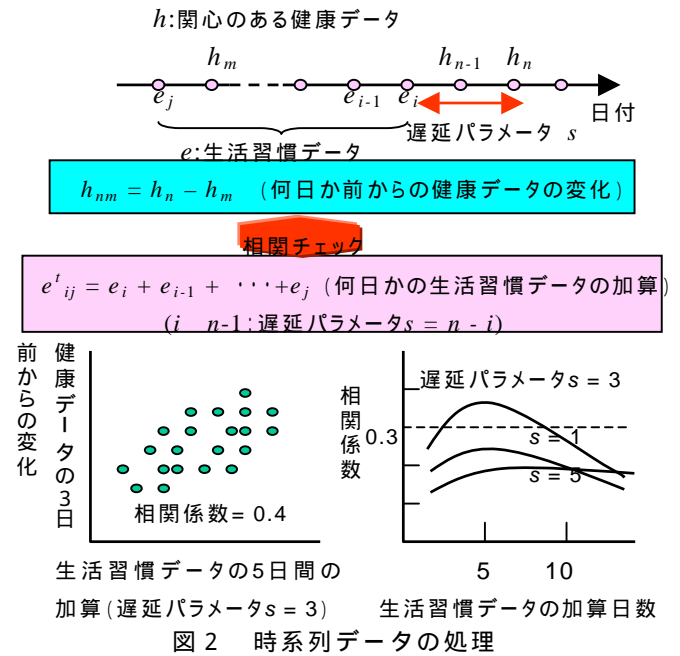


Fig.2 Actual Process of Time-Series Data Analysis

データはほぼ毎日取得されていることを前提とし、何日前からの健康データの変化 h_{nm} と何日かの生活習慣データの加算 $e^{t_{ij}}$ の相関の有無を、時系列データ間のピアソンの積率相関係数を評価してチェックする。ここで、何日かの生活習慣データの加算が遅れをもって健康データに反映されることがあることを想定し、遅延パラメータ $s=n-i$ を定義する。相関係数は、 $n-m$ 、 $i-j$ 、および s をパラメータとして変化させ、蓄積された時系列データをもとに(1)式で計算され、最大値が求められる。

$$r_{he} = S_{he} / (S_h \times S_e) \quad (1)$$

ここで、 S_h および S_e はそれぞれ h_{nm} の標準偏差、 $e^{t_{ij}}$ の標準偏差である。また、 S_{he} は h_{nm} と $e^{t_{ij}}$ の共分散である。

図2の例では、 $n-m=3$ (3日前からの変化)、 $i-j=4$ (5日間の加算)、 $s=3$ (3日遅れ)で相関係数が最大となり、ある閾値(ここでは0.3)を超える。すなわち、生活習慣データの5日間の加算が3日間の遅れをもって、3日前からの健康データの変化に影響を与えることが最も確からしいということになる。この場合には、「3日前から5日間の生活習慣データの加算」が相関

ルールマイニングの入力変数（フィールド）として採用される．なお、相関ルールマイニングは N 測度を評価する手法を用いる[8]．

3. 解析の実例

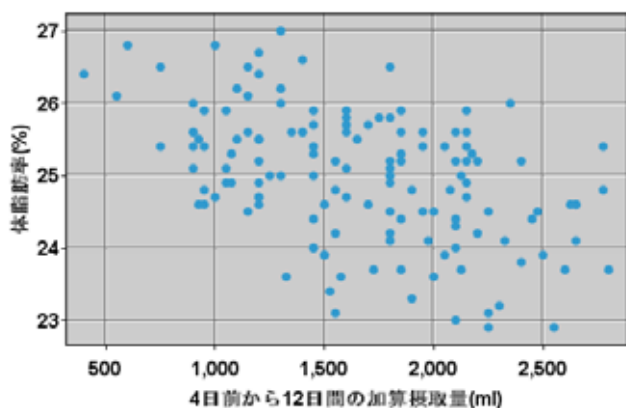
3.1. 大豆たんぱく質の摂取と体脂肪率

大豆たんぱく質は血中コレステロールを低下させ、特にその中に含まれる　　コングリシニンが中性脂肪を低下させ、さらには体脂肪率を下げる다고されている．あるシステムユーザ（22歳、女性）は約6ヶ月間に亘り豆乳により大豆たんぱく質を摂取し、その効果を検証した．体脂肪率の時系列データと豆乳加算摂取量の時系列データ間の相関係数は加算日数と遅延パラメータ

タの値によって大きく変化し、12日間の加算、4日間の遅延で相関係数の絶対値が最大(-0.470)となった[9]．このときの散布図を図3(a)に示す．豆乳の（継続的）摂取は遅延をもって明らかに体脂肪率を下げる事が判るが、他の条件では、例えば3日間の加算、遅延パラメータ1では図3(b)に散布図を示すように全く相関を示さない．時系列データ解析を適切に行わないと、豆乳の摂取は体脂肪率には影響を与えないという誤った結論になってしまうことが判る．

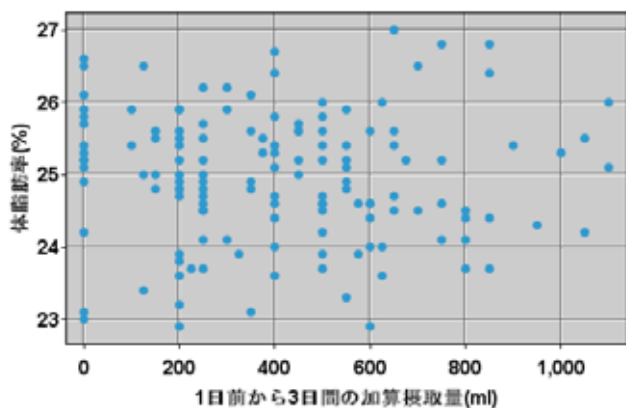
3.2. 運動による消費エネルギーと体脂肪率

別のシステムユーザ（59歳、男性）は運動による消



(a) 遅延パラメータ：4、加算日数：12(データ数：139 相関係数：-0.470) [9]

(a) Retardation parameter: 4, Summation-day number: 12 (data number: 139, correlation coefficient: -0.470)

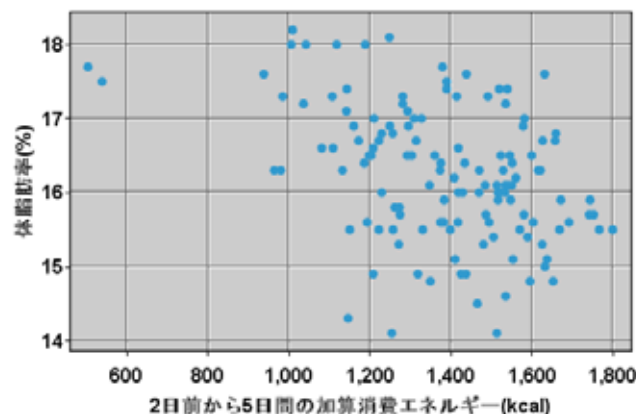


(b) 遅延パラメータ：1、加算日数：3(データ数：139 相関係数：-0.032)

(b) Retardation parameter: 1, Summation-day number: 3 (data number: 139, correlation coefficient: -0.032)

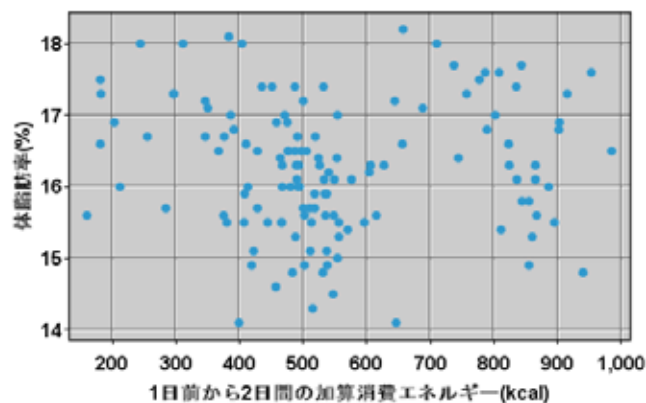
図3 豆乳加算摂取量と体脂肪率の散布図

Fig.3 Scatterplot of Body-fat Percentage versus Soybean Milk Intgestion



(a) 遅延パラメータ：2、加算日数：5(データ数：131 相関係数：-0.406)

(a) Retardation parameter: 2, Summation-day number: 5 (data number: 131, correlation coefficient: -0.406)



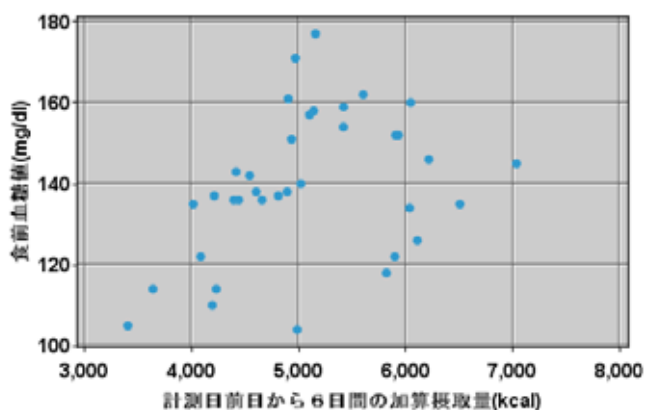
(b) 遅延パラメータ：1、加算日数：2(データ数：131 相関係数：-0.008)

(b) Retardation parameter: 1, Summation-day number: 2 (data number: 131, correlation coefficient: -0.008)

図4 運動による加算消費エネルギーと体脂肪率散布図

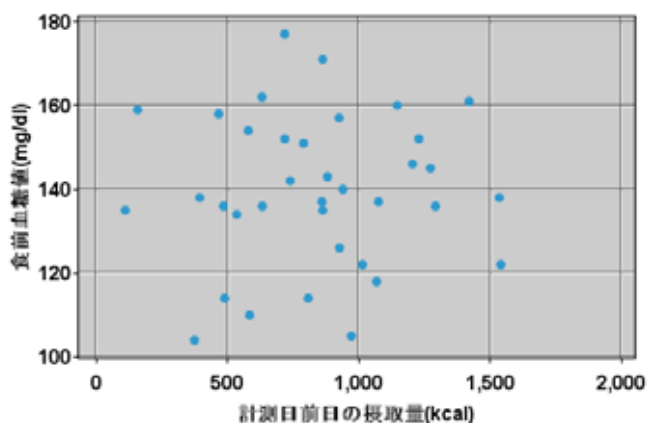
Fig.4 Scatterplot of Body-fat Percentage versus Energy Expenditure

費エネルギーと体脂肪率の時系列データを約6ヶ月間蓄積した。体脂肪率の時系列データと運動による加算消費エネルギーの時系列データ間の相関係数も加算日数と遅延パラメータの値によって大きく変化し、5日間の加算、遅延パラメータ2で相関係数の絶対値が最大(-0.406)となった。このときの散布図を図4(a)に示す。運動の(継続的)実施が遅延をもって明らかに体脂肪率を下げるのが判る。これは一般的に予想される当然の結果のようであるが、例えば、2日間の加算、遅延パラメータ1では図4(b)に示すように全く相関を示さない。やはり、時系列データ解析を適切に行わないと結論を誤ることになる。



(a) 遅延パラメータ：1、加算日数：6 (データ数：36 相関係数：0.367)

(a) Retardation parameter: 1, Summation-day number: 6 (data number: 36, correlation coefficient: 0.367)



(b) 遅延パラメータ：1、加算日数：1 (データ数：36 相関係数：0.032)

(b) Retardation parameter: 1, Summation-day number: 1 (data number: 36, correlation coefficient: 0.032)

図5 穀物の加算摂取量と食前血糖値の散布図

Fig.5 Scatterplot of Blood-sugar Level versus Grain Intgestion

3.3. 穀物の摂取と食前(空腹時)血糖値

定期的な外来診療を受けているある軽度の糖尿病患者さん(42歳、男性)の食生活を中心にした生活習慣と食前血糖値の約3年間に亘る時系列データの解析を行った。本例では、食前血糖値のデータは月に1回であり、食事内容はほぼ毎日記録されていた。そこで、食前血糖値の時系列データと食前血糖値計測日前の何日間かの穀物加算摂取量の時系列データとの相関をチェックした。

時系列データ間の相関係数はやはり加算日数と遅延パラメータにより変化し、6日間の加算、遅延パラメータ1で相関係数が最大(0.367)になった。このときの散布図を図5(a)に示す。穀物の(継続的)摂取が食前血糖値を上昇させることが判るが、この場合も他の条件、例えば計測日前日の摂取量との間では図5(b)に示すように相関を示さない。

4. 考察

4.1. 遅延パラメータと加算日数

4.1.1. 大豆たんぱく質の摂取

大豆たんぱく質に含まれる **コングリシニン**は「食べた油脂を完全には消化せず、一部を未消化で対外に排泄し」、かつ「肝臓内の中性脂肪を血中へと運び出す機能を低下させる」作用をもつとされている。さらに最近の研究では、**コングリシニン**は脂質の代謝に関する遺伝子に作用し、中性脂肪低減に効果を発揮することが示されている。このようなことから、

コングリシニンは体脂肪率の低下にも効果があるとされている。体脂肪率と豆乳摂取量の相関に関する図3の結果は、2週間近くに亘る継続的な摂取が4日程度の時間遅れをもってゆっくりと体脂肪率に影響することを示唆している。3日間程度の短期間の摂取で効果がないと判断してしまう可能性がある例である。上述の中性脂肪低減のメカニズムから、おそらく最大の相関を示す加算日数や遅延パラメータの値は、個人の遺伝的な体質によって異なると思われる。

4.1.2. 運動による消費エネルギー

消費エネルギーと摂取エネルギーのバランスが体脂肪率に反映されることは言うまでもない。しかし、体脂肪率との関係において時系列データを解析してみると興味深い結果が得られる。遅延パラメータ1および3では加算日数を変化させても相関係数の絶対値が0.3を超えることはないが、遅延パラメータを2にすると、加算日数5で相関係数の絶対値は一気に0.4を超える(図4(a))。このシステムユーザの場合には5日間程度の継続的運動が2日の遅延で体脂肪率に影響を強く与えるということが特徴になっている。またこの例においても、解析の条件によっては相関を全く示

さず(図4(b))、運動の効果について誤った判断をする可能性がある。

4.1.3. 穀物摂取と食前血糖値

糖尿病の指標の1つに食前(空腹時)血糖値がある。これは食後3時間以上経過して血糖値の変化が落ち着いた時点での値である。食前血糖値には、食生活や運動習慣が大きな影響を与えるとされている。食生活ではやはり糖(グルコース)の元となる穀物摂取の影響が大きいと予想される。実際に図5(a)に示すように、6日間程度の継続的穀物摂取が食前血糖値を押し上げるという解析結果が得られたが、食前血糖値計測日前日の穀物摂取量との間には相関がみられなかった。この結果はデータを提供した患者さんの食生活改善のヒントになると考えられる。

4.2. 時系列データ解析の一般化

生活習慣と健康状態に関する時系列データ解析の実例を3つ示したが、生活習慣データの加算(積分)と健康データの変化(微分)に着目し、遅延パラメータの概念を導入して解析する本手法は、代謝を含めた生体内の複雑な反応を最も単純にモデル化したものと言える。すなわち、「生活習慣の積み重ねが健康状態に変化をもたらし、その影響は時間遅れをもって現れることがある」という単純な発想に基づく解析手法である。解析の実例から明らかのように、遅延パラメータと加算日数を選ぶことにより、時系列データ間に潜む相関をあぶりだすことができた。

本研究では、1日きざみの時系列データを元に解析を行った(食前血糖値のデータは1月に1回で例外であるが)。すなわち、サンプリング周期1日でデータ解析を行ったが、この解析手法は必ずしも1日きざみである必要はなく、任意の時間間隔を設定できる。健康に関するデータでは、血圧や血糖値など日内変動が重要になる場合もあれば、内臓脂肪やコレステロール値など1週間毎、1ヶ月毎のゆっくりとした変化を見た

[差分値]

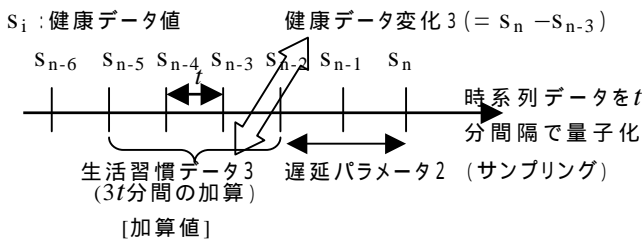


図6 一般化された時系列データ解析手法
(変化3、加算3、遅延2の場合)

Fig.6 Generalized Method for Time-Series Data Analysis
(In case of variation interval: 3, summation number: 3, and retardation parameter: 2)

い場合もある。図6に本研究の時系列データ解析の一般化された手法を示す。

現在、血糖値の変化と生活習慣(食事、運動などのイベント)の時系列相関をこの一般化された手法で解析中である。

4.3. 特定保健指導支援システムへの展開

特定保健指導においては、対象者への情報提供として

- (1) 生活習慣病の特性や生活習慣の改善に関する基本的な理解を支援する。
- (2) 健診結果の提供に合わせて、全員に個別のニーズ、生活習慣に即した情報提供を実施する。

とあり、対象者ごとの個別の計画を作成することが求められている。さらに、動機づけ支援として、生活習慣の改善に対する個別の目標を設定し、自助努力による行動変容が可能となるようにするとある。

本研究における時系列データ解析のアウトプットは、個人毎に生活習慣と健康状態の相関に関するエビデンスを提供するものであり、まさに対象者毎の生活習慣改善に対する個別の目標設定につながるものと期待できる。今後、特定保健指導を支援するツールとしての可能性を追究する。

5. まとめ

生活習慣病の予防を目的とした個人健康管理システムにおいて開発した生活習慣と健康状態に関する時系列データ解析手法の実例を3例示し、生活習慣の蓄積がある遅れをもって健康状態に反映することを想定して導入した遅延パラメータが非常に重要な役割を果たすことを示した。すなわち、遅延パラメータと生活習慣データの加算日数を選ぶことにより、生活習慣と健康状態の時系列データ間に潜む相関をあぶりだすことができた。この遅延パラメータは、代謝など生体内の複雑な反応が反映されているものと推察される。

また、本研究の時系列データ解析手法が任意の時間間隔サンプリングに一般化できること、2008年度からスタートするメタボリックシンドロームの予防を目的とした特定健診・特定保健指導を支援するツールとして展開できる可能性を示した。

謝辞

長期にわたる食前血糖値、生活習慣の時系列データを提供していただいた、竹内順氏に感謝いたします。また、糖尿病の運動・食事療法の効果に関してご指導いただいている本学健康福祉学部健康栄養学科の鶴見克則教授に感謝いたします。さらに、常に本研究を励まして頂いている本学学長須藤賢一博士に感謝いたし

ます .

文 献

- [1] 竹内裕之, 橋口猛志, 新谷隆彦, “日常の健康管理を目的とした個人対応動的データベース,” 医療情報学, vol.23, no.6, pp.497-502, Feb.2004.
- [2] H. Takeuchi, T. Hashiguchi, and T. Shintani, “Personal dynamic healthcare system utilizing mobile phone and Web technology,” Proc. 2nd Int. Conf. on Advances in Medical Signal and Information Processing, pp.304-307, Malta, Aug. 2004.
- [3] H. Takeuchi, N. Kodama, T. Hashiguchi, and N. Mitsui, “Healthcare data mining based on a personal dynamic healthcare system,” Proc. 2nd Int. Conf. on Computational Intelligence in Medicine and Healthcare, pp.37-43, Lisbon, Portugal, Jun.2005.
- [4] 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.
- [5] 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.
- [6] 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. Conf., pp. 3604-3607, New York, Sep.2006.
- [7] M. J. A. Berry and G. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Support, John Wiley & Sons, Inc., 1997.
- [8] 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.
- [9] H. Takeuchi, Y. Ikeda, and N. Kodama, “Time-series data analyses for healthcare-data-mining based on a personal dynamic healthcare system,” Proc. 12th World Congress on Medical Informatics, P309, Brisbane, Australia, Aug.2007.