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

黛 勇気　児玉 直樹†　竹内 裕之†　矢部 貴之‡

高崎健康福祉大学大学院　健康福祉学研究科　医療福祉情報学専攻　〒370-0033 群馬県高崎市中大類町 37-1
†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
‡ファーマシー中山株式会社 〒328-0061 栃木県栃木市新井町 1010-3
E-mail: 0710103@wm.takasaki-u.ac.jp, †{kodama, htakeuchi}@ takasaki-u.ac.jp, ‡yabe@p-nakayama.co.jp

あらまし　生活習慣と健康状態に関する時系列データを解析して両者の間の相関ルールを抽出する健康データマイニング技術をコアにし, 特定保健指導を支援するシステムの開発を行った. 特定保健指導対象者（動機付け支援, 積極的支援）の３ヶ月間を目処とした生活習慣と健康状態に関するデータを，所定のフォーマットでＥｘｃｅｌファイルとして時系列的に蓄積すると，ＰＣによるバッチ処理で個人毎に生活習慣と健康状態に関する相関ルールが出力される．保健師や管理栄養士はこの個人毎の相関ルールを根拠として保健指導を行うことができる．

キーワード　健康データマイニング，時系列データ解析，生活習慣，健康状態，相関ルール, 特定保健指導

System Supporting Health-Check for Protection Against Metabolic-Syndrome Using Healthcare-Data-Mining

Yuki MAYUZUMI　Naoki KODAMA†　Hiroshi TAKEUCHI†　and　Takayuki YABE‡

Post-Graduate Course of Healthcare Informatics, Graduate School of Takasaki University of Health and Welfare, 37-1 Nakaorui-machi, Takasaki-shi, Gunma, 370-0033 Japan

†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
‡Pharmacy Nakayama Co.,Ltd.　1010-3 Arai-cho, Tochigi-shi, Tochigi, 328-0061 Japan
E-mail: 0710103@wm.takasaki-u.ac.jp, †{kodama, htakeuchi}@ takasaki-u.ac.jp, ‡yabe@p-nakayama.co.jp

Abstract　A health-check system for protection against metabolic-syndrome started 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. We studied a system supporting the health-check using healthcare-data-mining technology we have developed. The system extracts association rules between lifestyle and health condition based on stored time-series lifestyle and health data for about three months. These personal data are subject to be stored in a time-series way with a prescribed format as the Excel file. Then association rules for each individual concerning lifestyle and health condition can be output through batch processing by PC. The health master and the national registered dietitian can guide each individual based on the extracted association rules.

Keyword Healthcare-data-mining, Time-series data analysis, Lifestyle data, Health data, Health-check, Metabolic-syndrome

1. はじめに

団塊世代の大量退職時代を迎え,少子高齢化が著しい現在おいて,国民医療費の増大を抑制するためには,国民一人一人が自分の健康を自分で守るという意識をもたなければならない.特に生活習慣に起因する疾病は，個人が日常の生活習慣と健康状態の管理をきめ細かく行うことで予防が可能である.平成２０年度よりスタートした「特定健康診査・特定保健指導」についても，厚生労働省健康局の標準的な健診・保健指導プログラムには,保健指導の第一の目的を,生活習慣病予備軍を生活習慣病に移行させないこととしたうえで,自身が健診結果を理解して身体の変化に気づき,生活習慣改善のための行動目標を設定,実践することを支援し,自分の健康に関するセルフケア（自己管理）ができるようになることとある[1].

特定健診では,まず基本的な健診の項目に則って診

査を行い,その結果により「情報提供」「動機づけ支援」「積極的支援」の３つの階層に分類（階層化）する[2].情報提供は健診受診者全員を対象に年１回,健診結果や健診時の質問票から,対象者個人に合わせた情報や健診結果の見方,健康の保持増進に役立つ情報を,健診結果と同時に提供する.動機づけ支援は,健診結果・質問票から生活習慣の改善が必要と判断され,生活習慣を変えるに当たって意思決定の支援が必要な者を対象に原則１回,生活習慣改善を動機づけるための支援を個別またはグループに対して行い,６か月後に評価する.積極的支援は,動機付け支援に加えて,定期的・継続的な支援を,専門職による継続的できめ細やかな支援が必要な者を対象に,３か月以上継続的に行う.

これらの保健指導を行う担い手は保健師や管理栄養士などであるが,対象者個人毎にきめ細かい指導を行うには全国的に人手不足であり,情報技術を用いた何らかの支援策が必要とされている.本研究は,この一助となることを目的とし,動機付け支援もしくは積極的支援対象者の３か月を目処とした生活習慣と健康状態に関する時系列データを基に,両者の相関ルールを抽出して生活習慣改善に関わる保健指導情報を提供することを狙いとしている.

2. 研究方法

2.1. 時系列データの処理

我々は,「生活習慣の蓄積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」という極めてシンプルなモデルで解析を行っている[3].図１に,日常の運動が体脂肪率に与える影響をみようとする場合を例に,このモデルを模式的散布図で判りやすく示した.この例では,Walking による消費エネルギーと体脂肪率が毎日取得されていると想定する.(a)は１日の消費エネルギーと翌日（遅延１日）の体脂肪率の関係である.全く相関が見られない.（b）は３日間の消費エネルギーの加算とその翌日（遅延１日）の体脂肪率の関係である.負の相関がありそうである.同様に（c）は５日間の加算とその３日後（遅延３日）の体脂肪率の関係で,明らかな負の相関がみられる.このようにして,相関係数の絶対値が最大となる生活習慣データの加算日数と遅延日数の条件を見つける.そして,この最大の相関係数がある閾値を超えている場合に,注目している健康状態にその生活習慣が有意な影響を与える因子の候補とする.図１の例では,(c)の条件で相関係数の絶対値が最大となり,その値が設定した閾値を超えていれば,「３日前から５日間の消費エネルギーの合計」がその日の体脂肪率に影響を与える因子の候補として採用され,相関ルールを抽出する健康データマイニングの入力変数のひとつとなる.

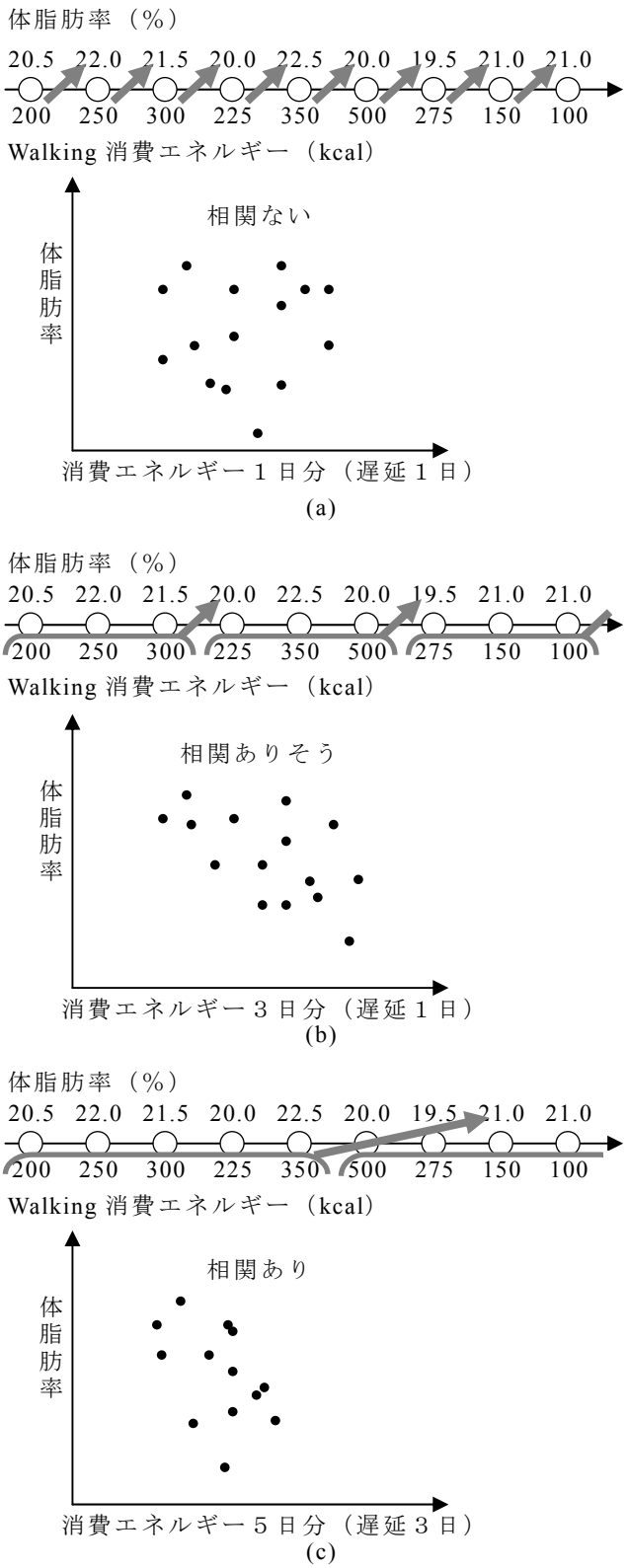


図 1 時系列データ処理の模式的説明図（例：体脂肪率と消費エネルギーの関係）

Fig.1 Concept of Time-Series Data Analysis (Example: Correlation Between Body-fat Percentage and Energy Expenditure)

2.2. 健康データマイニング

前項で述べた手法で生活習慣データ項目からなる

複数の入力変数 Y を決定し,関心ある健康データ項目の時系列データを「高い」「中間」「低い」の３つのシンボル値を持つ変数に変換したものを出力変数 X としてルールマイニングを行う[4,5].ここでは,情報理論に基づく相関ルール解析手法である ITRULE アルゴリズムを用いている[6].ITRULE アルゴリズムは,

$$\text{If } Y=y, \text{ then } X=x \text{ with probability } p \qquad (1)$$

という簡単なルールを生成する.ITRULE アルゴリズムでは,蓄積された多くのデータセットから有効なルールを生成するメカニズムとして J 測度を用いる[6].

J 測度は $Y=\mathrm{y}$ 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 $Y=y$ という前提がある場合とない場合で 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,$Y=\mathrm{y}$ 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ルールということになる.

2.3. PC 版スタンドアローンシステム

前項で説明した健康データマイニングのアルゴリズムを用いて,携帯電話から蓄積した個人の生活習慣と健康に関するデータの相関ルールを抽出する個人健康管理システムを現在試行中である[7].これを基に,健康データマイニングの特定健診・特定保健指導支援への応用の第一段階として,このシステムをスタンドアローンのＰＣ上の動作環境にカスタマイズした[8].システムへのログインや全体のデータフローなど,基本的なアーキテクチャに関する部分は先行システムより引き継ぎ,PC 環境に合わせた画面遷移の修正や,ルール生成でオンライン配信されるメイル機能の停止など,ネットワーク上の PC-UNIX からローカルの Windows PC へプラットフォームを移行するうえで必要と思われる部分を集中的に修正した.また,このシステムを用いて共同研究先で多数のユーザ（モニタ）を対象に解析を試みたところ,特にメタボリック症候群の予防を目的とした生活習慣の改善に関するルールを抽出することができた[8].特定保健指導においては,個人の生活習慣や健康状態の変化を基に,各ユーザにユニークな介入が求められることから,健康データマイニングで生成された相関ルールはこれを支援するコンテンツとしての有用であることが示唆された.しかし対象を個人から医療従事者に移す場合,システムはそれに合わせて変更する必要がある.そこで, 本研究では共同研究先での先行システムの試行体験をフィードバックし,健康データマイニングを応用した特定保健指導支援システムを開発した.

3. 結果

3.1. PC 版スタンドアローンシステムの評価

ＰＣ版スタンドアローンシステムを用いて共同研究先のデータを解析し,システムの使用感について評価した.この実験では,ＰＣ版のシステムを用いて同一の設定を行ったＰＣを２台用意し,男女合わせて３８名の生活習慣と健康状態に関する９０日分のデータを入力した.この作業において,データ入力１人分につき２～３時間が必要とされ,入力インターフェースが多数のデータの一括処理には非効率的であることが明らかとなった.また,このシステムではデータの履歴やルールを必ずブラウザ上で確認する必要があり,システムを特定保健指導へ応用する際の改善点として,帳票出力機能の追加などの意見が出された.

3.2. 対象ユーザのセグメンテーション

特定保健指導支援システムは,対象とする市場の原点を個人（ユーザ）から医療従事者（プレーヤー）に変更するため,システムの使用パターンについての見直しが必要となった.システムのターゲット自身が自分の健康管理をするためにシステムを利用するのではなく,医療従事者が生成されたルールを基にプレーヤーとしてユーザに保健指導を行う(図２)ため,システムを利用するプレーヤーはシステム内においてユーザより上位の権限を持ち,各ユーザのデータ入出力やルールの閲覧などを,ユーザ横断的に行えることが求められた.

また,併せて蓄積されたデータのグラフ化など,実際の指導を行う際にはルール以外の情報の可視化も求められた.

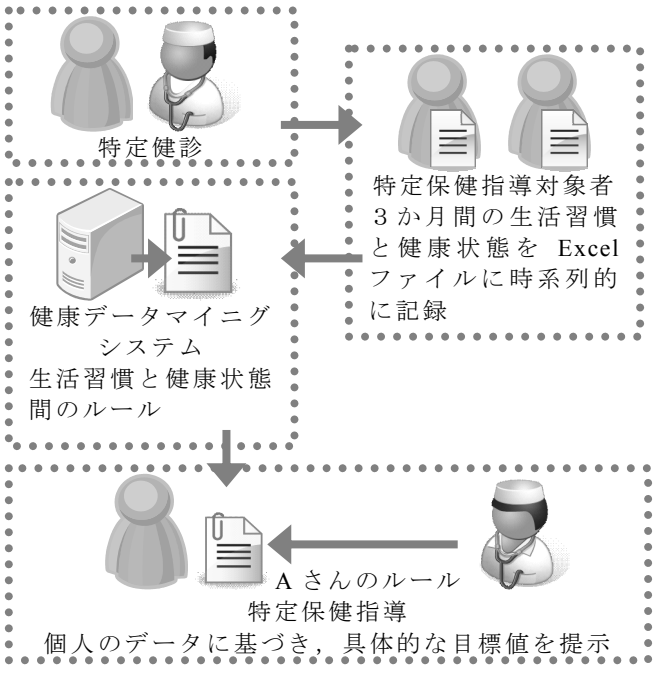


図２　特定保健指導を支援するシステムの概念図
Fig.2 Concept of Health-Check Supporting System

3.3. システムの変更点

特定保健指導支援システムでは,データの登録,閲覧,

修正の部分をリエンジニアリングし,Microsoft Excelで健康データ,生活習慣データなどを登録,閲覧,修正が可能なシステムを製作した.このシステムでは,日々のレコード時系列で入力できるシートを作成し,Excel VBAにより対象ユーザを選択して一括でデータを管理するようにした.

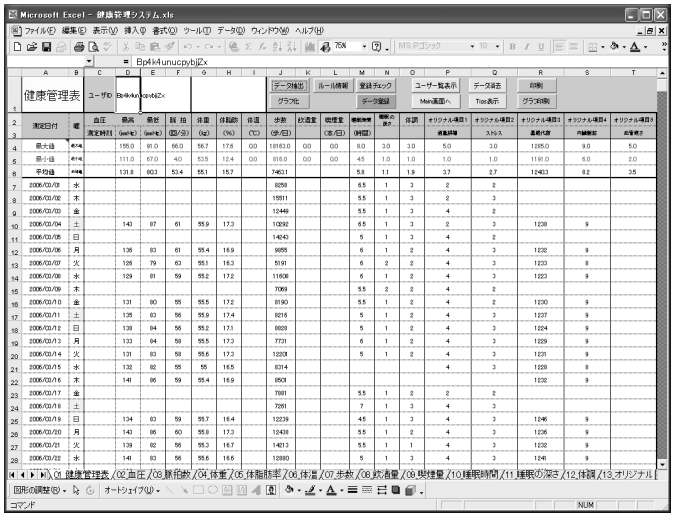
まず,先行システムはDBへのデータ登録を計測時点でブラウザより行う仕様であったのに対し,本システムはデータが入力された Excel Book を用いて,数日分のデータを一括で登録する方法に変更した(図３(ａ)).Excel シートを使うことで,画面上のスプレッドシートでデータを一気に閲覧でき,データの視認性,操作性ともに向上した.また,これらのデータを基にExcelシート上にグラフを生成し,Excel の印刷機能での帳票の出力を可能にした(図３(ｂ)).ルールについても,同手法により閲覧と帳票出力に対応した(図３(ｃ)).これらの変更により,データ管理はプレーヤーがユーザ毎に作成したファイルをユーザに渡し,ユーザが一定期間の記録を入力,データが入ったファイルをユーザがプレーヤーに送信するフローとなる.このため,プレーヤー側で必要だった入力作業をユーザに移行することが可能となり,プレーヤーのタスクを軽減することが可能となった.また,用紙で受け取ったデータをプレーヤーが入力する場合も,共同研究先での実験では１人９０日分を３０分程度での入力が可能となり,作業効率の飛躍的な向上が確認できた.

4. 考察

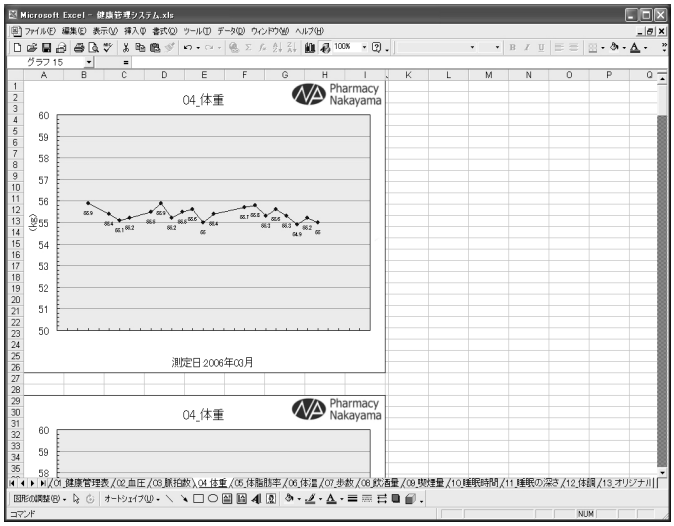
4.1. 保健指導に役立つシステムとは

生活習慣や健康状態の記録を履歴管理することだけが目的であれば,データの蓄積やグラフ化は市販の表計算ソフトで容易に行うことが可能である.しかし,蓄積したデータから生活習慣と健康状態の相関を知識として見つけ出すことは,単にグラフなどで可視化されたデータを閲覧するだけでは困難であることからも,特定保健指導において健康データマイニングは有用であると考えられる.

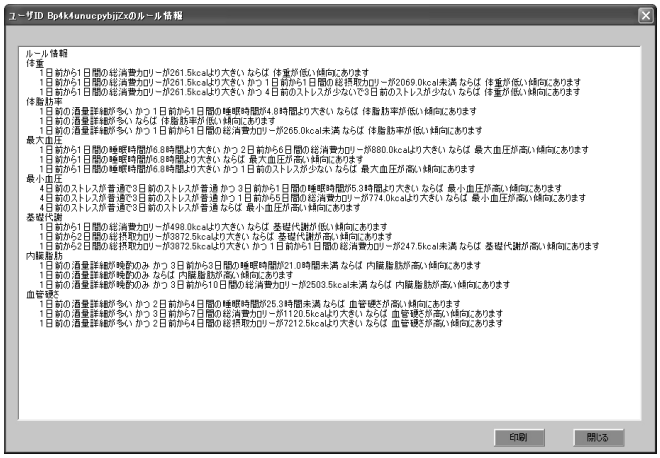
特定保健指導としての介入が必要な場合,対象者が自助努力を継続的に行うことを,保健師や管理栄養士が支援しなければならない.しかし,実際に従来の疫学研究のように集団を対象とした画一化された総論ではなく,多くの対象に対して個人の生活と健康の記録からユニークな傾向を割り出して指導に取り入れるためには,データ解析に関する深い知識と理解をもってしても,膨大な人的リソースが必要とされる.



(a) 健康管理表（データ入出力）



(b) グラフ表示（体重の例）



(c) ルール表示画面（健康データ項目毎に複数表示）

図３　開発した特定保健指導支援システムの PC 画面例

Fig.3 Examples of PC View of Developed Health-Check Supporting System

特定保健指導を支援するシステムに求められる性能とは,単にユーザの健康や生活に関するデータを蓄積するだけではなく,蓄積された膨大なデータの中からユーザに固有で有意義な知識を抽出できることであると考えられる.

4.2. 今後の課題

生成された相関ルールを,保健師や管理栄養士が保健指導に利用するにあたり,ルールの文章表現が難解であるとの指摘があった.また現在の時系列データ解析アルゴリズムでは加算期間中にデータ欠損があると,データエントリから除外されるので,周期的な欠損があると極端に入力変数のエントリ数が減り,算出される相関係数の信頼性がなくなるという問題が起きている.これによって理解不能な相関ルールが抽出されてしまい,ユーザが期待する結果が得られないという現象も確認されている.これらについて今後改善を行っていく.

5. まとめ

健康データマイニングは,特に特定保健指導における「自分の健康に関するセルフケア（自己管理)」という目標を実現する上において有効なアルゴリズムである.本研究により,このアルゴリズムを実装したシステムで,複数のエンドユーザのデータを容易に扱えるようになったため,ユーザ毎のグラフやルールを保健指導の具体的なツールとして利用することが可能となった.今後もさらに反復デザインを繰り返し，さらに改良された健康管理システムの構築を目指す.

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.

文　献

[1] 厚生労働省健康局，“標準的な健診・保健指導プログラム(確定版)“，Apr. 2008.

[2] 厚生労働省保険局，“特定健康診査・特定保健指導の円滑な実施に向けた手引き Ver1.7”，Jul. 2008.

[3] 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.

[4] 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.

[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相関ルール抽出アルゴリズム”，DBSJ Letters, vol.5, no.1, pp.25-28, Jun. 2006.

[6] 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.

[7] 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.

[8] 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.