

平成21年6月1日現在

研究種目：基盤研究（C）

研究期間：2007～2008

課題番号：19590562

研究課題名（和文）自己組織化マップ法を用いた健診業務支援システムの開発

研究課題名（英文）Development of health evaluation support system based on the self-organizing map technique.

研究代表者

市原 清志（ICHIHARA KIYOSHI）

山口大学・大学院医学研究科・教授

研究者番号：10144495

研究成果の概要：

大規模な健診結果に自己組織化マップ(SOM)法を適用し、種々の健常、発症前、疾患のパターンを識別するシステムの開発を行った。システムの評価には、某健診センターの8年間47,160件、26検査項目と基本身体計測値と生活習慣質問票よりなる健診結果情報(14,358名1人平均3.5回)を用いた。その結果、SOM法により健常人に現れやすい多様な病型とその発症前病型を分別できたが、その構成は男女で大きく異なった(男性では代謝症候群に属する病型が多いが、女性では少ない、等)。本研究では、得られた病型間の類似度から、その相互関連を病態系統樹の形に自動展開する機能を実装した。さらに(1)各病型群と健常典型群の生活習慣上の差異を多重ロジスティック分析で即座に解析・提示する機能、(2)受診者の病型の経年変化を個別に図示する機能を追加した。これらの改善により、効果的な健診後指導を可能とした。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,900,000	570,000	2,470,000
2008年度	1,600,000	480,000	2,080,000
総計	3,500,000	1,050,000	4,550,000

研究分野：医歯薬学

科研費の分科・細目：境界医学・病態検査学

キーワード：データマイニング法、自己組織化マップ、自己組織化病態系統樹法、予防医学、発症前診断、メタボリックシンドローム、健診業務支援システム

1. 研究開始当初の背景

多項目検査値よりなる大規模な健診データに自己組織化マップ(SOM: self-organizing map)法というデータマイニング手法を適用すれば、受診者の検査値パターンを自動分類(2次元図にマッピング)できることを確認した。しかし、分類初期状態に乱数を用いるため、マップの配置には一貫性がないという欠陥があった。そこで、分類後の検査パターンをその類似度の情報に基づいて、2次的に

樹形図の形に展開することで、再現性の問題を解決すると共に、各パターン間の連関を図示的に捉えられるようにした。

2. 研究の目的

大規模な健診検査の結果パターンをSOM法で自動分類し、各パターンの特性と病態意義を明らかにする。さらには、各病型パターン間の類似度を自動解析し、その相互関係を新考案の自己組織化系統樹法を利用してグラ

フィカルに提示できる健診支援システムを開発する。また、それに付加する機能として、個人の属するパターン（病型）の経時変化を樹形図上に提示したり、任意のパターン間で質問票から得た生活習慣情報にどのような差異があるかをリアルタイムに把握したりできるようにする。これらの機能を実装することで、従来、個々の検査値異常に着目して行われてきた健診後の指導を、全体的な健診結果パターンとしてとらえて、その病的意義を受診者にもわかりやすい形で伝えられるようにする。

3. 研究の方法

(1) 分析対象データ

システム開発のモデルデータとして、某大企業の同意を得て、その健診センターの過去8年間の精密健診結果 14,358名(男 7730名, 女 6628名、年齢：32歳～75歳)分、一人当たり平均 3.5回、延べ 47,160件のデータを使用した。

(2) 測定対象検査項目

末梢血検査 5項目：赤血球数(RBC)、ヘモグロビン(Hb)、ヘマトクリット(Ht)、白血球数(WBC)、血小板数(PLT)；生化学検査 21項目：グルコース(Glu)、中性脂肪(TG)、GGT、ALT、AST、LDH、ALP、総ビリルビン(TBil)、CRP、ZTT、総蛋白(TP)、アルブミン(Alb)、A/G比、総コレステロール(TCho)、HDLコレステロール(HDL)、コリンエステラーゼ(ChE)、尿酸(UA)、クレアチニン(CRE)、尿素窒素(UN)

(3) 個人特性情報

身体計測値：年齢、身長、体重、性別、血圧、BMI、体脂肪率(BF)；生活習慣調査項目：喫煙・飲酒習慣、ストレス、睡眠時間、食習慣、運動習慣、等

(4) データ解析方法

① 自己組織化マップ(SOM)法

健診時に得られた上下血圧、BMI、BFおよび26項目の検査成績をもとに、SOM法で8×8のパターンに自動分類した。その際にGlu、TG、GGT、ALT、AST、LDH、ALP、TBil、CRP、WBCは分布を正規化するため対数変換後のデータを用いた。この際全データを基準範囲を目安に平均値0標準偏差1となるよう基準化した。なお、各項目の基準範囲は、潜在異常値除外法によって基準範囲を計算するプログラム(市原清志 臨床検査 49増:1471-85, 2005)で設定した。なお、分類時の検査項目の並びは、血圧(SBP、DBP)、肥満度(BF、BMI)を左側に、脂質・肝機能(TG、GGT、ALT、AST)を中央左に、炎症マーカー(CRP、PLT、WBC)を中央に、栄養マーカー・腎機能・末梢血検査項目(RBC、Hb、Ht等)を右側に配置した。

② 自己組織化系統樹の作成

np個よりなる基準化された計測値 $Z_i [i=1, 2, \dots, np]$ を2乗して累和した値(偏差平方和に相当)の平方根に、偏差和 $\sum Z_i$ の符号を掛けた値を総合偏位度 D とした。

$$D = (\text{偏差和の符号}) \times \sqrt{\sum_{i=1}^{np} Z_i^2}$$

また、偏位の強い検査項目位置 P を

$$P = \sum_{i=1}^{np} (i - \frac{np}{2}) \times |Z_i|$$

として求め、それぞれ、Y軸、X軸に打点した。その結果、図1の中央に示すように、基線からのずれが最も少ないもの(基準パターン)を中心として、各パターンがマッピングされ、かつ類似度の強いパターンを線で結ぶことで樹形図の形となる。これを自己組織化系統樹と呼ぶ。

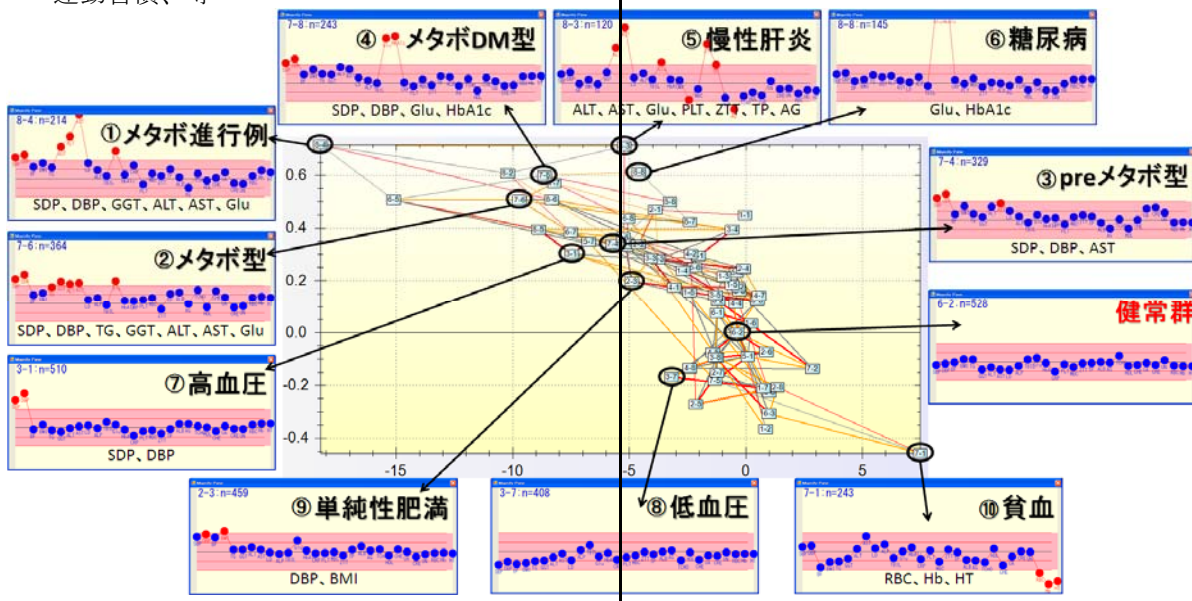


図1：SOM法で自動分類された8×8の検査値パターンの自己組織系統樹への展開

③疾患リスク学習システム

今回、各疾患パターンと健常パターン(基準パターンに近いもの5つを統合)との違いを、生活習慣の観点で比較する機能を作成した。各個人の生活習慣調査結果を説明変数として、多重ロジスティック分析4)を実施し、各疾患パターンと健常パターンの違いを分析した。

④個人の疾患パターンの経年変化

過去2回以上健診を受けた個人に対して、その検査プロフィールの推移とその属する疾患パターンの推移を、連続的にグラフ化して表示できる機能を作成した。

4. 研究成果

(1) SOM 自動分類結果から病態系統樹の作成

男女別に、健診結果情報(男性 26,611 件、女性 20,542 件)を SOM 法で 8×8 に自動分類し、それから系統樹を作成すると、図 1 中央に示すようになった。各パターンの病態意義を検査診断学的に判定し、名称を付けた。その代表的なパターンを図中に示す。また、主要な疾患パターンを整理し、色分けしたものを図 2 に示す。各疾患パターンと健常パターン(基準パターンに近い 5 つを統合)の違いを生活習慣の観点で比較するため、今回組み込んだ多重ロジスティック分析の自動実行機能を利用した。以下にその主要な解析結果を述べる

(2). 男性の疾患パターン

a) **メタボ型**: BMI が高いことを基準に、高血圧、肝機能障害、高血糖、高脂血症の程度からメタボ型(代謝症候群)と判断した。この範疇に属するパターンは 8 種類存在し、全体の 10.2% を占める。特にその程度が顕著なものをメタボ進行例(図 1-①)、メタボ型(図 1-②)、程度の軽いものをプレメタボ型(図 1-③)、メタボと糖尿病を合併したものをメタボ糖尿病型(図 1-④)、といったものが見られた。生活習慣特性については、健常群と比べ、メタボ型では食べるのが早く栄養バランスが悪い、メタボ糖尿病型は食べるのが早い、プレメタボ型では年齢が高めといった違いを認めたが、全体としては、健常者と明瞭に区別できる特徴的な傾向は見られなかった。

b) **肝機能障害型**(図 1-⑤): この範疇に属するパターンは 4 種類存在し、全体の 5.0% を占め、GGT、ALT、AST の高値を特徴とする。生活習慣特性については、健常群と比べ、間食はしないが塩辛いものを好み、多量に飲酒するという明らかな特徴が見られた。

c) **糖尿病型**(図 1-⑥): この範疇に属するパターンは 2 種類存在し、全体の 1.8% を占め、Glu、HbA1c の高値を特徴とする。生活習慣特性については、健常群と比べ、明瞭な差異は見られなかった。

d) **高血圧型**(図 1-⑦): この範疇に属するパターンは 4 種類存在し、全体の 21.2% を占める。生活習慣特性については、健常群と比べ、年齢が高めでよくお酒を飲むが、煙草は吸わず塩分を好まないといった特徴が見られた。

e) **低血圧型**(図 1-⑧): この範疇に属するパターンは 2 種類存在し、全体の 3.2% を占める。生活習慣特性については、健常群と比べ、食べるのが遅くあまり飲酒習慣がないという特徴が見られた。

f) **単純性肥満**(図 1-⑨): この範疇に属するパターンは 1 種類存在し、全体の 1.8% を占める。生活習慣特性については、健常群と比

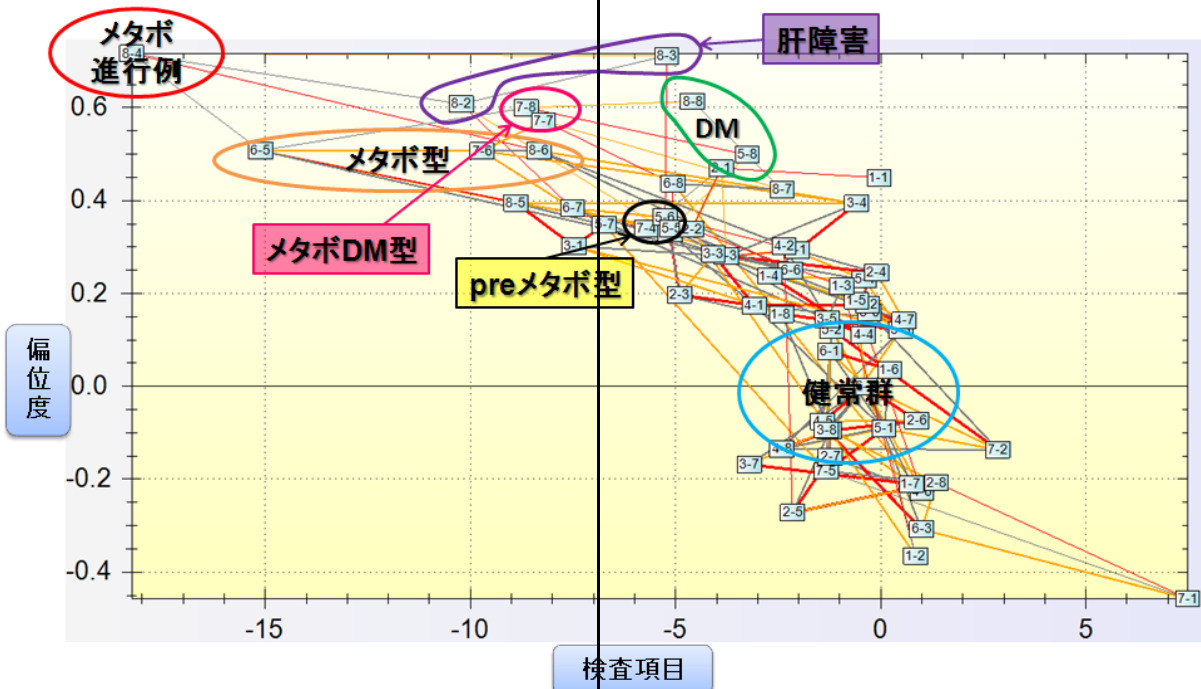


図 2 : 自己組織系統樹 : 8×8 の SOM パターンの相互関連と病態判定

べ、食べるのが早く良くお酒を飲むが、煙草はあまり吸わないといった特徴が見られた。
g) 貧血型(図 1-⑩)：この範疇に属するパターンは 1 種類存在し、全体の 0.9%を占める。生活習慣特性については、健常群と比べ、年齢が高めでよく間食をするが、食べるのは遅く煙草もあまり吸わないといった特徴が見られた。

(3) 女性の疾患パターン

女性に限定した SOM 分類では、メタボ型に属するパターンは少なく、単純性肥満、高グロブリン血症、高コレステロール血症、貧血を示すパターンが多かった (データ省略)。

(4) 個人の疾患パターンの経年変化

今回作成した個人のパターンの経年変化作図機能を使った代表的な出力例を図 3 に示す。この例では、ある個人の 7 年間の推移が示され、2 本の折れ線のうち実線がそのクラスの検査値、点線が各個人の検査結果である。この例の場合、1999 年では”健常”であったのに、”肝機能障害”を経て 2003 年には進行した”メタボ”となっている。その後おそらく 3 年間の努力を経て、”健常”に戻るといえる過程を読み取ることができる。

(5) 考察

自己組織化系統樹は、SOM 法の問題であった分類パターンの非再現性に対応したデータ解析法である。Y 軸は基線からの全体としての偏位度、X 軸は偏位の強い検査項目の位置を示すが、検査項目の配置に依存するため、この項目の順番を変えることで、特に注目したい疾患パターンの出現位置を調整しうる。逆に、検査項目の配置を誤ると、注目したい疾患パターンと

くくなる。このため、項目の配置に対して綿密な計画が必要である。

健診時の生活習慣調査票から、その個人がどの疾患になりやすいかを見るために、疾患リスク学習システムを作成した。しかし、今回、特に注目した代謝症候群については、生活習慣のみでその各サブタイプを分別するのは困難と判断された。これは、過去の健診後に指導を受け、すでに行動変容が起こっているためと思われる。従って、細かい生活習慣調査票の回答からはあまり適切な分析を行えず、特にメタボに関しては、BMI に生活習慣の偏り情報が集約されていると考えられた。

(6) 結語

本研究により、新考案の自己組織化系統樹法を取り入れた新しい健診業務支援システムを実用化すべく、そのユーザインターフェースも含めたプログラム開発を行うと共に、大規模な長期健診検査情報を利用して、その実用性の検証を行った。今回、新たに個人別経年変化の図式化機能を新たに導入し、かつ病型と生活習慣との関連をリアルタイムに示す機能を用意したことで、健診後の指導において、受診者の健診成績の経年変化とその病態意義を分かりやすく伝えることができるようになった。今後、本システムの普及により、健診業務が効率的・効果的に行われていくことを期待したい。

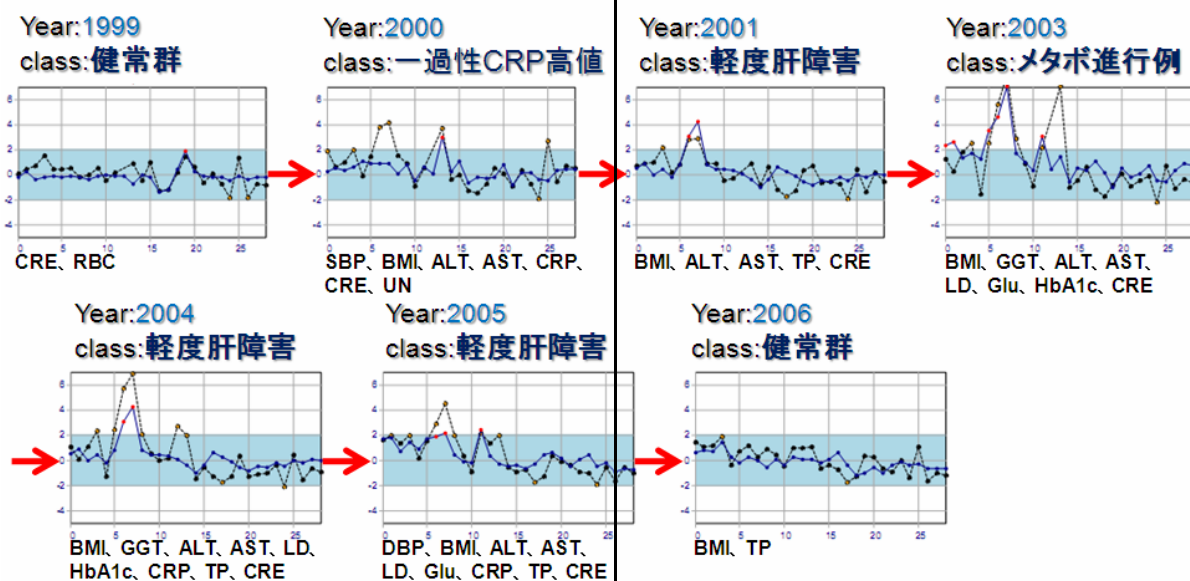


図 3：個人の病態パターンの経年変化チャート出力例 (各図の欄外記載項目は異常値を示した検査)

5. 主な発表論文等

〔雑誌論文〕（計 0 件）

〔学会発表〕（計 2 件）

1) Ichihara K. Practical application of data-mining methods in laboratory medicine. The 6th International Conference of Clinical Laboratory Automation and Robotics, Kobe, 2008 年 4 月 19 日

2) Ichihara K., Kataoka H, Nakajima N, Urano S. Clustering of health-check test results into a novel self-organizing tree of diagnostic categories. Clin Chem 54(S6): 59, 2008 年 7 月 29 日 (2008 Annual Meeting for Am Assoc Clin Chem, Washington DC).

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕

ホームページ等 なし

6. 研究組織

(1) 研究代表者

市原清志 (ICHIHARA KIYOSHI)

山口大学・大学院医学系研究科・教授

研究者番号：10144495

(2) 研究分担者 (2007 年度)

片岡浩巳 (KATAOKA HIROMI)

高知大学・医学部附属病院・

研究者番号：80398049

(3) 連携研究者 (2008 年度)

片岡浩巳 (KATAOKA HIROMI)

高知大学・医学部附属病院・助教

研究者番号：80398049