

多次元尺度構成法を用いた健診データの解析

Analysis of Health Checkup Data by Multidimensional Scaling

山本 けい子^{*1} 田村 哲嗣^{*1} 速水 悟^{*1} 紀ノ定 保臣^{*2} 浅野 昌和^{*3} 金川 誠^{*3}
Keiko Yamamoto Satoshi Tamura Satoru Hayamizu Yasutomi Kinoshita Masakazu Asano Makoto Kanagawa

^{*1} 岐阜大学工学部
Faculty of Engineering, Gifu University

^{*2} 岐阜大学大学院医学系研究科
Graduate School of Medicine, Gifu University

^{*3} 三洋電機(株)エコロジー技術研究所
ECO Technology Research Center, SANYO Electric Co., Ltd.

This paper presents a novel visualization analysis method that maps health check-up data of examinees using multidimensional scaling. New health check-up program to prevent lifestyle diseases started from April in Japan. The program uses a rather simple algorithm to classify the support level for each examinee. The proposed method provides visual information about more detailed condition of the examinee and its periodical change for monthly or half-yearly data. This graphical technique of multidimensional scaling helps intuitive understanding of the physical condition of the examinee for medical specialists or doctors. We tested this technique using 11 health check-up items for 648 examinees. The results show that groups for different support levels were plotted into the contrastive positions and to be helpful for health guidance plans.

1. はじめに

2008年4月から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・特定保健指導の実施に伴い、40歳から74歳までを対象とした大規模な健診データが収集されている。全ての健診受診者は、診断結果に応じて、情報提供・動機づけ支援・積極的支援の3段階の保健指導対象レベルと保健指導の対象外レベルの計4つに階層化され、保健指導対象レベルにおいては、保健指導担当者による保健指導が義務付けられる。

保健指導担当者には、一連の保健指導の中で、個々の対象者に適した指導内容の考案や定期的な観察・指導といった質の高い継続的な保健指導が要求され、特定健診関連の業務負担は明らかである。そこで、筆者らは、保健指導者が健診データの分析を行う際の支援を目的に、多次元尺度構成法を用いて健診データを2次元上にマッピングし、視覚的な情報として提示することを試みた。これにより、健診データで特徴付けられる受診者のメタボリックシンドロームに対するリスクを容易に把握することができ、円滑な保健指導につながると考えた。本稿では、これらの解析方法について述べ、結果から得られた知見や有効性について議論する。

2. 健診データの解析方法

2.1 特定健診・特定保健指導

健診項目(特定健診については、特定検診・特定保健指導プログラム[厚生省 HP]を参照)のうち、本研究で対象とする項目は、数値で表現可能な量的データであり、質的データは含まない。本研究の目的は、保健指導担当者が、膨大な健診データから階層化の手順に従って保健指導の対象者を選定し、個々の対象者に適した指導方針を容易に導くことができるような解析手法の確立である。対象者の選定に関しては、厚生労働省による「具体的な選定・階層化の方法」を用いて機械的に行うことができる。しかし、個々の対象者に合わせた保健指導を実現するためには、健診データを解析し、個々の詳細な特徴を捉える必要がある。そのため、本研究では、多次元尺度構成法を用いて、多変量の健診データで特徴付けられた健診受診者を2次元平面上にマッピングし、視覚的な情報として対象者の傾向を提示し、保険指導担当者の直感的な理解を支援する。表1に特定健診・特定保健指導の階層化に関する概略を記載する。

表1. 特定健診・特定保健指導の階層化の概略

階層レベル		男性: 腹囲 $\geq 85\text{cm}$ 女性: 腹囲 $\geq 90\text{cm}$	左記以外 BMI ≥ 25
中性脂肪 ≥ 150 or HDL コレステロール < 40 / 収縮期血圧 ≥ 130 or 拡張期血圧 ≥ 85 / 空腹時血糖 ≥ 100 or HbA1c ≥ 5.2 / 喫煙歴のリスク域に該当する数	0	情報提供	情報提供
	1	動機づけ支援	動機づけ支援
	2	積極的支援	動機づけ支援
	3	積極的支援	積極的支援
	4	積極的支援	積極的支援

連絡先: 山本けい子, 岐阜大学工学部, 岐阜市柳戸 1-1,
keiko@hym.info.gifu-u.ac.jp

先行研究としては、津本ら[津本 2005]が大規模データセットに対して多次元尺度構成法に基づく規則間の類似性、非類似性を示す可視化法を提案している。また、Bernatavičienė ら[Bernatavičienė 2007]は、医学データの可視化について議論し、クラスタリングと可視化を組み合わせた方法を提案している。

医療データの解析方法については、さまざまな統計的手法があるが、本論文では、保険指導担当者の支援という観点で、特に可視化の有効性について着目する。

2.2 多次元尺度構成法

多次元尺度構成法(MDS: multi-dimensional scaling)は、対象間の(非)類似度をもとに、その関係をできるだけ保つように対象を低次元空間の点で表す統計的な手法で、データの類似性を調べるための可視化によく用いられている。

保健指導担当者は、限られた時間内に多くの対象者に対して指導計画を作成しなければならない。人手や時間的な制限からも、個々の健診データの詳細な数値レベルでの分析は容易ではない。保健指導に必要なヒントとなる情報(指標)を健診データから読み取ることが重要なのである。本研究では、各健診受診者のデータを2次元上にマッピングすることで、指導対象者の状態に対して直感的な理解を支援しつつ、保健指導対象者の類似性を検査データの傾向とともに表現することを試みる。以下に解析手順を記す。

1. 特定保健指導の階層化手順に従い、健診データから対象外・情報提供・動機づけ支援・積極的支援の4レベルに該当する受診者を選定する。
2. 健診受診者の多次元の健診データから、受診者間の距離(類似度)を算出する。
3. 多次元尺度構成法を用いて、多変量の健診データを2次元に圧縮する。
4. 3の結果をもとに、マップを作成し、各受診者を2次元平面上にプロットする。

多くの検査データからなる多変量の健診データは、多次元空間上に配置された健診受診者の健康状態であると考えことができ、これらのデータを、各受診者間の健康状態(距離あるいは類似度)の関係を低次元(本研究では2次元)に写像することで、可視化を行うものである。

検査データの詳細な分析は、さまざまな多変量解析手法を用いればよいが、保健指導担当者が対象者がどの健康状態に位置しているのかといった概観を把握するためには、可視化が有効である。

3. 実験

特定健診は2008年4月からの実施のため、2008年4月時点で、特定健診の項目すべてが記載されているデータの入手が困難であった。そこで、本研究では、サンプルデータとして、より特定健診の項目に近い検査が行われていた三洋電機(株)によるヘルスケアサポートシステムの開発を通して収集されたデータを用いた。データの仕様を以下に示す。

- 受診者数:648名(うち、有効データ数574(男性168名/女性406名))
- 実験用データ(被験者1名につき、以下の11項目)
- 検査項目: BMI / 最高血圧[mmHg] / 最低血圧[mmHg] / 腹囲[cm] / 赤血球[万個/ μ l] / ヘマトクリット(Ht) [%] / 中性脂肪[mg/dl] / HDL コレステロール

[mg/dl] / LDL コレステロール[mg/dl] / 血糖(空腹時)[mg/dl] / ヘモグロビン[g/dl]

3.1 階層化の選定

表1に示した階層化に従い、保健指導対象者の選定を行った結果、得られた各レベルの人数と男女別の割合を表2に示す。

表2. 各レベルの人数

レベル	人数(割合)	男性(割合)	女性(割合)
0:対象外	379 (0.66)	92 (0.54)	287 (0.71)
1:情報提供	38 (0.07)	10 (0.07)	28 (0.07)
2:動機づけ支援	75 (0.13)	32 (0.18)	43 (0.11)
3:積極的支援	82 (0.14)	34 (0.21)	48 (0.12)
合計	574 (1)	168 (1)	406 (1)

特定健診・特定保健指導の階層化においては、約66%の人が保健指導の対象外と判定された。また、男性の方が女性よりも保健指導の対象者の割合が多いことがわかる。これらの階層レベルのうち、特定健診の結果を受けて、保健指導の対象となるレベルは、2:動機づけ支援と3:積極的支援である。しかし、保健指導対象外のレベル(0:対象外と1:情報提供)であっても、次回(翌年)の健診時に保健指導の対象となる可能性が高い予備軍が内在していることは否定できない。保健指導対象者だけでなく、予備軍についても、大まかな傾向をつかむことが重要である。

3.2 多次元尺度構成法によるプロット

本研究では、多次元尺度構成法とそのプロットを、オープンソースの統計解析システムRを用いて実行した。今回、可視化に重点をおき、用いた多次元尺度構成法の次数は2として、2次元平面上へのプロットを行った。また、距離行列はユークリッド距離を用いて計算し、各健診項目の数値は、正規化処理を行ったものを使用した。その結果を図1に示す。

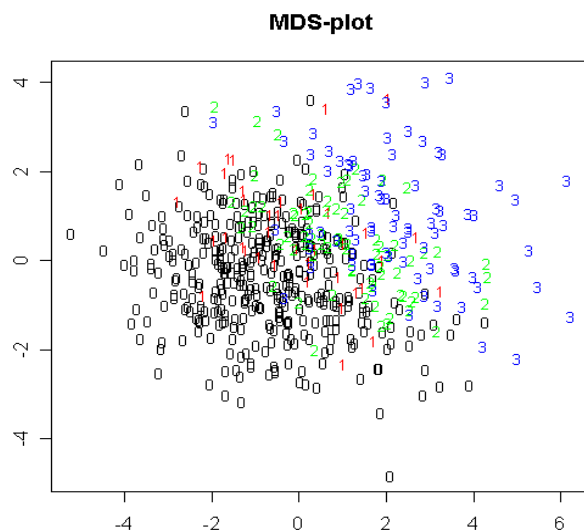


図1. MDSプロット

多次元尺度構成法の各プロットは、健診データで特徴付けられた受診者の相対的な位置関係を示している。図中の数字は、受診者の階層レベルであり、“0”は対象外、“1”は情報提供、“2”は動機づけ支援、“3”は積極的支援である。図の左側は階層化のレベルが低く、右側ほどリスクの高い対象者がプロットされていることがわかる。プロットは受診者 ID を用いてもよい。

MDS 上のプロットは、各レベルにおける詳細な位置を示す。新規の受診者データが得られたとき、プロットされた受信者データとの距離を、新規受診者のレベルに対する指標として使うことができる。

より詳細に検討するために、各レベルで MDS プロットしたものを図 2 に、レベルごとの MDS 上の密度分布を図 3 に示す。

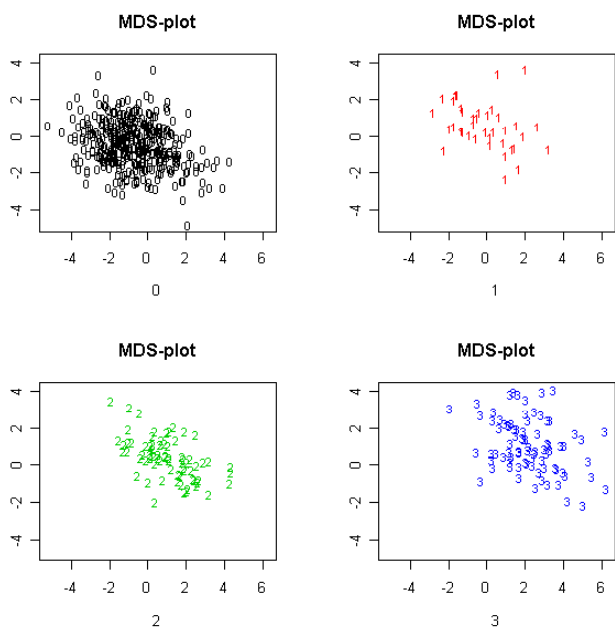


図 2. 各レベルの MDS プロット

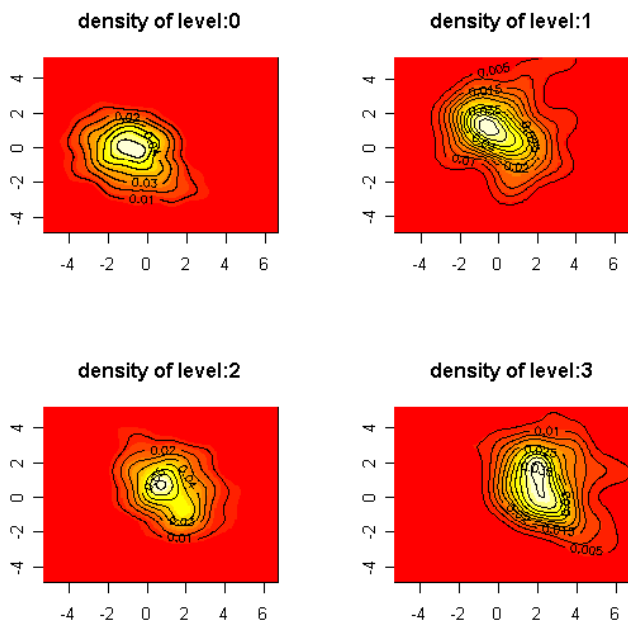


図 3. 各レベルの密度分布

図 3 において白い部分が密度分布の頂点であり、密度が高い状態を表す。頂点の位置が異なることから、それぞれのレベルの傾向を読み解くことができる。

レベル 0 とレベル 3 との傾向は明らかに対称的であることがわかる。またレベル 1 とレベル 3 との中間的な位置にレベル 2 がプロットされていることも読み取れる。レベル 1 に関しては、対象者が少ないこと、今回の階層化のレベルでは、腹囲か BMI のみ、あるいは腹囲と BMI の両方のみが基準値を満たさないグループであり、腹囲や BMI はそれ以外の検査項目と併発する機会が多いため、これだけでははっきりとした傾向がつかめないと考えられる。

以上の結果から、レベル 0 からレベル 2、そしてレベル 3 への推移の特徴が MDS のプロット上で示唆された。

3.3 多次元尺度構成法による詳細なプロット

次に、保健指導レベルの階層化に用いられている検査項目(腹囲[AC]・BMI[BMI]・血糖[BS]・脂質[L]・血圧[BP])と MDS プロットの傾向を調査するために、基準値を全て満たした受診者データ(対象外[0])と基準値を満たさなかった項目別に受診者データを MDS でプロットした。図 4 は検査項目の組合せのうち、特徴的な傾向を示したプロットである。

前述の図 2 の対象外([0])では、健診の階層化に従うレベルわけのため、腹囲あるいは BMI が基準値を満たしていれば、対象外としてプロットされていたが、図 4 では、より詳細に、血糖・脂質・血圧についても調査の対象とした。MDS の結果は、図 2 あるいは図 3 で示すように、階層化レベルの低い(問題のない)受診者は左側に、反対に階層化レベルの高い(言い換えれば、メタボリックシンドロームのリスク)の高い対象者ほど右側にプロットされる。

腹囲・BMI・血糖・脂質・血圧のそれぞれ、単独で基準値を満たさない場合は、問題のない対象外の範囲に近いプロットとなっている。しかし、腹囲と BMI の両方が基準値を満たしていない場合は、単独の結果より、右側へとシフトしている。また、基準値を満たさない検査項目の数が増えるに連れて(併発を意味する)、対象外よりも遠い位置、すなわち右側へのプロットへと移行している。これらの結果は、基準を満たさない検査項目の組合せによって細かくリスクが変化することも示唆している。

本研究で用いた MDS による可視化は、受診者の 1 ヶ月や半年などの定期的な変化を追跡することで、より詳細な状態(改善しているのか、あるいは悪化しているのかなど)についての視覚的な情報を与えることも可能である。

4. 考察

本研究では、多次元尺度構成法を用いて、サンプル健診受診者のデータを 2 次元上にプロットした。これにより、保健指導の対象外のグループとハイリスクのグループは対照的な位置にプロットされることが視覚的に確認できた。また、それらの中間的なリスクをもつグループはやはり中間的な位置にプロットされた。検査項目ごとのより詳細な分類によるプロットでは、併発することによるリスクの増加が読み取れた。これらの結果は、対象者の状態を直感的に知るだけでなく、その座標が、メタボリックシンドロームに対するリスクの 1 指標として有効であり、重点課題など保健指導計画の作成に活用できることを意味していると考えられる。

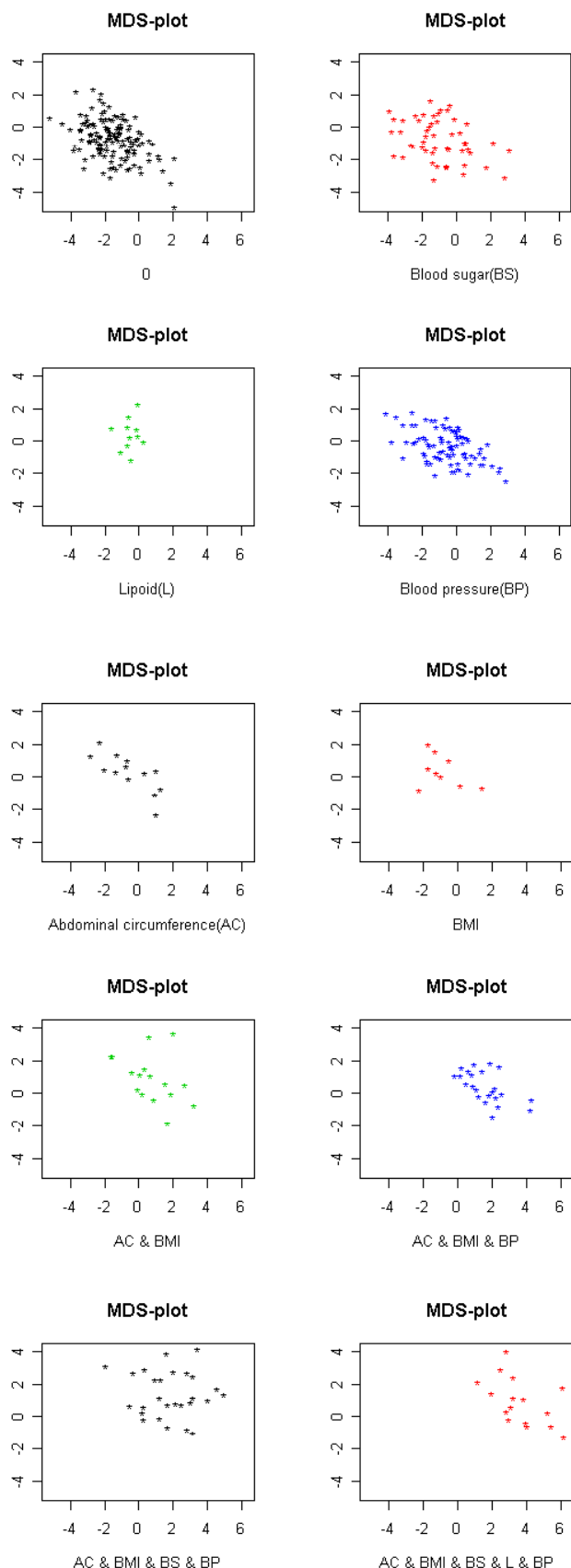


図 4. 詳細な MDS プロット

日本における特定健診の対象者は、5600 万人といわれている。このような大規模集団が同じ健診を受けるという試みは、前例がない。これらの結果の解析は、これからの生活習慣病予防への対策を決定するためにも非常に有効であると考えられる。

今後は、保健指導の担当者や医療従事者による結果の考察を通して、健診データの解析における可視化の有効性を検証していきたいと考えている。また、本研究で扱ったデータは、574 名という小規模なサンプルではあるが、特定健診で行われる検査項目のうち、11 項目を有するデータセットである。我々の調べた範囲では、そのような複数の検査項目に対する解析事例は少ない。実際に行われる特定健診の大規模データセットを用いることで、より詳細な傾向をマイニングし、健診データが表す対象者の健康状態や潜在的な疾患、症状への移行の程度を提示できる解析法への拡張が課題である。

謝辞 本研究の一部は、文部科学省「岐阜・大垣地域知的クラスター創成事業」の支援により行われた。

参考文献

- [Bernatavičienė 2007] Bernatavičienė J, Dzemyda G, Kurasova O, Marcinkevičius V and Medvedev V: The Problem of Visual Analysis of Multidimensional Medical Data. Optimization and Its Applications, 4:277-298, 2007.
- [津本 2005] Tsumoto S and Hirano S: Visualization of Similarities and Dissimilarities in Rules Using Multidimensional Scaling. Lecture Notes in Computer Science, 3488: 38-46, 2005.
- [Nakashima 2007] Nakashima N, Kobayashi K, Inoguchi T, Nishida D, Tanaka N, Nakazono H, Hoshino A, Soejima H, Takayanagi R, Nawata H: A Japanese model of disease management. Medinfo, 12(Pt 2) 1174-1178, 2007.
- [竹内 2008] 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, DEWS E1-5, 2008.
- [上村 2008] 上村勝美, 黛勇氣, 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, DEWS C5-4, 2008.
- [厚生労働省HP] 標準的な健診・保健指導プログラム[確定版], 厚生労働省ホームページ, <http://www.mhlw.go.jp/bunya/kenkou/seikatsu/>
- [R] The R Project for Statistical Computing. <http://www.r-project.org/>
- [金川 2007] 金川克子: 新しい特定健診・特定保健指導の進め方, 中央法規出版, 2007.
- [足達 2007] 足達淑子: 行動変容をサポートする保健指導パイタルポイント, 医歯薬出版, 2007.