

プロトタイプラインに基づく医療検査データの高次可視化

渡辺 健志[†] 鈴木英之進[†] 横井 英人^{††} 高林克日己^{††}

[†] 横浜国立大学 大学院工学研究院 〒240-8501 横浜市保土ヶ谷区常盤台 79-5

^{††} 千葉大学 医学部附属病院医療情報部 〒260-8677 千葉市中央区亥鼻 1-8-1

E-mail: †nabekun@slab.dnj.ynu.ac.jp, suzuki@ynu.ac.jp,

††yokoih@telemed.ho.chiba-u.ac.jp, takaba@ho.chiba-u.ac.jp

あらまし 本論文では不規則な多次元時系列データを確率的類型のシーケンスとして可視化する手法を提案し、医療データに適用する。慢性肝炎データは、典型的な病院検査データであり、種々の問題点を従来の機械学習手法に課している。われわれは以前、慢性肝炎データの検査結果を類型化する手法を提案し、医師が妥当で分かりやすいと認める類型を導出することに成功した。本論文では、この類型化手法に基づく可視化手法であるプロトタイプラインを提案する。プロトタイプラインは、情報量規準と色相認知を考慮し、各患者の病院検査データを分かりやすく変換して表示する。線維化程度に応じて患者を可視化した所、有望な結果が得られた。

キーワード プロトタイプライン, 可視化, 混合確率分布推定, 時系列データ, 医療検査データ

High-order Visualization of Medical Test Data with PrototypeLines

Takeshi WATANABE[†], Einoshin SUZUKI[†], Hideto YOKOI^{††}, and Katsuhiko TAKABAYASHI^{††}

[†] Faculty of Engineering, Yokohama National University 79-5 Tokiwadai, Hodogaya, Yokohama, 240-8501 Japan

^{††} Division of Medical Informatics, Chiba University Hospital 1-8-1 Inohana, Chuo-ku, Chiba, 260-8677 Japan

E-mail: †nabekun@slab.dnj.ynu.ac.jp, suzuki@ynu.ac.jp,

††yokoih@telemed.ho.chiba-u.ac.jp, takaba@ho.chiba-u.ac.jp

Abstract In this paper, we propose a method which visualizes irregular multi-dimensional time-series data as a sequence of probabilistic prototypes, and apply it to medical data. Chronic hepatitis data, which represent typical medical test data, pose various challenges to conventional machine learning methods. We have proposed a method which obtains prototypes from the data, and the prototypes were admitted valid and comprehensive by physicians. This paper proposes PrototypeLines which represent a visualization method based on the prototyping method. PrototypeLines consider information criterion and hue recognition, and display medical data of each patient in a comprehensive format. Promising results were obtained by experiments in which we visualized patients based on their degrees of fibrosis.

Key words PrototypeLines, Visualization, Estimation of Mixture Probabilistic Models, Time-Series Data, Medical Test Data

1. はじめに

病院検査データは患者の追跡調査 (follow-up study) などで採られるデータであり、高次元 (多属性)、不規則な計測、個人差、医師による測定バイアスなどの問題点を従来の機械学習手法に課している [11]。例えば、文部科学省科学研究費特定領域研究アクティブマイニングプロジェクトの共通データとして用いられている慢性肝炎データは、これらの問題点をすべて

含み、ECML/PKDD 2002 Discovery Challenge 国際ワークショップ [1] でのデータマイニングコンテスト [8] でも用いられた。

情報可視化は、その直観的な理解しやすさのために、ユーザとのインタラクションが必要な応用問題で重宝されており、データマイニングにおいても今後ますます重要となっていくと考えられる [2], [4]。データマイニングにおいて可視化は、データの前処理、知識の発見、および発見結果の提示などに用いる

ことができる。本論文では主にデータの前処理を想定し、病院検査データを理解・整理する方法を提案する。

情報可視化手法は一般に、データを可視化する手法とパターンを可視化する手法に分けられる。データを可視化する手法は、生データを可視化する手法、生データを変換して可視化する手法、および例を可視化する手法に分類できる。本論文では、生データを変換し、人間が分かりやすい形式に変換して可視化する手法を高次可視化手法と呼び、病院検査データに適する高次可視化手法を提案する。提案手法であるプロトタイプラインは、われわれが以前に発表した病院検査値の類型化 [12] に基づいている。論文 [12] においては、慢性肝炎データから発見された類型の多くが医師によって妥当で理解しやすいと認められた。プロトタイプラインは、これらの類型を用いてデータを抽象化し、情報量規準と色相認知を考慮して分かりやすく表示する。

2. 慢性肝炎データからの類型導出

2.1 慢性肝炎データ

慢性肝炎は、ウイルス感染により肝細胞が炎症を起こし傷害される疾患である。炎症が長期に持続した場合、肝硬変と呼ばれる終末像を呈する。肝硬変に至る道程においては線維化と呼ばれる指標が進行度を表し、線維化の全くない F0 から肝硬変を表す F4 の 5 段階がある。線維化は、直接肝臓に器具を挿入し肝組織を採取する肝生検によって診断するが、肝生検は出血などの危険性を伴い短期入院を必要とするため、頻回に検査できない。したがって、この検査と同等以上の線維化指標が、血液検査のような簡便な検査で代替できれば大きな意義がある。

図 1 に、慢性肝炎データから抜粋した患者 404, 702, 763, および 629 について、検査項目 GPT, CHE, T-BIL, および PLT の検査値に関する時系列データを示す。ただし、横軸の単位は日であり、0 は 1 回目の肝生検を受けた日を表す。水平方向の点線は検査結果のカテゴリ (UH: きわめて高い, VH: とても高い, H: 高い, N: 普通, L: 低い, VL: とても低い, UL: きわめて低い) が変化する閾値を表す。これらの患者は、線維化程度が全員 F4 であり、分類問題においては同じクラスに所属すると見なされる。図より、患者 404 と 702 はほぼ定期的に検査を受けており、前者は各検査結果のカテゴリが数回変化し、後者は検査結果が GPT 以外はほぼ安定していることが分かる。患者 763 は稀にしか検査を受けず、患者 629 は途中で検査を長い間受けなかったことが分かる。さらに図より、各患者は検査期間や検査値範囲が異なることも分かる。このように多様性に富むデータを単純な前処理で表形式データに変換し、従来の機械学習手法を適用することは、問題が多いと考えられる。病院検査データには、冒頭で述べた問題点があるため、個々のケースを検討して前処理した方が良く考えられる。したがって、このようなデータを効率良く整理・理解するための手法が重要である。

2.2 EM 法に基づく確率的類型の導出

図 1 の患者 629 の日付 0 付近より、各検査値が連動する場があることが分かる。実際、各種検査項目の検査値間には、片方が高ければもう片方も高い場合が多いなど、確率的な傾向

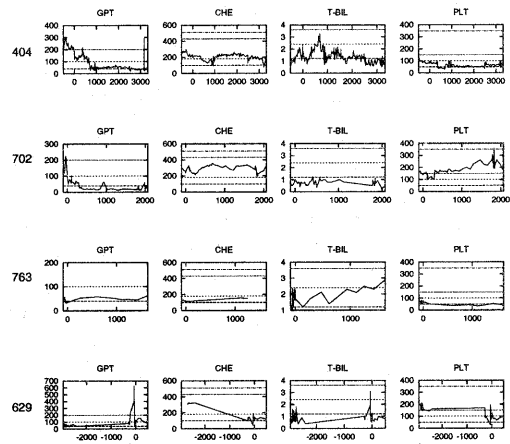


図 1 慢性肝炎データの一部

Fig. 1 Part of the chronic hepatitis data

表 1 慢性肝炎データの重要属性

Table 1 Important attributes in the chronic hepatitis data

検査項目	直観的説明	カテゴリ
GOT, GPT	壊れた肝細胞の量	N,H,VH,UH
ZTT, TTT	炎症の程度	N,H,VH,UH
T-BIL, D-BIL, I-BIL	免疫系の異常	N,H,VH,UH
ALB, CHE, T-CHO, TP	肝臓の予備能	VL,L,N,H,VH
WBC, PLT	炎症への血液の反応	UL,VL,L,N,H
RBC, HGB, HCT, MCV	血算検査	L,N,H

が存在する。すなわち、検査値の組合せには、起こりやすいものと、滅多に起こらないものがある。われわれは、検査結果をいくつかの類型 (prototype) で表せると考え、表 1 に示す検査項目の一部について確率的な類型を求めた [12]。ただし、医師のアドバイスに従い、検査値を表に示すカテゴリに変換して用いた。

患者 1 人が 1 日の特定回^(注1)に受けた検査結果の集合を 1 事例とする。検査項目集合を A で表し、 $A = \{a_1, a_2, \dots, a_{n(A)}\}$ とする。ただし $a_i, n(A)$ はそれぞれ検査項目、検査項目数である。この問題の入力はデータ集合 X であり、 $n(X)$ 個の事例 $X = \{x_1, x_2, \dots, x_{n(X)}\}$ で構成される。ここで x_i は属性値ベクトルを表し、 $n(A)$ 個の値 $x_i = \{v_i(a_1), v_i(a_2), \dots, v_i(a_{n(A)})\}$ で構成される。ただし各 $v_i(a_j)$ は事例 x_i における検査 a_j の値を示す。出力は集合 K であり、 $n(K)$ 個の類型 $K = \{k_1, k_2, \dots, k_{n(K)}\}$ で構成される。事例 x_i の生起確率は次で与えられる。

$$p(x_i) = \prod_{j=1}^{n(K)} p(x_i|k_j)p(k_j) \quad (1)$$

ただし $p(k_j)$ は類型 k_j の生起確率を表し、 $p(x_i|k_j)$ は k_j のときに x_i が起きる条件付き確率を表す。

(注1): データ集合 [1] において、同じ患者 ID、日付、回数をもつトランザクション集合

われわれは類型の推定に、EM法[3]を用いた。この手法は山登り法によって $p(k_j)$ と $p(x_i|k_j)$ の最尤値を推定する。これらの最尤値は負の対数尤度 ε を最小化する値として定義できる。

$$\varepsilon = - \sum_{i=1}^{n(X)} \ln \left(\sum_{j=1}^{n(K)} p(x_i|k_j)p(k_j) \right) \quad (2)$$

2.3 多重サイコロモデル

われわれは専門家を交えた反復的なデータ解析の結果、次に示す多重サイコロモデルを考案した。検査 a_l の値は、 $R(a_l) = \{r_1(a_l), r_2(a_l), \dots, r_{n(a_l)}(a_l)\}$ と離散化されるとする。ただし、 $n(a_l)$ と $r_m(a_l)$ は離散化区間数と m 番目の区間を表す。未検査値を区別するため、 $v_i(a_l)$ は右辺の1つの値をとることができる。

$$v_i(a_l) = \begin{cases} r_1(a_l) \\ r_2(a_l) \\ \vdots \\ r_{n(a_l)}(a_l) \\ - \text{(未検査)} \end{cases} \quad (3)$$

検査 a_l は $n(a_l)$ 面のサイコロを投げる行為としてモデル化されるため、 $n(A)$ 個の検査をモデル化するために $n(A)$ 個のサイコロを用意する。式(3)は、サイコロを投げない場合があるサイコロ投げを表す。

類型 k_j は次式で表される。

$$k_j = (k_{j1}, k_{j2}, \dots, k_{jn(A)}) \quad (4)$$

$$\begin{aligned} \text{ただし } k_{jl} &= (p(r_1(a_l)|k_j), p(r_2(a_l)|k_j), \\ &\dots, p(r_{n(a_l)}(a_l)|k_j)) \end{aligned} \quad (5)$$

$p(r_m(a_l)|k_j)$ は、類型 k_j において、検査 a_l の値が未検査でない場合 $r_m(a_l)$ となる確率を表す。この類型を用いた確率混合モデルを、多重サイコロモデルと呼ぶ。EM法において、条件つき確率 $p(x_i|k_j)$ は次式で与えられる、

$$p(x_i|k_j) = \prod_{l=1}^{n(A)} p(v_l(a_l)|k_j) \quad (6)$$

ただし $v_i(a_l) = -$ の場合は $p(v_i(a_l)|k_j) = 1$ とする。

条件つき確率 $p(x_i|k_j)$ は、次式を用いて更新される。

$$\begin{aligned} p^{new}(r_m(a_l)|k_j) &= \frac{\sum_{i=1}^{n(X)} p(k_j|x_i)p(k_j)\gamma_1(x_i, k_j, l, m)}{\sum_{i=1}^{n(X)} p(k_j|x_i)p(k_j)\gamma_2(x_i, k_j, l, m)} \end{aligned} \quad (7)$$

ただし

$$\gamma_1(x_i, k_j, l, m) = \begin{cases} 1 & (v_i(a_l) = r_m(a_l)) \\ 0 & (v_i(a_l) \neq r_m(a_l)) \end{cases} \quad (8)$$

$$\gamma_2(x_i, k_j, l, m) = \begin{cases} 1 & (v_i(a_l) \neq -) \\ 0 & (v_i(a_l) = -) \end{cases} \quad (9)$$

実験では、B型慢性肝炎患者を抗原と抗体の状態に分けて類

型を導出した。発見された類型はほぼ全て、医師らによって可読性に優れ妥当であることが認められた。さらに医師らは、この手法は患者の病態変化を把握することと例外的な患者を発見することに関してきわめて有効であるとコメントした。

3. プロトタイプライン

3.1 動機

2.1節と2.2節で述べた動機に基づき、病院検査データを整理・理解しやすいように患者毎に可視化することを考えた。可視化が適切であれば、患者や時期毎の特徴が把握しやすくなり、学習・発見アルゴリズムに入力するデータを準備しやすくなること期待される。

例えば慢性肝炎データには、問題を限定しても数百人の患者があり、主要な検査値だけで約17個ある。図1の形式では、A4用紙1枚に数人程度しか示せず、多数の患者を同時に分析することは難しい。この問題を解決するために、生データを2.3節で紹介した確率的類型を用いて変換し、事例を式(1)の形式で表示することにした。これは、検査結果を確率的タイプの線形和として表すことに相当する。

3.2 類型への色割り当て

前節の形式でデータを表示する場合、類型を分かりやすく表示することが重要である。データを直接可視化する手法では通常、色、形、および大きさなど種々の情報媒体を利用する[2],[4]。このような手法の表示結果は、習熟すれば分かりやすいと考えられるが、重要な情報を見落としてしまう場合も多いと考えられる。本論文では病院検査データの複雑さとユーザへの分かりやすさを考慮し、類型を色相(hue)[5]だけで表現することにした。類型は応用分野における尺度、例えば慢性肝炎データでは検査結果の善し悪しに応じて整列され、紫から赤の色相を割り当てられる。

類型を整列するためには、各検査の異なるカテゴリを評価する単一の指標が必要である。慢性肝炎データにおいては表1に示すように検査値にカテゴリが存在し、VLはLよりも悪いなどの情報が分かっている。さらに、例えばGOTの検査値は比較的HやVHの分類になりやすいが、T-BILは減多にならないなど、検査値結果のカテゴリは検査毎によって重要度が異なると考えられる。

これらの問題を解決するために、われわれは類型を平均類型からの逸脱度合いで表現する情報量規準を提案する。この規準において検査項目 a_l のカテゴリ $r_m(a_l)$ の重要度は、情報量 $-\log_2 p(r_m(a_l))$ で表される。われわれが実験した条件では、N以外のカテゴリの生起確率はNから遠いほど小さくなることと、 $p(r_m(a_l))$ が小さいほど情報量 $-\log_2 p(r_m(a_l))$ が大きいため、上記の要求は満たされる。類型 k_j の悪さを表す情報量規準 $I(k_j)$ は、次式で与えられる。

$$I(k_j) = \sum_{l=1}^{n(A)} \sum_{m=1, r_m(a_l) \neq N}^{n(a_l)} -p(r_m(a_l)|k_j) \log_2 p(r_m(a_l)) \quad (10)$$

k_j においてカテゴリNを無視するのは、Nは普通であるため

に注意されず情報量が 0 と見なせるためと、N の生起確率が H や L の生起確率より小さい場合があるためである。

HSV 表色系は、人間の色表現特性を考慮したカラー画像の表現法であり、色相、採度、および明度を用いる。色相は、赤→黄→緑→青→紫を基調色相とし、これらを中間色と共に円周上に並べた色相環に基づいている [5]。われわれは、暖色が寒色よりも目立つことを考慮し、類型を悪い順から赤 (0 度)、橙 (30 度)、黄 (60 度)、黄緑 (90 度)、緑 (120 度)、青緑 (150 度)、青 (210 度)、紫 (240 度) に対応させることにした。これらの色相の選択は、主に類型の区別し易さに基づく。類型が 8 種類よりも少ない場合には、適宜見分けにくい色相から省くことにする。

3.3 表示画面の設計

表示画面の設計に当たり、多数の患者を時系列的な変化が明らかになるように配置することを最も重視し、重要情報の表示と表示重なり回避を次の目標とした。

プロトタイプラインでは、患者を行として表示し、検査結果を表す類型の線形和を時間軸に沿って配置することにした。患者毎にあまりにも時間範囲が異なるため、患者毎の横幅を固定し、時間尺度を可変とした。この表示法により、最初の目標が達成されると考えた。次の 2 点に関しては、肝生検と IFN 情報を示し、1 検査結果の幅を 5 ピクセルとすることで対処した。このピクセル数は、類型の見やすさと検査結果の重なりトレードオフを考慮して決定した。

4. 慢性肝炎データの可視化

4.1 表示結果

慢性肝炎の性質より、肝生検の前 500 日と後 500 日で線維化程度が変化するのは稀である。したがって、この期間においては、各患者の線維化程度が比較的確実に把握できると考えた。肝生検を受けて線維化程度が分かっている全患者について、この期間の検査結果から 2. 節の手法で類型を導出した。検査項目は表 1 の 17 種類であり、EM 法のランダムリスタート数は 100 とした。類型数は、類型間の相違度と類型の生起確率を考慮し、8 とした。なお 2.3 節の手法を変更し、多数の検査を受けた患者の影響を減らす工夫を行った。

各カテゴリの平均値 (average) と学習された 8 個の類型 (1, 2, 4, ..., 8) を図 2 に示す。8 個の類型は左から右に、良いから悪い順に整列されている。図において平均値と各類型は 17 種類の検査項目で構成されており、各検査項目は縦方向の割合グラフで表されている。類型の後に示す括弧内の割合は、類型の生起確率である。各カテゴリの平均値より、GOT、GPT、TTT、ZTT、PLT および MCV は比較的異常なカテゴリになりやすく、これらの検査結果は情報量規準において軽視されることが分かる。実際、これらのカテゴリは各類型に頻出する。この点を考慮して各類型の特徴を類型が良い順に述べると、類型 1 と 2 では多くの検査項目が N である確率が高く、良い検査結果を表すと考えられる。類型 4 では WBC と ALB が低くなる確率が高く、類型 3 では GOT と GPT が VH となる確率が高いことが分かる。類型 5 では GOT、GPT、TTT、

ZTT、ALB、CHE、PLT、および MCV と多くの検査項目が異常なカテゴリになりやすい。これに比較すると、類型 6 は一見異常なカテゴリが少なく見えるが、D-BIL、I-BIL、および T-BIL が異常となる点が目立つ。類型 7 は血算関係の検査項目が、類型 8 はほぼ全ての検査項目が異常となる確率が高い。

図 3 に、F4 患者の肝生検の前後各 500 日におけるプロトタイプラインを示す。図では 1 行が 1 患者、1 本の棒が類型の線形和で表される 1 回の検査結果を表す。図より、各患者の類型の変化が検査時期と共に容易に把握できることが分かる。プロトタイプラインで特徴的な患者に関して検査値のグラフを検討した所、検査結果のカテゴリが総合的に判断されて可視化されていることが分かった。プロトタイプラインは特に、先入観や不注意による見落としを気づかせ、一見異なって見える類似患者を特定するのに有効であることが分かった。

4.2 多角的な分析

プロトタイプラインでは、既に発見された類型を用いて、任意の期間のデータを表示することが可能である。例えば図 4 は、図 2 の類型を用いた患者 602 の全検査結果に関するプロトタイプラインを表す。この図より、図 3 においては問題が少ないと推察される患者 602 の容体が、肝生検の約 2 年後から急速に悪化したことなどが分かる。

全ての検査項目を用いて類型を発見する方法は、多数の患者を同時に見られるが、検査項目をより細かく分けて多角的に分析したいとの要求もある。この要求に応えるため、図 5、6 に示すように検査グループ毎に類型を求めて患者を表示する、マルチプロトタイプラインも構築した。図 6 より、患者 763 は検査項目 TTT と ZTT が特に悪く、ALB、CHE、TP、および T-CHO なども悪いことが分かる。

5. 関連研究

これまでの可視化手法は、大量のデータを表示するために多種の情報媒体を用いたりズームに頼って来た。プロトタイプラインは、学習された類型を用いて情報を要約することにより、この問題に対処する新しいアプローチに属する。その設計哲学は、ユーザの負担を最小限とするために、学習手法によって得られた本質的な情報だけを、主に色相という情報媒体だけを用いて示すことにある。

シーケンス情報を表示する可視化手法としては、LifeLines [6] や InfoZoom [10] などが挙げられる。これらはいずれも、生データを表示する可視化手法に属し、主にズームによって大量情報に対処している。プロトタイプラインは、これらに比較して機能がきわめて限定されているが、上記の設計思想に基づき、生データを分かりやすく変換して表示する高次可視化手法に属する。

6. おわりに

「百聞は一見にしかず」という諺は真実を表していると思われるが、人間が見ることが出来る量には限りがある。われわれは情報可視化手法において、種々の情報媒体を用いてなるべく多くの情報を表示するアプローチよりも、情報を適切に抽象化

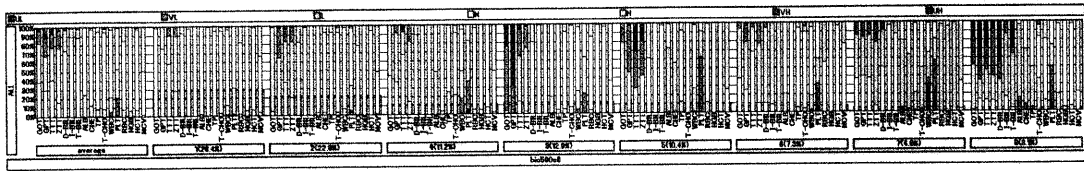


図2 肝生検の前後各500日のデータから発見された類型. 元のカラー画像は[9]

Fig.2 Prototypes discovered from data before 500 days and after 500 days from biopsy.
The original figure in color can be found in [9]

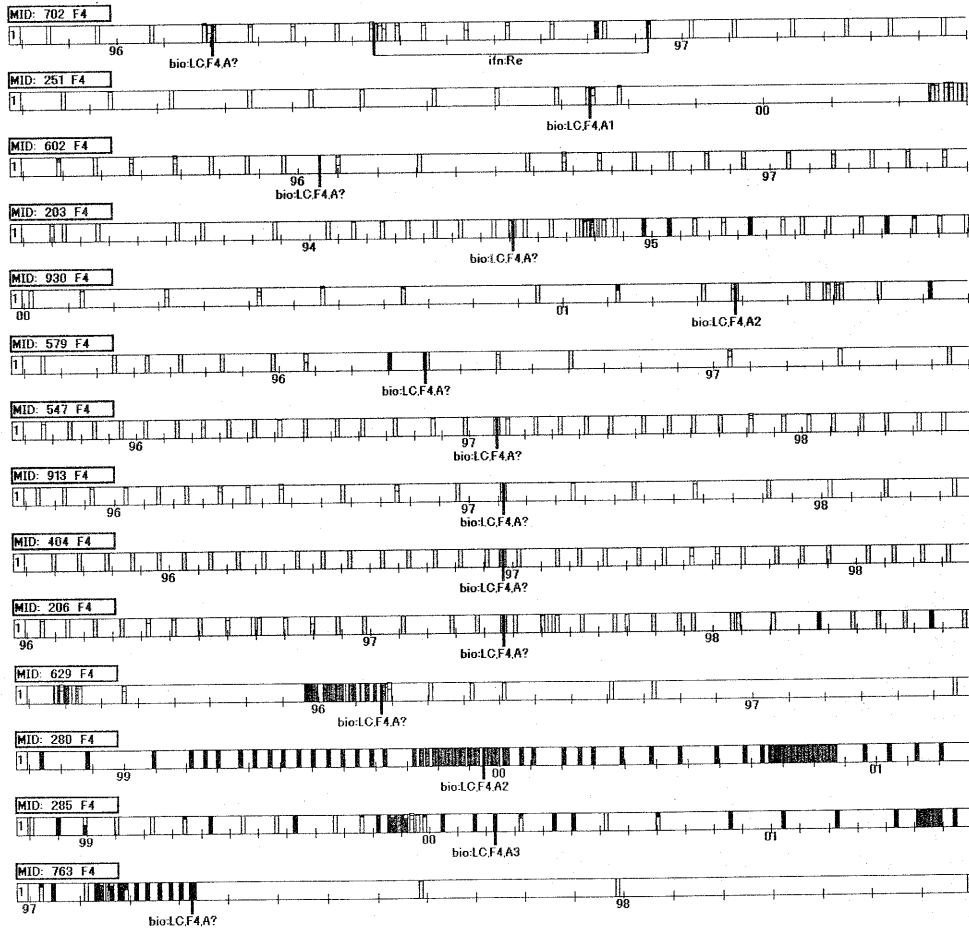


図3 肝生検の前後各500日におけるF4患者のプロトタイプライン. 元のカラー画像は[9]

Fig.3 PrototypeLines of F4-patients' data before 500 days and after 500 days from biopsy.
The original figure in color can be found in [9]

するアプローチが有望であると考えた。プロトタイプラインは、情報の抽象化をEM法に基づく確率的クラスタリングで行っている。このことに加えて、類型への適切な色相の割り当てや表示法、および医療データの解析に試行錯誤的に取り組んだ結果、4.2節に示す慢性肝炎データにおける成功もたらされたと考えられる。

データマイニングにおいて、パターン抽出の抽象度は事例の情報

を省略する度合いを表すと見なすことができる。直観的に言えば、パターンの抽象度が高ければ、そのパターンはより多くの事例を説明する。例えば属性数が少ないルールは、パターンの抽象度が高い。津本と鈴木らは事例発見[7]などを通して、専門家が理解しやすいパターンの抽象度は、属性数が少ないルールよりも低い場合があることに気づいた。さらに慢性肝炎データに関する解析では、医師はルールに比較して、グラフなど該

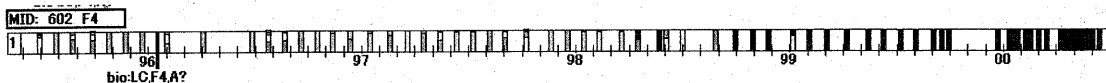


図 4 患者 602 の全検査結果に関するプロトタイプライン. 元のカラー画像は [9]

Fig. 4 Prototypelines of whole test results of the patient 602. The original figure in color can be found in [9]

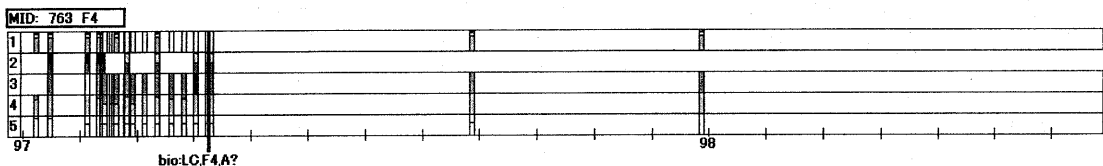


図 6 患者 763 に関するマルチプロトタイプライン. 元のカラー画像は [9]

Fig. 6 Multi-prototypelines of the patient 763. The original figure in color can be found in [9]

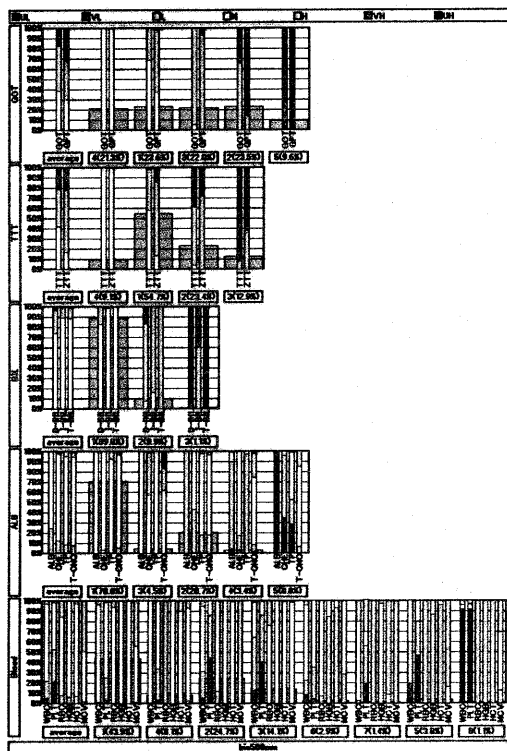


図 5 検査グループ毎の類型. 元のカラー画像は [9]

Fig. 5 Prototypes of grouped tests. The original figure in color can be found in [9]

当する患者が分かりやすい表現を好むことが観察された。プロトタイプラインは、ルールよりもパターンの抽象度が低いため該当する患者が分かりやすく、グラフよりもパターンの抽象度が高いため多くの情報を一度に見られるという利点がある。今後は、領域知識を併用した可視化、可視化手法の定量的評価、およびプロトタイプラインの他の問題への適用などに取り組んでいきたい。

謝 辞

本研究の一部は、文部科学省科学研究費特定領域研究「アクティマイニング」の援助を受けている。

文 献

- [1] P. Berka: ECML/PKDD 2002 Discovery Challenge, Download Data about Hepatitis, <http://lisp.vse.cz/challenge/ecmlpkdd2002/>, 2002 (current September 28th, 2002).
- [2] S. K. Card, J. D. Makinlay, and B. Shneiderman (eds.): *Readings in Information Visualization*, Morgan Kaufmann, San Francisco, 1999.
- [3] A. P. Dempster, N. M. Laird, and D. B. Rubin: "Maximum likelihood from incomplete data via the EM Algorithm", *Journal of the Royal Statistical Society B*, Vol. 39, No. 1, pp. 1-38, 1977.
- [4] U. Fayyad, G. G. Grinstein, and A. Wierse (eds.): *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann, San Francisco, 2002.
- [5] 平林雅英: C 言語によるプログラム事典第 2 巻, 技術評論社, 1992.
- [6] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman: "LifeLines: Visualizing Personal Histories", *Proc. CHI 1996*, pp. 221-227, 1996 (also in [2]).
- [7] S. Sugaya, E. Suzuki, and S. Tsumoto: "Instance Selection Based on Support Vector Machine for Knowledge Discovery in Medical Database", *Instance Selection and Construction for Data Mining*, pp. 395-412, Kluwer, Norwell, Mass., 2001.
- [8] 鈴木英之進, 津本周作: 「データマイニングコンテスト 編集にあたって」, 情報処理, Vol. 42, No. 5, pp. 443-444, 2001.
- [9] 鈴木英之進: 「プロトタイプラインに基づく医療検査データの可視化におけるカラー画像」, <http://www.slab.drj.ynu.ac.jp/paper/20030315/figs.pdf>, 2003 (current February 12th, 2003).
- [10] M. Spence: "Visualization and interactive analysis of blood parameters with InfoZoom", *Artificial Intelligence in Medicine*, Vol. 22, No. 2, pp. 159-172, 2001.
- [11] S. Tsumoto: "Rule Discovery in Large Time-Series Medical Databases", *Principles of Data Mining and Knowledge Discovery (PKDD)*, *LNAI 1704*, pp. 561-567, 1999.
- [12] T. Watanabe, E. Suzuki, H. Yokoi, and K. Takabayashi: "Prototyping Medical Test Results in Chronic Hepatitis Data with the EM Algorithm on Multi-Dice Models", *Proc. International Workshop on Active Mining (AM)*, pp. 45-51, 2002.