展　　望

離散 Bayes 識別則とその個別化医療への応用
A Discrete Bayes Decision Rule and its Application to Personalized Medicine

浜本　義彦
Yoshihiko HAMAMOTO

概　要

本稿では，まず人工知能の根幹に係る一般パターン認識の問題を明らかにするとともに，その解決に有用な統計的パターン認識とそれを発展させた離散 Bayes 識別則を概説する．次に離散 Bayes 識別則の個別化医療への応用として診断・予測問題と創薬における治験問題に言及する．

1. はじめに

認識の問題は，古くから哲学や認知科学における重要な研究課題であった．コンピュータの出現とともに，1950 年代からパターン認識の研究が開始され，これまで認識という知的能力をコンピュータによって実現しようとする多くの試みがなされてきた．昨今では，コンピュータ技術の急速な進歩により誕生した深層学習が注目され，画像認識を中心に研究が盛んに行われるようになった．しかし，これらの研究も一般パターン認識の問題に行き当たり，当初の期待通りとはならない状況に陥っている．

本稿では，この一般パターン認識の問題とは何かを解説し，それを解決する可能性を秘めた統計的パターン認識とその医学への応用に端を発して開発された独自の離散 Bayes 識別則を紹介する．

2. 統計的パターン認識とは

さて，パターン認識とは何かを定義することから始める．パターン認識とは，任意に与えられた認識対象を，それが属するとされる有限個のクラス（概念）の一つに対応づける機能である，とする．ここでクラスが未知の場合と既知の場合がある．前者は与えられた認識対象間の類似性からクラスを創造する場合であり，クラスタ分析とも呼ばれる．一方後者は，既知のクラスに関する知識に基づき認識対象をいずれかのクラスへ対応づけるもので，狭義のパターン認識と呼ばれ，本稿はこの場合に主眼をおく．

コンピュータのパターン認識系は，外界に存在する認識対象をコンピュータ内でデータの組により表現されたパターンと同一視する観測系，識別に有効な観測を特徴と呼んで選択する特徴選択系，特徴を用いてパターンをクラスへ識別する識別系からなる．

以上の準備のもと，前述の一般パターン認識の問題を解説する．この問題は，①クラスは存在するのか，②観測系の在り方は如何にあるべきか，③特徴選択は必要なのか，と 3 つある．

2.1　クラスは存在するのか

クラスとは何か，それは人間が価値のあると認めた概念である，と定義する．概念は，人間が定めた言葉（記号）で表現される．パターン認識は，英語では「Pattern Recognition」と書かれる．ここで「Recognition」を「Re」と「cognition」に分解すれば，パターン認識とはパターンを再び認識する，つまりパターン認識は 2 段階あって，まずパターンをクラスへ対応させる仕組みを学ぶ学習段階と学習後に新規のパターンをクラスへ対応させる識別段階からなる[1]．これは，対応先のクラスが存在しないとパターン認識はできないことを意味する．そのためコンピュータにクラスの教え方が課題となる．これは根本的な課題であるが，一般にはあまり意識されていない．例えば文字認識を行うコンピュータは，

認識対象となる数字の「2」を表す幾何学図形を観測し，それを「2」のクラスの一員とみなすために「2」のクラスをどのようにして獲得しているのであろうか．一方，人間は人間同士で「2」の概念をどのように共通理解しているのであろうか．

人間が行う概念の獲得は推論によると考えられる．推論は帰納的推論と演繹的推論に大別され，コンピュータに行わせる推論は帰納的推論である．具体的には事例となるサンプルとそれが属する正解のクラス（概念）との組をコンピュータに繰り返し提示して，事例間に存在する規則性を獲得させるのである．これを教師あり学習と呼ぶ．教師あり学習にはサンプルの選出とサンプルに付与されるクラスラベル名の信頼性という2つの課題がある他に、基礎となる帰納的推論には演繹的推論と異なって避けられない深刻な宿命がある．それは，①絶対に正しいことが保証されない，②学習結果がサンプルに依存する，ということである．つまり帰納的推論は，不確実性のある推論であり，誤りを原理的に避けられないのである．その打開策として，統計学は不確実性を数値化するアプローチを提案した[2]．

パターン認識に統計的手法を導入したのが統計的パターン認識[3),4),5)]である．統計的パターン認識は，1950年代に多変量解析を母体として誕生し，その後独自に発展して不確実性に対処できるパターン認識となった．統計的パターン認識では認識対象がどのクラスの一員であるかを識別する際に，認識対象を観測すること（パターンと同一視化）を条件として，それがクラスに属する条件付き確率を表す事後確率を用いて，事後確率が最大のクラスへ認識対象を識別する．これが統計的パターン認識のBayes識別則である．そこでは認識対象がどのクラスに属するかという不確実性は，事後確率で数値化されている．

2.2 観測系の在り方は如何にあるべきか

前述したようにコンピュータによるパターン認識では，認識対象は観測されたデータの組でパターンとして表現される．このとき，データはできるだけ多ければ良いとし，何も考えずに認識対象を観測すればパターン認識ができるのかと言えば，そうではない．

理想と現実の場合を対比させてみよう．まず理想的な場合である．認識対象を2次方程式，クラスを実根のクラスと虚根のクラスとする．このとき2次方程式の何を観測して如何なるデータを獲得すれば良いかを考える[6]．この問題は判別式を観測し，その値をデータとすれば，誤りなく100%の精度で2次方程式を識別できる．これができるのは，認識対象である2次方程式が解明されているからである．

しかし現実はそうではない．文字認識では技術者が試行錯誤的に創意工夫したデータを獲得しているが，それが良いという保証はない．深層学習はこれを自動的に行えるといわれるが，そこには事例となるに相応しいサンプルの選び方やネットワーク構成に人間の知恵が必要であり，しかも画像認識に限定される．更に深刻な帰納的推論の宿命もある．

この2次方程式の例を基に考えれば，観測の在り方は認識対象に関する研究に基づくべきという基本方針が見えてくる．医療問題の例では，認識対象は「患者」で，患者に関する研究とは「医学」である．クラスは「疾患名」，観測は「マーカー（検査項目）」，データは「測定値（検査データ）」となる．患者にどのような検査を行うかは，医師が患者を診察して，医学に基づく推論（臨床推論）によって決まる．

このように観測系が認識対象に関する研究を踏まえるため，必然的にパターン認識は認識対象に依存した個別論にならざるを得ない．例えば視覚情報では観測系をカメラとした画像認識，聴覚情報ではマイクロホンとした音声認識，等である．汎用化できるのは，前述の観測系，特徴選択系，識別系といったパターン認識系の中では識別系だけであろう．

この他に，多変量解析と同様にパターン認識においても，認識対象を観測してどんなデータを獲得すれば良いかは，やってみなければ分からない．従って試行錯誤は避けられず，より少ない回数で解に到達できるかは事前調査と経験次第である．

実際には何が有効かは不明なため，認識対象に関する研究から有望と見込まれる観測を可能な限り行ってデータを獲得している．例えば，癌は遺伝子異常の疾患であることから，癌に関する情報は遺伝子の中にあり，遺伝子に関する観測を行ってデータを獲得すれば良いと考える．しかし，人間の遺伝子の数は約2万個もある．このときパターンの次元数は2万となり，極めて高次元になってしまう．

2.3 特徴選択は必要なのか

3つ目の問題は，観測によって獲得されるデータは全て識別に用いられるのか，それとも選択されるのか，である．もし選択するとなると，高次元の場合では実行は極めて困難となる．

統計的パターン認識[7]によれば，観測には役に立つ観測とそうではない観測がある．役に立たない観測からのデータは，それが識別の際に何もしなければ放置しておけば良いが，悪さをするのである．そのため識別に有効な観測を特徴とみなし，特徴だけを選んで用いることになる．これを特徴選択という．

特徴選択においてどの観測が有効かを個別に調べて上位のみを特徴として選択する方法が考えられるが，事はそう簡単ではない．Elashoffら[8]は，互いに独立な観測を個別に評価してランキングし，最上位の2つの観測が必ず特徴となるとは限らない例を示した．更にCoverら[9]は理論的に全ての観測の組合せを調べないと最適な特徴の組合せを選択できないことを示した．遺伝子のように候補となる観測の数が万単位になれば，特徴選択は実行不可能となる．ただ，パターン認識の実務者から言えば，Coverらの結果は理論的興味があるだけである．それは，Coverらが論じたのは組合せの最適化法で，最適化すべき特徴の評価関数が現実には未知でその推定値を用いるしかないため，推定誤差のある不正確な評価関数をまじめに最適化してもそれほどの意味はないからである．注意すべきは，特徴の組合せである．

以上がパターン認識の実現を困難とする3つの問題である．次に，これらと密接に関連する，より根源的な価値観と認識との関係について論を進める．

2.4 価値観と認識

特徴選択の研究により観測には序列があり，それはパターン認識問題を定める認識主体者の価値観に基づいた重要性によって決まる．例えば，医師は診察時に患者が罹患していると想定される疾患に係る検査項目（観測）を考え，それ以外の検査項目には関心がない．また癌の診断を研究する分子生命科学者らは，ある種の癌に関する特定の遺伝子群（標的遺伝子群）があり，全ての遺伝子が等しく平等に当該の癌に関与するわけではない，と考えている．

この価値観は，人間のみにあり，コンピュータにはない．人間は，意味がある，価値があると認めたクラス（概念）を定義（創造）する．ところで価値観は，論理的ではなく，非論理的である．一見矛盾しているようであるが，「論理的なコンピュータに非論理的要素を入れなければパターン認識はできない」のである．これを初めて指摘したのが，渡辺　慧[1]である．渡辺は「みにくいアヒルの子定理」で，もし観測の重要性を決める価値観が人間によって与えられなければ，全ての認識対象は同じように類似してしまい，互いに区別できない，つまりパターン認識はできないと説明した．

渡辺は観測の在り方を考察したが，これは事例，すなわちサンプルの選択に対しても適用される．学習においてサンプルを選ぶ際にサンプルの質を考え，サンプルはただ単に多ければ良いわけではない．パターン認識の実際に携わった技術者は，学習に相応しい事例としてのサンプルの大切さを経験的に知っている．このようにパターン認識において価値観は必要不可欠なのである．

本章を終えるにあたり，統計学と統計的パターン認識との関係について私見を述べる．前述したように統計的パターン認識は帰納的推論の欠点を克服するために不確実性の数値化というアプローチを用いた．この数値化という点では両者はまったく同じ立場である．しかし決定的に異なるのは，価値観の扱いである．統計学は数学の一分野に属し，論理的である．原理的に論理（客観）と非論理（主観）は相いれない．統計学が数学の一分野であるためには主観を脱し，客観が絶対条件となる．統計学は，価値観の反映された質を意識して避け，全てのサンプルは質が同じで互いに平等とし，ただ数のみを論じる．一方パターン認識を可能とするために，統計的パターン認識は非論理的要素である人間の価値観を導入し，データやサンプルには優劣があるとする．

本章を総括すると，統計的パターン認識ではクラスは認識主体者の人間によって価値のあると認められた「こと」であり，観測では人間が認識対象の性質を踏まえて獲得されるデータがクラスを形成し得るように観測の候補を用意し，コンピュータがその候補の中から特徴選択により識別に有効な観測を特徴として選択してデータを獲得する．統計学と統計的パターン認識，両者は似て非なるものである．

3. 医学問題への応用

あらためて医学問題を想定して統計的パターン認識を定義すると，

①観測：認識対象（患者）から観測（マーカー）によってデータ（測定値）の獲得

②特徴選択：観測（マーカー）の中から認識に有効な観測（マーカー）を特徴（標的マーカー）として選択（探索）

③識別：特徴（標的マーカー）を用いて認識対象（患者）の識別（診断・予測）

となり，解くべき問題を対比させると

工学問題

（特徴選択の問題，認識対象の識別問題）

は，

医学問題

（標的マーカーの探索問題，患者の診断・予測問題）

に対応する．

解析前に確認することは，クラス，各クラスの利用できるサンプル数，観測（マーカー）の数と特性である．用いるデータの中に識別情報があるとは限らないため，情報の漏れがないようにマーカーを慎重に選定しなければならない．癌のように病態が医学的に未解明の場合，患者から予め何を検査してデータを獲得すれば良いかは不明である．そのため単純に可能な限りマーカーを多くすると，データ数（次元数）は増大し高次元となって「情報の爆発」が生じる．このため膨大な数のマーカーの中から有効なマーカー，すなわち標的マーカー探索が必須となる．

医学固有の問題としては，患者の背景因子を揃える等の条件があって，条件を満たすサンプルは一般に少ない．質の高いサンプルは限られ，質は医師によってのみ医学的に評価される．要するに高次元であり，かつサンプルは少数しかなく貴重である．更に，個人情報保護の問題もある．

3.1 統計的パターン認識による個別化医療

癌の克服は国民的課題である．それが困難な理由は，癌の多様性にある．同じ臓器の癌であっても患者個々によって異なる．そのため個々の患者に応じた最適な医療を行う個別化医療が特に癌治療において進められてきた．

前述したように，癌に関する情報は遺伝子の中にあると考え，発現データやメチル化データ等の様々な遺伝子関連データが試された．その中に遺伝子発現データを用いた統計的パターン認識による肝癌の術後再発予測[10),11)]がある．

もし肝癌が再発しないと予測されると，副作用のある抗癌剤治療は実施せず，また不要なCT等の検査も行わなくてもよく，患者の負担は大幅に低減される．一方再発が予測されると，再発の予兆を的確にとらえて効果的な先制医療を実施できる．このように予後（医学的な今後の見通し）の予測に基づく個別化医療によって，患者はこの上ない恩恵を得ることができ，また医療費も抑制できる．

この問題を医学的に定義すると，「肝癌の手術で癌を完全に切除した．切断面の病理検査でも癌が診られないことを確認して，手術は成功したと考える．手術で切除した癌組織から抽出された7000個の遺伝子に対してマイクロアレイ技術により遺伝子発現データを獲得し，手術後1年以内に早期再発する患者を予測する」となる．工学的には，患者は7000次元のベクトルで表され，学習に用いる訓練サンプルの数は再発・無再発併せて33例と極めて少ない．なお33例は複数の肝癌を専門とする医師により選定され，医学的に質の高いサンプルのみとした．

この研究では2つの工夫を行った．一つ目の工夫は，2段階の特徴選択である．前述したように特徴の組合せを評価すべきであるが，候補数が多いとそれをまともにはできない．また医学的にも肝癌再発に関与する遺伝子はごく一部と考えられる．そこで第1段階において期待できない多くの遺伝子候補を削除し，第2段階で見込みのある遺伝子候補のみを対象にそれらの中で組合せを調べることにした．

二つ目の工夫は，仮想的変動の活用である．訓練サンプル33例の中から仮想訓練サンプルをランダムに抽出（再標本化）し，仮想訓練サンプル上で特徴となれる標的遺伝子の候補を選択する．これを独立に繰り返して意図的に複数の仮想サンプル変動を起こし，それらの中で最頻出する，つまり最も多くの仮想サンプル変動に対して有効な標的遺伝子の候補を標的マーカーとして選択する．

その結果，7000個の遺伝子候補の中から12個の標的遺伝子を選択し，訓練サンプルとは独立の27例のテストサンプルに対して93%の精度で患者の早期再発を予測できた．この精度は，肝癌に熟練した専門医であっても70%程度の予測であることを考えれば，卓越したものと言える．

特筆すべきは，12標的遺伝子に対し医学的意味付けができたことであり，単にデータ駆動型で探索しただけではなく，結果が医学的に評価されている点にある．これにより，本成果は臨床系では世界的な論文誌であるLancetに採択された[10)]．

このように統計的パターン認識を用いれば，高次元と少数サンプルが困難とする個別化医療の実現に貢献できると思われた．しかし医学固有の数値

データと記号データの混在という大きな壁が立ちはだかった．医学データには，各種の陽性 / 陰性，性差，既往歴の有 / 無，遺伝子変異等の記号データがあり，それらは医学上無視できないマーカーなのである．しかし，統計的パターン認識は数値データを対象とし，記号データは扱えない．

3.2 離散 Bayes 識別則による個別化医療

数値データと記号データが混在した場合，多くは記号データを 1，0 の数値データに置き換えて解析を行う．しかし例えば平均や分散をとってもそれには何も意味がなく，人間は解釈できない．これは，本質的な問題である．前述したようにコンピュータの結果は原理的に誤りを避けられないから，特に医療においては患者の生命が係るため，医師によって結果の妥当性が是非とも評価されなければならない．

そこで逆に数値データを記号データ化するアプローチを採用した．検査で得られた数値データ自身に意味はない．数値データは，しきい値処理により陽性 / 陰性等の記号化がなされて医学的観点から意味が与えられる．数値データの記号化とは医学的な解釈づけ（概念化）であり，このような記号データを用いてコンピュータは診断・予測を行う．コンピュータの結果は，医学的意味のある記号の組合せであり，医師が解釈できる．この点が深層学習とは異なる．これを可能とするために，数値データを用いて認識対象の事後確率に基づく Bayes 識別則を記号データが扱えるように修正する．それが離散 Bayes 識別則 [12] である．続いて発表された 3 論文 [13),14),15)] は，いずれも離散 Bayes 識別則の臨床応用であり，結果に医学的意味があって，臨床系論文誌に掲載された．

3.2.1 肝癌の術後早期再発予測

上記論文 [12] は，前述の Lancet 論文と同様に肝癌の早期再発予測問題を論じた．しかし両者の相違は，Lancet 論文が保険のきかない遺伝子データを用いているのに対し，臨床応用を意識して，こちらは保険適用されている，通常診療の臨床データを用いている点で，より実際的である．遺伝子データの結果は，研究としては興味あるが，直ぐに診療には使えない．

肝癌を専門とする複数の医師により選定された 11 候補マーカーの中から 5 つの標的マーカー腫瘍数，腫瘍サイズ，ICG（いずれも数値データ），脈管侵襲，Liver damage（いずれも記号データ）を選択した．これらの標的マーカーは，専門医によって医学的妥当性が示されている．テストサンプルに対して感度 86%（癌の再発検出率），特異度 49% を得た．予後予測によく知られているステージを示す TM 分類や Modi fied JIS よりも高感度で，ROC 解析でも有効性を示した．

3.2.2 早期胃癌のリンパ節転移診断

今では早期胃癌は内視鏡で治療できるが，リンパ節転移が疑われると外科手術が必要になり，患者の負担は大きい．しかし実際にはリンパ節に転移がなくても手術が行われることがある．そこでリンパ節転移の診断を内視鏡治療直後に行えることが要望されている．胃癌を専門とする複数の医師によって選定した 8 マーカー候補の中から 3 つの標的マーカー深達度，リンパ管侵襲，静脈侵襲（全て記号データ）を選択し，テストサンプルに対し感度 100%（転移の見落としゼロ），特異度 86% を得た [13]．

3.2.3 進行大腸癌の再発予測

一般に早期癌の予後は良好で，進行癌の予後は不良である．しかしたとえ進行癌であっても予後良好の場合もあり，そのときは抗癌剤の投与など患者負担の大きい治療が必要とは限らない．

本研究では手術において治癒的切除の進行大腸癌の患者に対して予後良好な患者の識別を問題とする．大腸癌を専門とする複数の医師により選定された 10 マーカー候補の中から 3 つの標的マーカー CD4, FOXP3, Histologic grade を選択し，テストサンプルに対して感度 71%，特異度 67% の精度で進行大腸癌の術後再発を予測できた [14]．

3.2.4 抗うつ薬の投与効果の予測

これまでは癌治療への臨床応用であったが，本研究はうつ病の薬剤効果を予測する問題を扱う．薬剤の効果は個々の患者によって異なり，誰でも同じ効果があるとは限らない．一方で，副作用はある．副作用はないことが望まれるが，たとえあってもそれが軽微であれば薬効のある患者への抗癌剤治療は患者にとって有益である．

臨床研究として，うつ病薬剤反応性に関する網羅的遺伝子解析を行った．遺伝子発現データを用いて薬剤反応性の有無を識別するのである．なお，その信頼性には検体取り扱い等の施設側の問題とうつ病

に特有な正解クラス名の付与問題があり，一般に多施設間のうつ病遺伝子発現データ解析は困難とされている．本研究は多施設共同研究として広島大学から提供の訓練サンプルを用いて標的マーカー探索と識別器学習を行い，山口大学と徳島大学から提供のテストサンプルに対し，3つの遺伝子を標的マーカーとして感度91%，特異度75%の性能を達成した[15]．

3.3 今後の展望

新薬を開発するためには，安全性と有効性の両方を評価する治験が行われる．しかし，開発には多額の費用と長期間を要するにもかかわらず，治験の成功確率は極めて低い．そのため製薬会社にとって治験の成功確率を高めることは極めて重要となる．

治験は，開発薬剤を投与した治療群と偽薬を投与した対照群を対比させるランダム化比較試験（RCT：Randomized Control Trial）によって行われる．公平性を確保するために治験では群間でランダム化が必須となる．確かにランダム化によって患者の背景因子等を揃えられるが，それはサンプル（患者）が十分にある場合に限られる．実際にはサンプル数は少なく，統計学の要請には応えられていない．更に従来のRCTでは薬剤の効果があるとすればそれは誰にでも効くという前提であるが，患者によっては薬剤効果が異なるのが現実である．しかし，これも考慮されていない．仮に薬剤効果のない患者が治療群へランダムに混入されてしまうと，十分な薬剤効果が認められずに，治験は先に進めなく頓挫する．

これを回避するため，薬剤効果の期待される患者のみを対象とする患者層別化が治験の成功の鍵となる．患者層別化とは効く患者と効かない患者の識別であり，これは正にパターン認識の問題である．この問題にAI技術への期待が高まった．しかし，深層学習等のAI技術は主に画像診断に用いられ，主要な臨床データが非画像データの治験ではAI技術の活用は期待ほど進んでいない．また治験ではセンシング技術の発展により次世代シーケンサーからは医学上重要な遺伝子変異（膨大な規模の記号データ），その他にタンパク質等の各種オミックス情報のデータ等，多種多様な数値データと記号データが混在して用いられている．このため医療の現場からは，治験に特化した新しいAI技術が切望されている．この要請に応えるべく，治験に離散Bayes識別則を用いる．

ここで離散Bayes識別則の利点について整理する．

①数値データと記号データの混在への対処
数値データを記号データ化し，全て記号データ上で標的マーカー探索と識別を行う．

②データの「記号（こと）」化
医学的に意味のある記号（こと）に変換した記号データの組合せがコンピュータの結果となり，この組合せは医学知識として医師によって解釈できる．特に遺伝子変異は最も期待される記号データで，遺伝子変異の組合せを識別と関連づけて直接解析できる方法はこれまでに提案されていない．

③「次元の呪い問題」の回避
全て単一のマーカー毎の計算（スカラー計算）となり，次元の影響を受けない．そのため，特に高速計算機を用意する必要もなく，現実的な時間で処理できる．

④個人情報取り扱いの容易化
数値データは適切な離散化により復元不可となり，記号データも適当に置換すれば解釈不可となる．更に離散Bayes識別則では，データ自身は用いず，データを集計した統計情報を用いる．統計情報が個人情報ではないことは，留意すべきである．

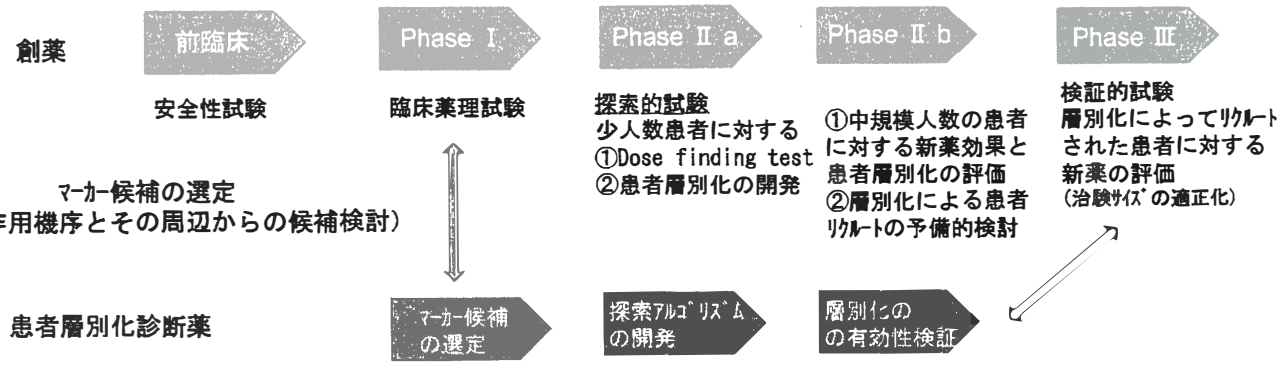

図1　治験の全体の流れ

図1に治験の第1相から第3相を示す．薬剤効果を調べる第3相の成否に直結するのが第2層である．第2層は，解析の専門家が手薄の探索的試験である．

この第2相で患者層別化を行い，効果の期待される患者のみをRCTの対象とし，第3相でそれを治療群と対照群にランダムに分けて薬剤効果を調べる．

第2相の詳細を図2に示す．まずPhase Ⅱ aとして小規模の患者群を収集し，標的マーカーの探索と標的マーカーを用いた離散Bayes識別則による患者層別化を行う．この解析はデータ駆動型であり，データとの会話を通して解に接近することになる．試行錯誤を経て医学的に見込みのあるマーカー群が得られたら，Phase Ⅱ bとして中規模の患者群に対する患者層別化を試行し，第3相へ進むべきか否かを評価する．このRCTでは，患者層別化により効果の期待される患者群のみを対象とし，効果の期待されない患者が試験に混入することを極力回避する．

患者層別化により治験サイズが適正化され，それによって成功確率が高まり，開発費の低減と開発期間の短縮が期待できる．また同時開発した患者層別化技術は，血液検査として薬剤投与の是非を判断する対外診断薬にも活用できる．これにより，薬剤の不要な投与が避けられ，医療費の抑制につながる．

4. おわりに

本稿では，AI技術の一つである統計的パターン認識を概説し，その個別化医療への応用として離散Bayes識別則までを紹介した．

最後に，AIによる診療について私見を述べる．AIつまりコンピュータは，診療においては医師が主であって，支援ツールに過ぎない．コンピュータによる支援とは，医師が気づかない点を指摘しヒントを与え，医師の臨床推論力や発想力を高めることである．この意味でコンピュータの支援は有益である．

ところで昨今，様々な解析用のソフトウエアが頻繁に利用されているが，そのソフトウエアがどのようなアルゴリズムによって構築されているのか，そのアルゴリズムが何を前提としているのか等に注意を向けずに，コンピュータによる結果を無批判に受け入れることは危険である．安易なソフトウエアの利用は，誤った解析の一因ともなる．医師は，原理的に誤りをゼロにすることができないコンピュータの結果を鵜呑みにせず，結果を必ず医学的に解釈すべきである．このために，研究デザインの段階から臨床目的をもつ医師とアルゴリズムを専門とするデータサイエンティストは一緒になって密にコミュニケーションを取り，相補的な解析を共に進めれば，この危険性を少しでも回避できるのではと思う．

また実際の経験が望まれるデータサイエンスやAIの大学教育についても，実務経験のない，あっても乏しい教員による理論，あるいはソフトウエアの操作技法のいずれかに偏った教育によって実践的人材の輩出が可能かと言えば無理がある．理想的には軽視され易い理論誕生までの背景や思想，その適用限界までを理解すべきである．このためには，手

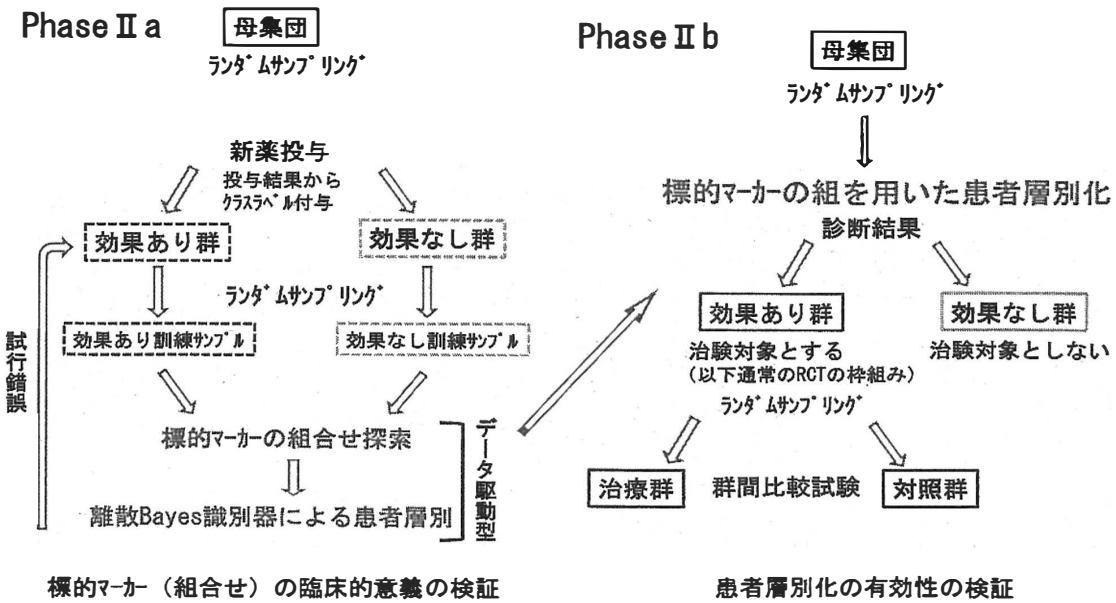

図2　第2層の標的マーカー探索と患者層別化

軽なノウハウ本やパワーポイントの講義ではなく，世界的な名著を繰り返し精読して，ときには経験に照らして理論を理解し直すことが必要である．このように実践的なパターン認識の習得には机上の勉強だけでは足りず，経験することでしか理論の正しい理解は深められないのである．

私自身，若いときには当たり前と思って軽視していたことが，数十年の時を経て，そんな意味であったのかと納得したことがある．離散 Bayes 識別則はこのような中で誕生した．パターン認識の理論と実際とは，車の両輪のごとく不可分の関係にある．パターン認識は奥が深い．

参考文献

1） 渡辺　慧：認識とパタン，岩波新書（1978）
2） C.R. Rao（藤越，柳井，田栗共訳）：統計学とは何か，丸善（1986）
3） 浜本義彦：統計的パターン認識入門，森北出版（2009）
4） 浜本義彦：パターン認識理論の最近の動向，電子情報通信学会誌，Vol.77, No.8, pp.853-864 (1994)
5） 浜本義彦：統計的パターン認識：過去・現在・未来，電子情報通信学会，信学技報，PRMU 2000-129（2000）
6） 上坂吉則：パターン認識と学習の理論，総合図書（1971）
7） A.K. Jain, R.W. Duin and J. Mao: Statistical Pattern Recognition : A Review, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.22, No.1, pp.4-37 (2000)
8） J.D. Elashoff, R.M. Elashoff and G.E. Goldman : On the Choice of Variables in Classification Problems with Dichotomous Variables, Biometrika, Vol.54, pp.668-670 (1967)
9） T.M. Cover and J.M. Van Campenhout: On the Possible Orderings in the Measurement Selection Problem, IEEE Trans. System, Man, and Cybernetics, Vol.7, No.9, pp.657-661 (1977)
10） N. Iizuka, M. Oka and Y. Hamamoto, et al.: Oligonucleotide Microarray for Prediction of Early Intrahepatic Recurrence of Hepatocellular Carcinoma after Curative Resection, Lancet, 361, pp.923-929 (2003)
11） 浜本義彦：統計的パターン認識による肝癌再発予測，電子情報通信学会，第 28 回信号処理シンポジウム（下関）（2013）
12） H. Ogihara N. Iizuka and Y. Hamamoto : Prediction of Early Recurrence of Liver Cancer by a Novel Discrete Bayes Decision Rule for Personalized Medicine, BioMed Research International, Vol.2016, (2016) doi.org/10.1155/2016/8567479.
13） A. Goto, J. Nishikawa, H. Ogihara and Y. Hamamoto, et al.: Lymph Node Metastasis can be Determined by just Tumor Depth and Lymphovascular Invasion in Early Gastric Cancer Patients after Endoscopic Submucosal Dissection : European Journal of Gastroenterology and Hepatology, Vol.29, No.12, pp.1346-1350 (2017) doi:10.1097/MEG.0000000000000987.
14） Y. Nakagami, S. Hazama, H. Ogihara and Y. Hamamoto, et al.: CD4 and FOXP3 as Predictive Markers for the Recurrence of T3/T4a Stage Ⅱ Colorectal Cancer : Applying a Novel Discrete Bayes Decision Rule, BMC Cancer, 22:1071 (2022) doi.org/10.1186/s12885-022-10181-7.
15） H. Yamagata, H. Ogihara and Y. Hamamoto et al.: Interferon Signaling and Hypercytokinemia-Related Gene Expression in the Blood of Antidepressant Non-Responders, Heliyon (2023) doi.:Heliyon, 9(1), art. no. e13059.

（はまもと　よしひこ／山口大学）

浜本　義彦

1983 年 3 月山口大学大学院修士課程修了．1983 年 4 月日本電気㈱入社，1986 年 12 月同社退職，1987 年 1 月山口大学工学部助手を経て，1998 年 4 月同大教授．2006 年 4 月から 2016 年 3 月まで同大医学系研究科教授，2016 年 4 月から同大創成科学研究科教授，2023 年 3 月定年退職．2023 年 4 月から山口大学名誉教授．統計的パターン認識の基礎と応用の研究に従事．博士（工学）．日本癌学会会員．