

医学における知識発見手法の可能性

津本 周作 (島根医科大学医療情報学講座)

tsumoto@computer.org

はじめに

計算機の能力の向上とデータベースソフトウェアの性能の向上により、遺伝子データベース、癌プロトコルデータベースをはじめとした医学データベースのみならず、検査データベース、検診データベースを含めた診療データベースも膨大な量のデータが蓄積されるようになってきた。このような膨大な量のデータの解析はすでに人の処理能力をはるかに越えるものとなっており、計算機による有効な使用方法の確立が医学・医療の分野においても急務とされている。医学においてはデータ解析は統計学的手法を用いることが伝統的であったが、本稿では医学における発見のプロセスが必ずしも頑健性に裏付けられていないことを示し、これまでのデータマイニングコンテストを開催してきた経験から、開発されてきた知識発見手法がいかに医学の発展に貢献し得るかについて論じる。

医学におけるデータ蓄積の現状

病院の情報を電子化するという試みは医事会計電算システムから始まった。1980年代には病院における伝票処理を自動化するオーダリングシステムの運用が開始され、これに付随して検査データをはじめとした診療情報のほとんどが電子化されるようになった。1990年代には画像データを電子化する仕組みであるPACS、そしてさらに診療録を電子化する試みが始まっており、今世紀前半にはすべての診療情報が電子化される時代が到来するといわれている¹⁾。病院情報システムによるデータの蓄積は1980年代以来、加速度的に発展している。1990年代前半には外来患者約1,000人、病床800床程度の大学病院クラスの病院で検査データだけでも、1年で数GB程度のデータが蓄積されていたが、ごく最近では、さまざまな診療データを含めて、1年で数十GBのデータが蓄積されるようになってきている。近い将来、遺伝子診断・

遺伝子治療が実施されるようになれば、これらのデータを含め、TB以上のデータベースが病院情報システムでサポートされなければならないといわれている。

驚くべきことに、現在までデータの蓄積の重要性が指摘されてきたにもかかわらず、蓄積されたデータを再利用し、診療支援に役立てることの重要性はほとんど指摘されていない。一方、ゲノムデータベースに関しては、膨大な遺伝子情報がデータベース化されつつあるが、早くからそのデータベースの再利用が叫ばれ、Bioinformaticsの中核として、さまざまな手法が提案されてきている²⁾。医療情報システムにおけるデータ解析の研究はまだ幼年期にも達していない。

以上のように、病院情報システム・ゲノムデータベースをはじめとして、医学・医療においても膨大なデータが蓄積されつつあり、今後蓄積されたデータの再利用が重要な課題となってきている。以下においては、病院情報システムからのデータマイニングの可能性について、著者が進めてきたデータマイニングコンテストの経験および医学における発見の例から論じていくことにする。

医学・医療データの特徴

病院情報システムのデータは属性数・標本数とも膨大であるが、欠損値が著しく多い時系列データであり、以下の特徴を持っている。

(1) 属性数がきわめて大きい：

1人のレコードあたりのデータはきわめて多くなるため、ある1属性の時間的推移を追跡することさえも困難であり、長期間にわたる各変量間の依存関係等を含む知識を抽出することは、かなり簡単なパターンでなければほとんど不可能である。

(2) それぞれの区間が不規則である：

ほとんどの外来患者は2週間～1か月ごとに来院するが、来院時に検査を行うわけではない。なかには1～3か月に一度の割合で定期的に検査を行う場合もあるが、時に偶発的な要因で容易に検査間の区間が不規則に並ぶ

ことが多い。慢性疾患の増悪による入院などでは、入院時の短期間に集中的な検査が行われる可能性もあるが、通常、容態に変化がなければ、その測定間の区間はきわめて不規則になる。以上のごとく、データ収集の方法がきわめてretrospectiveであり、科学的な目的で収集されたprospectiveなデータと比較して、次章で示すように解析が困難である。

(3) 欠損値が多い：

各測定日におけるデータは時間的に不規則な性質を持っているのみならず、各点における測定項目も不規則である可能性が高い。たとえば、糖尿病の経過観察においても、血糖 (BS)、尿糖 (US)、ヘモグロビンA1c等の項目を毎回検査しているとはかぎらない。

実際、専門家が上記のデータを扱う際には、厳密な意味で時系列分析を行わず、臨床上、重要である属性を選択して、データベースを作成し直すのが通例である。筆者が今まで「共通データからの知識発見の比較と評価」をはじめとしたデータマイニングコンテストで使用した髄膜炎データと細菌培養検査データは病院情報システムのデータから属性をあらかじめ選択し直して、抽出したデータである³⁾。したがって、この2つのデータ集合では、すでに時系列に関する情報はほとんど失われている。

では、解析困難な時系列データから、専門家はどのような知識の獲得を望んでいるのだろうか？ 主として、以下のようなものが挙げられる。

(1) 疾患の長期予後：

糖尿病では、治療のコントロールが悪いと、数年以後、末梢神経障害、白内障をはじめとした種々の合併症が現れてくる。ただし、どの時期に糖尿病の合併症が現れてくるかは明らかではない。生活習慣病の長期予後を推測できる知識をデータから発見できれば、生活習慣病の治療指針が決定できる。

(2) 薬剤投与における長期効果：

慢性関節リウマチを含めた自己免疫疾患、アレルギー疾患では、ステロイド剤が著効を示すが、本薬剤は長期投与によってさまざまな副作用を呈し、治療開始後、どの時点で合併症が現れてくるか明らかでない。もし、ステロイド剤による副作用が出やすい条件、また時期としていつごろが多いかが明らかになれば、ステロイド剤投与の治療計画にきわめて重要な指針を与えてくれる。

専門家でも、数年以上の経過で、さまざまな要因が重なって引き起こされる疾患、症状に関しての予測はきわめて困難である。長期予後、長期効果がデータの裏付け

を持った形で知識として抽出されれば、医学・医療に対するインパクトは非常に大きいものになると考えられる。ただし、残念ながら、これまで開催した膠原病データベースに関するコンテストでは期待された(1)のような時系列パターンに関する知識は得られなかった。

知識発見手法の可能性：データ収集の立場からみた医療データの制約

筆者はこれまでのデータマイニングコンテストにおいて、データマイニングの手法を仮説生成として利用すべきであることを主張してきた³⁾。これは疫学・公衆衛生で研究されてきたデータ収集方法に関する知見が医学において重要視されてきたからであり、以下にその要点を論じる。

病院情報システムから抽出された医療データにかぎらず、データマイニングで使用されている大半のデータは一般に、ある仮説に基づいて収集されたデータではない。データマイニングはとりあえず収集されたデータから有効な情報を抽出しようというのがその目的であると論じられてきた。したがって、自ずとそのデータの説明力には限界があることは知っておく必要がある。このようなデータの性質の違いは、疫学において論じられており⁴⁾、ある仮説に基づくデータ収集を元にしたデータ解析はprospective study (前向き研究)、とりあえず集められたデータを解析することをretrospective study (後向き研究)と呼ばれてきた。

Prospective studyは疫学ではcohort studyとも呼ばれ、観察開始時点(これはデータ収集開始時点に相当する)に対象の集団をいくつかの群に分類する(要因(+), (-))。その後、疾病の発生や疾病による死亡等を観察していくという方法であり、要因の有無を固定した上で、それによる結果の有無を観測する方法である。この情報収集の流れは、原因→結果の因果律に沿っており、前向き研究(prospective study)と呼ばれている。

他方、Retrospective studyはcase-control studyとも呼ばれ、疾病の発生した人を症例(case: positive examplesに相当する)に、症例に該当しない人を対照(control: negative examplesに相当する)として、両群の要因の有無について比較する方法である。この情報収集の流れは、結果→原因で因果律の逆向きとなり、後向き研究(retrospective study)と呼ばれている。

これら2つの方法の長所と欠点に関して、表-1に示した。現在データマイニングで用いられているデータベースはほとんどretrospective studyに相当し、一般にprospectiveなデータ収集によらないため、収集時におけ

	Prospective Study	Retrospective Study
別称	Cohort Study	Case-Control Study
データ収集の流れ	要因→結果	結果→要因
推定能力	さまざまな指標を正確に推定できる	正確さはデータからは推測することが困難
要因の解析能力	まれな要因でも分析可能	まれな要因についての分析は困難
解析の多重性	1つの要因に関して複数のTargetを分析可能	1つのTargetに関して複数の要因を分析可能
ノイズの影響	少ない	大きい
相対リスク	発生率の高いTargetの相対リスクの計算も可能	相対リスクの計算は不可能
調査コスト	調査対象の選別等, コストが高い	コストは低い
まれな疾患 (Target)	まれなTargetの調査は困難	まれなTargetの調査も可能
調査期間	Targetの発生の有無の確認が必要	調査期間は短くてよい
追跡調査 (Follow-Up)	追跡不能例の発生 (positive/negativeが明確でないことあり)	Targetのpositive/negativeはデータ収集時に確定

表1 Prospective Study (前向き研究) と Retrospective (後向き研究) の比較

るノイズとbiasの問題が発生する。仮説検定で最終的に信頼の高い結果を得るには、やはり prospective study の方が適している。両者の特徴から考えれば、収集されたデータから retrospective analysis で仮説を生成し、その仮説に基づき、prospective analysis, 特に統計的解析を試行するのが、データマイニングの手法を科学的な分野に適用する最良の形式であろう。有用な仮説を生成するためには、領域特有な知識による検証が不可欠である。たとえば、生成された規則に関して検証し、得られた結果が実際に適切な結果であるのかどうか、適切でなければ、その背後に何か知識が隠れているのか、隠れているとすれば、それが新たな発見につながるか、といった専門家を交えた分析が必要であり、後処理がきわめて重要なステップとなる。

著者のデータマイニングコンテストにおける経験からいえば、膠原病データベースの場合を除いて、これまで国内外の研究者が開発した手法は仮説生成の方法としては十分利用できるものであるという感触を得た^{3), 5)}。特に、鈴木によって開発された例外ルールの発見は仮説を構築する上で特に有用であった。ただし、上で論じたように、科学における発見では仮説の実証も非常に重要なプロセスであり、仮説生成は、発見プロセスのうち、初期の段階を支援するものであると考えられる。次章に示すように、医学における発見は他の分野同様、発見されたものが意外であればあるほど、それを実証するまでにきわめて長い時間が要求されることにも注意したい。データマイニングにおける医学の発見支援はまだ始まったばかりである。

医学における発見のパターン

医学部出身者はこれまでの発見の重要性とその難し

さについて徹底的に教育される。実際に、基礎・臨床医学の研究の場に立てば、発見の困難さは肌身に感じられ、著名な発見のプロセスが数多くの失敗や常識では考えられなかった事実裏打ちされていることに思いを新たにする。本章では、医学では半ば常識となっているいくつかの重要な発見のプロセスについて Scientific Discovery の立場から論じる。Scientific Discovery は発見プロセスを人工知能の立場から研究した Simon およびその弟子である Pat Langley や Jan Zytkow らによって提唱された分野⁶⁾で、彼らがデータマイニングの研究者の中心をなしてきたことを付記しておく。

抗生物質 (ペニシリン) の発見 - 失敗からの学習

19世紀～20世紀、医学の発展は細菌学の発展によるものであったといっても過言ではない⁷⁾。それまで、炎症性疾患とされてきたものの多くが細菌によるものであることが明らかとなり、その細菌の性質を研究することが、後に衛生環境の重要性を明確にさせることになった。しかしながら、細菌感染症の治療にはワクチンをはじめとした抗血清療法によるものがほとんどであり、細菌を死滅させる化学物質を発見あるいは人工的に生成するには至らなかった。ところが、抗生物質は以下のごとく、ある種の実験の失敗から偶然発見された。

1928年、フレミングはブドウ球菌を使った培養実験を行っていたが、培養皿の1つがアオカビが生えており、実験が失敗していることに気がついた。カビのせいで無駄になった培養皿を捨てるときに、フレミングはカビの周辺でブドウ球菌のコロニー (集団) が解けているのに気づいた。「ある生体反応には、必ずそれを引き起こす物質が存在する」という生物学の黄金則にしたがって、フレミングはアオカビ penicillium を培養し、培養液中に抗菌作用を示す物質があることを発見、その物質をペ

ニシリンと命名した。しかし、当時、この物質を精製できる化学者の協力がなかったため、この発見から抗生物質の研究は進まなかった。1940年にイギリスの病理学者フローリーと生化学者チェーンらがようやく、この発見の重要性に気づき、ペニシリンを精製・分離することに成功した。このペニシリンを実際の細菌に適用することで、細菌の増殖が抑制することが証明された。これがペニシリンの再発見であり、この後フローリーがペニシリンの大量生産を可能にするためにアメリカに渡り、ペニシリンのアオカビの培養による大量生産が可能になった(実用化)。

この話は医学の発見のみならず、発見というものの性質を語るにきわめて示唆的である。失敗からの発見、注目されない時期、12年後の再発見・実証。例外的な事実のみから発見されたのではなく、その例外的事実生物学の第一原理を忠実に適用することで生まれた発見は、知識を発見するという立場からも興味深い。

蛇足ながら、抗生物質は当時人工的に合成できず、カビというカビを探し、そこから抗生物質を抽出する努力が数十年続いた。筆者が大阪大学在学中に細菌学の講義で、新しい抗生物質を発掘するために、ヒマラヤ登山家のような生活を送った人が存在したという逸話を聞き、「もう、自然界から天然の抗生物質を発掘する作業はやり尽くされた。これからは人工的に合成する抗生物質の時代だ」と言われたことを印象深く記憶している。

以上を人工知能的な言葉でまとめれば、「失敗→失敗に対する観察→得られた所見に対する第一原理の適用→仮説の生成→仮説の検証」というプロセスがこの発見に重要で、特に上記の逸話を聞けば聞くほど、「失敗に対する観察→得られた所見に対する第一原理の適用」がきわめて重要であることを思い知らされるのである。

ピロリ菌の発見—例外的事象の発見

19世紀はすべての疾患は細菌によって起こると信じられてきた。その後、黄熱病をはじめとした細菌が検出できない疾患では、ウイルスが病原体であることが判明し、必ずしも細菌がすべての病原ではないことが明らかとなった。1940年～1980年代における医学の常識は「細菌・ウイルスが病因でない疾患も多く、細菌・ウイルスが検出されないような病理組織を持った疾患は感染症でない」ということに集約される。

胃潰瘍・胃癌もそのような疾患の代表であった。特に、胃内では強い酸が分泌され細菌にとっては厳しい環境であり、医学界においても長い間、胃内に常在する細菌は存在しないとされてきた。ところが、1979年にオーストラリアのRoyal Perth Hospitalの医師Warrenが偶然

内視鏡下である患者の胃粘膜生検の標本から細菌を発見、1983年に同WarrenとMarshallによって、Helicobacter Pylori (以下ピロリ菌)が胃内に常在されることが発見された。

面白い話が文献8)に記載されている。最初、WarrenとMarshallがある国際会議でピロリ菌が慢性胃炎の危険因子であるという仮説を報告したところ、馬鹿馬鹿しいと拒絶された。そこで、Marshallらは動物実験を施行したが成功しなかったため、Marshall自身が実際に培養したピロリ菌を自分で飲み、実際に内視鏡下で、自らが慢性胃炎に罹患、胃生検標本でピロリ菌の棲息を実証したそうである。まさにprospective studyである。

この菌は、胃内の強酸から身を守るため、胃内に湧出している尿素をアンモニアと二酸化炭素に分解し、周囲をアルカリ化することで胃粘膜上に棲息する。発見当初は注目されていなかったが、その後の研究で、消化性潰瘍患者の胃内に高率に(日本では90%以上)発見され、さらにこの菌を除菌することにより、潰瘍再発率が激減することが証明された。1994年にはWHOからピロリ菌は胃癌の危険因子として承認されている(ただし、ピロリ菌を保菌しているからといって、胃潰瘍・胃癌を発症するわけではない。つまり、必要条件に近いが、十分条件ではない)。

蛇足ながら、疫学上では、ピロリ菌は開発途上国で感染率が高く、これらの国では幼児期より感染率が高く成人では80～90%まで達する(衛生環境の悪い生水にピロリ菌が棲息するといわれており、そのような生水を摂取することで感染する)。一方、先進国では幼児期では低く、徐々に上昇し40～50歳で40～50%程度であり、我が国では幼児期では低いが、40歳以上になると約80%の人に感染を認める。これは第二次世界大戦後衛生環境の整備や生活の欧米化に伴い感染の機会が減ったことを反映しているが、それにつれ胃癌の発症率が低くなってきていることも、注目したい。

胃潰瘍・胃癌は1990年代までは感染症由来ではないと考えられてきたが、1983年に発見されて以来、1990年半ばには常在菌がきわめて重要な因子であることが確認されたわけである。ここでも、前節のプロセスに類似のプロセスが働いていることに注目したい。つまり、「意外な事実の発見→意外性に対する観察→得られた所見に対する第一原理の適用→仮説の生成→仮説の検証」というプロセスである。ここでも、「意外性に対する観察→得られた所見に対する第一原理の適用」がきわめて重要である。

平山病—症例データからの発見

著者は研修医時代に在籍していた千葉大学医学部附属病院神経内科の教授は運動ニューロン疾患が専門であり、その中で特異な疾患単位を発見した¹²⁾。この発見の形式は他の2つの発見のパターンとは異なり、最もデータマイニングから生まれる可能性の高いものである。運動ニューロン疾患とは通常、約数十年の経過で運動神経が変性するため、歩行困難、筋力低下の症状が徐々に進行し、最終的には寝たきり、あるいは筋力低下が呼吸筋にまで及んで呼吸困難で死亡するという予後の悪い疾患である。ところが、平山らはこのような運動ニューロン疾患の中で、途中で症状の進行が停止する症例があることを見出し、以下のような新たな疾患概念を確立した。

- (1) 通常、運動ニューロン疾患は40代以上で発症するが、当該疾患では、20代以下で発症する。
- (2) 運動ニューロン疾患では、数十年の経過で慢性的に筋力低下、筋萎縮が進行するが、当該疾患では、十年程度まで筋萎縮の進行が止まる。
- (3) 通常、運動ニューロン疾患では、両側性に筋萎縮が観察されるが、当該疾患では、片側性に症状が進行する。

これらの特徴から、平山らは若年性一側性筋萎縮症という疾患概念を提唱した。実際に、平山はこれらの症例データを実際の運動ニューロン疾患の症例データと比較し、その差異をデータ集合の裏付けから示した。発見当時から現在に至るまで、この疾患概念には異論も唱えられているが、実際の剖検例でも、平山病の病理像が運動ニューロン疾患のそれにきわめて類似していることが指摘されていること、実際に症例が蓄積されていることから、現在では、数十年の経過を経て、平山病の概念は定着してきている。

上記の発見過程を見て、お分かりのように、このパターンでは、データの差異から疾患概念を確立したという点で、手動ではあるが、データ集合からクラスタリング、ルール生成の手法を利用すれば、上記の発見過程が再現されるであろうことは予想される。医学においていまだよく分かっていない病態が存在することを考慮すれば、平山病のパターンはデータマイニングの手法が発見に直接つながる可能性を示唆している。

以上、3つのケースについて示したが、医学における発見では、このような例外的(非典型的)だと考えられた事象からそれが単なる例外ではなく、それらの症例と典型例との比較によって、非常に重要な示唆が与えられることが多い。また、これら3つのケースを含めて、医学上の発見は実際の発見、仮説の生成から、その仮説が実証・医学界で受け入れられるまでに、約10年以上の年

月を要しているのに注目したい。これらは極端な例かもしれないが、新奇性の発見が認められるには、それなりの年月が必要であることには注意しなければならない。上記のケースはデータを用いての発見ではないが、データベースから意外な知識を発見した場合にも、それが実証されるには、それなりの年月を要することは十分考えられる。

知識発見研究の展望

これまでの議論とデータマイニングコンテストから得た知見をカップリングすれば、コンテストの現状は以下のごとくまとめることができる。

(1) 欠損値、ノイズの少ない非時系列データに関しては、きわめて簡単で理解しやすいパターンを生成できる。属性が多く、専門家がすべての属性の性質を把握できていない場合、属性の組合せから意外な知識を発見する機会が得られる。

(2) 得られたパターンに関しては専門家の解釈のプロセスがなければ、仮説となり得るかどうかの判定は困難である。分類精度(正答率)を基準にパターンを生成すると、一般性の高いルールは得られるが、ごく当たり前の知識に対応したルールしか得られず、専門家の知識に一致しない属性選択が起こらないかぎり、発見に結びつく可能性は低い。

(3) 時系列データに関しては、ルール生成をはじめとしてパターン生成の方法では、きわめて理解しにくいパターンを生成するため、ほとんど仮説の生成にまで結びつかない。これは時系列にノイズが多く、ノイズに対応できなければならないこと、時系列のパターンが通常ルールなどでサポートできるパターンとは異なる形で表現されなければならないことを示唆している。

以上の指摘がパターン生成に関するものばかりであることは、コンテストが基本的にパターン生成指向であることに由来している。KohaviらがKDD-Cup2000の論文で、コンテスト参加者はパターンの生成・性能には興味はあるが、そこから発展する「洞察」問題に対する報告はきわめて少なかったと論じていることはきわめて示唆的である⁹⁾。データベースからの知識発見(KDD)の研究者であるFayyadはKDD Processを提唱した¹⁰⁾にもかかわらず、KDD-Cupで行われたコンテストはあくまでも生成されたパターンの予測精度の比較が中心であった。これはFayyadの図においてもパターン生成(データマイニング)が図上の高い位置になっていることから推察できる(図-1)。Fayyadによれば、データマイニングのプロセスがデータの獲得、データの選択、データの

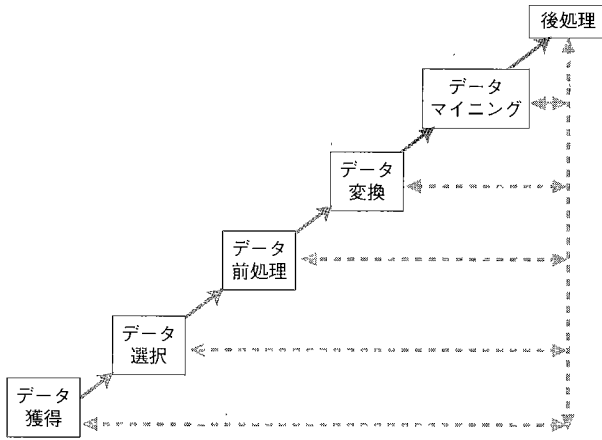


図-1 Fayyadによって提唱されたKDDプロセスの略図

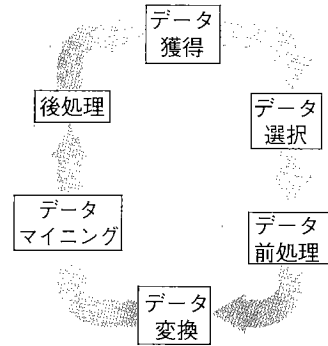


図-2 実際のKDDプロセスの流れ

前処理, データ変換, データマイニング, 後処理から構成され, 相互的に移り得る包括的なプロセスであると定義しているが, 提唱した図を見るかぎり, Mainstreamは右肩上がりのプロセスであるような誤解の有無を生じると思われる。Fayyadはこの図をしばしば改訂しているが, 基本的な構造は変化していない¹¹⁾。

データマイニングコンテストにおける限界の1つはデータマイニングのプロセスにおいて, 解析した結果結果に基づいてさらにデータを収集して, データマイニングを継続するというプロセスを実現できていないことにある。実際にはデータ解析はFayyadが提唱した形式ではなく, むしろ, サイクル上に動くプロセスであるとした方が実世界のデータ解析の性質を反映しているように思われる(図-2)。

医学の専門家の立場からみれば著者は次の3つの点がデータマイニングを知識発見のツールとして利用するためにはきわめて重要であると考えている: 第1に, 時系列データから自動的に理解しやすいパターンを生成できること。第2に, 時系列, 非時系列を問わず, 得られたパターンの解釈(後処理)を知識工学的なアプローチで支援してくれること。第3に, 仮説の生成にとどまらず, 次のデータ収集・選択のプロセスをサポートしてくれること。これら3つの事項は今まで統計的手法では十分にサポートしてくれなかった点であり, 病院情報システムのデータから仮説を生成するためにきわめて重要である。今後データマイニングの研究から, 上記の3つの課題を解決する研究が進むことを著者は望んでいる。

おわりに

医学研究の発展にデータマイニングそしてデータマイ

ニングコンテストがどう貢献し得るかについて概観した。総じていえば, 医学におけるデータベースからの知識発見においては, 得られた知識から再び専門家の知識がどのような性質を有しているかについて深い反省を促す絶好の機会を与えてくれるか否かが一番重要である。医師として立場から考えても, 専門家というものでさえも万能ではなく, 日常の「医学の常識」に流されていることを指摘したい。医師を含めたいずれの領域の専門家も膨大なデータの前では無力に近い。計算機に実装されたデータ解析ツールがなければ, 我々は仮説の生成もできなくなりつつあるのが現状である。データマイニングコンテストによって, データマイニングの手法の不十分な点が明らかになり, データベースからの「専門家による」発見を促進させるためのデータマイニング技術が確立されることを期待したい。

参考文献

- 1) Van Bommel, J. and Musen, M.: Handbook of Medical Informatics, Springer Verlag (1997).
- 2) 松原謙一, 柳 佳之: ゲノム情報生物学 BioinformaticsとInformation Biology, 中山書店(2000).
- 3) 鈴木英之進, 津本周作: 日本・アジアにおけるデータマイニングコンテスト, 情報処理, Vol.42, No.5, pp.457-461 (May 2001).
- 4) 稲田 裕, 野崎貞彦: 新簡明衛生公衆衛生, 南山堂(1994).
- 5) Berka, P.: ヨーロッパにおけるデータマイニングコンテスト, 情報処理, Vol.42, No.5, pp.454-456 (May 2001).
- 6) Langley, P., Simon, H. A., Bradshaw, G. L. and Zytkow, J. M.: Scientific Discovery: Computational Explorations of the Creative Processes, MIT Press (1987).
- 7) 天児和暢, 南嶋洋一(編): 戸田新細菌学, 南山堂(1997).
- 8) Thagard, P.: Explaining Scientific Change: The Formation and Development of the Bacterial Theory of Ulcers. In: Valdes-Perez, R., et al. (eds.), Proceedings of Symposium on Systematic Methods of Scientific Discovery, AAAI Press, pp.66-67 (1995).
- 9) Kohavi, R., Brodley, C. E., Frasca, B., Mason, L. and Zheng, Z.: KDDカップ2000主催者報告書: 本質をつかむ, 情報処理, Vol.42, No.5, pp.445-453 (May 2001).
- 10) Fayyad, U. M., et al. (eds.): Advances in Knowledge Discovery and Data Mining, AAAI Press (1996).
- 11) Fayyad, U. M.: Diving into Databases, <http://www.dbpd.com/vault/9803feat.htm> (1998).
- 12) 平山恵造: 神経症候学, 文光堂(1971).

(平成13年3月31日受付)