解 説

学習における計算論的アプローチ[†]
——機械学習に向けて——

有 川 節 夫[††] 西 野 哲 朗[†††]

1. はじめに

近年，エキスパートシステムなどの知識情報処理技術の実用化にともない，機械による学習や知識獲得への関心が高まっている．ここで**学習**とは，陽にプログラムとして与えられる以外の方法によって，知識を獲得することを指すものとする．このような学習に対して，各方面からさまざまな関心が寄せられているにもかかわらず，「機械がある概念を学習するのに，本質的にどれだけの計算資源（時間や領域など）を必要とするか？」という根本的な問題でさえ，理論的に十分解明されているとは言いがたい．実際，機械学習に関する理論的研究は，最近いくつかの重要な概念が提案され，今まさに発展期を迎えつつあるというのが現状である．

本特集で解説する**計算論的学習理論**は，形式言語理論や計算量理論などにおける形式的手法を用いて，機械学習の本質を明らかにしようとする理論計算機科学の新しい分野である．具体的には，

1. 新しい学習モデルの導入
2. 学習に要する計算量（時間，領域など）の評価
3. 学習可能性・不可能性の証明

などがその主な研究テーマである．

機械に学習させることは，少し大袈裟に言えば情報科学における究極の目標の一つであり，従来からさまざまなアプローチによって研究がなされてきた．人工知能の分野においては，おもに記号処理系におけるさまざまな学習の枠組みが提案されてきた．たとえば，例題や助言からの学習方式が提案され，それらの方式に基づくシステムが構築されてきた．また，ニューラルネットなどの神経回路網モデルにおいては，学習機能は理論的にも工学的にも大変重要な意味をもっており，従来から誤差逆伝播学習法などの学習アルゴリズムが盛んに研究されてきた．このほか，制御理論や情報理論においても，学習は重要な研究テーマとなっている．

計算論的学習理論が果たすべき一つの役割は，上で述べたような種々の学習方式の根底に共通に存在するであろう（おそらく簡潔な）数学的原理を解明し，その原理を研究者・技術者の共通の言葉として提供することで，学習の理論・応用研究の発展を加速することである．これは，形式言語理論がコンパイラ構成論に対して果たした役割のアナロジと考えると分かりやすい．コンパイラ構成論においては，文脈自由文法の概念が非常に重要な役割を果たしたことはよく知られている．現在の学習システムをコンパイラの発展段階になぞらえて考えると，まだまだ初期の段階にあると思われる．このような状況から抜け出すためには，新たな理論が必要であることを歴史は教えている．

本解説においては，計算論的学習理論がすでにそのような役割を果たし始めていることの例として，ニューラルネットの学習に関する問題の計算論的解法を紹介する．そこでは，(1)理論の側から提案された学習モデルが，実際問題の解析に対して本質的な示唆を与えていること，および，(2)理論において用いられている数学的道具が，実際的な結果を導くのに大変役立つことなどが示される．

計算論的学習理論は，近年，活発に研究が行われており，海外ではこの分野に関連した国際会議として，東ドイツの AII (International Workshop

† Computational Approaches to Machine Learning by Setsuo ARIKAWA (Research Institute of Fundamental Information Science, Faculty of Science, Kyusyu University) and Tetsuro NISHINO (College of Science and Engineering, Tokyo Denki University).
†† 九州大学理学部基礎情報学研究施設
††† 東京電機大学理工学部

☆ 通常，「編集にあたって」で述べられる，本特集の編集の意図や各解説の概要などについては，本解説の 1. と 5. をご覧ください．

on Analogical and Inductive Inference)[1),2)], 米国の COLT (Workshop on Computational Learning Theory)[3)~5)] などが開催されている．また，我が国でも 1990 年 10 月に東京で ALT '90 (International Workshop on Algorithmic Learning Theory)[6)] が開催され，日本におけるこの分野の研究が質・量ともに充実してきていることが確認されたばかりである．

この特集が一人でも多くの研究者，技術者，実務家および学生の方々の興味を引き，機械学習の数学的本質について考えるなんらかのきっかけとなれば幸いである．また，本特集を読まれて，多くの方がこの新しい分野に関連した研究に進まれることを希望する次第である．以下この解説では，本特集への導入として，計算論的学習理論全般に対する入門的解説を行うと同時に，最後の章では，本特集の構成および各解説の概要と，解説相互の関連についても整理して述べる．

2. なぜ計算論的学習理論なのか？

計算論的学習理論ということばが人々の注意をひくようになったのは，1988 年の MIT における COLT '88 のころからである．その前後に，ドイツにおける AII, 日本における ALT もスタートして今日の定着をみている．

計算論的学習理論のパラダイムとしては，この解説にもみられるようにいくつかのものがあるが，その基本となり，歴史的にも古くから研究され発展してきたものに帰納推論がある．帰納推論ということばは，論理学や哲学，確率的推論における文脈でも用いられ，その意味は必ずしも統一されていない．しかし，計算機科学においては，与えられたデータからそれを説明する規則を導き出す推論の総称として使われている．

計算機科学における帰納推論の出発点は，これも見方によって多少の違いはあるが，1950 年代の Moore[12)] における順序機械を入出力対から同定する研究に始まるといえる．この研究は，その後システム同定や文法推論などへと発展していった．このうち，文法推論の研究は，人間による言語習得のプロセスのモデルという観点から，Chomsky による各言語階層や，各種のオートマトン，特に有限オートマトンを対象にして活発に研究された．そうした研究は，パターン認識の分野で応用され，同時にその後の言語を対象にした帰納推論の展開の契機となった．

幾人かの人々による試行的研究の後で，1967 年に Gold は，極限における同定 (identification in the limit) という帰納推論の成功基準を考え，言語と関数の帰納推論の理論的基礎を築いた．

言語の帰納推論は，1950 年代からのいわゆる言語理論とオートマトン理論の豊かな成果を利用して急速に発展していった．同様に関数，すなわち帰納的関数 (recursive functions) の帰納推論も，1930 年代以来の成果を利用して急速に発展していった．したがって，そのころの帰納推論は，いわゆる計算理論に基づく例題からの学習として研究されたのである．

初期のころの重要な理論的成果として，まず，Gold による正データからの帰納推論がある．これから同定（学習）する言語に属するサンプルを正データ (positive data) といい，属さないサンプルを負データ (negative data) という．Gold は，正データだけからの極限における同定という基準の下での帰納推論の能力は，正負両データを用いる場合よりも真に小さいという結果を示している．具体的には，すべての有限言語と少なくとも一つの無限言語を含む言語族（超有限言語族という）が，正データからは帰納推論できないことを簡単に証明している．これによって，その当時の言語理論において最も下位に位置づけられていた正則言語の族でさえ，正データからは帰納推論できないことになり，この方面の理論および実際に関係した研究者に大きな失望を与えた．

しかし，1979 年に Angluin により，正データから帰納推論が可能であるための必要十分条件および使いやすい十分条件が与えられ，同時にパターン言語族という基本的な言語族が正データから帰納推論可能であることが示され，帰納推論の研究が復活した．その後，1982 年に Shinohara によって，多項式時間内による正データからの帰納推論可能な言語族とその具体的応用例が示され，これによって帰納推論と計算量理論とが明確に関連付けられた．この間の事情やその後の発展については，本特集における横森・篠原による解説を参照されたい．

1967 年の論文において，Gold はまた正負両データからの帰納推論の可能性について基本的な

結果を導き出して，極限における同定という意味での言語の帰納推論を体系化し，その基礎を築いた．

一方，計算理論の中心である帰納的関数の理論と関連した関数の帰納推論の研究は，1965年からやはり Gold によって始められ，次章で述べるように，数多くの帰納推論の方式とその能力について，詳細な研究が展開されてきた．この枠組みに関する研究は，現在に至るまで，特に米国やドイツにおいて盛んに続けられ，重要な成果が次々に得られている．たとえば，Blum & Blum によって，1975年に，h 従順性という概念が導入されている．これは，帰納推論の各ステップにおいて計算量を考えることの必要性を主張したもので，1982年の Shapiro によるモデル推論システムにおいても，仮説の検証に要するステップ数のモニタ役として使われている．

こうした帰納推論の研究のほかに，1960年代の主要な機械学習のパラダイムとしてパーセプトロンの研究がある．これは，いわゆるパラメータ学習に分類されてきたもので，パターン学習の分野において重要な貢献をしてきた．パーセプトロンの研究は，1969年に Minsky と Papert によって体系化されたが，同時に能力の限界が明確に示されたこともあって，やがて下火になった．しかし，周知のように，長年にわたる基礎研究と実際的な研究の成果が実り，今日のニューロ・コンピュータへと発展している．このニューラルネットの学習は，4. で示すように，現在の計算論的学習理論の主要パラダイムの一つである PAC 学習とも関連した，重要な学習方式となっている．

以上，今日の「計算論的学習理論」が登場するに至るまでの経緯と，なぜ「計算論的」と呼ばれるかについて手短かに述べてきた．詳細については,本特集のほかの解説を参照していただきたい．

3. 計算論的学習理論における主な学習パラダイム

計算論的学習理論の主な学習パラダイムとしては，前章で述べた帰納推論と PAC 学習, EXACT 学習と呼ばれる三つのものがあり，そのほかにニューラルネットの学習などがある．これらのパラダイムのうち，帰納推論については，「形式言語の学習」と「モデル推論」という2編の解説が本特集に用意されており，PAC 学習, EXACT 学習についてもそれぞれ1編の解説があるので，本章では，PAC 学習と EXACT 学習の基本的アイディアのみを説明する．また，帰納推論に関しては，ほかの解説に含まれていない関数の帰納推論について少し詳しく説明する．なお，ニューラルネットの学習については，次章で説明する．

3.1 PAC 学習

PAC 学習は，1984年の Valiant による研究に始まるもので，命名は Angluin による．PAC 学習とは，Probably Approximately Correct Learning の略であり，それまでの帰納推論を中心とした学習モデルが，厳密な正確さを要求したのに対して，名前が示すとおり，確率的近似的に正しい学習のモデルである．正確な定義は,「PAC 学習」の解説に譲るが，大雑把にいうと，「ほどほどの例から，ほどほどの正確さで，ほどほどの時間内に学習できる概念クラス」について研究するものである．次の主婦（夫）の買物の例[22]が，この学習パラダイムを分かりやすく表している．

たとえば，魚屋で魚を選ぶとき，彼女（彼）は，色，艶，弾力感，目の透明感，匂いなどから，その魚が，その値段に値するかどうかを判断しようとする．そして，実際に調理し食べてみて，事実を知る．幾度となくこのような買物を経験するうちに，商品の判別法を習得し，「賢い買物」という概念を学習することになる．

その際の学習に，30年も40年もかかるようでは夫婦関係が破綻する可能性もある．ほどほどの時間内に学習しなければなるまい．また，あらゆる魚に対して，完璧な判別法を期待することは無理である．たとえば，よく使う鰯に対してはよい判定法を得ても，まず買うことのない河豚については，まったく分からず，鰯の判定法がそのまま適用できるかどうかも分からない．しかし，日常よく買う典型的なものについては，ほぼ正しく判断できなければならないだろう．また，魚によっては，その習得した判別法が役に立たないこともあろう．しかし，そのようなことは稀であって欲しい．

このような学習の過程においては，間違いが基準以下であるような判別法を得るために必要な実例の個数とその判別法を作る手間が重要な問題となる．PAC 学習は，正確には，概念の族 C，例の

確率分布 P，学習の精度を表す数値 ε，それを出力する確率に関係した数値 δ，学習時間を表す多項式などを用いて定義され議論される．

3.2 EXACT 学習

未知の対象を近似的に学習する PAC 学習に対して，未知の対象を正確に学習するという学習パラダイムもある．これをEXACT学習と呼んでいる．この学習方式は，1987年の Angluin による MAT (Minimaly Adequate Teacher) 学習に始まるものである．

MAT 学習は，文字どおり最小限のことを教えてくれる教師を仮定したもので，学習における教師の役割を計算論的に考察した最初の試みでもある．この学習パラダイムが扱うのは，主として形式言語やオートマトンである．MAT 学習における学習システムは，未知の言語を $L\subseteq\Sigma^*$ とするとき，

（1） 所属性質問，すなわち $x\in\Sigma^*$ に対して，$x\in L$ か否かに答えることができ，

（2） 等価性質問，すなわち，システムが生成した仮説 H に対して，$L=H$ ならば yes と答え，$L\neq H$ ならば L と H の対称差集合の要素を反例として示す．

MAT 学習のシステムは，このような教師を内部にもち，さらに仮説としての文法Hを生成する機能をもつ．MAT 学習における学習の結果は，教師により「正しい」と判定されたものであり，その意味で EXACT である．

MAT 学習においても，学習に要する計算量の問題が重要であり，たとえば，Angluin の論文では，有限オートマトンの学習が，オートマトンの状態数と反例の最大長とに関する多項式時間で可能であることが示されている．有限オートマトンの学習は，正負両データから帰納推論可能であるが，計算の非常に困難な問題として知られている．したがって，上記のような最小限の教師による教示によって学習が容易になることが示されたわけである．

MAT 学習だけでなく，一般の EXACT 学習においては，許される質問の種類によって，学習能力に差違が生じる．質問の種類としては，所属性，等価性のほかに，部分性 (subset)，包含性 (superset)，非交差性 (disjointness)，全体性 (exhaustiveness) が考えられている．

Angluin によって始められた EXACT 学習は，学習を容易にするには，正負のデータのほかに，いかなる付加情報が有用であるかという一般的な問題意識も提起して，現在多くの研究者によって活発に研究されている．

3.3 帰納推論

帰納推論は，前章でも述べたように，計算論的学習理論を歴史的に支えてきた重要な学習パラダイムである．帰納推論の研究は，言語の帰納推論，モデル推論，関数の帰納推論に大別される．はじめの二つについては，本特集に独立した解説があるので，ここでは，関数の帰納推論について簡単に説明しておく．

帰納推論の成功基準として一般的に使われている「極限において同定する」ことを，EX 推論するということにする．また，EX 推論可能な関数族の全体も同じ記号 EX を使って表す．帰納推論システムでは，原理的には，仮説を次々に枚挙して，それまでに受け取ったデータを説明できる仮説で，最初に見つかったものを出力するという戦略で十分である．この戦略を枚挙手法 (enumerative method) といい，枚挙手法が適用できる関数族の全体を NUM で表す．Gold[13] によって，次の定理が得られている．

定理 3.1 $NUM\subseteq EX$

枚挙手法という最も単純で素朴な手法が，帰納推論にいつでも使えるだろうか．この問題は，Barzdin[24] によって否定的に解決されている．

定理 3.2 $NUM\subsetneq EX$

この定理は，枚挙手法の適用できない EX 推論可能な帰納的関数の族が存在することを主張している．

帰納的関数の全体をRで表す．Rは，枚挙できないことが知られている．したがって，$R\notin NUM$ である．しかも，EX 推論もできない．つまり，$R\notin EX$ であることが Gold[13] によって示されている．そこで，帰納推論の成功基準をどこまで弱めるとRを推論できるようになるかという重要な理論的問題が生じる．この問題は次のようにして解決されている．

時点 n における帰納的関数 f のデータ $f(0), f(1), \cdots, f(n)$ が与えられたときの，帰納推論機械 M の出力する仮説を $M(f[n])$ で表し，$M(f[n])$ が（f を計算する）あるプログラム p に収束した

とする．p が最大 m 個の例題を除いて($m=*$ のときは，有限個の例外を除いて) 正しく f を計算するとき，M は f を EX^m 推論するという．

基準をさらに弱めて，仮説の無限列が収束しなくても，ある時点以降はプログラムの動作として正しければよい (behaviorally correct) という基準を考え，この基準の下で，$M(f[n])$ が最大 m 個の例外を除いて ($m=*$ のとき，有限個の例外を除いて) f を正しく計算するプログラムであるとき，M は f を BC^m 推論するという．そうすると，次の定理が成立する．(Case & Smith[25])

定理 3.3 (1) $R\in BC^*$

(2) $EX=EX^0\subsetneq EX^1\subsetneq\cdots\subsetneq EX^*\subsetneq BC^0\subsetneq BC^1\subsetneq\cdots\subsetneq BC^*$

以上のほかにもさまざまな基準や推論方式が考案され，帰納的関数の理論の枠組みで多くの成果が得られている．1985年ごろまでの重要な成果については，文献 26) を参照されたい．また，その後の展開については，本特集の安倍らによる解説「計算論的学習理論の米国における現状と動向」の **4.** を参照されたい．

本章では，計算論的学習理論における代表的な三つのパラダイムについて，ごく簡単に説明した．こうしたパラダイムは固定されたものではなくて，実際の機械学習の研究からの提起を受けて新しいパラダイムが次々に生じるであろう．たとえば，機械学習における重要な学習パラダイムである EBL や類推なども，計算論的立場から研究する必要があり，それによって新しい分野が開けるものと考えられる．また，計算論的学習理論から実際の機械学習の分野へのフィードバックも不可欠である．そうした，実際と理論との連携によって，機械学習は健全に発展していくであろう．

4. ニューラルネットにおける学習への応用

最近，エキスパートシステム構築などの多くの実際的問題に対して，ニューラルネットが応用されている．そのような応用においては，まずトレーニング例のデータベースを用意し，次に適当なネットワークを選んで，誤差逆伝播方式などの学習アルゴリズムにより，そのネットワークをなるべく多くのトレーニング例に適応させる．そこでの目標は，トレーニング後に入力された同種の例に対し結果を正しく予測するように，ネットワークのノード間の適当な重み付けを得ることである．

このような場合に重要となるのは，m 個のトレーニング例が与えられたときに，どのくらいのサイズのネットワークが正しい一般化を得るのに適当かという問題である．この問題を定式化するために，以下のような仮定をおく．トレーニング例は，任意のある固定された確率分布のもとで選ばれるものとする．学習アルゴリズムはある正確さのパラメータ ε を事前に与えられ，トレーニング後に同じ確率分布のもとで選ばれたテスト例を，少なくとも $1-\varepsilon$ の確率で正しく分類するようなフィードフォワード・ニューラルネットワーク (後述) を高い確率で生成することをその目標とする．

上記の仮定は，L. G. Valiant によって提案された PAC 学習 (解説 5 参照) という学習方式における仮定に基づいている．本章では，計算論的学習理論とほかの学習パラダイムとの関係を示す一つの例として，計算論的手法を用いて最近得られた，ニューラルネットの学習に関するいくつかの理論的結果を紹介する．以下で紹介する結果は，E. B. Baum ら[7]によるものである．

まず，幾つかの概念を定義する．ln で自然対数を，log で底 2 の対数を表し，また e を自然対数の底とする．例とは，対 (x,a) のことと定義する．ただし，$x\in\mathbf{R}^n$ かつ $a\in\{-1,+1\}$ とする．ランダムサンプルとは，$\mathbf{R}^n\times\{-1,+1\}$ 上のある確率分布 D の下で独立に選ばれた例の列と定義する．f を $\mathbf{R}^n$ から $\{-1,+1\}$ への関数とする．D に関する f のエラーとは，ランダムな例 (x,a) に対して，$a\neq f(x)$ となる確率のことと定義する．

F を $\mathbf{R}^n$ から $\{-1,+1\}$ への関数のクラスとし，S を $\mathbf{R}^n$ の m 個の点からなる集合とする．$f\in F$ によって導かれる S の二分割 (dichotomy) とは，以下の条件を満たす二つの互いに素な集合 S^+ と S^- への S の分割のことをいう：$x\in S^+$ に対しては $f(x)=+1$ となり，$x\in S^-$ に対しては $f(x)=-1$ となる．$d_F(S)$ で，関数 $f\in F$ によって導かれる S の異なる二分割の個数を表す．また，$d_F(m)$ で，要素数 m のすべての $S\subset\mathbf{R}^n$ に関する $d_F(S)$ の最大値を表す．$d_F(S)=2^{|S|}$ のとき (すなわち，S のすべての二分割が F に属する関数によって導かれるとき)，S は F によって細分される

(shattered) という．

F の VC 次元 (Vapnik-Chervonenkis dimension) $VC(F)$ とは，F によって細分される $S\subset\mathbf{R}^n$ の最大要素数のことをいう．すなわち，$VC(F)$ とは，$d_F(m)=m$ となるような最大の m のことである．この VC 次元は，PAC 学習可能性と大変密接な関係にあることが知られている．一方，Baum らは，VC 次元がニューラルネットの学習容量に対する尺度として適当であることを示している[7]．

$\mathbf{R}^n$ の点を入力とするフィードフォワード・ネットとは，n 個の順序付けられた入力ノード（入ってくる辺をもたない）と，1 個の出力ノード（出ていく辺をもたない）をもつサイクルを含まない有向グラフ G のこととする．G のノードのうち，入力ノードでないノードを計算ノードと呼び，入力ノードでも出力ノードでもないノードを隠れノードと呼ぶ．各計算ノード n_i には，関数 f_i: $\mathbf{R}^{IN(n_i)}\rightarrow\{-1,+1\}$ が対応している．ここで，$IN(n_i)$ はノード n_i に入ってくる辺の本数を表す．ネットそれ自体には，各 f_i を合成して得られる関数 f: $\mathbf{R}^n\rightarrow\{-1,+1\}$ が対応している．

ニューラルネットが，ある確率分布のもとでランダムに選ばれた m 個のトレーニング例から一般化できるのは，どのような場合かについて考えよう．前にも述べたように，トレーニング後のテストで用いられる例も，同一の確率分布のもとで選ばれるものと仮定する．次の二つの補題が知られている（証明に興味のある読者は，原論文を参照されたい）．

補題 4.1[7] G を E 本の辺と $N\geq 2$ 個のノードをもつ固定された基底グラフとする．ただし，G の各ノードはある線型しきい値関数を計算するものとする．G 上で定義されたフィードフォワード・ネットで計算できる関数全体のクラスを F とする．いま $W=E+N$ としよう．このとき，すべての $m\geq W$ に対し，$d_F(m)\leq(Nem/W)^W$ である．

上の補題において，W は各辺に重み 1，各ノードにしきい値 1 を割り当てたときのネットワーク内の重みの合計になっていることに注意する．

補題 4.2[9),10)] F を $\mathbf{R}^n$ 上の関数のクラスとし，$0<\gamma\leq 1, 0<\varepsilon,\delta<1$ とする．S を確率分布 D のもとで独立に選ばれた m 個の例からなるランダムサンプルとする．F の中にある関数が存在して，S の例のうちの高々 $(1-\gamma)\varepsilon$ の割合のものとしか

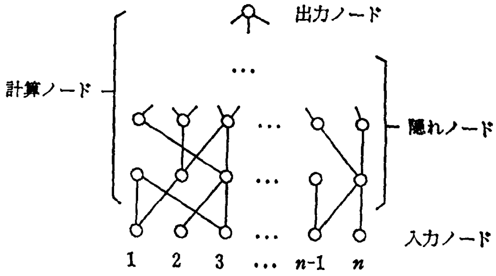


図-1 フィードフォワード・ネット

矛盾しないが，その関数が D に関して ε より大きなエラーをもってしまう確率は，

$$8d_F(2m)e^{-\gamma^2\varepsilon m/4}$$

よりも少ない．

上の補題 4.1 と 4.2 を用いて，妥当な一般化が得られるための次のような十分条件を導くことができる．

定理 4.3[7] G を E 本の辺と $N\geq 2$ 個のノードをもつ固定された基底グラフとする．ただし，G の各ノードはある線型しきい値関数を計算するものとする．（すなわち $W=E+N$ がネットワーク内の重みの合計である．）さらに，$0<\varepsilon\leq 1/2$ を固定された定数とし，m 個のトレーニング例が与えられたとする．ただし，

$$m\geq\frac{32W}{\varepsilon}ln\frac{32N}{\varepsilon}$$

とする．ネットワークの重みを適当に選択して，少なくとも確率 $1-\varepsilon/2$ でトレーニング例を正しく分類するようにできるならば，そのネットは少なくとも確率 $1-8e^{-1.5W}$ で，トレーニング例と同じ確率分布のもとで選ばれたテスト例を誤り率 ε で正しく分類する．

証明 $\gamma=1/2$ として，補題 4.2 を適用する．その際，$d_F(m)$ に対する上界は，補題 4.1 の結果を用いる．簡単な計算により，エラーが ε より大きく，かつトレーニング例のうちの少なくとも割合 $1-\varepsilon/2$ のものと矛盾しない関数を定義するような重みの選択が存在する確率は，高々

$$8(2Nem/W)^W e^{-\varepsilon m/16}$$

であることが分かる．特に，

$$m=\frac{32W}{\varepsilon}ln\frac{32N}{\varepsilon}$$

のとき，この値は

$$8\left(2e\frac{\varepsilon}{32N}ln\frac{32N}{\varepsilon}\right)^W$$

となり，これは $N\geq 2$ かつ $\varepsilon\leq 1/2$ のとき，$8e^{-1.5W}$ よりも小さい．（証終）

また，Baum らは妥当な一般化が得られるための必要条件として，次の定理も証明している（証明略）．

定理 4.4[7]　一つの隠れ層をもつ完全に接続された (fully-connected) フィードフォワード・ネットに対しては，W/ε のオーダより少ない個数のトレーニング例しか使わない学習アルゴリズムは，（妥当な重み選択と矛盾しないような例のある確率分布に関して，）テスト例を ε より小さい誤り率で正しく分類するような重みを選択することに，少なくともある固定された確率で失敗する．

定理 4.3 と 4.4 から，ネットのトレーニングに適当な例の個数は，およそ W/ε 個程度であると結論できる．これは，Widrow[11] によって経験的に指摘されている結果とほぼ一致する．

上で述べた結果のほかにも，Baum はフィードフォワード・ネットが学習対象の実際的な表現としては，ある意味で完全である（最も汎用性がある）ことを理論的に示している[8]．また，冒頭でも紹介した米国で開催されている国際会議 COLT でも，ニューラルネットに関する結果が毎年報告されている[3)~5]．今後，この二つの分野の研究者が互いに触発し合い，それぞれの研究を発展させていくことが大いに期待される．

5. 本特集の構成と概要

本特集では，計算論的学習理論の研究の現状と動向，および将来の課題について解説していく．本特集の構成は次のとおりである．まず計算論的学習理論の歴史的背景が良く理解できるように，この理論の源流である形式言語の帰納推論・学習から解説を始める．さらに，本分野誕生の契機を作った一時期を画する研究として，モデル推論についても解説する．そして最近，精力的に研究されている学習モデルとして EXACT 学習と PAC 学習について詳しく解説する．また，本理論のデータエントリシステムやエディタなどへの応用についても紹介する．最後にこの分野を米国で研究している研究者に，米国における最近の研究動向を中心として，本分野の現状と動向を詳しく解説してもらう．各解説の内容は，概略以下のとおりである．

解説 2：「形式言語の学習」では，その歴史的背景から，最近の研究動向にいたるまでを解説する．解説される形式言語のクラスとしては，最も基本的な正則言語のクラスから始まり，文脈自由言語のクラス，さらには文脈依存言語のクラスまでを対象とし，それらに対する効率的・一般的な学習モデルおよびアルゴリズムについて，特に正の例からの学習という話題を中心に解説する．

解説 3：「モデル推論」では，E. Shapiro によって考案されたモデル推論のメカニズムを解説すると同時に，それに対する学習の立場からの問題点および，それらの問題点に対する最近のアプローチについて概観する．Shapiro による一般理論や，Prolog 上で実現されたシステム MIS について解説した後，理論項の取り扱いを中心に，Shapiro 以降の最近の研究についても述べる．

解説 4：「EXACT 学習」では，学習者が教師への質問を用いて学習するという機械学習の一つの枠組みを取り上げる．D. Angluin は，このような教師と学習者のパラダイムにおける概念学習の問題を形式的かつ自然に定義することに成功し，さらに多項式時間学習という概念学習における計算効率の基準を導入した．この解説では，このような機械学習の問題における質問の種類と多項式時間学習可能性との関係を，最新の肯定的・否定的結果を織りまぜながら解説していく．また極限同定（解説 2 参照）や PAC 学習（解説 5 参照）などのほかの学習モデルとの関連についても述べる．

解説 5：「PAC 学習」では，L. G. Valiant が提案した PAC (Probably Approximately Correct) 学習の枠組みについて解説する．冒頭でも述べたことだが，ある概念を学習することが，本質的にどれほどの時間や領域などの計算資源を必要とするかという問題は，従来，理論的にはあまり解明されていなかった．PAC 学習可能性の概念を用いることにより，多項式時間で例から学習可能な概念について研究することが可能となる．また，PAC 学習の枠組みとそれに関連する概念（VC 次元など）は，前章で述べたように，最近ニューラルネットにおける学習の問題を理論的に研究する道具としても用いられている．

解説 6：「計算論的学習理論の応用システム」で

は，計算論的学習理論で解明されたアルゴリズムを実際のシステムに応用した例を紹介する．具体的には（1）データエントリシステムへの応用，（2）テキストエディタへの応用，（3）構造エディタへの応用の三つの応用について解説する．また，理論を応用する際の利点と問題点についても論じる．

解説 7：「計算論的学習理論の米国における現状と動向」では，計算論的学習理論の現状と動向を，近年の米国における研究活動および成果を中心に概観する．特に最近の学習理論の主流をなす三つのパラダイム，極限同定（解説2参照），質問による学習（解説4参照），PAC 学習（解説5参照）における学習可能性に関する結果および証明手法などを紹介する．また，一階述語論理のモデル推論（解説3参照）への応用など，近年におけるこれらのパラダイムの拡張についても述べる．

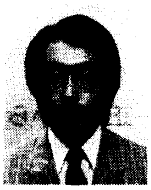
参 考 文 献

1) Jantke, K. P. (ed.): *Analogical and Inductive Inference*, Lecture Notes in Computer Sciences, Vol. 265, Springer-Verlag (1987).
2) Jantke, K. P. (ed.): *Analogical and Inductive Inference*, Lecture Notes in Artificial Intelligence, Vol. 397, Springer-Verlag (1989).
3) Haussler, D. et al. (eds.): *Proceedings of the 1988 Workshop on Computational Learning Theory*, August 3-5, 1988, MIT, Morgan Kaufmann (1988).
4) Rivest, R. et al. (eds.): *Proceedings of the Second Annual Workshop on Computational Learning Theory*, University of California, Santa Cruz, July 31-August 2, 1989, Morgan Kaufmann (1989).
5) Fulk, M. et al. (eds.): *Proceedings of the Third Annual Workshop on Computational Learning Theory*, University of Rochester, Rochester, New York, August 6-8, 1990, Morgan Kaufmann (1990).
6) Arikawa, S. et al. (eds.): *Proceedings of the First International Workshop on Algorithmic Learning Theory*, Tokyo, October 8-10, 1990, Ohmsha (1990).
7) Baum, E. B. and Haussler, D.: What Size Net Gives Valid Generalization?, *Neural Computation* 1, pp, 151-160 (1989).
8) Baum, E. B.: Complete Representations for Learning from Examples, in Abu-Mostafa, Y. S. (ed.) *Complexity in Information Theory*, Springer-Verlag (1988).
9) Blumer, A., Ehrenfeucht, D., Haussler, D. and Warmuth, M.: Learnability and the Vapnik-Chervonenkis Dimension, *Journal of the ACM* 36, pp. 929-965 (1989).
10) Vapnik, V. N.: *Estimation of Dependences Ba-sed on Empirical Data*, Springer-Verlag (1982).
11) Widrow, B.: ADALINE and MADALINE—1963, Plenary Speech, Vol. I: *Proc. IEEE 1st Int. Conf. on Neural Networks*, San Diego, CA, pp. 143-158 (1987).
12) Moore, E. F.: *Gedanken-experiments on Sequential Machines*, in Shannon, C. E. and Macasthy, J. editors, Automata Studies, pp. 129-153, Princeton Univ. Press (1956).
13) Gold, E. M.: *Limiting Recursion*, J. Symbolic Logic 30, pp. 28-48 (1965).
14) Gold, E. M.: *Language Identification in the Limit*, Inform. & Control. 10, pp. 447-474 (1967).
15) Angluin, D.: *Finding Patterns Common to a Set of Strings*, Proc. 11th Annual Symposium on Theory of Computing, pp. 130-140 (1979).
16) Angluin, D.: *Inductive Inference of Formal Languages from Positive Data*, Inform. & Control. 45, pp. 117-135 (1980).
17) Shinohara, T.: *Polynomial Time Inference of Pattern Languages and Its Application*, Proc. 7th IBM Symp. Math. Found. of Comp. Sci., pp. 191-209 (1982).
18) Blum, L. and Blum, M.: *Toward a Mathematical Theory of Inductive Inference*, Inform, & Control. 28, pp. 125-155 (1975).
19) Klette, R. and Wiehagen, R.: *Research in the Theory of Inductive Inference by GDR Mathematicians—A Survey*, Information Sciences 22, pp. 149-169 (1980).
20) Minsky, M. and Papert, S.: *Perceptrons—An Introduction to Computational Geometry*, MIT Press (1969).
21) Valiant, L. G.: *A Theory of the Learnable*, CACM, 27, pp. 1134-1142 (1984).
22) 篠原 歩，宮野 悟：PAC 学習とは，LA 会誌第13号，pp. 1-11 (1989).
23) Angluin, D.: *Learning Regular Sets from Queries and Counterexamples*, Inform. & Computation, 75, pp. 87-106 (1987).
24) Barzdin, J. M. and Freivald, R. V.: *On the Prediction of General Recursive Fundations*, Soviet Math. Dokl. 13, pp. 1224-1228 (1972).
25) Case, J. and Smith, C.: *Comparison of Identification Criteria for Machine Inductive Inference*, Theoretical Computer Science 25, pp. 193-220 (1983).
26) 有川節夫，篠原 武，宮原哲浩：帰納推論の理論，大須賀，佐伯編「知識の獲得と学習」，pp. 147-197，オーム社 (1987).

（平成2年11月22日受付）

有川 節夫（正会員）

昭和 39 年九州大学理学部数学科卒業．現在，九州大学理学部基礎情報学研究施設教授．大学院総合理工学研究科情報システム学専攻教授を兼務．理学博士．現在の主要研究テーマは，計算理論，情報検索論，知識情報処理，特に各種推論機構，計算論的学習理論，第 11 回丹羽賞学術賞，第 1 回人工知能学会論文賞受賞．日本数学会評議員，人工知能学会理事．ソフトウェア科学会，LA 各会員．



西野 哲朗（正会員）

昭和 34 年生．昭和 57 年早稲田大学理工学部数学科卒業．昭和 59 年同大学院理工学研究科博士前期課程修了．同年日本アイ・ビー・エム(株)入社．昭和 62 年東京電機大学理工学部情報科学科助手．この間，属性文法，自然言語処理，計算論的学習理論の研究に従事．オートマトンと形式言語理論，計算量理論に興味をもっている．電子情報通信学会，日本ソフトウェア科学会，CAI 学会各会員．