氏　　　　名	瀧(たき)　本(もと)　英(えい)　二(じ)
授　与　学　位	工　学　博　士
学位授与年月日	平成 3 年 3 月 28 日
学位授与の根拠法規	学位規則第 5 条第 1 項
研究科，専攻の名称	東北大学大学院工学研究科 （博士課程）情報工学専攻
学位論文題目	**学習可能性に関する研究**
指　導　教　官	東北大学教授　丸岡　章
論文審査委員	東北大学教授　丸岡　章　　東北大学教授　木村　正行 東北大学教授　野口　正一

論　文　内　容　要　旨

1．まえがき

Valiantが提案したPAC学習モデルは，帰納的推論の妥当なモデルの1つとして，計算論的学習理論の分野で広く研究されている。本論文では，3層パーセプトロンなどに見られる学習能力を増大させる前処理の効果や，Occam's Razorと呼ばれる科学の方法論における理念を，それぞれPAC学習モデルの下で厳密に定式化し，学習のメカニズムとの関係について論じる。

2．PAC学習モデル

PAC学習モデルでは，学習対象である関数を十分高い精度で近似する仮説が，十分高い確率で得られるという条件に基づいて学習可能性が定義される。以下に厳密な定義を与える。

学習領域とは，学習の対象となっている関数の定義域のことで，これをXと表す。関数のクラスFを，学習領域Xから｛0，1｝への関数の集合とする。$f \in F$を目標関数とするとき，対$(x, f(x))$をfの例題という。例題は，X上のある確率分布Dに従って生成される。すなわち，例題$(x, f(x))$が生成される確率（密度）は$D[x]$で与えられる。確率分布Dに関しては，時間によらず不変であるということ以外，何の仮定も置かない。fのm個の例題の系列をfのmサンプル，または長さmのサンプルと呼ぶ。

Hを関数のクラスとするとき，HによってFを学習するアルゴリズム$\mathcal{A}$とは，任意の$f \in F$の任意のmサンプルが入力として与えられたとき，関数$h \in H$（の表現）を出力するアルゴリズムである。

仮説hが目標関数fのε近似であるとは，

$$\int_{f(x)\neq h(x)} D[x]\,dx \leq \varepsilon$$

が成り立つこととする。

［定義］Fが学習可能であるとは，次のような性質を持つ学習アルゴリズム（PAC学習アルゴリズム）$\mathcal{A}$が存在することである。すなわち，任意の$0<\varepsilon$，$\delta<1$，任意の$f\in F$に対して，整数$m_{\mathcal{A}}(\varepsilon, \delta, f)$が存在して，$X$上の任意の確率分布$D$に対して，長さ$m\geq m_{\mathcal{A}}(\varepsilon, \delta, f)$の$f$のサンプルを$\mathcal{A}$に与えたとき，$\mathcal{A}$は$1-\delta$以上の確率で$f$の$\varepsilon$近似である仮説$h$を出力する。

上の定義において，$m_{\mathcal{A}}$が目標関数fに依存しない場合，Fは一様に学習可能であるという。Blumer, Ehrenfeucht, Haussler, Warmuth は，Fが一様に学習可能であることの必要十分条件が，Fの Vapnik-Chervonenkis 次元（VC次元）が有限であるということを示した。

［定義］$S\subseteq X$が関数のクラスHによって全識別されるとは，任意の$T\subseteq S$に対して，ある$h\in H$が存在して，$h^{-1}(1)\cap S=T$となることである。

［定義］Hの VC次元とは，Hによって全識別される集合$S\subseteq X$の要素数の最大値である。

多項式時間学習可能性を議論する際には，関数のクラスFは表現のクラスとして定義されているものとし，さらに，次のようにパラメータづけを行なう。すなわち，$F=\cup_n F_n$，$F_n=\cup_s F_{n,s}$. ここで，F_nはFに属するn変数関数のクラス，$F_{n,s}$はF_nに属し，表現の記述の長さがs以下の関数のクラスとする。また，X_nをF_nに属する関数の定義域とし，$X=\cup_n X_n$とする。

［定義］Fが多項式時間学習可能であるとは，次のような性質を持つFの学習アルゴリズム$\mathcal{A}$が存在することである。すなわち，任意の$0<\varepsilon$，$\delta<1$，任意の$n, s\geq 1$に対し，整数$m_{\mathcal{A}}(n, s, \varepsilon, \delta)$，ただし，$m_{\mathcal{A}}$は$(n, s, 1/\varepsilon, 1/\delta)$の多項式，が存在し，任意の目標関数$f\in F_{n,s}$，$X_n$上の任意の確率分布$D$に対し，長さ$m\geq m_{\mathcal{A}}(n, s, \varepsilon, \delta)$の$f$のサンプルを$\mathcal{A}$に入力したとき，$\mathcal{A}$は入力の記述の長さの多項式時間以内に停止し，$1-\delta$以上の確率で$f$の$\varepsilon$近似である仮説$h\in F_n$を出力する。

3．学習アルゴリズムの変換

計算の理論において，ある問題を他の問題に帰着させるという概念は基本的なものである。本章では，この概念を学習アルゴリズムに適用して，関数のクラスGの学習アルゴリズムを別の関数のクラスFの学習アルゴリズムに変換した場合の多項式時間学習可能性について論じる。

$\{h_n, n\geq 1\}$を多項式時間で表現可能かつ計算可能な関数のクラスとし，ある多項式ℓが存在して，任意の$n\geq 1$，任意の$x\in X_n$に対し，$h_n(x)\in X_{\ell(n)}$が成り立つとする。

［定理3．1］Gを多項式時間学習可能な関数のクラスとすると，関数のクラス$F=\cup_n F_n$，ただし，$F_n=\{gh_n : g\in G_{\ell(n)}\}$も多項式時間学習可能である。ここで，$gh_n$は$h_n$と$g$を合成した関数を表す。

前処理$\{h_n\}$を学習領域間の変換として捉えたとき，一般に高次元空間への変換が学習可能な関数のクラスを拡張する効果を持つと考えられる。実際，神経細胞の学習モデルとして古典的な3

層パーセプトロンは，前処理として単に高次元空間へのランダム写像を行なう中間層を導入することによって，学習可能な関数のクラスを拡張しようと試みたものである。次の定理では，この種の前処理の効果をVC次元によって特徴づける。

［定理３．２］Gを多項式時間学習可能な関数のクラスとする。F_nの学習問題を$G_{\ell(n)}$の学習問題に帰着させるために，前処理で高めなければならない次元$\ell(n)$が$G_{\ell(n)}$のVC次元にのみ依存するような関数のクラスFが存在する。

一方，前処理としてある固定された写像hを用いるかわりに，前処理の効果を表すブール関数のクラスHを考え，Hと学習可能なクラスGを結合することによって得られるクラス$G \circ H$の多項式時間学習可能性について考える。この場合，前処理は目標関数に応じてクラスHから適当に選ばれることになるが，これを選択することはクラスGとHが非常に単純な場合でも，一般には困難となる。その典型的な例が一般化されたしきい値関数のクラスTHである。

［定義］所属質問とは，任意の$x \in X_n$を引き数として呼び出されたとき，目標関数の値$f(x)$を返すオラクルである。

［定理３．３］所属質問を学習アルゴリズムに与えたとき，関数のクラスに関するある仮定の下で，$G \circ H$は多項式時間学習可能となる。

特に，THは定理３．３における仮定を満たすので，所属質問のもとで多項式時間学習可能となる。

４．学習仮定と情報圧縮過程

計算論的学習理論の分野でこれまで提案された，さまざまな学習モデルに共通する原理が存在することが指摘されている。この原理は，Occam's Razorと呼ばれる科学の方法論における理念に基づくもので，具体的には，これまで提示されたデータをよく説明する複数の仮説候補が存在するときには，それらの候補の中でより単純なものがより良い仮説となると主張するものである。

PAC学習モデルにおいても，この理念が有効であることが確認されている。

［定義］FのOccamアルゴリズムとは，入力mサンプルに無矛盾で，実効仮説クラスのVC次元が$o(m)$であるようなFの学習アルゴリズムである。

［定理４．１］FのOccamアルゴリズムは，Fの無矛盾なPAC学習アルゴリズムでもある。

これは，情報圧縮が学習の本質であることを示唆している。なぜなら，こうして得られた仮説は受け取ったサンプルに無矛盾であるという意味で，サンプルの情報を失わずに保持していると考えられ，さらにその複雑さがサンプル長に比べて十分小さいということから，サンプル情報をある意味で圧縮した形式で表現しているということができるからである。本章では，この学習過程と情報圧縮過程との関係について議論する。

［定理４．２］Fの無矛盾なPAC学習アルゴリズムは，必ずしもFのOccamアルゴリズムではない。

Occamアルゴリズムの条件をゆるめた，弱Occamアルゴリズムを新しく定義する。弱Occamアルゴリズムも Occam's Razorの定式化の１つとみなすことができる。

［定義］Fの弱 Occam アルゴリズム$\mathcal{A}$とは，無矛盾で，かつ，任意の$\gamma>0$，任意の$f\in F$，任意の$\beta>0$に対し，整数$m_0(\gamma, f, \beta)$が存在して，任意の$m\geq m_0$，X上の任意の確率分布Dに対してVC次元が高々βmであるような関数のクラスH'が存在し，$\mathcal{A}$にfのmサンプルを与えたとき，$\mathcal{A}$は$1-\gamma$以上の確率でH'の要素を仮説として出力する。

［定理４．３］Fの弱 Occam アルゴリズムは，Fの無矛盾な PAC 学習アルゴリズムでもある。

［定理４．４］Fの無矛盾な PAC 学習アルゴリズムは，必ずしもFの弱 Occam アルゴリズムではない。

学習アルゴリズムは，例題を１つずつ入力するたびに仮説を更新する逐次的なアルゴリズムであるとして一般性を失わない。学習アルゴリズムが保存的であるとは，仮説の更新が起こるのは，それまで得られた仮説に矛盾する例題が現われたときに限るような学習アルゴリズムである。

［定理４．５］ある条件の下で，Fの無矛盾で保守的な PAC 学習アルゴリズムは，Ｆの弱 Occam アルゴリズムでもある。

５．学習可能性と情報圧縮可能性

４章では，学習過程と情報圧縮過程との等価性について議論したが，本章では関数のクラスの学習可能性と情報圧縮可能性，すなわち関数のクラスFの学習過程が存在するということと，Fの情報圧縮過程が存在するということの等価性について議論する。

Board, Pitt は，ある条件のもとで，必ずしも Occam アルゴリズムとは限らない多項式時間学習アルゴリズムから確率的多項式時間 Occam アルゴリズムを構成することによって，確率的多項式時間アルゴリズムによる定式化の下で，関数のクラスの情報圧縮可能性と学習可能性が等価であることを示している。本章では，この結果を決定性多項式時間アルゴリズムによる定式化の下でも成立するように拡張した。

［定理５．１］Board, Pitt の結果と同じ条件の下で，Fが多項式時間学習可能ならば，Fの（決定性）多項式時間弱 Occam アルゴリズムを構成することができる。

審 査 結 果 の 要 旨

学習は，大量のデータからその背後にある規則を推論（学習）する過程とみなすこともできる。これまで，この推論メカニズムを解明しようとする試みが続けられているが，そのメカニズムの本質はいまだ明らかにされていない。本論文は，この推論過程を，推論に要する時間も考慮に入れて，計算論の立場からPAC学習としてモデル化し，3層パーセプトロンなどにみられるような，学習能力を増大させる前処理過程の効果や，学習過程と情報圧縮過程の関連を論じたもので，全編6章より成る。

第1章は序論である。第2章では，PAC学習モデルや学習可能性の定義を与え，学習過程が学習アルゴリズムや学習関数として定式化できることを述べている。

第3章では，3層パーセプトロンなどにみられるように，前処理により学習能力が増大する現象に着目し，これを帰着の概念に基づき定式化し，能力の増大がVappnik-Chervonenkis次元を用いて説明できることを示している。また，一般に2段階処理に対応して，2つのクラスを結合することによって得られる関数のクラスを定義し，これが多項式時間で学習可能となるための条件を与えている。

第4章では，PAC学習モデルのもとで，情報圧縮過程をOccam関数や弱Occam関数として定式化し，それと学習関数との関係を論じている。すなわち，Occam関数は直ちに無矛盾なPAC学習関数となるが，逆は必ずしも成立しないことを示している。さらに，保存的という，学習アルゴリズムに対する概念を新しく導入し，保存的なアルゴリズムが，直ちに弱Occamアルゴリズムとなるための条件を与えている。これらは，優れた結果といえる。

第5章では，上で述べた逆の命題が成立するように，PAC学習関数からOccam関数を構成する問題について論じている。すなわち，ある条件のもとでは，PAC学習関数から弱Occam関数が構成できることを，同様に，多項式時間学習アルゴリズムから弱Occamアルゴリズムが構成できることを示している。

第6章は結論である。

以上要するに本論文は，Vapnik-Chervonenkis次元を用いて学習能力の増大するメカニズムを明らかにすると共に，学習と情報圧縮がある意味で等価な概念であることを示すなど，計算論の立場から学習のメカニズムの一端を解明したもので，情報工学の発展に寄与するところが少なくない。

よって，本論文は工学博士の学位論文として合格と認める。