ベイズ統計学に基づくOn-line学習モデルと学習可能性

3AH－1

中澤 真 †　松嶋 敏泰 †　平澤 茂一 †

† 早稲田大学理工学部経営システム工学科

1 はじめに

計算論的学習理論のパラダイムの１つであるOn-line学習モデルでは，重みを用いた学習アルゴリズムについて多くの研究がなされてきた．WINNOW [3] はその代表的なアルゴリズムの１つであり，ブール関数のクラスにおいてリテラルや節などに対応する重みを用いて学習を行なう．一方 Weighted Majority Algorithm (WMA) [4] は重みを個々の概念に対応させて学習するアルゴリズムである．これらのアルゴリズムの長所として，概念クラスから任意の１つの仮説を選択して予測するアルゴリズムよりも優れた性能を持つことが，ヒューリスティックな方法で示されている．

この性質はベイズ統計学の枠組を用いることにより，厳密に証明することができる．D.Haussler ら [1] は計算論的学習理論の立場からベイズ的学習モデルを研究し，ベイズアルゴリズムの sample complexity についての解析を行なっている．また情報理論の考え方を取り入れた解析結果も明らかにしている [2]．このベイズアルゴリズムはベイズ統計学に立脚した最適性を保証するという優れた性能を持っているが，問題の領域が大きくなると事後確率の更新や平均損失を最小にする決定を探索する計算量が膨大になるという問題点もある．

計算量の問題は，もう１つの学習パラダイムである Probably Approximately Correct (PAC) 学習モデル [6] では，多くのクラスについての学習可能性が明らかにされているが，On-line 学習モデルの予測可能性についての議論は不十分である．また PAC 学習モデルにおいても，従来の学習可能性の議論のみではネガティブな結果しか得られず，現実問題との乖離が大きい．

本稿はこれらの点を考慮し，ベイズアルゴリズムを考える場合に sample complexity だけでなく time complexity も考えた学習モデルの枠組を提案する．このモデルではベイズ統計学と計算量理論の両方の視点からモデル化を行ない，PAC 学習モデルだけでなく On-line 学習モデルについても同様に扱うモデルであることを明らかにする．さらにベイズ予測可能性という新しい概念を提案し，典型的な概念クラスについて予測可能性を明らかにし，その性質についても述べる．

なお，本文ではブール関数を対象とするが，q元，$q > 2$，の事例空間，概念表現空間のクラスへの拡張は可能である．

2 準備

まず最初に，学習モデルを定義する．事例空間を X，概念表現のクラスを Rで表す．これらは，それぞれアルファベット，Σ_E, Σ_C の有限系列の集合であり，$X \subseteq \Sigma_E^*$，$R \subseteq \Sigma_C^*$ となる．$r \in R, x \in X$ に対し，rがxを受理するか否かを計算する関数を $A(r,x) : R \times X \rightarrow \{0,1\}$ とする．計算量を考える場合，当然この関数が $|x|, |r|$ の多項式のオーダにならなくてはならない．ここで，$|\cdot|$ は文字列の長さとする．

$\mathcal{P}$ を R 上の確率分布族，$\mathcal{G}$ を X 上の確率分布族とする．概念クラス $\mathcal{C}$は概念表現 Rとその空間上の確率分布族 $\mathcal{P}$の対によって構成され，$\mathcal{C} = (R, \mathcal{P})$ と表される．概念クラスをこのように定義するのは，PAC 学習モデルと大きく異なる点である．

学習アルゴリズム L は 確率分布 G に従って生起する x とそのラベル l の対の形で情報を受け取る．ラベル l は真の概念表現 r^*に依存して決定され，$l = A(r^*, x)$ となる．この対をサンプルとよび，i 番目のサンプルを (x_i, l_i) で表す．また m 個のサンプルを $S^m =< (x_1, l_1), (x_2, l_2), \ldots, (x_m, l_m) >$ で表記する．ただし,i, m は任意の正整数とする．

3 ベイズ学習可能性

従来の計算論的学習理論において，計算可能性は sample complexity が多項式でバウンドされることを要求していた．しかしベイズアルゴリズムでは任意の時点における最適性，ここでは平均損失最小の意味，が保証されているため，sample complexity についての要求は不要になる．これはベイズ流の枠組で学習可能性を考えるとき，サンプル当たりの time complexity が計算可能であるベイズアルゴリズムの存在を要求することによって，従来の sample complexity についての条件をより厳しく設定していることに他ならない．そこでベイズ学習可能性を以下のように定義する．

定義 3.1 概念クラスを $\mathcal{C} = (R, \mathcal{P})$，損失関数を $l(r,h), r, h \in R$ と定義する．このとき任意の $p \in \mathcal{P}$に対し，事後確率による平均損失を最小にし，その計算時間が n の多項式でバウンドされるアルゴリズムが存在するとき，概念クラス $\mathcal{C}$がベイズ学習可能であるという．特に損失関数が $l(r,h) = \sum_{x \in X} |A(r,x) - A(h,x)|$ の場合ベイズ予測可能という．

3.1 重みと事前分布

WINNOW や WMA における重みは概念表現クラス上の部分集合にヒューリスティックな重みを対応させている．一方ベイズ統計学に基づいてモデル化すると，重みは概念表現クラス上の確率分布のパラメータと解釈できる．

この重みが事前分布の complexity を決定する要因となる．

厳密に述べると，この概念表現のクラス上の部分集合の集合族を $\mathscr{R}$で表す．すなわち $\mathscr{R} \subseteq 2^R$である．この $\mathscr{R}$によってベイズアルゴリズムの計算量は大きく影

---

† " Based on Bayes Statistics the On-line Learning Model and its Learnability "
M.Nakazawa, T.Matsushima and S.Hirasawa
Department of Industrial and Management Systems Engineering, School of Science and Engineering, WASEDA University
e-male: nakazawa@hirasa.mgmt.waseda.ac.jp

響することになる．本稿のモデルはこの2つの空間に依存して complexity が決定される．

4 代表的クラスのベイズ予測可能性

4.1 単調連言形・選言形

定義 4.1 Γ_iを i 番目のリテラル x_iを含むすべての概念表現の集合とし,$\forall i, \Gamma_i \in \mathscr{R}$となるように $\mathscr{R}$を構成する．集合Γ_iの中に真の概念が含まれる確率を $P(\Gamma_i)$ で表すと，概念表現のクラスの確率分布族はΓ_iの積集合上で定義される．この分布族を $\mathcal{P}_l$で表し，リテラルに依存する分布族という． □

定理 4.1 事例空間を $X = \{0,1\}^n$とするとき，概念クラス $C = (R_{MC}, \mathcal{P}_l)$ はベイズ予測可能である． □

補題 4.1 概念クラス $C = (R_{MC}, \mathcal{P}_l)$ に対し，サンプル S^mが学習者に与えられたもとで事後確率を計算する手続きは n の多項式時間で上界される． □

補題 4.2 概念クラス $C = (R_{MC}, \mathcal{P}_l)$ に対し，事後確率のもとで平均損失を最小にする仮説を出力するのに必要な計算時間は n の多項式で上界される． □

系 4.1 定理 *4.1*と同様な結果が単調選言形についても成り立つ． □

4.2 k-CNF・k-DNF

定義 4.2 リテラルの数が k個の節に対し，インデックス i を与え，その節をψ_iで表す．
$\Gamma_i = \{r | r \in R_{k-CNF}, r \rightarrow \psi_i\}$ と定義し，
$\forall i, \Gamma_i \in \mathscr{R}$となるように $\mathscr{R}$を構成する．集合Γ_iの中に真の概念が含まれる確率を $P(\Gamma_i)$ で表すと，概念表現のクラスの確率分布族はΓ_iの積集合上で定義される．この分布族を$\mathcal{P}_{clause}$で表し，節に依存する分布族という． □

定理 4.2 事例空間を $X = \{0,1\}^n$とするとき，概念クラス $C = (R_{k-CNF}, \mathcal{P}_{clause})$ はベイズ予測可能である． □

系 4.2 定理 *4.2*と同様な結果が *k-DNF* についても成り立つ． □

4.3 k-clause CNF・k-term DNF

このクラスについては $\{0,1\}$-損失のもとでの，ベイズ学習可能性に関する結果 [5] とは異なる結果が得られる．

定理 4.3 事例空間を $X\{0,1\}^n$とするとき，概念クラス $C = (R_{k-cCNF}, \mathcal{P}_{clause})$ はベイズ予測可能である． □

系 4.3 定理 *4.3*と同様な結果が *k-term DNF* についても成り立つ． □

以上の結果からベイズ予測可能性についてのいくつかの性質を示すことができる．以下の2つの系は，$\{0,1\}$-損失のもとでのベイズ学習可能性と共通の性質を示したものである．

系 4.4 任意の概念表現クラス R_A, R_Bが $R_A \subseteq R_B$という関係を満足し，概念クラス $(R_A, \mathcal{P})$ がベイズ予測可能であるならば，概念クラス $(R_B, \mathcal{P})$ もベイズ予測可能である． □

系 4.5 任意の概念表現クラス Rに対し，$||R||$ が n の多項式オーダ，$\mathcal{P}$ が一様分布であるとする．このとき概念クラス $(R, \mathcal{P})$ はベイズ予測可能である． □

次の系は On-line 学習モデルに依存した性質を導いたものである．

系 4.6 以下の条件を満たす分布族 $\mathscr{R}$はベイズ予測可能である．

$$||\mathscr{R}|| = polynomial\ (n)\ , \tag{1}$$

$$\forall \gamma_t \in \{\Gamma_t, \Gamma_t^C\}, \bigcap_{t \in all} \gamma_t \neq \phi\ . \tag{2}$$

□

5 まとめ

本稿はベイズ統計学の立場から最適性と計算量を考慮したベイズ学習モデルが，On-line 型の学習モデルを一般的に扱うことが可能であることを示した．このモデルではアルゴリズムが事前確率を利用するため，分布族の複雑さと仮説の複雑さの両方によって complexity が決定され，従来の学習可能性を一歩押し進めたものとなっている．

このベイズ予測可能性という新しい概念は既に提案したベイズ学習可能性の特殊な場合に対応し，代表的クラスの計算可能性については異なる場合があることを導いた．これは学習の最終的ゴールを決定する損失関数が計算可能性を大きく左右することを示している．一方ベイズ学習可能性と共通の性質についても明らかにしている．

今回代表的な概念クラスについてのベイズ予測可能性について明らかにし，その性質についても述べたが，より詳細に概念表現のクラスと分布族の関係について明らかにする必要がある．概念クラスが予測不可能となるその境界の部分を今後明らかにしていくことで，多くの問題が解決されるであろう．

参考文献

[1] Haussler,D., Kearns,M. and Schapire,R., "Bounds on the sample complexity of Bayesian learning using information theory and VC dimension," *Proc. of the Fourth Workshop on Computational Learning Theory*, pp.61-74, 1991.

[2] Haussler,D. and Barron,A., "How well do Bayes methods work for On-line prediction of {±1}values?," *Proc. 1992 NEC Symp. on Computation and Cognition*, Chapter 4, 1992.

[3] Littlestone,N., "Learning quickly when irrelevant attributes abound:A new linear-threshold algorithm," *Machine Learning*,vol.2, No.4, pp.285-318, 1987.

[4] Littlestone,N. and Warmuth,M., "The weighted majority algorithm," *Info. and Comput.*, vol. 108, pp.212-261, 1994.

[5] Nakazawa,M., Matsushima,T., Hirasawa,S., "Based on Bayesian Statistics the Computational Learning Model and its Learnability," 電子情報通信学会 技術研究報告, COMP97, No.157, pp.71-78, 1997.

[6] Valiant,L., "A theory of learnable," *Commun. ACM*, vol.27, No.11, pp.1134-1142, 1984.