確率的考察に基づくブール関数学習則の最適化 †

4 J − 7

中澤 真 ††　　松嶋 敏泰 ††　　平澤 茂一 ††

†† 早稲田大学理工学部工業経営学科

1 はじめに

Littlestone は線形分離可能なブール関数（特に単調連言形，単調選言形）を対象として，リテラルの重み更新による学習アルゴリズム WINNOW1[3] を提案した．しかし計算論的な立場が強いため，重みの更新法についての理論的根拠が希薄であった．これに対し野村らは統計的決定論の立場から最適性を考慮した予測アルゴリズム，重み更新アルゴリズム [1] を提案した．

本研究はこの立場に基づき，学習対象を非単調連言形・選言形のクラスまで拡張した場合の学習アルゴリズムを最初に提案する．さらに学習対象をブール関数全体のクラスに対応するアルゴリズムを示し，これらのアルゴリズムの最適性についてベイズ統計学の立場から明らかにする．最後に概念学習の枠組みに適用した場合について論じる．

2 準備

事例空間	:	$X = \{0,1\}^n$ n は命題変数の最大数
事例	:	$\boldsymbol{x} = (v_1, v_2, \cdots, v_n) \in X$ v_i は命題変数
出力空間	:	$Y = \{0,1\}$
概念クラス	:	$\mathcal{F}$
ブール関数	:	$f(\boldsymbol{x}) \in \mathcal{F} : X \to Y$
真の真理値	:	$y = f(\boldsymbol{x})$
学習者の予測	:	$\lambda(\boldsymbol{x}) : X \to Y$
サンプル	:	$s = (\boldsymbol{x}, y) \in X \times Y$

3 On-Line learning model

3.1 学習パラダイム

この学習パラダイムは Littlestone[3] によって提案されたもので，その流れは以下のようになる．

1. 学習者はランダムに選ばれた事例を受け取る．ただし n は固定である．
2. 学習者は事例の真理値について予測を行う．
3. 教師から真の真理値を教えてもらい，自分の仮説を修正する．

3.2 評価基準

このモデルにおける評価基準として Littlestone は mistake bound を用いているが，予測の正確さについての最適性を論じるためには，野村らが用いた平均予測誤り確率で評価することが望ましい．本研究ではこの立場により議論を進める．

† 本研究の一部は文部省科学研究費 試験研究 B 07558168, 早稲田大学特定課題研究 95A-267 の助成による．

" Optimum Learning Rules for Boolean functions based on Bayesian Statistics "

M.Nakazawa, T.Matsushima and S.Hirasawa
Department of Industrial Engineering and Management, School of Science and Engineering, Waseda University
e-male: nakazawa@hirasa.mgmt.waseda.ac.jp

3.3 非単調連言形

野村らのアルゴリズムは概念クラスとして単調な連言形･選言形を対象としていたが，本研究はこれをさらに一般的なクラスへと拡張する．

3.3.1 概念クラス

定義 3.1 (非単調連言形) このクラスは任意の正のリテラル（否定記号を持たない）と負のリテラル（否定記号を持つ）の連言で表現される関数の族で以下のように定義する．

$f(v_1, \ldots, v_n) = v_{\eta_1} \wedge \cdots \wedge v_{\eta_j} \wedge \neg v_{\eta_{j+1}} \wedge \cdots \wedge \neg v_{\eta_{j+k}}$
ただし，$\eta_i \in \{1, \ldots, n\}$

3.3.2 重み

Littlestone の重みは理論的な意味付けがなされていなかったが，これに対し野村らは統計学に基づいて重みを定義した．この定義による重みは各リテラルが真の概念に含まれている確率を表している．

野村らの方式では正のリテラルに対応する重みのみを定義したが，本研究では負のリテラルに対応する重みも定義する．

定義 3.2

P_i : i 番目の正のリテラル v_i に対応する重み
Q_i : i 番目の負のリテラル $\neg v_i$ に対応する重み
ただし，$P_i, Q_i \in [0,1]$

3.3.3 予測アルゴリズム

$$\lambda(\boldsymbol{x}) = \begin{cases} 1 : \prod_{i \in \{j|v_j=0\}}(1-P_i) \prod_{i \in \{j|v_j=1\}}(1-Q_i) > \frac{1}{2} \\ 0 : \prod_{i \in \{j|v_j=0\}}(1-P_i) \prod_{i \in \{j|v_j=1\}}(1-Q_i) < \frac{1}{2} \end{cases}$$

ただし，$\prod_{i \in \{j|v_j=0\}}(1-P_i) \prod_{i \in \{j|v_j=1\}}(1-Q_i) = \frac{1}{2}$ の場合は確率 $\frac{1}{2}$ のランダム決定．

3.3.4 重み更新アルゴリズム

	v_i	y	更新結果 $P_i^{(t+1)}$
①	0	0	$\dfrac{P_i^{(t)}}{1-\prod_{i \in \{j|v_j=0\}}(1-P_i^{(t)}) \prod_{i \in \{j|v_j=1\}}(1-Q_i^{(t)})}$
②	0	1	0
③	1	0	$P_i^{(t)}$
④	1	1	$P_i^{(t)}$

3.4 ブール関数全体 (CNF 表現)

この節ではブール関数全体を概念クラスとしたときの学習アルゴリズムを示す．このアイデアは Valiant[4] の考え方を利用している．

3.4.1 重み

定義 3.3 重複を許さず n 個のリテラルを持つ全ての節を作り，各節を A_i とする．このとき節 A_i と事例 $\boldsymbol{x}$ の関係を表す関数を次のように定義する．

$$\alpha_i(\boldsymbol{x}) = \begin{cases} 0: & \text{事例}\boldsymbol{x}\text{によって節}A_i\text{が充足されない} \\ 1: & \text{事例}\boldsymbol{x}\text{によって節}A_i\text{が充足される} \end{cases} \tag{1}$$

また，R_i を節 A_i に対応する重みとする．

3.4.2 予測アルゴリズム

$$\lambda(\boldsymbol{x}) = \begin{cases} 1 & : \prod_{i\in\{j|\alpha_j(\boldsymbol{x})=1\}} R_i > \frac{1}{2} \\ 0 & : \prod_{i\in\{j|\alpha_j(\boldsymbol{x})=1\}} R_i < \frac{1}{2} \end{cases}$$

ただし，$\prod_{i\in\{j|\alpha_j(\boldsymbol{x})=1\}} R_i = \frac{1}{2}$ の場合は確率 $\frac{1}{2}$ のランダム決定

3.4.3 重み更新アルゴリズム

$\alpha_i(\boldsymbol{x})$	y	更新結果 $R_i^{(t+1)}$
0	0	$\frac{R_i^{(t)}}{1-\prod_{i\in\{j|\alpha_j(\boldsymbol{x})=0\}}(1-R_i^{(t)})}$
0	1	0
1	0	$R_i^{(t)}$
1	1	$R_i^{(t)}$

3.5 性質

定理 3.1 重み更新アルゴリズムによって更新された P_i は，新たな事例 $\boldsymbol{x}$ が与えられた下でリテラル v_i が真の概念に含まれる事後確率となっている．

系 3.1 Q_i, R_i の更新則についても P_i と同様のことがいえる．

定理 3.2 上述の重み更新アルゴリズムと予測アルゴリズムによる予測は平均予測誤り確率を最小にする．

4 Concept Learning Model

計算量的学習理論で古くから研究されているこのモデルは On-Line Learning のように予測の正確さを追求するのではなく，いかに真の概念を導出するかということに主眼が置かれている．このモデルの中でも例からの学習と呼ばれているモデルは，ランダムに得られたサンプルに基づいて学習を行う．統計的決定理論の見地からこの 2 つのモデルを見た場合，その差違は損失関数の違いによるものであることが分かる．詳細は文献 [2] を参照されたい．

4.1 評価基準

真の概念を導出したか否かについての平均誤り確率により評価する．

4.2 学習アルゴリズム

Concept Learning Model では事後確率最大の概念を導出すればよいので，すべての重み更新を必ずしも必要としない場合がある．計算量的学習理論での効率性は概念クラスの構造のみに依存していたが，本研究モデルでは事前確率の構造にも依存する．この構造を利用した効率的なアルゴリズムを以下に示す．

定義 4.1 (seed) 節 A の中に含まれている任意のリテラルを 1 つだけ除いて構成される節 B を節 A の *seed* と呼ぶ．また，節 A を節 B の *parent* と呼ぶ．

仮定 1
R_i を節 A_i に対応する重みとする．
このとき各節の重みの初期値（事前確率）は，節 A_i が節 A_j の *seed* であるならば，$R_i > R_j$ であると仮定する．

```
procedure CNF-LEARNING;
begin
    S+ :=正例の集合; S- :=負例の集合;
    while OPEN の中の節が S+ に関して矛盾 do
    begin
      矛盾する節 A を OPEN から取り除き，
          CLOSED に入れる;
      if A の親となる節 A' の seed が
          すべて CLOSED に存在する
        then  OPEN に A' を加える
    end;
    repeat すべての s ∈ S- について  do
      repeat すべての Ai ∈OPEN について  do
            Ri についての重み更新;
      until
    until
    repeat すべての Ai ∈OPEN について  do
      if (Ri > 1/2) then  add Ai to CNF;
    until
    output  CNF
end.
```

4.3 性質

補題 4.1 仮定 1 の条件の下で，更新される事後確率は次の式を満足する．
$R_i^{t+1} > R_j^{t+1}$

定理 4.1 CNF-LEARNING アルゴリズムは事後確率最大の概念を導出する．

5 まとめ

本稿では，従来のブール関数学習則をより広いクラスに対応するアルゴリズムへ拡張し，そのベイズ的最適性を示した．また概念学習モデルにおける効率的なアルゴリズムを示し，その最適性を証明した．

参考文献

[1] 野村，松嶋，平澤，“単調連言形の学習アルゴリズム”，*Proc. of the 14th Symp. on Information Theory and Its Applications*, 709-712, Ibusuki, Japan, 1991.

[2] 松嶋，稲積，平澤，“不確実性をもつ論理式の帰納推論に関する一考察”，情処学会論文誌，1461-1475, Vol.33, No.12, 1992.

[3] Littlestone,N., “Learning quickly when irrelevant attributes abound:A new linear-threshold algorithm”, *Mach. Learning*, 2, 285-318, 1987.

[4] Valiant,L., “A theory of learnable”, *Comm. ACM*, vol. 27 (11), 1134-1142, 1984.