

研究速報

アンサンブル学習における符号を保つ重み正規化手法に関する検討

前原 洋祐^{†*}(学生員) 内田 真人^{††}(正員)

塩谷 浩之^{†††}(正員) 大森 隆司^{††††**}(正員)

Weight-Vector-Regularization Methods without Change of Sign for Ensemble Learning

Yousuke MAEHARA^{†*}, Student Member, Masato UCHIDA^{††}, Hiroyuki SHIOYA^{†††}, and Takashi OMORI^{††††**}, Members

[†] 室蘭工業大学大学院情報工学専攻, 室蘭市

Graduate School of Computer Science and Systems Engineering, Muroran Institute of Technology, 27-1 Mizumoto, Muroran-shi, 050-8585 Japan

^{††} 九州工業大学ネットワークデザイン研究センター, 北九州市

Network Design Research Center, Kyushu Institute of Technology, 3-8-1 Asano, Kokurakita-ku, Kitakyushu-shi, 802-0001 Japan

^{†††} 室蘭工業大学工学部情報工学科, 室蘭市

Department of Computer Science and Systems Engineering, Muroran Institute of Technology, 27-1 Mizumoto, Muroran-shi, 050-8585 Japan

^{††††} 北海道大学大学院情報科学研究科, 札幌市

Graduate School of Information Science and Technology, Hokkaido University, Sapporo-shi, 060-0814 Japan

* 現在, 室蘭工業大学大学院生産情報システム工学専攻

** 現在, 玉川大学学術研究所

あらまし 指数型混合で表現されるアンサンブル予測器の確率モデルでは, 予測器の重み係数は正値に制約されない. そこで, 正負が混在した重み係数のための正規化手法を提案する. 更に, 数値実験を通じ, 提案手法の有効性について検証する.

キーワード アンサンブル学習, 確率モデル, 指数型混合, 正規化手法.

1. ま え が き

複数の予測器を混合することにより最終的な予測器を構成するというアンサンブル学習は, 汎化能力の向上に有効な学習手法として知られており, その代表的な手法には Bagging と Boosting が挙げられる. Bagging の典型的な手続きでは, 混合される予測器群に対し均一な重みが割り当てられる [1]. これは混合される予測器群の算術平均をとることに相当し, このような重み付けによるアンサンブル学習は単純アンサンブル [3] と呼ばれる. 一方, 混合される予測器群に対する重み付き平均をとるというアンサンブル学習は一般アンサンブル [3] と呼ばれ, 次のように定義される.

入力 $x \in \mathbb{R}^m$ に対して望ましい出力 $y \in \mathbb{R}$ の組 $z = (x, y)$ がある確率分布 $p^*(x, y) (= q(x)p^*(y|x))$ から独立に $\sum_{i=1}^M n_i$ 個得られたとし, それを $D_i =$

$\{z_1^i, \dots, z_{n_i}^i\}$ とおく ($z_j^i = (x_j^i, y_j^i) \sim p^*(x, y)$, $i = 1, \dots, M$, $j = 1, \dots, n_i$). また, 入力 x に対し $f_{\theta_i}(x)$ を出力とするものを予測器とする. 更に, $\theta_i \in \mathbb{R}^{k_i}$ を予測器 $f_{\theta_i}(x)$ の修正可能なパラメータとすると, 予測器の学習は

$$\hat{\theta}_i = \arg \min_{\theta_i \in \mathbb{R}^{k_i}} \sum_{(x, y) \in D_i} (y - f_{\theta_i}(x))^2 \quad (1)$$

と表され, 学習済み予測器を $f_{\hat{\theta}_i}(x)$ と表す. このとき, M 個の予測器による一般アンサンブルは, 重み係数 $\beta = \{\beta_1, \dots, \beta_M\} \in \mathbb{R}^M$ を用いて次のように表される.

$$\bar{f}_{\hat{\theta}, \beta}(x) = \sum_{i=1}^M \beta_i f_{\hat{\theta}_i}(x) \quad (2)$$

$$\beta_i > 0, \quad \sum_{i=1}^M \beta_i = 1 \quad (3)$$

ここで, $\bar{f}_{\hat{\theta}, \beta}(x)$ が混合された予測器を表す ($\hat{\theta} = \{\hat{\theta}_1, \dots, \hat{\theta}_M\}$). なお, $\beta_i = 1/M$, ($i = 1, \dots, M$) の場合が単純アンサンブルに相当する.

重み係数の符号に着目すると, 以上の定義より, 一般アンサンブルは式 (3) を制約条件としているため重み係数が負値をとることはないことが分かる. これに対し, Boosting の代表的手法である AdaBoost [2] では, 訓練データに対する学習精度の悪い予測器に対し負の重みが割り当てられることがある.

ところで, 上述の一般アンサンブルを特別な場合を含む形式として, 指数型混合形式の確率モデルを導入した枠組みがある [4]. この枠組みでは, 確率モデルとしての特徴 (体積が 1 になるという特徴) を根拠とし, 式 (3) における非負条件が課されない. また, この枠組みを用いることで, 予測精度の悪い予測器を含むアンサンブル学習では, 重み係数の非負条件を満たさない場合があることも示すことができる [5]. ところが, 同じく負の重みを許容する AdaBoost とは異なり, [4] の枠組みを背景とした重み係数の決定手法についての検討は不十分である. そこで, 新たに負の値を許した重み係数の決定手法について検討する. 特に, この枠組みは, 与えられた予測器群の (重み付き) 平均を最終的な出力とする学習法を一般化したものであるということを考慮し, 負の値を許した重み係数のための正規化手法について検討する. 更に, 数値例を用いた実験を通じて重み係数の符号について考察する.

2. 指数型混合形式によるアンサンブル学習の概要
指数型混合による確率モデルを用いたアンサンブル学習 [4] の概要を示す.

集合 \mathcal{Z} ($\subset \mathbb{R}^d$) 上の確率分布の集合を

$$\mathcal{P}(\mathcal{Z}) \stackrel{\text{def}}{=} \left\{ p \mid \mathcal{Z} \rightarrow \mathbb{R}, p(z) \geq 0 (\forall z \in \mathcal{Z}), \int_{\mathcal{Z}} p(z) dz = 1 \right\}$$

とおく. また, あるパラメータ θ_i ($\in \mathbb{R}^{k_i}$) をもつ確率分布 p_{θ_i} ($\in \mathcal{P}(\mathcal{Z})$) を用いて ($i = 1, \dots, M$), 新たな確率分布 $\bar{p}_{\theta, \beta}$ ($\in \mathcal{P}(\mathcal{Z})$) を

$$\bar{p}_{\theta, \beta}(z) \stackrel{\text{def}}{=} \frac{\prod_{i=1}^M p_{\theta_i}(z)^{\beta_i}}{\int_{\mathcal{Z}} \prod_{i=1}^M p_{\theta_i}(z)^{\beta_i} dz} \quad (4)$$

と定義する. ただし

$$\int_{\mathcal{Z}} \prod_{i=1}^M p_i(z)^{\beta_i} dz < \infty, \quad \sum_{i=1}^M \beta_i = 1$$

である (重み係数の非負条件が課されないことに注意). このとき

$$\hat{\theta}_i \stackrel{\text{def}}{=} \arg \min_{\theta_i \in \mathbb{R}^{k_i}} D(p^* \| p_{\theta_i}) \quad (5)$$

$$\bar{p}_{\hat{\theta}, \beta}(z) \stackrel{\text{def}}{=} \arg \min_{p \in \mathcal{P}(\mathcal{Z})} \sum_{i=1}^M \beta_i D(p \| p_{\hat{\theta}_i}) \quad (6)$$

$$\hat{\beta} \stackrel{\text{def}}{=} \arg \min_{\beta \in \mathbb{R}^M} D(p^* \| \bar{p}_{\hat{\theta}, \beta}) \quad (7)$$

と定義される, Kullback divergence $D(\cdot \| \cdot)$ に関する 3 段階の最小化操作を考える. ここで, 最終的に推定された確率分布は $\bar{p}_{\hat{\theta}, \hat{\beta}}$ で表される. なお, 式 (6) は式 (4) を用いて定義可能である. この最小化操作は, 式 (1), (2) で表される一般アンサンブルを特別な場合に入れている. このことは, 以下のようにして分かる.

まず, ある関数 $g(x)$ ($\in \mathbb{R}$) を用いて, x に対する y の条件付確率分布を

$$p_G(y|g(x)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-g(x))^2}{2\sigma^2}\right\}$$

と定め, $p_{\theta_i}(z) = p_G(y|f_{\theta_i}(x))q(x)$ とおく. ただし, σ^2 は正定数である. ここで, n_i ($i = 1, \dots, M$) が十分大きいとすると, 式 (1) は式 (5) に対応する. また, $\bar{p}_{\theta, \beta}(z) = p_G(y|\bar{f}_{\theta, \beta}(x))q(x)$ が成り立つことが

ら, 式 (2) は式 (4), (6) に対応する. 更に, 式 (7) は重み係数 β を次のような最小二乗法により学習することに対応している.

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^M} \sum_{(x, y) \in D_0} \left(y - \sum_{i=1}^M \beta_i f_{\hat{\theta}_i}(x) \right)^2 \quad (8)$$

ここで, $D_0 = \{z_1^0, \dots, z_{N_0}^0\}$ ($z_i^0 = (x_i^0, y_i^0) \sim p^*(x, y), i = 1, \dots, N_0$) は重み係数 β を学習するための例題である.

3. 重み係数の正規化

3.1 目的

式 (8) により得られる $\hat{\beta}$ は, $\sum_{i=1}^M \hat{\beta}_i = 1$ を満たす保証がない. このとき, あらかじめ用意された各予測器の (重み付き) 平均をとるといふ, アンサンブル学習における本来の意味を反映するためには, 重み係数の総和が 1 になるよう $\hat{\beta}$ の値を正規化する必要がある.

よく用いられる正規化手法としては, 正規化される値の総和で各値を割るという操作がある. しかし, この操作ではすべての重み係数が正の値である場合には合理的であるものの, 負の値の重み係数が含まれる場合には重み係数の符号が正規化後に反転してしまう場合がある. 負の値が割り振られている予測器は, アンサンブルを行う予測器の中で否定的に扱うことで全体の予測精度の向上に貢献していると考えられるため, 正規化により符号が反転することは, 予測器を肯定的, 若しくは否定的に扱うという意味も反転してしまうということになる.

以上を考慮し, 本論文では, 重み係数の符号が正規化の前後で反転しない手法を提案する (3.3, 3.4). なお, それらとの比較・検証を行うために, 重み係数の符号が正規化の前後で必ずしも保存されない手法ではあるが, ラグランジュの未定乗数法を用いた正規化手法を示す (3.2).

3.2 ラグランジュの未定乗数法による正規化

$\|\cdot\|$ をユークリッドノルムとする. 制約条件 $\sum_{i=1}^M \bar{\beta}_{L,i} = 1$ のもとで, 目的関数 $\|\bar{\beta}_L - \hat{\beta}\|^2$ を最小化する $\bar{\beta}_L$ をラグランジュの未定乗数法を用いて求める.

まず

$$F(\bar{\beta}_L, \lambda) = \|\bar{\beta}_L - \hat{\beta}\|^2 - \lambda \left(\sum_{i=1}^M \bar{\beta}_{L,i} - 1 \right)$$

とおく．このとき

$$\frac{\partial F}{\partial \beta_{L,i}} = 0, \quad (i = 1, \dots, M), \quad \frac{\partial F}{\partial \lambda} = 0$$

を $\bar{\beta}_L$ と λ の連立方程式として解くと

$$\bar{\beta}_{L,i} = \hat{\beta}_i + \frac{1 - \sum_{j=1}^M \hat{\beta}_j}{M}, \quad (i = 1, \dots, M) \quad (9)$$

が得られる．このようにして得られた $\bar{\beta}_L$ は $\sum_{i=1}^M \bar{\beta}_{L,i} = 1$ を満たす．しかし，この手法により正規化された重み係数は更新前と更新後の重み係数に関して，符号が保たれない場合がある．正規化前後で符号が保たれない条件は，式 (9) より

$$0 < \hat{\beta}_i < \frac{\sum_{j \neq i} \hat{\beta}_j - 1}{M - 1} \quad (10)$$

若しくは

$$0 > \hat{\beta}_i > \frac{\sum_{j \neq i} \hat{\beta}_j - 1}{M - 1} \quad (11)$$

を満たす場合であることが分かる．なお， $\hat{\beta}$ ， $\bar{\beta}_L$ ，任意の β ($\sum_{i=1}^M \beta_i = 1$) に関して次のような性質がある（導出については付録参照）．

$$\|\beta - \hat{\beta}\|^2 = \|\beta - \bar{\beta}_L\|^2 + \|\hat{\beta} - \bar{\beta}_L\|^2 \quad (12)$$

このことから， $\hat{\beta}$ ， $\bar{\beta}_L$ ， β をそれぞれ結ぶと， β と $\hat{\beta}$ を結ぶ辺を斜辺とする直角三角形となり， $\bar{\beta}_L$ は， $\hat{\beta}$ から $\{\beta\} \sum_{i=1}^M \beta_i = 1$ への正射影であることが分かる．

3.3 符号が保たれる正規化 1

正規化後の重み $\bar{\beta}_{N_1}$ ($\in \mathbb{R}^M$) を次のように表す．

$$\bar{\beta}_{N_1,i} = \begin{cases} \hat{\beta}_i \left(1 - \sum_{\hat{\beta}_j < 0} \hat{\beta}_j\right) \left(\sum_{\hat{\beta}_k \geq 0} \hat{\beta}_k\right)^{-1} & (\text{if } \hat{\beta}_i \geq 0) \\ \hat{\beta}_i & (\text{if } \hat{\beta}_i < 0) \end{cases}$$

この手法では，正規化前後で符号が保たれ，かつ正の値をもつ重み係数についてのみが更新される．なお， $\hat{\beta}_i > 0$ のとき， $\sum_{j=1}^M \hat{\beta}_j < 1$ ならば $\bar{\beta}_{N_1,i} > \hat{\beta}_i$ であり， $\sum_{j=1}^M \hat{\beta}_j > 1$ ならば $\bar{\beta}_{N_1,i} < \hat{\beta}_i$ であるという関係がある．

3.4 符号が保たれる正規化 2

正規化後の重み $\bar{\beta}_{N_2}$ ($\in \mathbb{R}^M$) を

$$T \sum_{\hat{\beta}_j \geq 0} \hat{\beta}_j + T^{-1} \sum_{\hat{\beta}_k < 0} \hat{\beta}_k = 1 \quad (13)$$

を満たすような $T > 0$ を用いて次のように表す．

$$\bar{\beta}_{N_2,i} = \begin{cases} T \hat{\beta}_i & (\text{if } \hat{\beta}_i \geq 0) \\ T^{-1} \hat{\beta}_i & (\text{if } \hat{\beta}_i < 0) \end{cases}$$

この手法では，正規化前後で符号が保たれ，かつすべての重み係数が変更される．なお， $\sum_{i=1}^M \hat{\beta}_i > 1$ のとき， $T < 1$ すなわち $\bar{\beta}_{N_2,i} < \hat{\beta}_i$ ($i = 1, \dots, M$) となる．また， $\sum_{i=1}^M \hat{\beta}_i < 1$ のとき， $T > 1$ すなわち $\bar{\beta}_{N_2,i} > \hat{\beta}_i$ ($i = 1, \dots, M$) となる．

4. 数値実験

本章では，数値実験を通して，各正規化手法により得られる重み係数が学習特性とどのようにかかわっているのかということについて検討する．また，一般アンサンブルにおける制約条件である式 (3) を満たす重み係数の一例として

$$\tilde{\beta}_i = \frac{\exp(-E_i)}{\sum_{j=1}^M \exp(-E_j)} \quad (14)$$

$$E_i = \sum_{j=1}^N (y_j^0 - f(x_j^0; \hat{\theta}_i))^2$$

を数値実験の際の比較に用いる．それに加え，よく用いられている手法である AdaBoost との比較も行う．

以下では，簡単な関数識別問題に加え，実問題の例として近年深刻化するスパムメール [6] の識別問題を取り扱う．関数識別問題は， $x \in [0, 1]$ ， $y \in [-1.5, 1.5]$ の組 (x, y) を入力とし， $y > \sin(2\pi x)$ にはラベル 1， $y \leq \sin(2\pi x)$ にはラベル 0 を割り当てるものとする．また，スパムメールの識別問題に用いるデータは The UCI Repository of Machine Learning Databases [7] のものを用い，入力データは，電子メールの本文中における，48 種の特定の語が含まれる割合，6 種の特定の文字が含まれる割合，大文字の文字列の平均長と最大長，大文字の総数の計 57 入力である．なお，各アンサンブル学習における個々の予測器には，多入力 1 出力の 3 層パーセプトロンを利用した (表 1 参照)．

4.1 実験 1

重み係数の符号に着目した比較実験を行う．予測

表 1 実験条件
Table 1 Experimental conditions.

		関数識別	スパムメール
入力数 (m)		2	57
中間素子数		5	7
予測器数 (M)		10	5
データ数	データ総数	2601	4601
	予測器の学習	1000	500
	重み係数の学習	100	1000
	予測精度の評価	1501	3101

表 2 実験 1 の結果 (各学習法における予測精度)
Table 2 Result of experiment 1.

	関数識別	スパムメール
各予測器の平均	82.95 %	85.49 %
$\bar{\beta}$ を用いた場合	88.27 %	89.34 %
AdaBoost	91.63 %	90.26 %
$\hat{\beta}$ を用いた場合	93.20 %	90.94 %
$\bar{\beta}_L$ を用いた場合	93.00 %	90.94 %
$\bar{\beta}_{N_1}$ を用いた場合	93.00 %	90.94 %
$\bar{\beta}_{N_2}$ を用いた場合	93.00 %	90.94 %

器の学習・重み係数の学習・予測精度の評価のためのデータは、それぞれ母集団となるデータから重複を許さず無作為抽出されたものである (表 1 参照)。また、各予測器の学習の際には、パラメータの初期値を変えた上で、予測器の学習のためのデータを共通に用いた。なお、AdaBoost については予測器の学習・重み係数の学習を行うためのデータを合わせて用いて学習を行った。表 2 の実験結果より $\hat{\beta}$ を用いたアンサンブル学習の予測精度は、各予測器の予測精度の平均、 $\bar{\beta}$ を用いたアンサンブル学習の予測精度よりも高くなっていることが分かる。また、 $\hat{\beta}$ を更新して得られた $\bar{\beta}_L$, $\bar{\beta}_{N_1}$, $\bar{\beta}_{N_2}$ については、 $\hat{\beta}$ の要素の和が 1 に近い値であったため (関数識別では 0.952, スパムメールでは 1.006), 正規化の前後で予測精度が保たれる結果となった ($\bar{\beta}_L$ については正規化の前後で重み係数の符号も保たれた)。一方、AdaBoost では両問題において、すべての予測器に正の値が割り当てられていたが、 $\hat{\beta}$ には負の値が含まれていた。これらの実験結果より、負の重み係数を利用したアンサンブル学習が予測精度の向上に貢献し得るということが分かった。

4.2 実験 2

実験 1 では、各正規化手法を用いたアンサンブル学

習に関して精度の違いは見られなかった。そこで、 $\bar{\beta}_L$ について正規化後に符号が反転する要素が存在する場合について、表 1 の条件で計算機実験を行った。まず、ある一つの予測器の学習に用いるデータについてのみ、ノイズデータ (出カラベルが反転したデータ) の混合率を 10 % ずつ増やし、そのデータを学習させた 10 パターンの予測器を用いることとした。また、意図的に $\bar{\beta}_L$ について正規化後に符号が反転する要素が存在するように、 $\hat{\beta}$ について次のような処理を行ってから正規化することとした。

[Step 1] 式 (10), (11) を満たすような $\hat{\beta}_i$ ($i = 1, \dots, M$) が存在する場合、すなわち、 $\bar{\beta}_L$ について正規化後に符号が反転する要素が少なくとも一つ存在する場合は処理を終了する。逆に一つも存在しない場合は Step 2 へ進む。

[Step 2]

$$c = \arg \min_{1 \leq i \leq M} \left| \hat{\beta}_i - \frac{\sum_{j \neq i} \hat{\beta}_j - 1}{M - 1} \right|$$

として、 $\hat{\beta}_c \geq 0$ かつ $\sum_{j \neq c} \hat{\beta}_j > 1$, または $\hat{\beta}_c < 0$ かつ $\sum_{j \neq c} \hat{\beta}_j < 1$ であれば

$$\hat{\beta}_c = \frac{\tau \left(\sum_{j \neq c} \hat{\beta}_j - 1 \right)}{M - 1}, \quad (0 < \tau < 1) \quad (15)$$

と更新し処理を終了する。これにより、少なくとも $\hat{\beta}_c$ の符号は正規化後に反転する。

[Step 3] $\hat{\beta}_c \geq 0$ かつ $\sum_{j \neq c} \hat{\beta}_j < 1$, または $\hat{\beta}_c < 0$ かつ $\sum_{j \neq c} \hat{\beta}_j > 1$ であれば、 $\hat{\beta}'_i$ ($i \neq c$) を

$$\hat{\beta}'_i = \hat{\beta}_i + \mu \left(1 - \sum_{j \neq c} \hat{\beta}_j \right), \quad (\mu > 0)$$

と求め、 $\hat{\beta}'_i \hat{\beta}_i > 0$ ($i \neq c$) を満たしていれば (すなわち、 $\hat{\beta}'_i$ と $\hat{\beta}_i$ の符号が同じであれば) $\hat{\beta}_i = \hat{\beta}'_i$ とする。このとき、 $\hat{\beta}_c \geq 0$ かつ $\sum_{j \neq c} \hat{\beta}_j > 1$, または $\hat{\beta}_c < 0$ かつ $\sum_{j \neq c} \hat{\beta}_j < 1$ であれば式 (15) を用い $\hat{\beta}_c$ を更新する。そうでなければ、再び Step 3 を実行する。

$\tau = 0.1$, $\mu = 0.1$ とし、実験を行った結果 (図 1 参照), $\bar{\beta}_L$ について正規化前後で符号が反転する要素が存在する場合には、正規化後に符号が反転しない $\bar{\beta}_{N_1}$ と $\bar{\beta}_{N_2}$ を用いたアンサンブル学習の方が、 $\bar{\beta}_L$ を用いたアンサンブル学習よりも若干ではあるが予測精度が高いことが分かる。したがって、本実験においては、

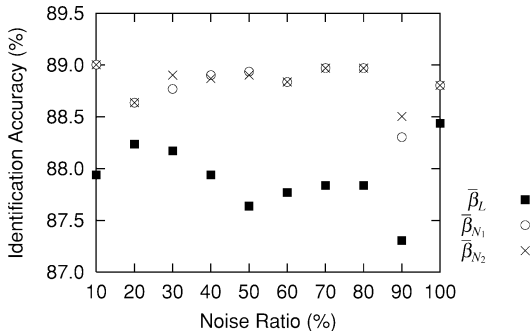


図1 実験2結果: ノイズ混合率 vs. 識別精度
Fig.1 Result of experiment 2: Noise ratio vs. identification accuracy.

実験1とは異なり, $\bar{\beta}_L$ を用いたアンサンブル学習について正規化後に符号が反転する要素が存在する場合には, $\bar{\beta}_{N_1}$, $\bar{\beta}_{N_2}$ を用いたアンサンブル学習よりも $\bar{\beta}_L$ を用いたアンサンブル学習の方が精度が低くなるという結果が得られた。

ところで, 式(12)の解析結果より, 正規化後の重みについて, $\bar{\beta}_L$ は $\bar{\beta}_{N_1}$, $\bar{\beta}_{N_2}$ よりもベクトル間の距離という点においては $\hat{\beta}$ に近いということが分かっている。しかしながら, 以上に示した実験2の結果は, ベクトル間の距離を近づけることよりも, 正規化により符号が保存されることの方がアンサンブル学習においては重要になる場合があるということを示唆している。これは, 正規化により符号が保存されることがアンサンブル学習における各予測器の役割を保存することの重要性を意味すると考えられる。

5. むすび

指数型混合形式のアンサンブルモデルに基づいた重み付きアルゴリズムと, それに伴う負の値を含む正規化手法の提案・解析, 及び数値実験を行った。その結果から, アンサンブル学習として負の値を含む重み付き平均をとる場合, 正規化により符号を保つことが重要であることが分かった。これは, 正規化の前後で重み係数の符号を保存することが, アンサンブル学習における各予測器の役割を保存することを意味するためであると考えられる。

一般に符号を保つ正規化は多数考えられるが, 本論文で示した正規化1及び正規化2は, ベクトル $\hat{\beta}$ の要素和を非負の要素の和と負の要素の和に分け, $\hat{\beta}$ に依存した数をそれぞれに乗算することで, 全体の和を1に調整するという意味で共通な手法である。更に, どちらの正規化においても, $\hat{\beta}$ の要素がすべて非負であ

る場合には, 要素の総和で除算する通常の正規化に帰着される。よって, 正規化1及び正規化2の手法は, そのような通常の正規化手法の拡張として位置づけられる。

アンサンブル学習において扱われる負を許容する重みベクトルの空間は, 遺伝的アルゴリズムにおける遺伝子型分布遷移の力学系において扱われており [8], 拡張された遺伝子型分布と, アンサンブル学習における重みパラメータとの関連性は興味深い。今後の課題としては, このような関連性を軸に, 拡張された混合配分のデータ学習における数値, 及び汎化誤差等との関係などが挙げられる。

文 献

- [1] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol.24, pp.123-140, 1996.
- [2] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol.55, no.1, pp.119-139, 1997.
- [3] 上田修功, 中野良平 "アンサンブル学習における汎化誤差解析" 信学論 (D-II), vol.J80-D-II, no.9, pp.2512-2521, Sept. 1997.
- [4] 内田真人, 塩谷浩之, 伊達 惇, "アンサンブル学習の解析と拡張" 信学論 (D-II), vol.J84-D-II, no.7, pp.1537-1542, July 2001.
- [5] 内田真人, 塩谷浩之, "アンサンブル学習モデルにおける重み付けに関する考察" 信学論 (D-II), vol.J86-D-II, no.7, pp.1131-1134, July 2003.
- [6] 総務省, "平成17年版情報通信白書" pp.107-109, 2005.
- [7] C. Black and C. Merz, "The UCI repository of machine learning databases," <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [8] M.D. Vose, "Modeling simple genetic algorithms," *Evolutionary Computation*, vol.3-4, pp.453-472, 1996.

付 録

式(12)の導出

式(9)より

$$\|\hat{\beta} - \bar{\beta}_L\|^2 = \frac{\left(1 - \sum_{j=1}^M \hat{\beta}_j\right)^2}{M} \quad (\text{A-1})$$

が成り立つ。また

$$\begin{aligned} \|\beta - \bar{\beta}_L\|^2 &= \sum_{i=1}^M \left(\beta_i - \hat{\beta}_i - \frac{1 - \sum_{j=1}^M \hat{\beta}_j}{M} \right)^2 \end{aligned}$$

$$= \|\beta - \hat{\beta}\|^2 + \frac{\left(1 - \sum_{j=1}^M \hat{\beta}_j\right)^2}{M} - 2 \left(\frac{1 - \sum_{j=1}^M \hat{\beta}_j}{M}\right) \sum_{i=1}^M (\beta_i - \hat{\beta}_i) \quad (\text{A}\cdot 2)$$

が成り立つ．ここで

$$\sum_{i=1}^M (\beta_i - \hat{\beta}_i) = 1 - \sum_{j=1}^M \hat{\beta}_j$$

を式 (A・2) に代入し，式 (A・1) を用いて整理すると式 (12) が得られる．

(平成 18 年 2 月 14 日受付，5 月 1 日再受付)