重み付き最近隣法における重み学習について*

3J－1

佐藤 健　岡本 青史
（株）富士通研究所

1. はじめに

これまで，事例ベース推論における類似度関数を学習する実験的な枠組は多く提案されているが，その評価は，ベンチマークを用いて行なわれることが多く，これらの手法の比較やどのような条件で効果的なのかに関する一般的な結論を導くには十分ではない．このため，類似度関数学習の理論的解析が必要である．

しかしながら，事例ベース推論における理論的解析は，類似度関数を固定して行なれており，類似度関数の学習に関する理論的解析はほとんど存在していない．そこで，事例ベース推論の類似度関数学習解析の準備として，われわれは，[2] で類似度の重みに関する PAC 学習の枠組を提案したが，事例ベースが設定されたときの重みの学習ではなかった．本論文では，事例ベース推論でよく使われている重み付き最近隣法 (Weighted Nearest Neighbor Method) における重み学習を [2] と似た方法で定義し，それに必要なデータ数と計算時間を解析する．

2. 3つ組法に基づいた重み学習

$\mathbf{P}$ を n 次元ユークリッド空間 E^n 上の任意の確率分布とする．教師は，重みベクタ W^* を $[0,\infty)^n$ から取り，任意の異なった2点 C_1, C_2 を E^n からとる．

次に，教師は，$\mathbf{P}$ にしたがって，N 個の点をとり，以下の相対距離情報を学習アルゴリズムに与える．各点 A に対して

If $dist_{W^*}(A,C_1) \le dist_{W^*}(A,C_2)$,
$(A,C_1) \le_{W^*} (A,C_2)$ を与える．

If $dist_{W^*}(A,C_1) < dist_{W^*}(A,C_2)$,
$(A,C_1) <_{W^*} (A,C_2)$ を与える．

ここで，$dist_{W^*}(A,C)$ は以下のように定義される：

$$\sum_{i=1}^{n} W^*_{(i)}(A_{(i)} - C_{(i)})^2.$$

学習アルゴリズムは，重みベクタ W^* を相対距離情報から有限時間で近似する．そのようにして近似された $[0,\infty)^n$ 上の W と真の重みベクタ W^* の相違点の集合 $diff_{\langle C_1,C_2\rangle}(W,W^*)$ を

$$\{A \in E^n | different_{W,W^*}(A,C_1,C_2)\}$$

と定義する．ここで，

$different_{W,W^*}(A,C_1,C_2) \stackrel{def}{=}$
$((A,C_1) \le_{W^*} (A,C_2) \wedge (A,C_1) >_W (A,C_2)) \vee$
$((A,C_1) >_{W^*} (A_i,C_2) \wedge (A_i,C_1) \le_W (A_i,C_2))$.

W が，分布 $\mathbf{P}$ と $\langle C_1,C_2\rangle$ において W^* との相違点の集合に関して ϵ- 近似であるとは，誤り率 (error rate) $0 < \epsilon < 1$ に対し，確率 $\mathbf{P}(diff_{\langle C_1,C_2\rangle}(W,W^*))$ が，たかだか ϵ であることをいう．

以下の定理は，この問題が PAC- 学習可能であることを示している．

定理 1. E^n 上の任意の確率分布 $\mathbf{P}$ と任意の $(0,1)$ 内の実数 ϵ と δ に対して，以下の学習アルゴリズムが存在する．

1．教師は，任意の $[0,\infty)^n$ 上のベクタ W^* と E^n から任意の異なった2点 C_1, C_2 を選ぶ．
2．教師は，$\mathbf{P}$ にしたがって N 点を E^n からとり，W^* によって，相対距離情報を学習アルゴリズムに与える．
3．アルゴリズムは，$[0,\infty)^n$ 上のベクタ W を出力し，以下が成立する．
・W が分布 $\mathbf{P}$ と $\langle C_1,C_2\rangle$ において W^* との相違点の集合に関して ϵ- 近似でない確率は，たかだか δ である．δ を確度 (confidence) と呼ぶ．
・学習に必要な相対距離の個数 N は，n, $\frac{1}{\epsilon}$, $\frac{1}{\delta}$ の多項式のオーダでおさえられ，その実行時間は，相対距離の個数の多項式時間でおさえられる．

証明： [2] と同様に，上の問題を，超平面によって分割される半空間の学習に変換し，[1] の結果を利用する．□

3. 重み付き最近隣法における重み学習

2章では，ペアに対する重みの学習を述べた．以下では，このペアを拡張して事例ベース $\mathcal{CB}$ に対する重みの学習に関して述べる．

2章と同様に $\mathbf{P}$ を n 次元ユークリッド空間 E^n 上の任意の確率分布とする．教師は，重みベクタ W^* を $[0,\infty)^n$ から取り，任意の2つ以上の点からなる有限集合 $\mathcal{CB}$ を E^n からとり，事例ベースとする．

さて，学習アルゴリズムは，さきほどと似たような枠組で，重みベクタ W を学習するのだが，この W と W^* の事例ベース推論における相違 $diff_{\mathcal{CB}}(W,W^*)$ を先に以

---

*Learning Weight in Weighted Nearest Neighbor Method
Ken Satoh, Seishi Okamoto,
Fujitsu Laboratories Ltd.
email: ksatoh@flab.fujitsu.co.jp

下のように定義する．

$$\{A \in E^n | min_{W^*}(A, \mathcal{CB}) \neq min_W(A, \mathcal{CB})\}$$

ここで，$min_W(A, \mathcal{CB})$ は以下のように定義される：

$$\{C \in \mathcal{CB} | \text{for every } C' \in \mathcal{CB}, (A, C) \leq_W (A, C')\}.$$

上の式は，W^* の重みで重み付き最近隣法で選ばれる最近隣事例の集合と W を使った最近隣事例の集合との相違を意味している．

W が，分布 $\mathbf{P}$ と事例ベース $\mathcal{CB}$ において W^* との相違に関して ϵ- 近似であるとは，確率 $\mathbf{P}(\mathit{diff}_{\mathcal{CB}}(W, W^*))$ が，たかだか ϵ であることをいう．

以下の定理は，3つ組法による重み学習を $\mathcal{CB}$ の任意の異なった事例のペアおのおのに適用させて出てきた重みが重み付き最近隣法での誤り率を確率的に押えていることを示している．

定理 2. E^n 上の任意の確率分布 $\mathbf{P}$ と任意の $(0,1)$ 内の実数 ϵ と δ に対して，以下の学習アルゴリズムが存在する．

1．教師は，任意の $[0,\infty)^n$ 上のベクタ W^* と E^n から任意の点の有限集合 $\mathcal{CB}$ を選ぶ．点の数を $|\mathcal{CB}|$ とする．
2．教師は，$\mathcal{CB}$ の中の異なった事例のペアおのおのに対して，$\mathbf{P}$ にしたがって N 点を E^n からとり，W^* によって，相対距離情報を学習アルゴリズムに与える．
3．アルゴリズムは，$[0,\infty)^n$ 上のベクタ W を出力し，以下が成立する．
・W が分布 $\mathbf{P}$ と $\mathcal{CB}$ において W^* との相違に関して ϵ-近似でない確率は，たかだか δ である．
・学習に必要な各ペアに対する点の個数 N は，n, $\frac{1}{\epsilon}$, $\frac{1}{\delta}$, $|\mathcal{CB}|$ の多項式のオーダでおさえられ，その実行時間は，点の個数の多項式時間でおさえられる．（したがって，学習に必要な点の全個数は，たかだか，$N \cdot \binom{|\mathcal{CB}|}{2}$ となる）

証明： 図1の学習アルゴリズムを用いると，そこで得られた W は，定理1と[1]の定理2.1より，$\mathcal{CB}$ からとられた各ペア $C_i, C_j (1 \leq i < j \leq |\mathcal{CB}|)$ に対して

$$Prob(\mathbf{P}(\mathit{diff}_{\{C_i,C_j\}}(W, W^*)) \geq \epsilon') < \delta'$$

を満たしている（ϵ', δ' は図1参照）．上は，

$$\sum_{i<j} Prob(\mathbf{P}(\mathit{diff}_{\{C_i,C_j\}}(W, W^*)) \geq \epsilon') < \delta$$

を意味し，これは，

$$Prob(\textstyle\bigvee_{i<j}(\mathbf{P}(\mathit{diff}_{\{C_i,C_j\}}(W, W^*)) \geq \epsilon')) < \delta \qquad \text{(a)}$$

を意味する．

Learn_from_relative_distance(ϵ,δ,n,$\mathcal{CB}$)
ϵ: 誤り率，δ: 確度，n: 次元数, $\mathcal{CB}$: 事例ベース
begin
　$\epsilon' := \dfrac{\epsilon * 2}{|\mathcal{CB}| * (|\mathcal{CB}| - 1)}$
　$\delta' := \dfrac{\delta * 2}{|\mathcal{CB}| * (|\mathcal{CB}| - 1)}$
　for every pair $(C_i, C_j) \in \mathcal{CB} (1 \leq i < j \leq |\mathcal{CB}|)$
　begin
　　$max(\frac{4}{\epsilon'} log_2 \frac{2}{\delta'}, \frac{8n}{\epsilon'} log_2 \frac{13}{\epsilon'})$ 個以上のペアと
　　相対距離情報を教師から得る．
　　for every point A
　　　if $(A, C_i) \leq_{W^*} (A, C_j)$
　　　　then $dist_W(A, C_i) \leq dist_W(A, C_j)$
　　　　　を制約集合に加える．
　　　if $(A, C_i) >_{W^*} (A, C_j)$
　　　　then $dist_W(A, C_i) \geq dist_W(A, C_j) + 1$
　　　　　を制約集合に加える．
　end
　上の制約集合を矛盾しない W と線形プログラミングによって得る．
end

図 1: 重み付き最近隣法に対する重み学習アルゴリズム

W が分布 $\mathbf{P}$ と $\mathcal{CB}$ において W^* との相違に関して ϵ- 近似でないときは，

$$\mathbf{P}(\bigcup_{i<j} \mathit{diff}_{\{C_i,C_j\}}(W, W^*)) \geq \epsilon$$

を意味するから，これは，

$$\bigvee_{i<j}(\mathbf{P}(\mathit{diff}_{\{C_i,C_j\}}(W, W^*)) \geq \epsilon') \qquad \text{(b)}$$

を意味する．

(a),(b) により，

$$Prob(\mathbf{P}(\mathit{diff}_{\mathcal{CB}}(W, W^*)) \geq \epsilon) < \delta$$

となる．□

参考文献

[1] Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K., "Learnability and the Vapnik-Chervonenkis Dimension", *JACM*, **36**, pp. 929 – 965 (1989).

[2] Satoh, K., and Okamoto, S., "Toward PAC-Learning of Weights from Qualitative Distance Information", *Proceedings of AAAI'94 Workshop on CBR*, pp. 128 – 132 (1994).