D-027

SVMを用いたメンバーシップ関数による多種識別
Multiclassification by SVM Membership Function

田中 秀俊 †
Hidetoshi TANAKA

1. SVMを用いた多種識別

サポートベクターマシン (Support Vector Machine, SVM) は、構造リスク最小化という規準において予測性能最良となる識別方式で、近年適用成功例が多く報告されている。一般のSVMは1対1の2種識別法であり、多種識別を行うには、多種用の特殊なSVMを採用することもできるが、一般的な1対1のSVMを反復適用する方法もある [1]。N種類の識別問題を種類相互の $N(N-1)/2$ 個の1対1識別問題に分解してSVMを算出、識別時にはSVMをもとに各種類へのメンバーシップ関数を定義し、総合評価して種類を回答する。

2. ペアワイズメンバーシップ関数

本稿ではメンバーシップ関数を2種類扱う。それぞれの関数の説明のために、以下、集合A対集合Bのサポートベクターをそれぞれ $\boldsymbol{a}$ と $\boldsymbol{b}$ のみとする。$\boldsymbol{a}$ と $\boldsymbol{b}$ を用いて、SVM$(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x})$ は式 (1) のように書ける。

$$\mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x})=\frac{2}{||\boldsymbol{a}-\boldsymbol{b}||^2}(\boldsymbol{a}-\boldsymbol{b})(\boldsymbol{x}-\frac{\boldsymbol{a}+\boldsymbol{b}}{2}) \quad (1)$$

点 x が集合Aと集合Bのどちらに属するかを示すメンバーシップ値 $a_b(\boldsymbol{x})$ は、式 (2) のように定義できる。

$$a_b(\boldsymbol{x})=\begin{cases}0 & (\mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x})\le 0)\\ \mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x}) & (0<\mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x})<1)\\ 1 & (\mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x})\ge 1)\end{cases} \quad (2)$$

また、帰属しないケースにペナルティを課すソフトマージンとの整合性をとり、式 (3) の $a'_b(\boldsymbol{x})$ のように下限を設けない定義も可能である。

$$a'_b(\boldsymbol{x})=\begin{cases}\mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x}) & (\mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x})<1)\\ 1 & (\mathcal{S}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{x})\ge 1)\end{cases} \quad (3)$$

参考として、1次元で $\boldsymbol{a}=1,\boldsymbol{b}=0$ の場合を図1に示す。

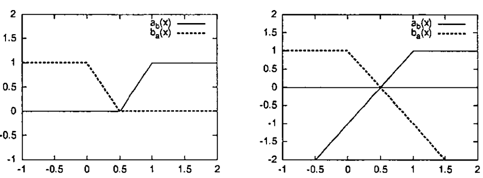

図 1: 1次元での例 $(\boldsymbol{a}=1,\boldsymbol{b}=0)$

3. 論理積と肯定値加算

サンプル $\boldsymbol{x}$ についての種類Aと種類Bの間のAらしさ $a_b(\boldsymbol{x})$ と、種類Aと種類Cの間のAらしさ $a_c(\boldsymbol{x})$ といったペアワイズのメンバーシップ関数をベースに、総合的なAらしさを評価する、種類毎トータルメンバーシップ関数を定義する。その候補を挙げる。

共通集合を求める問題とみなして、ペアワイズメンバーシップ関数を式 (3) の $a'_b(\boldsymbol{x})$ とし、トータルメンバーシップ関数をそのファジィ論理積、式 (4) とする。

$$\hat{a}(\boldsymbol{x}) \;=\; \min[a'_b(\boldsymbol{x}),a'_c(\boldsymbol{x}),\ldots] \quad (4)$$

これが最大となる種類を回答するというトータルメンバーシップ関数が提案されている [2]。この方法では、複数種類の可能性が均衡する領域の大部分において、適当と思われる一つの回答を返すことができる。これは同時に、ペアワイズ境界と総合評価の境界に不整合が生じる場合があることを示している。すなわち、$\hat{a}(\boldsymbol{x})>\hat{b}(\bar{\boldsymbol{x}})$ となる $\bar{\boldsymbol{x}}$ で、$a'_b(\bar{\boldsymbol{x}})>a'_c(\bar{\boldsymbol{x}})$ や $b'_a(\bar{\boldsymbol{x}})>b'_c(\bar{\boldsymbol{x}})$ の場合が存在し、このとき $a'_b(\bar{\boldsymbol{x}})<b'_a(\bar{\boldsymbol{x}})$ の場合がありうる。

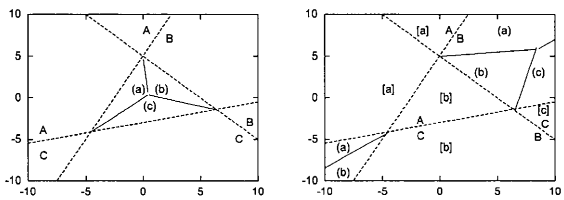

図 2: 最大論理積法による3種類の判別

図2に、A,B,Cの3種類の場合の種類判別を示す。AB,AC,BCの3つの境界超平面は、2次元断面を考えたとき、3枚が並行でないなら3角形を形成する。A,B,C各領域の存在形態は、その3角形の内側の種類が全部異なる場合と2つが一致する場合の2パターンである。それぞれについて、(a)(b)(c)は複数種類の可能性が均衡する領域であり、論理積法では、(a)からA、(b)からB、(c)からCが回答される。ただし、例えば(a)はAB境界のB側、あるいはAC境界のC側であるにも関わらずAと回答する領域でもあり、ペアワイズ境界と総合評価の境界に不整合が生じていることがわかる。

ペアワイズメンバーシップ関数を式 (2) の $a_b(\boldsymbol{x})$ とし、トータルメンバーシップ関数を式 (5) のような代数和とすることも考えられる。

$$\tilde{a}(\boldsymbol{x}) \;=\; a_b(\boldsymbol{x})+a_c(\boldsymbol{x})+\ldots \quad (5)$$

この定義では、トータルメンバーシップ関数による境界すなわち $\tilde{a}(\bar{\boldsymbol{x}})=\tilde{b}(\bar{\boldsymbol{x}})$ となるような $\bar{\boldsymbol{x}}$ と、ペアワイズメ

† 三菱電機（株）情報技術総合研究所

ンバーシップ関数による境界すなわち $a_b(\bar{x}) = b_a(\bar{x})$ となる $\bar{x}$ とにずれが生じる。

$$\tilde{a}(\bar{x}) - \tilde{b}(\bar{x}) = [a_b(\bar{x}) - b_a(\bar{x})] + [a_c(\bar{x}) - b_c(\bar{x})] \quad (6)$$

式 (6) によると、$a_c(\bar{x}) = b_c(\bar{x})$ のときに両者は一致するが、これは例えば $\bar{x} = \frac{a+b}{2}$ を考えた場合、なりたつとは限らないことがわかる。

4. 二重否定値加算

サンプル $\boldsymbol{x}$ における種類 A のトータルメンバーシップ関数 $a(\boldsymbol{x})$ を作成するにあたり、種類 A の可能性の単純な代数和ではなく、種類 A の可能性を陽に否定する $b_a(\boldsymbol{x}), c_a(\boldsymbol{x}), \ldots$ を除いた代数和、いわば二重否定値を用いて式 (7) のように定義すると、トータルメンバーシップ関数による境界とペアワイズメンバーシップ関数による境界のずれが顕在化しない。

$$a(\boldsymbol{x}) = \sum_{S \neq T} \sum_{T \neq a} S_T(\boldsymbol{x}) \quad (7)$$

式 (7) は以下の式 (8) と等価、すなわち種類 A に関する否定値の加算 $\sum S_a(\boldsymbol{x})$ を比較することによって評価できるので、計算量は肯定値加算法と等しい。

$$a(\boldsymbol{x}) = \sum_{S \neq T} \sum_{T} S_T(\boldsymbol{x}) - \sum_{S \neq a} S_a(\boldsymbol{x}) \quad (8)$$

4.1 境界の整合性: 1 次元の場合

図 3 は、単純な一次元の例を用いて境界の整合性を示している。集合 A,B,C のサポートベクターをそれぞれ $\boldsymbol{a} = 0, \boldsymbol{b} = 1, \boldsymbol{c} = 2$ のみとする。肯定値加算法 (左上) では $a(\boldsymbol{x})$ と $b(\boldsymbol{x})$ の交点がペアワイズ境界である 0.5 からずれており、$b(\boldsymbol{x})$ と $c(\boldsymbol{x})$ の交点がペアワイズ境界である 1.5 からずれていることが見てとれる。これが二重否定値加算法 (右上) では一致している。同様に、集合 A,B,C,D のサポートベクターがそれぞれ $\boldsymbol{a} = 0, \boldsymbol{b} = 1, \boldsymbol{c} = 2, \boldsymbol{d} = 3$ のみの場合、肯定値加算法 (左下) では $a(\boldsymbol{x})$ と $b(\boldsymbol{x})$、$c(\boldsymbol{x})$ と $d(\boldsymbol{x})$ の各交点が本来の位置とずれているのに対し、二重否定値加算法 (右下) ではすべて一致していることがわかる。

4.2 可能性の均衡領域: 3 種類の場合

A,B,C の 3 種類の場合の種類判別を図 4 に示す。図では、AB,AC,BC の 3 つの境界超平面の 2 次元断面において形成される境界 3 角形の 2 パターン、内側が 3 種類になる場合 (左) と 2 つが一致する場合 (右) を示している。3 種類の可能性が均衡する領域を X で示している。二重否定値加算法の場合、X 領域では $a(\boldsymbol{x}) = b(\boldsymbol{x}) = c(\boldsymbol{x})$、すなわち 3 種類の可能性が均衡したままであり、3 種類が回答されることになる。ここが論理積法と異なる。マージン領域、例えば集合 A と集合 B の境界にあるマージンでは、B の側で X に隣接する部分は A、A の側で X に隣接する部分は B になる。すなわち、$a(\bar{x}) > b(\bar{x})$ において $a_b(\bar{x}) < b_a(\bar{x})$ となるのは $0 < b_a(\bar{x}) < 1$ でかつ $c_a(\bar{x}) > c_b(\bar{x})$ の領域にとどまる。この部分の判定は論理積法と一致する。

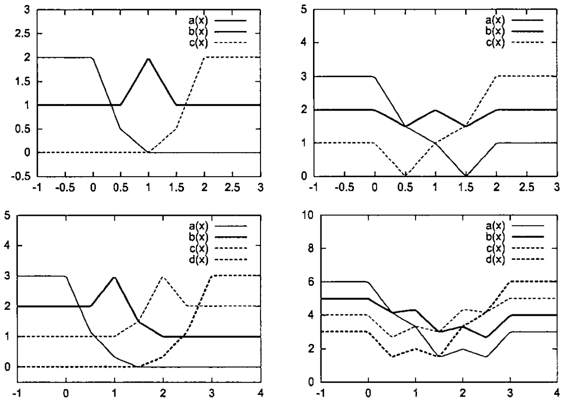

図 3: 1 次元の例

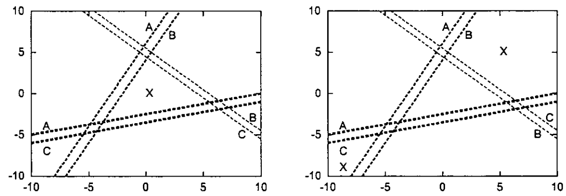

図 4: 2 次元 3 種類の均衡領域

5. まとめと課題

一般的な 1 対 1 の識別を行う SVM を用いて多種識別を行う方式として、SVM 関数を 0 と 1 の間だけ用いるペアワイズメンバーシップ関数を二重否定値加算によって統合する、二重否定値加算法を提案した。否定側無制限のペアワイズメンバーシップ関数を論理積により統合する方法、二重否定値加算ではなく肯定値加算によって統合する方法を提示して、比較検討を行い、二重否定値加算法の特徴を示した。二重否定値加算法は論理積法と類似している。論理積法ではペアワイズ境界とトータル境界の整合性を犠牲にして種類を基本的にひとつ回答するのに対し、二重否定値加算法では整合性を犠牲にしない範囲で種類をひとつ、同点であれば複数回答する。

課題として、UCI リポジトリを用いて数値実験を実施し、否定側無制限のペアワイズメンバーシップ関数を論理積により統合する既存方法との識別精度の差の例を示す必要があると考える。

参考文献

[1] Kreissel, U.H.-G., "Pairwise Classification and Support Vector Machine," Scholkopf, Burges and Smola (eds.), Advances in Kernel Methods, pp.255-268. (1999).

[2] Inoue, T. and Abe, S., "Fuzzy Support Vector Machines for Pattern Classification," Proc. Int. Joint Conf. on Neural Networks, pp.1449-1454, (Jul. 2001).