A-18

レーダパルス識別へのSVMの適用
Radar pulse recognition using SVM

川上 かおり† 田中 秀俊† 三石 彰純†
Kaori Kawakami† Hidetoshi Tanaka† Akitoshi Mitsuishi†

1．まえがき

現状，レーダの機種をレーダパルスから識別する試みは既に行われている．例えば，パルスの間隔や搬送周波数など，レーダパルスに関するいくつかの特徴量を抽出し，それら特徴量の範囲の組をレーダの機種別辞書として用意しておくと，その辞書と照合することによって，完全ではないにせよ，識別を試みることができる．一方、同一機種のレーダの個体識別は，著者らの知る範囲では試みられていない．そこで，同一機種２機のパルスデータを多数用意して，識別が可能か否かをまず検討した．識別には音声認識の手法を採用することとし，簡易な隠れマルコフモデル(HMM)に基づいてパルスモデルを構築し識別を試みたところ，搬送周波数によっては識別可能との結果が得られた[1]．本稿で述べる実験は，第一段階実験で識別不能となったケースに対する追試を目的としている．前述の簡易 HMM の各状態を対象として，サポートベクターマシン(SVM：Support Vector Machine)による識別を試みた．以下に，その手法と結果について述べる．

2．隠れマルコフモデル

形式的には d 次 HMM，すなわち過去 d 個のデータが有限の数の状態を確率的にとり，その次のデータはとった状態毎に固有の確率で決まる状態遷移モデルとした．実質的には，d 個のデータにウォルシュ変換を施して得られた w 次元ベクトルに対して，1次 HMM を構築する方法を採用した．ただし，現段階ではモデル適合度に状態遷移確率を算入せず，出力分布のみを用いている[1]．

3．サポートベクターマシン（SVM）

SVM にはカーネルという仕組みが存在し，その基本である線形カーネルは，境界付近の学習データのみに注目し，2 クラスの学習データ間の最小幅（以下，マージン）を最大化する線形判別器である．線形判別不能，すなわち分離不可能な学習データが存在する場合には，そのデータへのペナルティを表す変数（以下，スラック変数）が導入される(ソフトマージン)．さらに，非線形カーネルと呼ばれる，高次空間での線形判別境界を算出する仕組みにより，適用可能な問題が増える．本稿の実験では，線形カーネルと，非線形カーネルの一つである多項式カーネルとを採用した．

4．解析手法

4.1) 前処理

前処理として「窓」を導入し，データをベクトル化する．以下，時間単位を Tu とする．レーダパルス信号に対し時間幅 d [Tu]の「窓」を，全測定時間にわたって 1[Tu]ずつずらすことにより，測定時間数程度の窓データを生成する．レーダパルス信号から局所的周期性を特徴として捉えるために，得られた各窓データについて w 次までのウォルシュ係数を求め，w 次元の中間データを生成する．

4.2) 状態分類

各レーダのデータから p_1 本ずつパルスを選び，w 次元の中間データに対し，レーダの区別をせずに，部分母集団の平均，標準偏差，及び混合比を求める．これらの値の算出には期待値最大化法を用いている．

4.3) 状態遷移列の算出

各中間データがどの状態に属するかを，帰属確率を比較して決定する．この中間データの帰属状態の時系列から，各パルスの状態遷移列が得られる．

4.4) SVM 用状態別データの抽出

状態遷移列に基づき，各状態のデータを抽出する．各パルスを構成するウォルシュ変換結果データの中から，個々の状態毎に，属するウォルシュ変換結果データを抽出し，クラスを表すレーダ識別子と組み合わせたものを SVM 用状態別データとする．最終的には，２レーダの複数本のパルスに対するデータを１ファイルにまとめたものが SVM 用状態別データとなる．その際，２レーダのウォルシュ変換結果データ数が等しくなるように調整する．

4.5) SVM による学習及び判別

SVM 用状態別データに対して学習を行い，判別境界を算出する．算出された判別境界に基づき，判別対象データの判別を行う．

5．線形カーネルによる識別実験

5.1) データ生成処理

同一機種のレーダ２機について，まず[1]の実験で識別可能であった周波数Ａにおいて解析実験を行った．約13万[Tu]の検波後のビデオ信号データであるパルスデータを，各レーダ／各周波数で32 本ずつ用いた．16 次12 状態 HMM(N=12, d=16)を想定し，ウォルシュ変換により８次元の中間データ(w=8)を作成した．状態分類には，各レーダ／各周波数のパルス 10 本(p_1=10)の計 20 パルスを対象とし，さらにそれらのウォルシュ変換結果データを SVM 用状態別データに変換した．

なお，今回使用したパルスデータは，個々のレーダの移動平均に若干の差があり，この差が識別精度に大きく影響することが[1]より明らかになっている．そこで移動平均の差が直接的に影響しないよう，正規化による補正（以下，レーダ間正規化）を行った上で，ウォルシュ変換を施した．

5.2) 学習／判別処理

SVM 用状態別データを各状態において学習後，自己識別テストを実施した．オープンソースのフリーソフトウェアである SVM^{light}[3]を用いた．状態数は 12 とし，パルス間の空状態とみなせる状態（状態 ID が 5〜7）は対象外とした．カーネルは線形カーネルを採用した．

事前にデフォルトのパラメータ設定における予備実験を行ったところ，全状態の正解率は 49.67〜52.00%の範囲で，平均正解率は 50.89%であった．そこで識別性能向上のため，以下に示すデ

ータ加工や SVM^{light} が提供しているパラメータの変更を行った．

（１） SVM 用状態別データの成分間正規化

SVM では学習データの各成分の振幅（絶対値の最大値）のばらつきが大きく影響すると考えられている．そこで， ウォルシュ変換後のデータを加工して成分間の正規化を行った．具体的には，学習対象となるデータの振幅の平均値を求め，平均振幅がほぼ均等になるよう，成分毎に異なる定数を乗じて正規化を行った．

（２） スラック変数の導入（パラメータ c）

SVM^{light} のパラメータ c は， 線形分離が不可能な場合における，最小化を行う目的関数のペナルティ項の係数（すなわちスラック変数の係数）の意味合いを持つ．パラメータ c を大きくすると，分離不可能な学習データへのペナルティが厳しくなり（＝修正コストが大きくなる），学習データの正解率の向上が期待できる．しかしペナルティを厳しくするとマージンは狭くなる傾向にあるため，そのトレードオフの調整が重要となる．

5.3 判別結果（線形カーネル）

成分間正規化後，及びパラメータ c 導入後の各状態での正解率を表 1に示す．なおパラメータ c は，ほぼ全状態において，実用時間内で収束可能な最大値として，試行実験で得られた $c=10^{-6}$ を用いた．

表 1：(1)成分間正規化／(2)パラメータ c 導入後の正解率[%]

状態	0	1	2	3	4	8	9	10	11
(1)	56.50	59.00	52.33	51.67	53.83	53.33	55.33	51.33	53.17
(1)＋(2)	61.83	59.33	51.33	56.17	54.33	59.33	56.83	53.83	56.33

ほぼ全ての状態で成分間正規化及びパラメータ c の導入により正解率が向上している．しかし，正規化のみの場合が 51.33～59.00[%]（平均 54.05[%]），パラメータ c の導入で 51.33～61.83[%]（平均 56.59[%]）と，全体で 11[%]程度の向上であり，識別可能とは言い難い結果であった．

6．多項式カーネルによる識別実験

6.1 多項式カーネル

多項式カーネルは 式 1のように定義される．今回の実験では，線形カーネルの際に用いたパラメータ c と，式 1 を構成するパラメータのうち，次元を表すパラメータ D のみを変更対象とし，その他のパラメータ r,s はデフォルト設定のままとした．また対象とする状態は任意に抽出した状態（状態４）のみとした．

$$K(x_i, x_j)=(r+s\, x_i\, x_j)^{D} \quad (式１)$$

6.2 判別結果（多項式カーネル）

各状態で得られた最良の正解率 R[%]を表 2に示す．なおパラメータ c は，各状態において 1～3 時間程度で収束可能な最大値とした[1]．なお，成分間正規化データによる学習では，演算上の誤差が生じたため，学習データ全体の振幅の平均値を使って，各成分の最大振幅が 1.0 前後になるように正規化したデータを用いた．

[1]各状態において，パラメータ c が大きいほど正解率が向上する傾向が見られ，また c 及び D が大きいほど実行時間は増大した．

表 2：多項式カーネルにおける正解率 R[%] (状態 4)

D	2	3	4	5	6	7	8	9
c	50	10	0.2	10^{-2}	4×10^{-4}	10^{-5}	10^{-6}	10^{-8}
R	62.5	70.3	72.3	76.0	77.8	79.5	83.5	80.2

$c=10^{-6}$, D =8(8 次元)の場合に最良の正解率 83.5[%]が得られ，また全状態でも 62.5～83.5[%]，平均 75.3[%]の正解率となった．なお，D =10 では上記時間内で収束不可能であった．以上より，多項式カーネル導入による識別率向上の効果が明らかになった．

上述の結果は全て学習データを対象とした自己識別の結果であり，「8 次元多項式カーネル程度の個体差」があると言える．しかし，実際に識別できるかは今後の大きな課題である．本実験での最良の結果が得られたパラメータの組み合わせを用いて，非学習データを対象に実験を行ったところ，平均正解率は 52.2 [%]となり，まだ非学習データに対して識別可能といえる結果は得られていない．また，そもそも今回のデータは工場内で採取した非常にきれいなデータであって，フィールドで採取したものではない．

非学習データに対する正解率が大幅に低下した原因の一つとして，パラメータ c を，収束可能な範囲内で最大にした点も考えられる．c が大きくなるほど，ペナルティがきつくなり，学習データでの正解率が向上する反面，マージンが狭くなり，非学習データでの正解率低下につながった可能性が高い．その他にも多項式カーネルの次元数や，カーネルそのものの適性等，識別性能に影響する要因は多数考えられる．非学習データでの正解率向上のために，これらパラメータの調整，手法の改良などを今後検討する．

7．むすび

本報告では，レーダパルスを識別する一方法を提案し，評価結果を示した．（1）パルスデータ分割（2）状態分類（3）状態内での SVM による判別 の手順で，隠れマルコフモデルをベースとした状態分類を行い，各状態内での SVM による判別精度を検証した．実験から，SVM の線形カーネルによる正解率は 50[%]前後と低く，各データの成分間の正規化や，スラック変数の導入によって 11[%]程度の向上は見られたが， 識別可能といえる結果は得られなかった．次に多項式カーネル導入による実験では，学習データを対象に，8 次元の場合に最良の正解率 83.5[%]が得られ，多項式カーネル導入の効果を示すことができた．しかし非学習データでは，平均正解率は 52.2 [%]となり，識別可能といえる結果は得られなかった．非学習データでの正解率低下に関して，スラック変数の設定方法，カーネルの次元数，カーネルの選択方法等，多数の要因が考えられるため，今後は，これらのパラメータの調整や手法の改良等を検討する予定である．

参考文献

[1] 川上 他，“隠れマルコフモデルによるレーダパルス識別，”信学技報，SANE2001-118，pp.39-44，2002．

[2] C .Burges, “A Tutorial on Support Vector Machines for Pattern Recognition, ”Data Mining and Knowledge Discovery, vol.2, no.2, pp. 121-167, 1998.

[3] T.Joachims, “Making large-Scale SVM Learning Practical. Advances in Kernel Methods – Support Vector Learning, ”B.Schölkopf, C.J.C.Burges, and A.J.Smola, edits, pp.169-184, Cambridge, MA, MIT Press,1999.