Journal of National Fisheries University 59 (4) 253－256 (2011)

サポートベクトルマシンを用いた有明海ノリ養殖生産定量モデル構築へ向けての一試論

青木邦匡*†，楫取和明*

An essay on a construction of quantitative models of cultured nori prodution in Ariake Sea using SVM

Kunimasa Aoki*†, Kazuaki Kajitori*

Abstract : Many ecological modelings have been done about relationships between nori cultivation and environments in Ariake Sea. In this paper, we try to use the support vector machines (SVMs) to model a relation between nori production and oceanic conditions. SVMs are flexible and robust tools to model non-linear phenomena. We find that the number of data records are not enough but SVM models show some abilities to keep track of the data.

ASFA keywords : Data, Analysis, Models

目　的

有明海のノリ養殖に関して環境との関係が論じられてきた。定性的には植物プランクトンとの栄養素に関する競合関係が推測され[1)]，その定量的なモデリングも試みられてきた[2)]。栄養素と環境の関係のモデル化に当たっては，低次層を中心とした生態系モデリングの研究が多くなされている[2,3)]。本小論の目的は，生態系を構成するパーツの関係を明示する生態系モデリングの方法ではなく，一般の非線形な関係が支配していると思われる現象に汎用的に用いられるサポートベクトルマシン（以下SVM）[4)]というモデリング手法を海況データとノリ生産の関係に適用してみることである。

SVMは，いわゆるカーネル法[5)]の一種で，データを非線形な関数により高次元空間に写像してから線形モデルを適用するもので，従来の線形モデルに比べ柔軟に現象をとらえる潜在力を持っていると見られている。また，SVMは外れ値の影響を受けにくいというロバストな特徴を持っている。この様々な方面で効果を発揮し出している[8,9,10,4)]比較的新しい手法を有明ノリと環境の関係に対して適用し，その可能性を検討する。

方法と結果

ノリの生産データとしては，有明海沿岸の諸漁協の各期の共販における生産枚数を使用する。生産枚数をノリの生産量とすることについては，ノリ養殖業者へのヒヤリングから生産枚数が養殖場における生産量を反映しているという感想を得ていることから妥当であると判断した。しかし極度の不作であった2000年度については検討の余地があると考えている（この点については「結語」で触れる）。

海況データは全漁連提供の以下のものを使用する。有明海湾奥部の11地点（Fig. 1）における水深（m），プランクトン（ml/m^3），採水層（m），水温（℃），塩分（‰），DIN（μg-atm./L），NO_3（μg-atm./L），NO_2（μg-atm./L），NH_4（μg-atm./L），PO_4（μg-atm./L），Si（μg-atm./

2010年12月6日受付．Received December 6, 2010.

＊　水産大学校水産流通経営学科（Department of Fisheries Distribution and Management）
†　別刷り請求先（corresponding author）: aoki@fish-u.ac.jp

L）を各月の数日において記録したものである。我々はこのうち，プランクトン，水温，塩分，NO_3，NO_2，NH_4，PO_4，Siを使う。

ノリの生産量データと海況データの地域的な対応は，海況調査地点６，７，11と福岡県有明海漁連の柳川大川地区（Fig.１のA），大和高田地区（同，B），大牟田地区（同，C）の３地区の25漁協が対応するとした（Fig.１）。期間的な対応は，ノリの枚数データは年度計をとることにし，海況データは各年度の考慮する期間での平均値をとって，年度毎に対応データを作ることにした。ここで，考慮する期間とは，ノリの漁期としては９月から翌年３月までと設定した。

すなわち，福岡県有明海漁連の年度毎の生産枚数と調査地点６，７，11の各年度の（考慮期間の）海況平均値を対応させてその関係を見ることにした。

このようにして対応させた海況，生産枚数のデータが1988年度から2001年度まで14年間分できた。これを前半９年分をモデルの訓練データとし，後半５年分をモデルを試すテストデータとした。データ全体が14年分と少なく，モデルの学習を優先して訓練データに多く割り振った。

SVMの実装系としては，定評のあるLIBSVM[6]を用いた。SVMには回帰（SVR）と分類（SVC）があるので両方試した。

サポートベクトル回帰

サポートベクトル回帰（SVR）は，データを高次元空間に写像したもの（特徴ベクトルとカーネル法では呼ぶ）に，回帰式を当てはめる手法である。

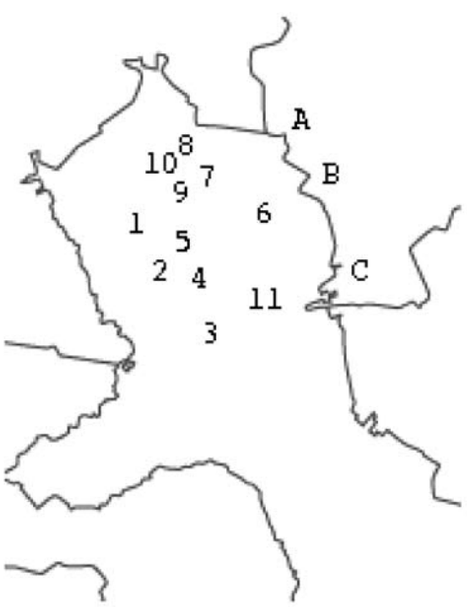

Fig. 1. The observation stations in Ariake Sea

SVMを含むカーネル法では，データを特徴ベクトルとして処理するためのカーネル関数を選ばねばならないが，genericに用意されたものがいくつかあるほか，問題に合わせてカーネル関数を作ることもできる。ここでは，genericなカーネルのうちreasonable first choice[7]とされるRBFカーネルを使用する。

また，サポートベクトル回帰では，誤差関数として従来の二乗誤差ではなく，a-不感応関数という，データのはずれ具合に対する損失の増え方が緩やかな関数を使うので，外れ値に対するロバスト性が出てくる。

こうした柔軟性とロバスト性という長所がある半面，SVMでは，いくつかのパラメータを決定する必要がありその決定法に必然性がないことと，高次元に写像しているため元のデータに即した結果の（従来の回帰ではできる）統計的な解釈が難しいという短所もある。

ここでは，パラメータはLIBSVMのデフォルトを試してから少しずつパラメータをずらして結果を改良していった。

LIBSVMの実行結果をまとめたものは以下のとおりである。

９月から翌年３月までの海況データと生産枚数（単位10億枚）の関係を見ると，以下のようになる。

```
訓練情報：
 訓練オプション：-n 0.4 -c 1.4 -s 4 -g 0.1
 nSV = 9, nBSV = 0
 Mean squared error = 0.00123455 (regression)
 Squared correlation coefficient
        = 0.983315 (regression)
    年：1988 1989 1990 1991 1992 1993 1994 1995 1996
 予測値：1.32 1.39 1.32 1.47 1.42 1.50 1.48 1.46 1.32
 データ：1.29 1.35 1.28 1.51 1.46 1.54 1.52 1.49 1.28
予測情報：
 Mean squared error = 0.122518 (regression)
 Squared correlation coefficient
        = 0.620297 (regression)
    年：1997 1998 1999 2000 2001
 予測値：1.36 1.36 1.38 1.32 1.38
 データ：1.46 1.44 1.27 0.58 1.55
```

訓練オプションの意味は，SVMとして，損失関数を最適化できるν-SVRを使ったこと（-s4），νを0.4，コストパラメータCを1.4，カーネルパラメータγを0.1に設定したことを意味する。

訓練データを使って得られたモデルをテストデータの海況データに当てはめて計算した予測生産枚数と実データの枚数のR^2乗値は0.62（相関係数0.79）とある程度の正の相関が認められた。ちなみに同じデータに従来の重回帰を適用すると，予測値とデータの相関係数は0.39と低く，予測値の平均は0.07（７千万）で（マイナスの予測値もあり），データからは乖離していた。

しかし不作であった2000年度のSVMの予測値は，データ（5.8億）よりかなり高めの13億であった。これは，訓練データでは生産枚数は最も少ない年でも12億８千万あったのに対し，2000年の枚数は桁が違うほど小さく，訓練データではカバーできない状況であったためと思われる。

カーネル法では，訓練データを100％再現するモデルを構築すること（訓練データへの当てはまりR^2乗値が１になること）ができるが，通常これは過学習と呼ばれ，訓練データに特化した追随の結果であるため，汎化能力（訓練データ以外のデータを正しく予測する能力）が高いことを意味しない。訓練データを忠実に反映しつつそのことが汎化能力を高めることになるのがよい学習である。

SVMでは訓練データへの過度の追随（過学習）を防いでモデルの汎化能力を確保するため，正則化項と呼ばれる項を誤差関数に加えた上で最小化を行う。この正則化項の重みを決める係数がコストパラメータCであり，これが大きいほど過学習になりやすい。本データではデータ数が少ないため訓練データへの当てはまりはよいが，Cが小さいため過学習とは認められない。

SVMでは，サポートベクトルと呼ばれる特徴ベクトルが，最終的にモデルを決定するのに使われる（SVMの名の由来）。サポートベクトルの割合が少ないほど計算量は少なくて済むし（訓練データ数が十分なら）汎化能力も高いとされる。

上の結果ではサポートベクトルの数は９で特徴ベクトル全体の数と同じであるから最大となっている。パラメータを変えればサポートベクトルの数を減らすことはできる。例えば，C，gは上の例と同じでパラメータνを0.4から0.1にすると，サポートベクトルの数は６まで減る。しかし，R^2乗値は0.14に低下してしまう。これは，訓練データの数（９）が少ないので，無理にサポートベクトルの割合を減らすと十分な学習ができないことを示している。

本データの場合，データ数が少ないので，十分な学習をしつつ汎化能力を上げることがむずかしいようである。

サポートベクトル分類

サポートベクトル分類（SVC）は，回帰と違い目的変量は数値ではなくクラス（基本では０か１）であり，したがって目的は分類である。

我々の９-３月分の訓練データとテストデータにおいて，枚数に代えてクラスを，枚数が14億未満か以上かによって，０か１に設定した。

SVCは誤識別関数として２乗誤差に比べ誤差の増え方が緩やかな関数を使うので，やはり外れ値に対するロバスト性を持つことはSVRと同様である。

使ったカーネルは前節と同様RBFカーネルであり，パラメータの選定も前節と同様に行った。

９月から翌３月までの海況データとクラス分類の関係は以下のとおりである。

訓練情報：

訓練オプション：**-c 2.0 -g 2.0**

nSV = 9, nBSV = 0

Accuracy = 100% (9/9) (classification)

年	1988	1989	1990	1991	1992	1993	1994	1995	1996
予測値	0	0	0	1	1	1	1	1	0
データ	0	0	0	1	1	1	1	1	0

予測情報：

Accuracy = 60% (3/5) (classification)

年	1997	1998	1999	2000	2001
予測値	1	1	1	1	1
データ	1	1	0	0	1

訓練オプションは，SVMとしてC-SVCを使ったこと（デフォルトなので明示せず），コストパラメータCを2.0，カーネルパラメータγを2.0に設定したことを意味する。

クラス（不作０，豊作１）の予測精度は60％で，不作の2000年度の予測クラスは０であった。

データ数が少ないため汎化能力の点で難があると見られることはSVRの場合と同様である。

結　語

データ数が少ないため前節のモデルの信頼度は低いが，SVMは従来の線形モデルに比べデータなりの結果を導き出す可能性があることは示せたのではないかと思う。

重回帰などでは変数の数が増えると結果の信頼度が落ちるとされるが，海況データでは本小論の使った項目以外にも降雨量や日照時間，経営判断など追加を検討すべきものもある。SVMでは変数の数を抑えるのではなく，正則化によって汎化能力を確保するという方法をとる。この点で考えるべき変量が多い本課題に適している。

SVMは汎用モデリング手法であるため気軽に試すことができるので，生態系モデルを組むに当たっても何らかの知見を得るために活用できるのではあるまいか。

データ数を確保するには，調査年数を増やすことと，調査値域を複数設けることが考えられる。長年に渡ってのデータを使うときは，生産規模など生産の構造の変化を取り入れる必要があるかもしれない。複数の地区のデータを使う場合は，生産指数を扱うなどの標準化が必要になる。

生産量データとして，福岡県有明海漁連の枚数データをそのまま使ったが，2000年度の予測値は現実のデータを大きく外れるものであった。前章では訓練データ内に極端な不作の年がなかったためにモデルが十分に汎化されなかった可能性を述べた。2000年度の不作（の程度）だけは海況データだけからは導けないのかもしれない。一つの可能性として色落ちが激しいため製品化されなかったノリがかなりの部分あったのではないかということも考えられよう。

SVMモデル構築に当たっては適切なパラメータの選択が重要課題である。本来SVMではcross validationという訓練データをいくつかに分けて訓練とテストを繰り返して最適なパラメータを探る方法を採るのであるが，本データではデータ数が少ないためcross validationでは低いaccuracyしか達成できないので採用しなかった。また，本小論では，カーネルはgenericなものを使用したが，問題に適したカーネルを設計することもできる。

SVMモデルにおける変数選択と変数の予測への影響のしかたを探るのは一般に簡単ではないが，我々のノリモデルに即してその方法を開発しなければならないので課題としたい。

本論では，ノリ漁期全体の海況データを使ったが，漁期前の６～９月期などもっと多くの期間を試して期間の影響を調べることも課題である。

事例数の大きなデータを揃えた上でこのような方向でSVMによるモデリングをさらに試していきたい。

参考文献

１）農林水産省農林水産技術会議事務局，有明海の海洋の変化が生物生産に及ぼす影響の解明，研究成果432，（2005）

２）九州環境管理協会有明研究会，有明海環境の定量的評価の研究「漁業生産の回復に向けて」，九州環境管理協会，（2007）

３）横山佳裕，有明海におけるノリ養殖場海域の窒素収支の定量評価，九州環境管理協会会報「環境管理」，36，38-43（2007）

４）Support Vector Machines（SVMs），http://www.svms.org/

５）赤穂昭太郎，カーネル多変量解析，岩波書店，(2008)

６）C.C. Chung, C.J. Lin, LIBSVM 3.0, http://www.csie.ntu.edu.tw/~cjlin/libsvm/, (2010)

７）C.C. Chung, C.J. Lin, LIBSVM guide, http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, (2010)

８）E. Byvatov, G.Schneider, Support vector machine applications in bioinformatics, Appl Bioin fomatics, ２（２）, 67-77 (2003)

９）Financial Applications, http://www.svms.org/finance/ (2006)

10）小野田崇 他，特集　サポートベクターマシン：その仕組みと応用，オペレーションズリサーチ，46（５），（2001）