

新たな統計解析手法による原産地・原料原産地判別マニュアルの見直し

山川 義正, 中村 哲

Yoshimasa Yamakawa, Satoru Nakamura

要 約

農林水産消費安全技術センターにおいて食品の原産地判別のために線形的な判別モデルを採用している6品目について、過去のデータを活用して非線形の判別モデルの採用を検討し、再解析を実施した。その結果、一部の品目について判別能力の向上した判別モデルや検査コストを削減し得る判別モデルが得られた。

1. はじめに

食品表示法（平成25年法律第70号）に基づき定められた食品表示基準（平成27年内閣府令第10号）により、生鮮食品には名称及び原産地表示が義務付けられ、輸入品以外の加工食品においても名称、原材料名などの他に、対象原料について原産地名の表示が義務付けられている。また、特色のある原料として特定の原産地を表示している等、原産地に関する表示がなされている食品は多く、また消費者の関心も高まっている。そのため食品の原産地及び原料原産地表示の真正性を客観的に検証する科学的な技術の開発が求められている。事実、食品の産地判別のための科学的技術として食品中の元素濃度や、軽元素・重元素の安定同位体比などの分析による手法が研究されており^{1)~9)}、農林水産消費安全技術センター（以下「FAMIC」という。）においてもこれらの技術を用いた食品の産地判別法の検討・開発・運用を行っているところである^{10)~19)}。しかしながら、現代においては計算機的能力上昇や計算手法そのものが新たに登場するなど技術の進歩は著しく、加えて近年の食品の原産地への関心の高さから判別技術への要求は高まっており、より高精度な判別のために測定などの分析技術のみならず、それらの結果を解析する手法も見直す必要がある。

FAMICでは、食品の原産地判別において食品中の元素濃度と重元素安定同位体比の単独又は組み合わせにより判別するものがある。これらの判別において、平成28年度以前に検討されたものはその説明可能性や計算速度の問題からほとんどが線形分類器である線形判別分析(LDA)及び線形カーネルのサポートベクターマシン(SVM)を用いた解析を行ってきたところである。しかしながら、例えば外国産が複数産地から構成されている場合、直線（平面）的な判別を行う線形関数による判別が不向きな場合がある。こういった場合においては複数の判別関数を構築することで対応していたが、必ずしも判別検査の特異度・感度の向上にはつながらないことがある。また、複数の判別関数を用いた場合には最適な判別基準を設定するのが困難という問題もある。そのため新たに作成する判別法については線形的な判別に加え、二次判別分析(QDA)や非線形カーネルのSVMを導入して解析を行っており、判別精度の向上に寄与している。そこで、以前に作成された古い判別法の判別モデルに対して従前利用していたこれらの線形的な解析のみならず、非線形の解析と人工ニューラルネットワーク(ANN)を加え、総合的により適した手法を検討した。

2. 実験方法

2.1 解析データ

すでに FAMILIC において元素濃度分析のみ又は元素濃度分析にストロンチウム安定同位体比分析を併せて利用する産地判別を行っている品目において、その開発段階で線形での判別を行っているもので過去にモデル作成用に使用した元素濃度及びストロンチウム安定同位体比データが現在の解析に利用可能なものを対象とした。品目は6品目（ショウガ・アサリ・ゴボウ・乾燥ひじき（以下「ひじき」という。）・タマネギ・はちみつ）を対象として解析をした。

2.2 統計手法の探索と実装

解析に当たっては基本的な処理は中村らの方法¹⁾に従ったが、従前から導入されていた LDA、線形カーネル SVM に加え、QDA、非線形カーネル（ラジアル、多項式）SVM、ロジスティクス回帰及び ANN を導入、実装し解析に用いた。

非線形カーネルの SVM は線形での分離が不可能または困難なデータについて、より高次元の空間に射影して超平面により分離を試みるものであり、比較的説明が容易で入り組んだデータにも適用可能性があることから機械学習においても頻用される手法である。一方でパラメータによっては過適合によりモデルデータしか判別精度がでない等の状況が起きる汎用性の低下の恐れも有り、汎化性能に注意が必要である。

ANN は線形分離不可能な問題に対して、比較的小さい計算量で結果が良好なことが多いことから、近年データマイニング等で広く応用されている方法である。その特徴は活性化関数などにより決まるが、実装するにあたり、特異度や感度をカーネル密度推定から算出する現在の方法との整合性を検討し、活性化関数に Rectified Linear Unit (ReLU) を使用した ANN を実装した。

また、説明変数探索のための手法として前進（後進）ステップワイズが用いられていたが、新たに遺伝的アルゴリズム (GA) を実装した。

説明変数の探索においては Receiver Operating Characteristic (ROC) 曲線を描くことにより、最適となる判別モデルを選択した²⁾。ROC 曲線を描くに当たっては、国産試料を陰性試料、外国産試料を陽性試料として扱い、ROC 曲線の特定の接線の傾きの値に対して接線が最も左上になるよう、変数選択を行う。接線の傾き m は (偽陽性の損失) / (偽陰性の損失) $\times (1-p)/p$ (p : 事前確率) である。このとき、損失指標を $\alpha + \beta/m$ (α : 偽陽性率、 β : 偽陰性率) で定義でき、この値が最小になるとき ROC 曲線の接線が最も左上になる。

前進（後進）ステップワイズは一つずつ変数を追加（削除）し損失指標が最小になるものを選択していくため、ローカルミニマムに達し、変数の追加（削除）をしても損失指標の減少が得られない場合にそれ以上の探索ができなくなる。しかしながら、GA ではランダムに選ばれた変数群から損失指標の低いものから順にいくつか選び、交差や変異といった遺伝的な変化を加えて複数世代を重ねることで、最適な説明変数の探索を行い、これにより局所解に落ち込んでしまうことを防ぐことができる。

解析には統計解析向けのプログラミング言語である R 3.5.1²⁾を使用し、LDA 及び QDA においては MASS パッケージに同封された `lda` 及び `qda` 関数を、SVM においては `e1071` パッケージに同封された `svm` 関数を、また、ロジスティック回帰においては `nnet` パッケージに同封された `multinom` 関数により解析した。

2.3 データの解析

既往モデル試料における元素濃度及びストロンチウム安定同位体比のデータを変数として前述の解析手法を用いて網羅的に解析した。このとき、変数の標準不確かさから、モンテカルロ法により判別得点の不確かさを推定した。この方法では変数の標準不確かさが正規分布に従うと仮定して、乱数により濃度を変動させた時の判別得点の分散から判別得点の標準不確かさを推定している。また、Sheather & Jones の方法²²⁾によるカーネル密度推定のバンド幅及び判別得点の標準不確かさを合成したものを最終的なバンド幅として、カーネル密度推定により判別得点分布を推定した。これらから、偽陽性率と偽陰性率を元にした損失指標を算出し、これが最小となるように GA による変数選択を行った。

3. 結果及び考察

3.1 遺伝的アルゴリズムの導入効果

GA は世代を重ねることで最小の値を探索するアルゴリズムであるが、その世代ごとの値の変化を見ると（図 1）、ある世代において一度低い値を取った後にしばらく値の変化が見られなくとも、世代を進めるにつれてランダムな変異によってより低い値の組み合わせが現れる場合がある。また、前進ステップワイズにおいて計算回数は変数の数の 2 乗に比例した量となるが GA においては変数の数の定数倍に比例するため、GA の採用により大量の変数を解析する際の計算量を削減することができるようになった。

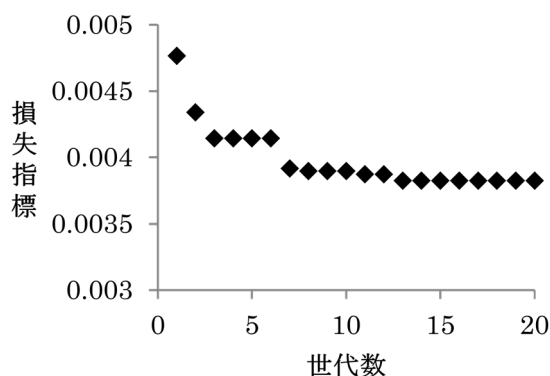


図 1 世代ごとの損失指標

ラジアルカーネル SVM におけるタマネギの損失指標（事前確率 1%）。第 3～6 世代の頃に変数の変更による変動のないローカルミニマムとなっているが、その後の世代で変異によってより低い損失の状態を発見できている。

3.2 解析方法による損失指標の違い

解析については、7 種の方法で判別式を探索した結果、各品目の損失指標は表 1、図 2 のとおりとなった。はちみつを除く全てでラジアルカーネルの SVM が最小の損失を示し、線形分類器よりも良い成績を示している。ANN においては、いくつかの品目において線形の判別よりも小さい損失を示したが、一方で品目によっては線形の判別よりも損失が大きく、必ずしも線形分類器よりも良い方法となるとは限らない。はちみつにおいては、全ての方法で一様に高い損失であり、現行の線形分類器と比べ特異に成績が向上することはなかった。

本研究においてはほとんどの品目においてラジアルカーネル SVM が好成績であったが、他の方法の成績を比較すると、その優劣は一概ではない。これは、モデル試料のデータの複雑さや構造の違いによるものであると考えられる。それぞれの解析方法は適したデータ構造があり、例えば LDA では一様な分布を前提としているなど十分な性能を得るためには最適な方法がある。複雑な判別曲線を描くことのできる非線形の SVM や ANN に関しても、様々なデータに対応するために各種カーネルや活性化関数が存在し、また、直線的な分離が十分可能なデータについては線形分類器が最適な場合も存在しうる。したがって、新たな品目を検討する場合には全ての方法に

ついて引き続き網羅的に検討することが必要である。

表 1 解析方法ごとの各品目の損失指標

解析方法	品目					
	ショウガ	ひじき	アサリ	ゴボウ	タマネギ	はちみつ
線形カーネルSVM	0.0173	0.0061	0.0058	0.0006	0.0075	0.0141
ラジアルカーネルSVM	0.0058	0.0027	0.0048	0.0003	0.0054	0.0149
多項式カーネルSVM	0.0071	0.0061	0.0105	0.0049	0.0070	0.0128
LDA	0.0101	0.0055	0.0072	0.0006	0.0072	0.0136
QDA	0.0082	0.0048	0.0070	- ^{a)}	0.0075	0.0160
ロジスティック回帰	0.0074	0.0046	0.0061	0.0034	0.0084	0.0158
ANN	0.0062	0.0051	0.0066	0.0016	0.0081	0.0148

注)損失指数は全て事前確率 2%を想定して計算した。

a)システム上解析できなかったため欠損値とした。

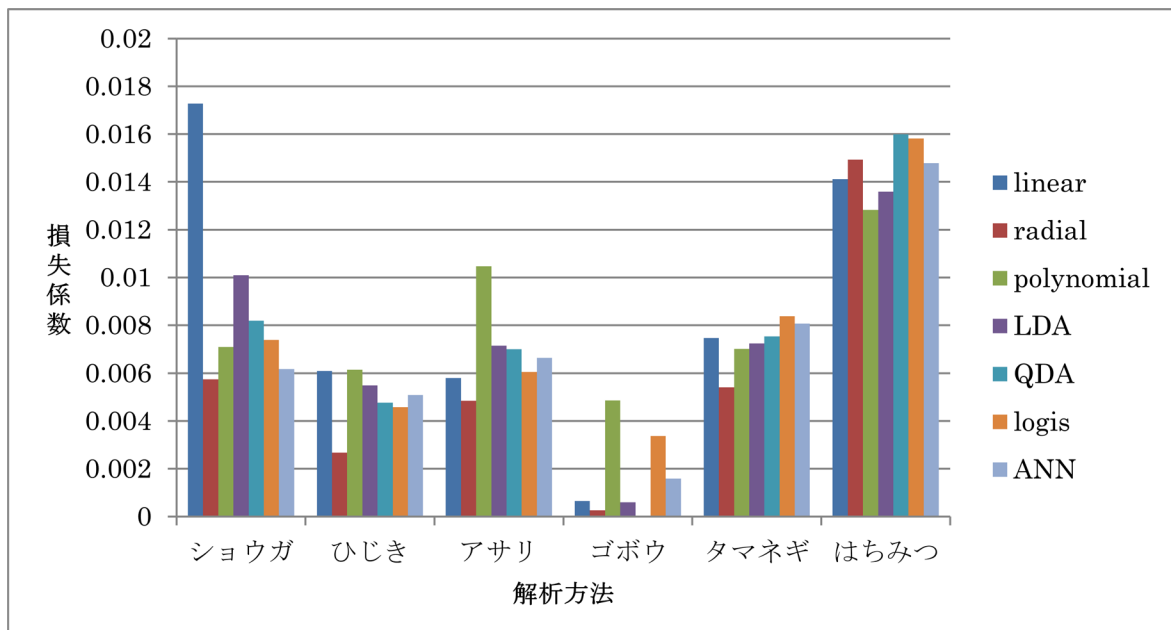


図 2 解析方法ごとの各品目の損失指標

線形カーネル SVM(linear)、ラジアルカーネル SVM(radial)、多項式カーネル SVM(polynomial)、線形判別分析(LDA)、二次判別分析(QDA)、ロジスティック回帰(logis)及び人工ニューラルネットワーク(ANN)を解析方法としたときの各品目の損失指標。

3.3 判別モデルの作成

ほとんどの品目で最小の損失指標を示したラジアルカーネル SVM を用いて作成した各品目の判別モデルについて、得られた新しいモデルを使った判別（新規法）及び各品目の従来の判別方式（旧法）の現行の運用における事前確率においての特異度、感度、説明変数の数、判別式の数及びストロンチウム安定同位体比分析値(SrIR)の利用の有無について表 2 に示す。

損失指標の比較でショウガ及びひじきにおいて、ラジアルカーネルの SVM が線形分類器の結果と有意な違いがあったにもかかわらず、旧法と比較してほとんど違いがないか、感度の低下を引き起こしている。これは旧来の方法では、判別精度を向上させるために二つの判別式を組み合わせる手法を採用しているためである。こういった判別モデルの組み合わせの場合にはその判別の基準値の設定は複雑となるため、モデルが現在の試料に適合するかの評価とモデル修正が難し

表2 新旧の方法での判別式の性能

品目	新/旧 ^{a)}	特異度	感度	変数の数	判別式の数	SrIR
ショウガ	新	99.96	75.9	2	1	有
	旧	99.93	76.1	4	2	有
アサリ	新	97.8	95	20	1	
	旧	97.8	94.2	6	1	
ひじき	新	99.99	87.34	13	1	
	旧	99.94	88.5	7	2	
ゴボウ	新	99.99	99.08	11	1	有
	旧	99.9	71.2	6	1	
はちみつ	新	98.7	71.8	4	1	
	旧	98.9	77.1	7	1	
タマネギ	新	99.96	69.5	15	1	
	旧	99.94	62.8	9	2	

a)ラジアルカーネルSVMを使用した方法を新規法(新)、線形的な判別を用いているこれまで用いていた方法を旧法(旧)とした。

くなる。一方で新規法ではわずかな感度低下はあるものの単一モデルでの判別を可能としたためモデルの検証及び修正が効率的に可能となる。

ゴボウとタマネギにおいては判別精度の向上が見られた。ゴボウは元素分析による判別とストロンチウム安定同位体比による判別の二つの方法があったが新規法においては単一モデルとして運用できる。タマネギも二つの判別式により判別していたところを単一のモデルとできるが、ゴボウと比較し判別精度の向上は小さい。

アサリは性能の向上がほとんどなく変数が増大しており、はちみつは従来の線形での判別の方が有利となっている。これらの品目については、非線形の判別法が線形の判別に比べて優位とはならなかった。

3.4 採用の検討

特異度や感度の向上した品目（ゴボウ及びタマネギ）については現在の流通品とモデルの整合性についての追加検討が必要ではあるが、今回のモデルを利用すれば判別精度の向上が期待できる。また、ショウガとゴボウについては、元素分析とストロンチウム安定同位体比分析の二種類の分析を単一モデルにできている。これらについては、そのうち一種類の分析値（通常はコストの低い元素分析）の変数を抜き出したモデルを作成することが可能である。このモデルは元のモデルから大きく判別性能が低下することは免れないが、目的によって感度あるいは特異度に大きく特化した判別基準を採ることによりスクリーニング判別として利用が可能なため検査の時間削減とコスト削減に大きく寄与しうる。ショウガでは判別精度の向上はないものの、これらの運用上の利点から新たな判別モデルの採用は十分に利点がある。

4. まとめ

FAMIC においては多くの品目で元素濃度分析及びストロンチウム安定同位体比分析による産地判別法を運用してきたが、新しい解析手法の導入がされ、効果を上げていることから、さらなる手法の導入と過去に作成されてきた判別法についても新たな解析手法による性能向上が期待されていた。このため、新たに変数探索のために遺伝的アルゴリズムを、また解析法として人工ニ

ューラルネットワークを導入した。これによってより多くの解析手法により判別モデルの検討が可能になった。また、過去のデータの解析から、ラジアルカーネル SVM による判別によってゴボウとタマネギについて判別精度の向上が見込まれる判別モデルが得られ、またショウガとゴボウについては検査コスト低減等の利点のある判別モデルが得られた。

文 献

- 1) 井上博道, 梅宮善章, 中村ゆり: ウメ干しの仁の微量元素濃度による日本産と中国産の判別, 日本土壤肥料学雑誌, **76**(6), 875-880(2005)
- 2) 中下留美子, 鈴木彌生子, 赤松史一, 小原和仁, 伊永隆史: 安定同位体比解析による国産・豪州産・米国産牛肉の産地判別の可能性, 日本食品科学工学会誌, **55**(4), 191-193(2008)
- 3) 中下留美子, 鈴木彌生子, 一宮孝博, 伊永隆史: 生元素安定同位体比解析による養殖ウナギの産地判別の可能性, 日本食品科学工学会誌, **56**(9), 495-497(2009)
- 4) 篠崎美由起, 有山薫, 川崎晃, 平田岳史: 単一検出器型高分解能誘導結合プラズマ質量分析装置を用いる穀物のストロンチウムと鉛の同位体比分析, 分析化学, **59**(11), 1003-1011(2010)
- 5) 小泉鏡子, 中下留美子, 鈴木彌生子: 安定同位体比分析によるしらす干しの原料原産地判別の可能性, 日本食品科学工学会誌, **58**(6), 259-262(2011)
- 6) 鈴木彌生子, 中下留美子, 河邊亮, 北井亜希子, 富山眞吾: 炭素・酸素安定同位体比分析による青森県産および中国産リンゴの産地判別の可能性, 日本食品科学工学会誌, **59**(2), 69-75(2012)
- 7) 鈴木彌生子, 國分敦子, 絵面智宏, 中山和美: 炭素・窒素・酸素安定同位体比分析による湯通し塩蔵ワカメの産地判別の可能性, 日本食品科学工学会誌, **60**(1), 1-10(2013)
- 8) 石田悦基, 中村健治, 有山薫, 川崎晃: 微量元素濃度及び重元素同位体比を利用する小麦の産地判別, 分析化学, **63**(3), 255-261(2014)
- 9) 絵面智宏, 國分敦子, 阿部洋俊, 濱田真子, 加藤栄一, 鈴木彌生子: わかめの加工による微量元素組成変動と産地判別の可能性, 日本食品科学工学会誌, **62**(10), 484-491(2015)
- 10) 森田先恵, 伊澤淳修, 申基澈, 中野孝教: 元素分析及びストロンチウム安定同位体比分析による冷凍ほうれんそうの原料原産地判別法の検討, 農林水産消費安全技術センター食品関係等調査研究報告, **39**, 31-37 (2015)
- 11) 中村哲, 法邑雄司, 豊田正俊: ゴボウの原産地判別の試料調製法の再検討, 農林水産消費安全技術センター食品関係等調査研究報告, **37**, 1-10 (2013)
- 12) 服部賢志, 木村康晴, 船木紀夫, 法邑雄司: 無機元素組成によるさやえんどうの産地判別, 日本食品科学工学会誌, **56**(10), 529-532(2009)
- 13) 船木紀夫, 服部賢志, 木村康晴, 佐藤耕一, 塚田政範, 津村明宏, 法邑雄司: 無機元素分析によるゴボウの原産国判別, 日本食品科学工学会誌, **57**(2), 70-77(2010)
- 14) 一色摩耶, 中村哲, 鈴木彌生子: 多元素同時分析によるアカシアはちみつの原料原産地判別, 日本食品科学工学会誌, **62**(5), 257-262(2015)
- 15) 石井修人, 箱田晃子: 原木栽培乾しいたけの元素分析及びストロンチウム安定同位体比分析の組み合わせによる原料原産地判別法の検討, 農林水産消費安全技術センター食品関係等調査研究報告, **40**, 9-16(2017)
- 16) 後藤祐之介, 伊澤淳修, 申基澈, 中野孝教: ネギのストロンチウム安定同位体比分析による原産地判別法の検討, 農林水産消費安全技術センター食品関係等調査研究報告, **40**, 17-24

(2017)

- 17) 渡邊彩乃, 寺田昌市: 炭素・酸素安定同位体比分析によるりんごジュースの原料りんごの原産地判別法の検討, 農林水産消費安全技術センター食品関係等調査研究報告, **40**, 25-32(2017)
- 18) 小岩智宏, 伊澤淳修, 後藤祐之介: かき類の元素分析及び軽元素安定同位体比分析による原産地判別法の開発, 農林水産消費安全技術センター食品関係等調査研究報告, **41**, 15-21(2017)
- 19) 松野和久, 井伊悠介, 寺田昌市, 太田民久, 申基澈, 陀安一郎: ショウガの元素分析及びストロンチウム安定同位体比分析による原産地判別法の検討, 農林水産消費安全技術センター食品関係等調査研究報告,**41**,8-14(2017)
- 20) Zweig, M. H. and Campbell, G., Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clin. Chem.*, **39**, 561-577(1993)
- 21) R Foundation (<http://www.r-project.org/>)
- 22) Sheather, S. J. and Jones, M. C. : A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society series B*, **53**, 683-690 (1991).