

## 元素分析によるネギの原産地判別マニュアルの検証

中村 哲

Satoru Nakamura

### 要 約

元素分析によるネギの原産地判別において、主に Na の濃度分布が過去の開発時に使用した試料と検査試料で異なっており、その影響で誤判別を引き起こしている可能性があると考えられたため、その原因を確認した。その結果、標本抽出におけるかたより（選択バイアス）である可能性が高いと判明した。選択バイアスが発生している条件下における判別モデルの修正方法として半教師付き分類と教師付き分類の組合せを採用し、過去の開発時に使用した試料及び検査試料の両方に対して適切に判別できる判別モデルを構築した。

### 1. はじめに

元素分析によるネギの原産地判別については有山（Ariyama）ら<sup>1)~3)</sup>から報告され、一色ら<sup>4)</sup>から再検証の結果が報告されている。しかしながら、検査試料として購入した市販のネギの元素濃度分布を確認すると、判別モデルに使用した試料の元素濃度分布と異なる例が見られた。特に問題となっていた元素として Na がある。ネギの Na の Mg 濃度比の常用対数の分布を図 1 に示す。図 1 には、乾燥試料中 Mg 濃度を 1000 mg/kg としたときの乾燥試料中 Na 濃度換算の結果も示している。従来原産地判別で用いられてきた判別分析やサポートベクターマシンといった教師付き分類（教師あり学習）は、検査試料の分布が国産の分布と外国産の分布の混合分布として表現可能であることを前提としている。一方、図 1 を見ると、Na の Mg 濃度比の常用対数が -0.5（乾燥試料中 Na 濃度換算で約 300 mg/kg）以上の領域は混合分布による再現が不可能であり、前提条件を満たしていない。そのため、検査において誤判別を起こしている可能性が否定できなかった。

母集団と標本集団の分布のずれを発生させるかたより（バイアス）の原因を大別すると、選択バイアス（標本選択バイアス）、情報バイアス及び交絡バイアスの 3 種類となる<sup>5)</sup>。選択バイアスは標本抽出におけるかたよりであり、多変量解析における共変量シフトと呼ばれるものも同種の問題である<sup>6)7)</sup>。情報バイアスは標本の情報を得る際に発生するかたよりであり、化学分析においては分析時の未知の不確かさが存在する場合に発生する可能性がある。交絡バイアスは実験計画における問題の一種で、目的変数と説明変数の間に関連性があるときに両者に影響する未知の要因が存在することにより発生するかたよりである。このうち、交絡バイアスについては原産地判別にはほぼ無関係であることから、選択バイアスか情報バイアスのいずれかが原因となる。

選択バイアスが原因の場合、教師付き分類のみで対応することは困難であり、別の方法を採用する必要がある。報告されている方法としては、標本に重み付けをする方法<sup>6)7)</sup>、半教師付き分類（半教師あり学習）による方法<sup>8)9)</sup>等が挙げられるが、いずれの方法もラベル（原産地判別の場合、国産、外国産等）のある標本だけでなく、ラベルのない未知標本も用いることになる。食品の原産地判別において選択バイアスを議論した報告はほとんど見られないが、選択バイアスの発生頻

度については"rather frequently"と表現している例もあり<sup>7)</sup>、食品の原産地判別においても発生している可能性は否定できない。一方、情報バイアスが原因の場合、分析方法の修正か未知の不確かさの補正のいずれかが必要となる。情報バイアスについては分析化学における基本的な問題として、食品の原産地判別においても頻繁に議論されている。

本研究では、元素分析によるネギの原産地判別において発生していた分布のかたよりの原因を確認するとともに、判別モデルの修正を行ったので報告する。

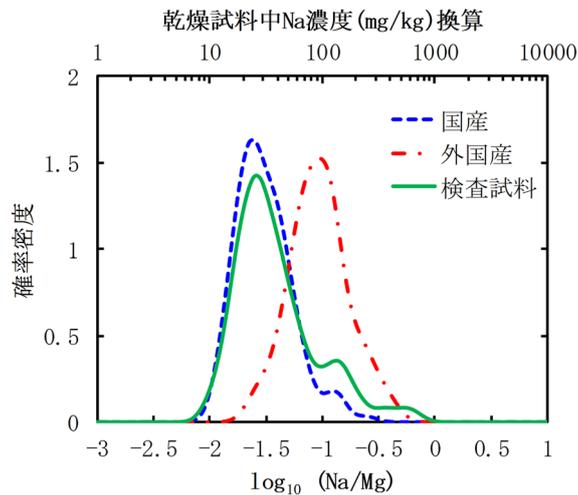


図 1 ネギ試料の Na の Mg 濃度比の常用対数の分布

データの詳細は 2.2.1 参照

## 2. 実験方法

### 2.1 元素分析の検証

#### 2.1.1 試料

市販の群馬県産と表示された同一ロットのネギ 22 個体を使用した。

#### 2.1.2 試薬

実験に使用した水は、超純水製造装置 (Milli-Q Integral 3、メルクミリポア) で製造した超純水 (比抵抗値  $18 \text{ M}\Omega \cdot \text{cm}$  以上) を用いた。試料の酸分解には、61%硝酸 (高純度電子工業用、関東化学) 及び 70%過塩素酸 (TAMAPURE-AA-100、多摩化学工業) を用いた。元素濃度測定用の検量線用混合標準溶液は、単元素標準溶液 (メルク、シグマアルドリッチ及び AccuStandard) を適宜希釈して調製した。また、In の単元素標準溶液を内標準溶液として希釈調製し、検量線用混合標準溶液及び試料溶液に添加した。

#### 2.1.3 装置

試料粉碎時の Na 汚染の検証には、セラミック刃のミキサー (B-400 : 日本ビュッヒ) を用いた。元素濃度の測定には、誘導結合プラズマ質量分析装置 (ICP-MS) (Agilent 7500cx : Agilent Technologies) 及び誘導結合プラズマ発光分光分析装置 (ICP-OES) (iCAP 6300 : Thermo Fisher Scientific) を使用した。

#### 2.1.4 試料粉碎時のNa汚染の検証

試料粉碎時の Na の汚染を検証するため、ミキサーに超純水約 50 mL を加え、混合したものを回収し、超純水で 50 mL に定容したものを ICP-OES 分析に供した。本操作は 4 点併行で実施した。

また、これとは別に、ネギ 2 個体について超純水で表面を洗浄後、セラミック包丁で基部から 1~10 cm の部位を採取し、更に縦に 4 分割して 2 個体分の対角を採取する縮分を行った。一方は等量の超純水を添加してミキサーで粉碎し、酸分解及び固形分測定に供した。もう一方は凍結乾燥し、ハンマーで粉碎したものを酸分解に供した。

#### 2.1.5 試料調製

試料 10 個体については超純水で表面を洗浄後、セラミック包丁で基部、基部から 0~1 cm 及び基部から 1~10 cm の 3 点の部位に分割し、更に基部から 1~10 cm の部位については表面に付いた傷のある部位とそれ以外の部位に分割した。各部位について凍結乾燥し、ハンマーで粉碎して 10 個体分を混合したものを酸分解に供した。

時間経過に伴う Na の濃度変動を確認するため、試料 5 個体については購入当日に、別の試料 5 個体については購入から 21 日冷蔵保存した後で、超純水で表面を洗浄後、セラミック包丁で基部から 1~4 cm、4~7 cm 及び 7~10 cm の各部位を採取した。各個体、各部位について凍結乾燥し、ハンマーで粉碎したものを酸分解に供した。

#### 2.1.6 酸分解

乾燥試料約 0.5 g 又は粉碎試料約 5 g を採取した樹脂製ヒータブルビーカーに 61 %硝酸を 10 mL 加え、樹脂製時計皿でふたをし、120 °C のホットプレート上で加熱した。褐色のガスが発生する激しい反応が収まった後に放冷し、70 %過塩素酸を 2.5 mL 加え、徐々に温度を上げながら 180~230 °C の範囲で加熱し、分解液が透明かつ無色または薄い黄色になるまで分解した。その後時計皿を外し、約 230 °C で加熱して乾固した。ビーカーに 1 %硝酸を 5~10 mL 程度加え、ビーカー内の残留物をホットプレート上で加熱・溶解し、乾燥試料は 50 mL 容、粉碎試料は 25 mL 容の樹脂製全量フラスコに移す操作を 3 回繰り返す。内標準として In を 5 µg/L になるように加え、1 %硝酸で定容後、元素濃度測定用試料溶液とした。時間経過に伴う Na の濃度変動確認試料は 1 点で、ミキサーで粉碎した試料は 3 点併行で、それ以外の試料は 2 点併行で分析した。

#### 2.1.7 固形分測定

粉碎試料約 5 g をアルミ箔容器に採取して、恒温乾燥器を用いて 70 °C で 24 時間乾燥し、乾燥前後の重量差により、粉碎試料中の固形分含有量を求めた。固形分含有量は、各元素の粉碎試料中濃度から乾燥重量あたりの濃度（以下「乾燥試料中濃度」という。）を求めるのに用いた。固形分測定は 3 点併行で実施した。

#### 2.1.8 機器測定

ICP-MS を用いて内標準法により Mn、Fe、Co、Ni、Cu、Zn、Rb、Mo、Cd、Cs、La、Ce 及び Tl の 13 元素の元素濃度を測定した。また、ICP-OES を用いて検量線法により Na、Mg、P、K、Ca、Sr 及び Ba の 7 元素の元素濃度を測定した。

### 2.2 選択バイアスへの対応

#### 2.2.1 データ

Ariyama らの収集したネギ 101 試料（国産 65 試料、外国産 36 試料）<sup>3)</sup>、平成 25 年度に収集したネギ 45 試料（国産 33 試料、外国産 12 試料、一部は一色ら<sup>4)</sup>と共通）及び平成 27 年度に収集したネギ 72 試料（国産 46 試料、外国産 26 試料）並びに農林水産消費安全技術センターが平成 22 年度から平成 25 年度に分析した検査試料 398 試料の 20 元素（Na、Mg、P、K、Ca、Mn、Fe、Co、Ni、Cu、Zn、Rb、Sr、Mo、Cd、Cs、Ba、La、Ce、Tl）の濃度の結果を用いた。ただし、La 及び Ce については半教師付き分類及び教師付き分類の計算には使用しなかった。また教師付き分類の実施に当たっては、検査試料 398 試料のうち 144 試料のデータをランダムに抽出し、検査試料以外のデータと組み合わせた。

なお、平成 22 年度から平成 24 年度までの検査試料の結果については各元素の Mg 濃度比のデータ、平成 25 年度の検査試料の結果については各元素の粉砕試料中濃度のデータのみであり、乾燥試料中濃度のデータは存在しなかったが、半教師付き分類及び教師付き分類の計算に用いた変数は各元素の Mg 濃度比の常用対数であるため、以降の計算への影響はほとんど無視できる。Horwitz 式の修正式<sup>10)</sup>による標準不確かさの計算に当たっては、検査試料以外のデータから得た乾燥試料中 Mg 濃度の分布から得た乱数又は粉砕試料の固形分を 5 % に固定した場合の値を用いた。

## 2.2.2 半教師付き分類

半教師付き分類の解析は Grandvalet ら<sup>11)</sup>が報告したエントロピー最小化の概念を人工ニューラルネットワークに応用した方法を採用した。ラベルのある標本から計算される交差エントロピー

$$-y_i \cdot \ln p_i - (1 - y_i) \cdot \ln (1 - p_i) \quad \dots \text{(式 1)}$$

$y_i$  : 試料  $i$  のラベル (0 又は 1)  
 $p_i$  : 試料  $i$  のラベルの予測値

及びラベルのない未知標本から計算されるエントロピー

$$-p_i \cdot \ln p_i - (1 - p_i) \cdot \ln (1 - p_i) \quad \dots \text{(式 2)}$$

の総和を計算した後、その総和（実際には  $L_1$  正則化項及び  $L_2$  正則化項を追加）を最小化することで判別モデルの係数の最適化を行った。

なお、Grandvalet らはロジスティック回帰及びカーネルロジスティック回帰に適用しているが、本研究では中間層（隠れ層）が 1 層の人工ニューラルネットワークに適用した。中間層の活性化関数には ReLU<sup>12)</sup>を使用した。また、Grandvalet らはエントロピーの総和にラグランジュ係数を掛けて正則化を行っているが、本研究では係数を使用しなかった。

式 2 は式 1 の  $y_i$  を  $p_i$  に置換したものとなっているが、これは試料のラベルの代わりに試料の事後確率を使用したと解釈することが可能である。実際、出力層の活性化関数をシグモイド関数

$$1 / \{1 + \exp(-x)\} \quad \dots \text{(式 3)}$$

としたとき、 $p_i$  は平均 0.5、分散  $1^2$  の正規分布及び平均 -0.5、分散  $1^2$  の正規分布の 2 種類の分布に対する、事前確率を 0.5 (50 %) としたときの事後確率となっている。一方、本研究で使用した未知標本は大半が国産と考えられ、事前確率は 0.5 (50 %) と大きく異なっている。そこで、

式1の $y_i$ に対応している $p_i$ の代わりに、事前確率の値を $p_0$ としてシグモイド関数の $x$ を $\ln \{p_0/(1-p_0)\}$ だけずらした値（ずらす方向は場合による。）を使用することにより、事後確率の値が適切な値になるよう補正を行った。 $p_0$ はパラメータの一種として扱い、判別結果から推定される事前確率の値と $p_0$ がおおむね一致するまで繰り返し計算を行った。

カーネル密度推定の結果を安定なものにするため、判別得点は出力層の結果そのものではなく、出力層の活性化関数を適用しない値とした。また、ReLUは負の値を0にする性質があることから、そのまま使用すると判別得点が特定の値に集中し、カーネル密度推定の結果が不安定なものになる。そのため、ある1つの中間層にReLUを適用しない状態で判別得点の計算を行う操作を全ての中間層に対して行い、その中から判別結果に影響しない結果となる特定の順序統計量を採用した。

以上の計算に当たっては、Leave-one-out cross-validation (LOOCV)を行った。その他の解析については中村ら<sup>13)</sup>の方法に準じたが、判別得点の不確かさの計算にはモンテカルロ法を採用するとともに、カーネル密度推定のバンド幅の計算はSheather & JonesのDirect Plug-in法<sup>14)</sup>に変更し、モンテカルロ法の試行回数分バンド幅の計算を繰り返した後、その中央値を採用した。また、カーネル密度推定に当たっては、モンテカルロ法による判別得点の出力結果全てを用いて計算を行った。モンテカルロ法の試行回数は2000とした。

計算にはR 3.5.1 (<https://www.r-project.org/>)を使用した。

### 2.2.3 教師付き分類

解析は中村ら<sup>13)</sup>の方法に準じたが、判別得点の不確かさの計算にはモンテカルロ法を採用するとともに、カーネル密度推定のバンド幅の計算はSheather & JonesのDirect plug-in法<sup>14)</sup>に変更し、モンテカルロ法の試行回数分バンド幅の計算を繰り返した後、その中央値を採用した。また、カーネル密度推定に当たっては、原産地（検査試料にあつては表示産地）の出荷量に応じた重みを付け、モンテカルロ法による判別得点の出力結果全てを用いて計算を行った。モンテカルロ法の試行回数は3500とした。

計算にはR 3.5.1を使用し、中間層（1層）の活性化関数がReLUの人工ニューラルネットワーク、RのパッケージMASS 7.3-50に同封されたlda関数による線形判別分析（LDA）及びqda関数による二次判別分析（QDA）、e1071 1.7-0に同封されたsvm関数によるサポートベクターマシン（SVM、線形カーネル及びガウシアンカーネルを適用）並びにnnet 7.3-12に同封されたmultinom関数によるロジスティック回帰で解析した。

## 3. 結果及び考察

### 3.1 元素分析の検証

試料粉碎時のNaの汚染を検証した結果、ミキサーからは最大で13 mg/kg相当（乾燥試料中濃度換算）のNaが抽出された。検出限界は10 mg/kgなので、分析結果に多少の影響を与える可能性がある結果となった。ただ、判別で問題となっているのは国産及び外国産のいずれの分布からも再現困難となっている300 mg/kg以上のものであり、ミキサーによる粉碎の判別結果への影響は小さいと考えられた。また、ミキサーで粉碎した試料の乾燥試料中Na濃度は15.6 mg/kg、凍結乾燥した試料の乾燥試料中Na濃度は15.2 mg/kgであり、ほとんど差はなかった。

ネギの部位別の元素濃度の結果を表1に示す。FAMICのネギの原産地判別検査における試料採取部位は基部から1~10 cmであるが、傷の有無によって特にLa及びCeの濃度が大きく異なっ

表1 ネギの部位別元素濃度(乾燥試料中、mg/kg)

元素	部位				検出限界 実測値
	基部	0~1 cm	1~10 cm		
			傷	傷以外	
Na	254	29.3	11.2	12.9	2
Mg	2049	1178	857	1020	2
P	5272	3290	2045	2577	4
K	15284	11200	9823	13936	141
Ca	6377	4191	8426	3616	8
Mn	11.6	6.88	8.54	8.04	0.004
Fe	106	17.0	22.3	13.8	2
Co	0.0511	0.0178	0.0361	0.0178	0.0002
Ni	0.278	0.191	0.221	0.155	0.002
Cu	8.39	3.83	1.36	2.88	0.04
Zn	24.8	20.0	16.3	17.8	0.2
Rb	6.03	4.05	3.59	5.12	0.006
Sr	23.5	14.7	26.6	12.5	0.05
Mo	0.266	0.132	0.187	0.147	0.001
Cd	0.120	0.0715	0.112	0.0595	0.0002
Cs	0.00775	0.00209	0.00276	0.00236	0.00003
Ba	4.91	3.05	5.96	2.94	0.05
La	0.0203	0.00128	0.00657	0.00076	0.00006
Ce	0.0357	0.0014	0.0087	0.0005	0.0001
Tl	0.0145	0.0084	0.0110	0.00734	0.00004

ていた。La は国内の土壤中に 16.5 mg/kg、Ce は 30.2 mg/kg (いずれも中央値) 程度含まれており<sup>15)</sup>、ネギと比較して高濃度であることから、傷に付着した土壌の影響を受けやすいものと考えられた。なお、有山ら<sup>16)</sup>も La 及び Ce が土壌の影響を受けると報告しているが、今回の結果ほど顕著な差は出ていない。一方、Na を含むその他の元素については傷の有無による濃度差が小さかった。Na は Na<sub>2</sub>O 換算で国内の土壤中に 2.16 % (中央値、Na として 16000 mg/kg) 程度含まれており<sup>15)</sup>、ネギ本体と比較すると高濃度であるが、Na 塩が水溶性であることから土壌の影響は超純水で洗浄することで十分低減できているものと考えられた。有山ら<sup>16)</sup>も、Na については土壌の影響が小さいと報告している。La 及び Ce 以外の元素は傷の有無により多少の影響はあるものの、傷のある部位の量は少なく、試料調製に伴いほぼ平均化されることから、傷の影響は小さいと考えられた。一方、La 及び Ce は産地よりも傷の有無に左右されてしまい、また、分析の品質管理においても問題を生じていることから、判別には適さないと判断した。

Na は基部付近に高濃度に含まれるため、時間経過に伴い上部に移動する可能性がないか確認した。その結果を表 2 に示す。部位間による差はいずれの個体でもほとんど見られないことから、保管期間を要因 1、個体間を要因 2、部位を残差とした 2 段完全枝分かれ実験と考へ、Na 濃度の常用対数に対して分散分析により確認したが、3 週間の保管前後で有意な差は見られなかった ( $p = 0.162$ )。仮に影響があったとしても図 1 のような分布を再現するほどのものではないと考えられた。

その他の分析上の問題、例えば標準溶液の調製ミスのような問題についても考察したが、図 1 のような分布は検査年度 (2.2.1 で使用しなかった平成 26 年度以降を含む。) や試験室にほぼ関係なく生じており、また、図 1 の低濃度領域の分布は国産の分布と一致していることから、元素分析における影響だけでは図 1 のような分布を再現するのは困難であり、情報バイアスによる分布への影響は小さいと考えられた。

表2 ネギの時間経過に伴うNa濃度の変動(乾燥試料中、mg/kg)

部位	購入直後					3週間保管				
	No. 1	No. 2	No. 3	No. 4	No. 5	No. 1	No. 2	No. 3	No. 4	No. 5
1~4 cm	14.8	10.7	16.5	17.2	14.9	16.3	13.8	22.1	16.3	39.5
4~7 cm	13.2	9.4	15.0	15.9	17.1	17.0	13.9	21.0	14.5	35.1
7~10 cm	14.2	10.6	16.0	15.8	15.7	16.4	13.9	21.9	13.4	31.4

### 3.2 選択バイアスへの対応

3.1 では Na の選択バイアスが問題の主要因であると結論付けたが、Na を除く元素で判別モデルを構築しても図2のようになり、検査試料の判別得点分布が国産の判別得点分布と外国産の判別得点分布の混合分布として表現できなかつた。これは、Na 以外の元素も選択バイアスの影響を受けていることを示唆している。また、選択バイアスの影響を受けている元素を除外すると、判別精度が実用レベルに到達しなくなるという問題もある。

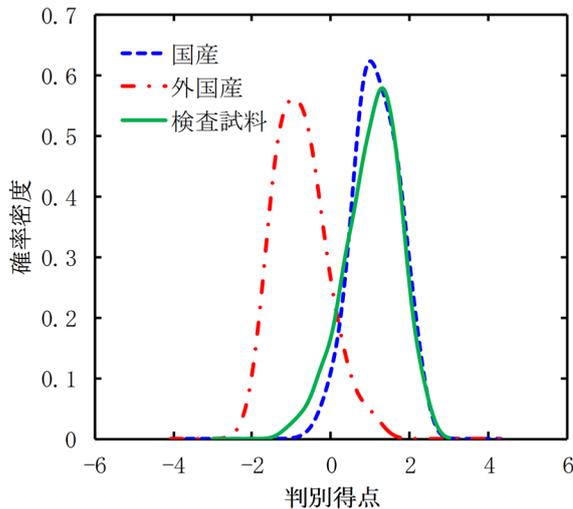


図2 Na を除外したネギの教師付き分類の結果

選択バイアスにより誤判別が起きる原理を図3に示す。図3では、未知標本の分布が陰性試料の分布と比較してy軸方向にずれたものになっている。図3には人工ニューラルネットワークで判別した結果も掲載されているが、判別得点が0となる線が未知標本の分布上に引かれている。ただ、図3を見る限り、未知標本は陽性試料の分布より陰性試料の分布に近いことから、本来は全て陰性として判別されるべきものである。

2.2.2 の半教師付き分類を図3のデータに適用した結果を図4に示す。図3と異なり、未知標本が全て陰性側に判別されており、選択バイアスによる影響を低減できていることが分かる。なお、事前確率を0.5 (50%) のままとすると図5のようになるが、図3は本来  $x=0$  で判別されるように設計されたデータであり、図4と比較するとやや不安定な結果となっている。

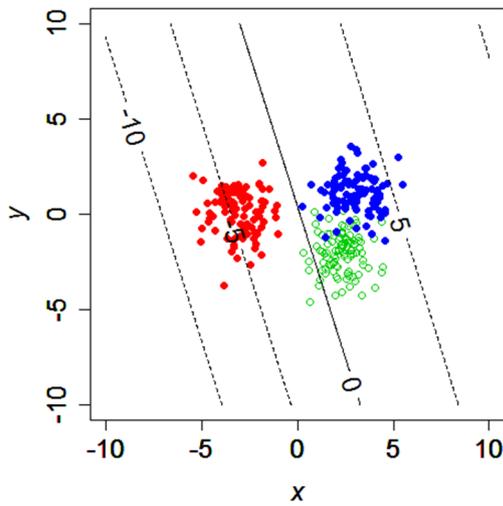


図 3 選択バイアスが存在するデータに教師付き分類を適用した例

青：陰性、赤：陽性、緑：未知標本（以下同じ）  
線上の値は判別得点

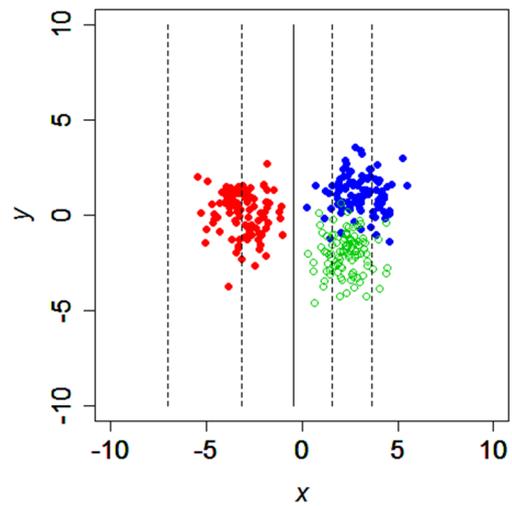


図 4 選択バイアスが存在するデータに半教師付き分類を適用した例

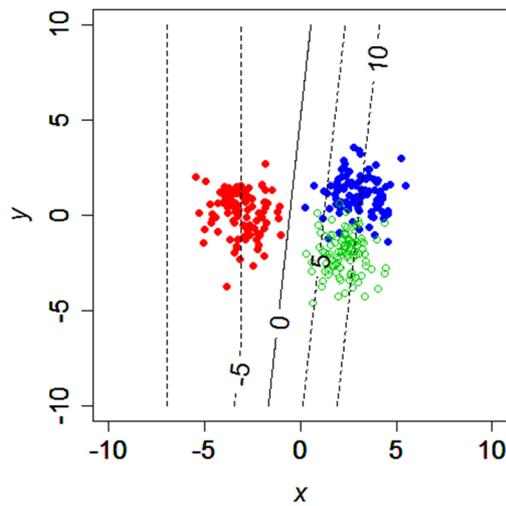


図 5 事前確率  $p_0$  を 0.5 としたときの半教師付き分類

2.2.2 の半教師付き分類をネギのデータに適用した結果を図 6 に示す。外国産の事前確率  $p_0$  は 0.004~0.011 (0.4~1.1%) に収束しており、また、検査試料（未知標本）の判別得点の分布は国産の判別得点の分布と一致していることから、検査試料の大半は国産と考えられた。ただ、図 6 の判別モデルについて、判別得点が正の場合に国産、負の場合に外国産としたとき、国産試料の 100.0% (144/144)、外国産試料の 78.4% (58/74) を正しく判別し、検査試料の 98.3% (391/398) を国産と判別したが、2.2.2 の半教師付き分類を判別モデルとしてそのまま使用するには不十分と考えられた。そこで、検査試料の大半は国産と考えられることから、検査試料を検査試料以外の国産試料と同数含める形で教師付き分類を行った。ただし、全ての検査試料を含めると標本数のかたよりによる別の選択バイアスが発生する可能性があることから、検査試料の一部のみを使用することにより選択バイアスの影響を最小化した。

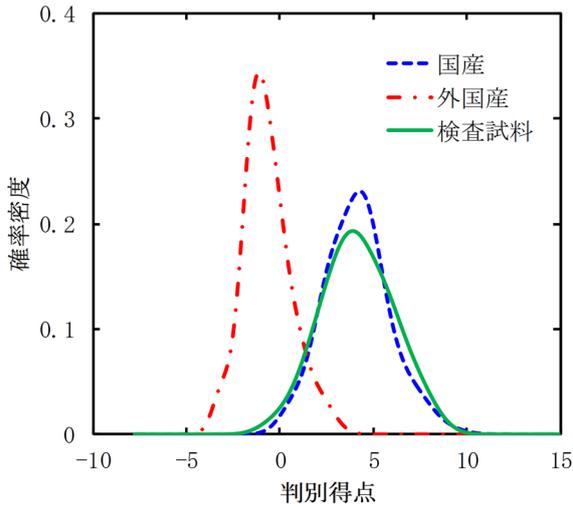


図 6 ネギの半教師付き分類の結果

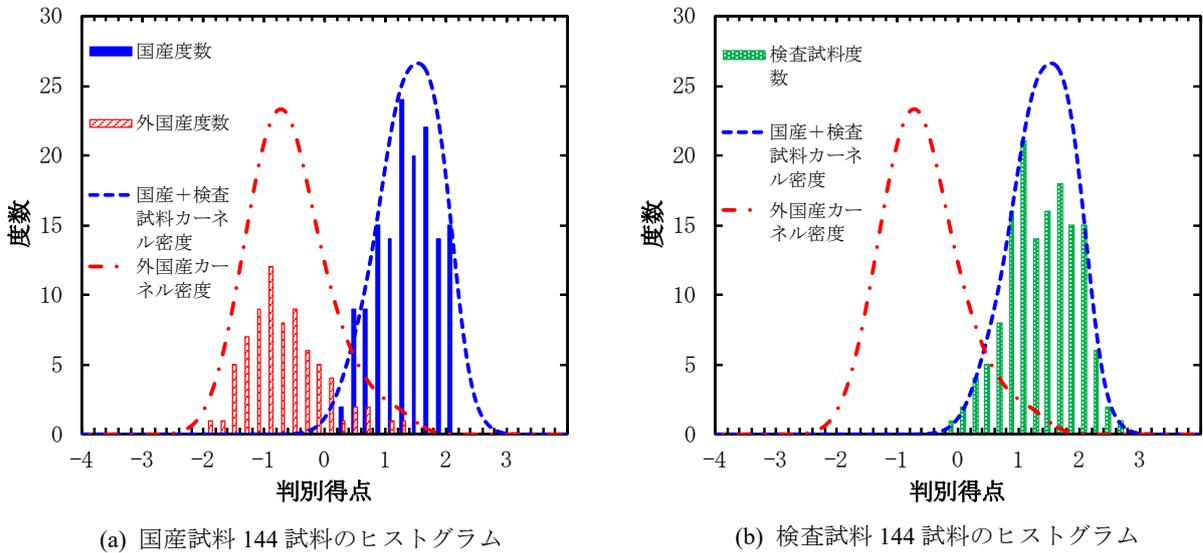


図 7 検査試料を含めたネギの教師付き分類の結果

2.2.3 の教師付き分類をネギのデータに適用した結果を図 7 に示す。図 7 の判別モデルについて、判別得点が正の場合に国産、負の場合に外国産としたとき、国産試料の 100.0 % (144/144)、外国産試料の 85.1 % (63/74) を正しく判別し、検査試料の 99.3 % (143/144) を国産と判別した。また、2.2.3 の教師付き分類に含めなかった検査試料 254 試料については、99.6 % (253/254) を国産と判別した。検査試料の判別得点の分布は国産試料の判別得点の分布とおおむね一致する結果となっており、半教師付き分類により得た事前確率  $p_0$  の推定値と比較しても問題のない結果となった。これにより、過去の開発時に使用した試料及び検査試料の両方に対して適切に判別できる判別モデルを構築することができた。

Na が高濃度となる検査試料の表示産地には海沿いの都道府県が多いといった傾向が見られたものの、ネギにおいて選択バイアスの原因が何かを確定することは困難であった。ただ、原産地判別全般において選択バイアスが発生する原因は複数考えられる。例えば、以下のようなものが挙げられる。

- ・各都道府県の出荷量等に応じた収集数とした場合に、市町村レベルの出荷量等の調査が不可能であることが原因で、特定の市町村からの収集に偏ってしまう
- ・原産地が確かな試料を収集する際、収集難易度の問題（収集のしやすさ、収集に協力的でない業者等の存在）により、収集先が偏ってしまう
- ・気象状況や天変地異によって年変動や季節変動が生じる
- ・畜産物では飼料原料の入手先（原産地）や飼料の組成が年によって変わる可能性があり、その影響で年変動が生じる

標本抽出の段階でこれらの問題に対処しようとする、原産地の確かさが失われたり原産地判別法の開発が著しく遅れたりするといった別の問題が発生するが、これらの問題はトレードオフの関係にある。そのため、標本抽出を是正するよりも、開発段階又は検査段階においてラベルのある標本とラベルのない未知標本の両方を収集しておき、半教師付き分類を用いて解析を行うといった手順の方が有効と考えられる。半教師付き分類の元々の目的は、ラベルのある標本の収集の困難さやコストの高さに対応するために未知標本を活用するというものであり<sup>10)</sup>、選択バイアスへの対応以外の面でもメリットがある。ただし、教師付き分類と比較するとまだ一般的ではなく、統計解析のハードルが上がるという欠点もある。また、本研究で使用した未知標本は大半が国産と考えられる検査品に偏っているが、このようなデータセットは半教師付き分類との相性が悪いという問題もある。これらの問題への対処は今後の課題と思われる。

## 文 献

- 1) 有山薫, 堀田博, 安井明美: ネギの産地判別のための無機元素測定法の確立と予備的検討, 分析化学, **52**(11), 969-978 (2003)
- 2) Ariyama, K.; Horita, H., Yasui, A.: Chemometric Techniques on Inorganic Elements Composition for the Determination of the Geographic Origin of Welsh Onions. *Anal. Sci.*, **20**, 871-877 (2004).
- 3) Ariyama, K.; Horita, H., Yasui, A.: Application of Inorganic Element Ratios to Chemometrics for Determination of the Geographic Origin of Welsh Onions. *J. Agric. Food Chem.*, **52**, 5803-5809 (2004).
- 4) 一色摩耶, 箱田晃子, 寺田昌市: ネギの原産地判別法の再検討, 農林水産消費安全技術センター調査研究報告, **38**, 1-5 (2014)
- 5) 対馬栄輝: 研究デザインと統計解析の基礎, 理学療法学, **44**(6), 463-469 (2017)
- 6) 杉山将, 山田誠, ドウ・プレシ マーティヌス・クリストフェル, リウ ソン: 非定常環境下での学習: 共変量シフト適応, クラスバランス変化適応, 変化検知, 日本統計学会誌, **44**(1), 113-136 (2014)
- 7) Huang, J., Smola, A. J., Gretton, A., Borgwardt, K. M., Schölkopf, B.: Correcting Sample Selection Bias by Unlabeled Data. *Adv. Neural Inf. Process. Syst.*, **19**, 601-608 (2007)
- 8) Yamada, M., Sugiyama, M., Matsui, T.: Semi-supervised Speaker Identification under Covariate Shift. *Signal Processing*, **90**(8), 2353-2361 (2010)
- 9) 藤野昭典, 上田修功, 永田昌明: ラベルありデータの選択バイアスに頑健な半教師あり学習, 情報処理学会研究報告, **2010-MPS-80**(8), 1-6 (2010)
- 10) Thompson, M.: Recent Trends in Inter-laboratory Precision at ppb and Sub-ppb Concentrations in Relation to Fitness for Purpose Criteria in Proficiency Testing. *Analyst*, **125**, 385-386 (2000)

- 11) Grandvalet, Y., Bengio, Y.: Semi-supervised Learning by Entropy Minimization. *Adv. Neural Inf. Process. Syst.*, **17**, 529-536 (2005)
- 12) Glorot, X., Bordes, A., Bengio, Y.: Deep Sparse Rectifier Neural Networks. *JMLR Workshop Conf. Proc.*, **15**, 315-323 (2011)
- 13) 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調査研究報告, **37**, 1-10 (2013)
- 14) Sheather, S. J., Jones, M. C.: A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *J. Royal Stat. Soc. Ser. B*, **53**(3), 683-690 (1991)
- 15) 海と陸の地球化学図, 国立研究開発法人産業技術総合研究所, 2019-06-06, <<https://gbank.gsj.jp/geochemmap/setumei/setumei-heikin.htm>>
- 16) Zhu, X.: Semi-Supervised Learning Literature Survey, University of Wisconsin-Madison, 2019-06-06, <[http://pages.cs.wisc.edu/~jerryzhu/pub/ssl\\_survey.pdf](http://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf)>