

野菜類の産地判別システムにおける効率的な微量元素含有量のコード化

佐藤 永欣[†] 上原 稔[‡] 下村 謙一郎[†] 山本 浩文[†] 上條 賢一[†]

E-mail: jju@prrc.itakura.toyo.ac.jp, uehara@cs.toyo.ac.jp, {shimomur,yamamoto,kamijo}@itakura.toyo.ac.jp

[†]東洋大学植物機能研究センター [‡]東洋大学工学部情報工学科

農産物の産地偽装を防止するため、我々は、野菜類の微量元素含有量を蓄積、比較することでその産地を判別する分散型産地判別システムを提案した。本システムは、各地の出荷団体に配したデータベースに微量元素含有量の分析結果を蓄積し、流通経路上の野菜類の微量元素含有量と比較することで野菜類の産地を判別し、トレーサビリティシステムの情報を検証することを目的としている。微量元素含有量は密ベクトルで表現されるため1対1照合を必要としたが、Similarity Preserve Hash (SPH) を利用して絞り込み、1対1照合を不要とした。しかし、効率が悪いため絞り込みを数段階に分けるなどして効率化した。しかし、力技的な手法であるため、限界がある。そこで、本論文では微量元素含有量をコード化する方法を検討し、効率的な検索を実現する。

Efficient Encoding of Trace Element Compositions in Geographical Origin Identification System of Vegetables

Nobuyoshi Sato[†] Minoru Uehara[‡] Koichiro Shimomura[†]

Hirobumi Yamamoto[†] Kenichi Kamijo[†]

[†]Plant Regulation Research Center, Toyo University

[‡]Dept. of Information and Computer Sciences, Toyo University

We proposed a distributed geographical origin identification system for vegetables by using trace element compositions to prevent geographical origin camouflaging. The system compares trace metal element compositions of vegetables which stored in databases and distribution channel, and verifies food traceability information. Since trace element compositions are expressed as dense vector, we employed SPH (Similarity Preserve Hash) to let one-to-one comparison be needless. However, it was not enough. In this paper, we discuss on coding of trace element compositions and realize efficient retrieval of trace element data from database.

1 はじめに

ここ10年ほど、産地偽装などの食品の安全や信頼に関わる事件が多発し、食品トレーサビリティシステムを導入が進みつつある。食品トレーサビリティシステムでは、流通用パッケージなどにIDを割り当て、バーコードやRFIDタグに記録し、追跡する。また、流通課程での食品の加工や小分けによりIDが再割り当てされることもあるため、パッケージの中身のすり替えや、加工時の不正行為の可能性がある。

一方、農産物自体の分析により産地や品種を特定する試みもある。遺伝子解析による品種の特定は実用に達しているが、産地の特定は遺伝子解析では不可能である。一般に、植物は土壌中に微量に含まれる金属イオンを生育中に吸収する。土壌の組成は産地や圃場によって違うため、土壌中の微量元素の含有量とこれらを吸収した植物体内の微量元素含有量は産地や圃場によって異なった特徴を示し、産地判別に利用できる。そこで我々は、野菜などの農産物にごく微量に含まれる元素の含有量を測定、蓄積、比較することで農産物の産地を判別するシステムを提案した[1]。本システムは、6種類の元素を産地判別に用いた場合で、概ね10万ヶ所の産地を判別できる。日本の農家は約200万軒あり[2]、同一作物・品種を栽培する農家数を考えると

必要十分と考えられる。本システムは農家からの野菜の出荷時に微量元素含有量を測定し、各産地に配したデータベースに蓄積する。産地偽装が疑われる野菜がある場合、該当する野菜の微量元素含有量を測定し、データベースに蓄積された微量元素含有量と比較し、類似する微量元素含有量の産地を本当の産地を判断する。この際、産地偽装が疑われる野菜類のトレーサビリティによる産地情報を用いる。トレーサビリティによる産地が類似する微量元素含有量の産地の集合に含まれる場合、産地偽装はないと判断し、そうでない場合は産地偽装の疑いがあると判断する。

微量元素含有量は実数値を各要素に持つ密ベクトルとして表現される。提案システムはクエリと検索対象の微量元素含有量のベクトルを、相関係数の計算という形で1対1照合する必要があり、スケーラブルではない。スケーラビリティを実現するための戦略としては、1対1照合を行う範囲を絞り込む、1対1照合以外の技法を開発するの2者が考えられるが、後者は現実的でない。そこで、我々はSimilarity Preserve Hash (SPH) を利用し、ある程度の効率化を実現した[3]。この手法は、クエリのベクトルと類似したベクトルをデータベース中から検索し、1対1照合を行う範囲を絞り込む。SPH自体が力技的な手法であり、効率が良いと

は言えないが、若干の工夫をすることにより検索所要時間を数秒程度まで短縮した [4]。SPH は密ベクトルに対する一般的な絞り込み手法として実用可能であるが、検索所要時間のこれ以上の短縮は困難と考えられる。そこで、本論文では微量元素含有量の値の分布に着目し、コード化することで判別所要時間を短縮する方法について検討する。

本論文の構成は以下の通りである。まず 2 章で関連研究を述べ、3 章で我々が開発した野菜類の分散型産地判別システムの概要と動作を述べる。4 章では微量元素含有量のコード化に必要な要件を検討した後、コード化する手法を述べ、5 章で評価を示す。最後にまとめを述べる。

2 関連研究

農産物などの植物体内、植物製品の微量元素含有量を用いた産地判別の試みは、ワイン [5][6]、コーヒー [7]、茶 [8] 等の産地や生産年が重要な製品の他に、オレンジジュース [9]、じゃがいも [10]、長ネギ [11]、玄米 [12] 等で農芸化学分野の研究者を中心に行われてきた。この他に、安定同位体比 [13] や有機物 [14] を用いた研究がある。

サンプル収集と分析の時間的・経済的コストの問題により、広範囲から継続的にサンプルを収集し、農産物の微量元素含有量の地理的・時間的分布を明らかにした研究は存在しない。産地により微量元素含有量が異なること、微量元素含有量の傾向は大陸、国、州県等の単位で大まかに分類できることは示されているが、農場・圃場レベルの産地判別は全く議論されていない。日本国内で行われた研究 [11][12] は原産地表示の適正化を視野にいれているが、特定国からの輸入品と国産品の識別が可能であることを示したに過ぎない。我々は、論文 [1] において産地判別を行う計算手法を提案し、細粒度の産地判別が可能であることを示した。

Similarity Preserve Hash (SPH)[15] は類似した入力に対して類似した出力を与えるベクトル空間モデルに基づくハッシュ関数である。SPH の入力は全て任意の固定長であり、 m 次元のベクトルを入力する必要がある。SPH では n ビットの出力を得るために n 個のランダムベクトルを用意する。ランダムベクトルを変更すると SPH の値も変わる。

m 次元空間内の n 個のランダムベクトル $R_n = (r_{n,1} \cdots r_{n,m})$ を生成し、ハッシュ値計算の規準とする。ハッシュ値を計算する対象のベクトル $S = (s_1 \cdots s_m)$ とする。 R_n と S のなす角 d_n は以下のように計算で

きる。

$$d_n = \frac{\sum (s_m - \bar{s})(r_{n,m} - \bar{r}_n)}{\sqrt{\sum (s_m - \bar{s})^2} \sqrt{\sum (r_{n,m} - \bar{r}_n)^2}}$$

ただし、 \bar{s} は $s_1 \cdots s_m$ の平均、 \bar{r}_n は $(r_{n,1} \cdots r_{n,m})$ の平均である。 d_n を用いて、 S のハッシュ値の n ビット目 h_n は、以下ようになる。

$$h_m = \begin{cases} 1 & d_n \geq 0 \\ 0 & d_n < 0 \end{cases}$$

SPH を検索に用いるときは、クエリのベクトル Q の SPH によるハッシュ値 $\text{SPH}(Q)$ を用いて検索する。しかし、 Q と等しいベクトルを検索する場合以外は、ヒットすべきベクトル R の SPH によるハッシュ値 $\text{SPH}(Q)$ が必ずしも同じ値を持たないことに注意する必要がある。すなわち、SPH を検索に利用する場合、 $\text{SPH}(Q)$ から数ビット程度以内のハミング距離にある全ての $\text{SPH}(R)$ が検索すべき範囲となる。ハッシュ値が k ビット幅のとき l ビットハミングすると、 l ビット以内のハミング距離にある全ての $\text{SPH}(R)$ の場合の数 a は、 $a = \sum_{i=1}^l k C_i$ となる。

3 産地判別システム

まず、微量元素含有量を用いた農産物の産地判別手法について述べる。我々が提案した手法は、出荷した農産物と流通経路上の農産物の微量元素含有量のベクトルを比較し、両者のなす角がほぼ 0 であれば産地が一致すると判定する。微量元素含有量はあらかじめ規準化する。ベクトルの比較は相関係数を計算することで行う。よって、なす角がほぼ 0 であるかの判定は、相関係数が適切な閾値を越えるかどうかの問題と等価である。

m 種類の元素の規準化した測定値のベクトル $S_i = (s_{i,1} s_{i,2} \cdots s_{i,m})$ で出荷された農産物の微量元素含有量を表す。 i は出荷された農産物の識別番号とする。流通経路上で入手した農産物の規準化した微量元素含有量をベクトル $U = (u_1 u_2 \cdots u_m)$ とすると、二つのベクトルの相関係数 $r_{S_i, U}$ は、

$$r_{S_i, U} = \frac{\sum (s_{i,m} - \bar{s}_i)(u_m - \bar{u})}{\sqrt{\sum (s_{i,m} - \bar{s}_i)^2} \sqrt{\sum (u_m - \bar{u})^2}} = \cos \theta_{S_i, U}$$

となる。 $r_{S_i, U}$ が 1 に近い閾値を超えたとき、 S_i と U は同一の圃場で生産されたと判断する。ただし、閾値を超える S_i は複数ある可能性がある。また、 $r_{S_i, U}$ が最大となる S_i が U の産地とは限らないため、前述のように食品トレーサビリティによる産地情報を検証する形で産地判別を行う。

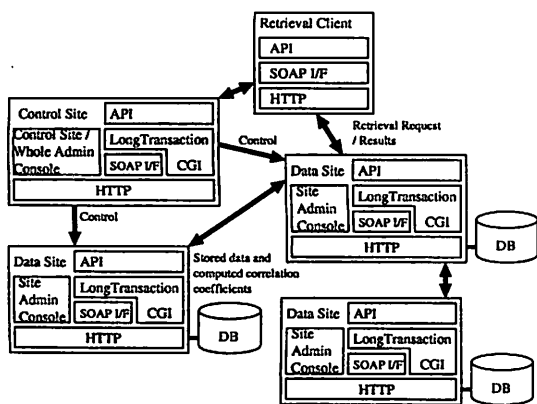


図1. 分散産地判別システムの概要

流通経路上での不正を最小限に抑えるため、微量元素の含有量は各産地の農家や出荷団体、農業試験場などにおいて測定し、蓄積することが望ましい。そこで、微量元素分析結果を蓄積するデータベースを各地の計算機に分散配置し、相関係数の計算も各地の計算機で並列に実行することで、計算時間の短縮を図る。図1にこの構成図を示す。微量元素分析結果データベースは、各地の農業試験場など、測定機器を持つ機関に Data Site として分散配置する。また、各産地での不正に対応するため、Control Site を出荷団体の全国組織などに1台配置する。各構成要素間の通信は SOAP によるメッセージ交換を用いる。また、この他に検索クライアントがある。

まず、蓄積データの平均値と標準偏差を計算する手順を述べる。この平均と標準偏差を用いて微量元素含有量のデータを規準化する。

1. 各データの平均の計算

サイト i における元素 e のデータの平均値 $\bar{x}_{i.e}$ は、

$$\bar{x}_{i.e} = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ije}$$

となる。ここで、 x_{ije} はサイト i における j 番目に蓄積されたサンプルの元素 e の含有量である。 e の各サイトの全てのサンプルの平均値 $\bar{x}_{..e}$ は、

$$\bar{x}_{..e} = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n m_i \bar{x}_{i.e}$$

となる。 $\bar{x}_{i.e}$ は各サイトローカルのデータのみで計算可能である。各サイトで計算した $\bar{x}_{i.e}$ と m_i を Control Site が収集し、 $\bar{x}_{..e}$ を計算する。

2. 各データの標準偏差の計算

全サイトの元素 e の標準偏差 s_e は、

$$s_e = \sqrt{\frac{1}{\sum_{i=1}^n m_i - 1} \sum_{i=1}^n \sum_{j=1}^{m_i} (x_{ije} - \bar{x}_{..e})^2}$$

$$= \sqrt{\frac{1}{\sum_{i=1}^n m_i - 1} \sum_{i=1}^n d_{ie}}$$

となる。ただし、 $d_{ie} = \sum_{j=1}^{m_i} (x_{ije} - \bar{x}_{..e})^2$ はサイト i における元素 e の残差平方和である。各サイトで d_{ie} を計算したのち、Control Site がこれを収集し、 s_e を計算する。

SPH を用いて相関係数の計算対象を絞りこむ場合は、この後、各サイトが各データの SPH を計算し、データベースに格納する。また、各サイトが持つ SPH の値を Control Site に送り、サイトの絞りこみに用いる。

次に、産地判別の動作を述べる。

1. 検索クライアント c は産地判別をしたい農産物の微量元素分析データのベクトル $U = (u_1 u_2 \dots u_l)$ 、作物名、品種と出荷日時を Control Site に送り、産地判別を依頼する。なお、 U はあらかじめ正規化されている。SPH を用いる場合は、 c は SPH をあらかじめ計算してから送る。
2. Control Site は、全てのサイトの集合 S に c から受け取った U を送り、相関係数の計算を依頼する。SPH を用いる場合は Control Site は c から受け取った SPH と、その SPH から数ビット以内のハミング距離にある SPH を持つサイトのみに関係数の計算を依頼する。
3. $s \in S$ は自身のデータベースから蓄積データの微量元素含有量 Y_{ij} を得る。SPH を用いる場合は、SPH の値をキーとしてデータベースから Y_{ij} を取り出す。
4. s は蓄積データと産地を判別したい農産物の微量元素含有量データの相関係数を以下のように計算する。

$$r_{(Y_{ij}, U)} = \frac{\sum_{e=1}^l (u_e - \bar{u})(y_{ije} - \bar{y}_{ij.})}{\sqrt{\sum_{e=1}^l (u_e - \bar{u})^2} \sqrt{\sum_{e=1}^l (y_{ije} - \bar{y}_{ij.})^2}}$$

5. s は適当な閾値 t よりさらに低い閾値 t' を越えた $r_{(Y_{ij}, U)}$ をその品種、産地、出荷日時などの情報とともに Control Site に返す。
6. Control Site は適当な閾値 t を越えた $r_{(Y_{ij}, U)}$ を持つ i, j を産地判別の結果とする。複数ある場合、品種、産地、日時が一致していれば正当なものとする。一致しない場合は全てを産地の候補として扱う。なお、 t は高い値 ($t = 0.95$ 程度) から、トレーサビリティによる品種、出荷日時、産地など

が一致する候補が見つかるまで徐々に下げる。下限 ($t = 0.80$ 程度) に達したら、一致する候補がないと判断する。

7. Control Site は上記の産地判別の結果を c に返す。
8. c は産地判別結果をユーザに提示する。

4 微量元素含有量のコード化手法

まず、クエリ対象の絞り込みについて述べる。本論文で述べる微量元素を用いた分散型産地判別システムは、類似した微量元素含有量データを見つけ出す情報検索システムの一つとみなすこともできる。一般的な情報検索では語と、語を含む文書やサイトは疎ベクトルであり、これらの組合せは疎行列として表現される。データが存在しない部分は全て0か帯域外の値であり、クエリの対象から除いてよい。このため、インデックスを作成する等による高速な検索手法が知られている。しかし、提案システムでは密なベクトルである。あるサンプルのある元素の値が0であることは、その元素が含有されていない、又は測定できないほど含有量が小さいことを意味する有効なデータである。したがって、当該サンプルをクエリの対象から単純に除くことはできない。一方、帯域外の異常な値を含む場合は、当該サンプルが産地判別対象として有効なデータでは無いことを示す場合もある。

以下では、密ベクトルの検索を効率的に行うため、密ベクトルを適当な長さのビット列または整数値にコード化する手法を検討する。

微量元素含有量の密ベクトルの効率的な検索を行うため、まず、微量元素含有量のデータそのものについて検討する。微量元素含有量のデータの特徴として、以下が挙げられる。

1. 各元素の値は独立した0または正の実数値である。
2. 各産地の微量元素含有量の平均は、概ね正規分布に従うと考えられる。ただし、片側の裾が切れているかも知れない。
3. 各産地の微量元素含有量自体も正規分布に従うと考えられる。

この模式図を図2に示す。

これらの特徴から、産地 d の h 番目の野菜類のサンプルの元素 A の測定値 x_{Adh} が、元素 A の含有量の分布のどのあたりに属するかを用いて分類する手法が考えられる。 x_{Adh} の全体の標本平均を \bar{x}_A 、標本標準偏差を s_A とする。 x_{Adh} は以下により規準化でき、規準化した微量元素含有量 y_{Adh} は標準正規分布に従う。

$$y_{Adh} = \frac{x_{Adh} - \bar{x}_A}{s_A}$$

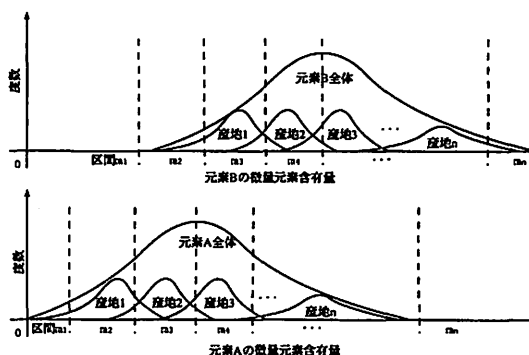


図2. 微量元素含有量の分布の概念図

y_{Adh} の分布範囲 $-\infty \leq y_{Adh} \leq \infty$ を b 個の区間に分割し、 $g (g = 1, 2, \dots, b)$ で呼称する。 y_{Adh} がどの区間に属するかは以下の式、

$$F(z_{Adh}) = \int_{-\infty}^{z_{Adh}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_{Adh}^2} \leq \frac{g}{b}$$

を満たす最小の g を見付ければ良いことになる。この方法を用いれば、 y_{Adh} が全ての区間 g に属する確率が等しくなる。実装においては、 $\int_{-\infty}^{z_{Adh}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_{Adh}^2}$ と x 軸により囲まれる部分の面積を b 等分するような y_{Adh} をいくつか計算し、大小の比較を行うだけで済む。

以上により、微量元素含有量 x_{Adh} を b 種類の符号 g に分類できる。以下では、分類した g をコード化する手法を述べる。図2を見てもわかるように、同一産地の野菜類が異なる区間に分類されてしまう可能性があることに留意する必要がある。すなわち、検索の際には、クエリ Q として与えられた野菜類の微量元素含有量の元素 A の含有量が属する区間 g_{QA} の前後の区間 $g_{QA} \pm 1$ にも、相関係数を計算するとクエリの微量元素含有量と高い相関を持つ野菜類のデータが格納されていることがある。ここで、 l は小さな正の整数であり、典型的には1であるが、クエリの微量元素含有量のばらつきや、 b の大きさにも依存する。隣接する区間に分類されてしまうことがある原因は、区間の境界が全産地を通じた微量元素含有量に基づいて決定され、クエリや各産地の野菜類の微量元素含有量とは直接関係していないためである。検索洩れが発生すると本システムの目的は達せられないため、全て検索する必要がある。また、 b を大きくし過ぎ、余りにも細かく分類すると検索すべきコードの数がが増えてしまう。また、隣接区間のさらに隣の区間も探す必要があることが考えられる。逆に b を小さくし過ぎると、絞りこみの効率が悪くなる。

次にコード化の手法を述べる。まず、元素 A のためのコードを b ビット幅とる。このコードをビット列 $b_b \dots b_2 b_1$ とし、全ての元素のためのコードを連結し

て特定の野菜類の微量元素含有量を表すコードとする。そして区間 g に分類された場合は、該当する b_g を 1 にし、それ以外は 0 としてコード化する。クエリとして与えられた野菜類 Q の元素 A の含有量が区間 g_{QA} に分類された場合、当然、該当するビット $b_{g_{QA}}$ のみを 1 にして検索すれば良い。また、 $g_{QA} \pm l$ の範囲に分類される全ての微量元素含有量を検索する場合は、該当するビットが全て 1 にして検索するが、絞り込み結果得られる範囲が広がってしまう。データベースに蓄積された元素 A のコードの集合を D_A 、クエリの元素 A のコード部分のビット列を b_{qA} で表すと、

$$\text{eachof } d \in D_A \& b_{qA} \neq 0$$

と表せる。この場合、産地判別に用いる元素数を e 種類とすれば、検索対象 1 つあたり高々 e 回の計算で済む。SPH の場合、概ね ${}_{16}C_3 = 560$ 回の計算が必要であった。また、分類された区間の隣接する区間を検索する必要が無い場合は単純にコードが一致しているかどうかを調べれば良く、この場合は高々 1 回で済む。

ここで、微量元素含有量のデータを絞りこむべき件数について議論する。本システムでは、最終的には相関係数の計算という形で 1 対 1 照合を必要とする。本システムは、日本の農家の数がほぼ 200 万軒あるため [2]、同一作物、同一品種をほぼ同じ時期に出荷する農家は 10 万軒程度であろうと想定している。以前、提案した方式による産地判別の有効性をシミュレーションにより検討したが、その際、クエリと類似している、産地の候補と考えられるデータの数は、各クエリあたり 100 個程度以内であった。余りに絞りこみ過ぎると検索洩れが増えてしまう。提案方式による絞りこみの結果、100 個程度が得られるのがよい。

微量元素含有量のデータをいくつの区間 p に分るべきかは、検索時に微量元素含有量のデータを絞りこむべき件数 d と、産地判別に利用する元素の数 e に依存する。本システムが蓄積する微量元素含有量の件数を t とすると、

$$d' = \frac{t}{p^e} \approx d$$

を満たす p を採用すればよい。 $t = 100000$ 、 $e = 6$ 、 $d = 100$ の場合で $p = 2$ とすると $d' = 1562$ 、 $p = 3$ とすると $d' = 137$ 、 $s = p$ とすると $d' = 24.4$ である。

5 評価

SPH の評価に用いた 10 万件の微量元素含有量が蓄積されたデータベースに対し、コード化の計算と結果の格納を行ったときの実行時間を表 1 に示す。コード化

表 1. コード化実行時間

分類区間数	実行時間 [m:s]
3	8:14.5
4	8:12.4

表 2. 絞りこみ所要時間

分類区間数	実行時間 [s]	実行時間 (隣接)[s]
3	0.533	1.18
4	0.643	0.760

の計算を行うプログラムは Perl により記述した。測定環境は、Pentium4 2.8GHz、1GB Memory、FreeBSD 5.5-RELEASE、PostgreSQL 7.4.13、Perl 5.8.8、Perl DBI 1.51 を使用した。SPH による絞りこみの所要時間の評価と同じ計算機を持ちいた。3 区間、4 区間に分類した場合の双方で概ね 8 分強であり、10 万件のデータベース中の蓄積データに対する操作としては十分早い。新規にデータを追加する場合には、データの追加と同時にコード化する計算を行えばよいが、この所要時間は無視してよいほどに短い。

次に、クエリ 1 件を絞りこむのに要する時間を表 2 の実行時間の列に示す。全ての場合で 1 秒以内に終了している。ただし、Perl で記述した絞りこみを行うプログラムの起動時間を含んでいる。以前用いていた SPH による手法では、3 秒強であったため [4]、提案方式は大幅な効率化を実現したことになる。クエリの微量元素含有量が分類された区間の隣接した区間にも、検索すべき微量元素含有量データが存在する場合の検索所要時間を表 2 の実行時間 (隣接) の列に示す。6 元素全てにおいて、両隣を検索したため、この結果は絞りこみに要する時間の最悪値と考えてよい。いずれも所要時間が延びている。また、3 区間に分類した場合は、絞りこみが不十分となり、データベースから取り出す件数が増えたため、4 区間に分類した場合よりも明らかに長い時間を要している。

提案方式による絞りこみの結果、得られる微量元素含有量の件数を表 3 に示す。どちらの場合でも絞りこみの結果としては十分小さくなっている。

表 4 に両隣の隣接区間に分類される微量元素含有量も含めて絞りこんだ場合の件数を示す。最悪の条件であるが、4 区間に分類した場合はある程度絞りこめていいる。1 対 1 照合の所要時間にもよるが、隣接区間を

表 3. 絞りこみの結果得られる件数

分類区間数	最大	最小	平均
3	331	35	167
4	107	3	33.4

表4. 絞りこみの結果得られる件数 (隣接)

分類区間数	最大	最小	平均
3	62510	5807	13708
4	11408	1504	3639

考慮しなければ3区間に分類、考慮するならば4区間に分類するのが効率がよい。

6 まとめ

本論文では、微量元素を用いた野菜類の分散型産地判別システムにおいて、産地判別のためのデータベースの検索を高速に実行する手法について述べた。従来、産地判別の計算は1対1照合によるしかなかったが、10万件程度のデータの蓄積が必要な本システムにはスケラビリティが不足していた。我々は過去に、汎用的な密ベクトルの検索高速化が可能なSPHによる手法を適用し、ある程度の高速度を実現したが、数秒程度までの短縮が限度であった。本論文では、野菜類の微量元素含有量のデータの特徴に基づき、各元素の微量元素含有量を数種類の区間に分類し、各区間をビット列の特定のビットに代表させるようにコード化した。本方式によって微量元素含有量を分類しても、隣の区間に分類されてしまうことがあるが、コード化方式を工夫することにより、複数の区間を高々1回で検索することができる。これにより、大幅な検索時間の短縮と負荷の軽減を実現した。

今後の課題として、以下の点が考えられる。まず、農芸化学分野の研究者の課題として、野菜類の微量元素含有量の地理的な分布の広範囲な調査である。また、同一の圃場内での微量元素含有量の安定性や、施肥の影響などの調査が必要である。つぎに、情報関連分野の課題として、本システムの分散システムとしての動作の効率化や、食品トレーサビリティシステムとの連携の実現などがあげられる。

謝辞

本研究は文部科学省産学連携研究推進事業(平成15年度~平成19年度)により実施された。

参考文献

- [1] 佐藤永欣, 上原稔, 玉岡迅, 下村謙一郎, 山本浩文, 上條賢一, 微量元素含有量を用いた野菜類の分散型産地判別システム, 情報処理学会論文誌, Vol.47, No.7, pp.2151-2159.
- [2] 総務省統計研修所(編): 日本の統計, 総務省統計局(2006). <http://www.stat.go.jp/data/nihon/>.

- [3] Sato, N., Uehara, M., Tamaoka, J., Shimomura, K., Yamamoto, H. and Kamijo, K.: Target Selection by Similarity Preserve Hash in Distributed System for Geographical Origin Identification of Vegetables, *Proc. 20th Int. Conf. on Advanced Info. Networking and Applications (AINA2006)*, pp. 173-178 (2006).
- [4] 佐藤永欣, 上原稔, 下村謙一郎, 山本浩文, 上條賢一, 野菜類の分散型産地判別システムにおけるSPHの効率化, マルチメディア, 分散, 協調とモバイルシンポジウム, pp. 917-920 (2006).
- [5] Baxter, M. J., Crews, H. M., Dennis, M. J., Goodall, I. and Anderson, D.: The Determination of the Authenticity of Wine from Its Trace Element Composition, *Food Chemistry*, Vol. 60, No. 3, pp. 443-450 (1997).
- [6] Kallithraka, S., Arvanitoyannis, I., Kefalas, P., El-Zajouli, A., Soufleros, E. and Psarra, E.: Instrumental and Sensory Analysis of Greek Wines; Implementation of Principal Component Analysis (PCA) for Classification According to Geographical Origin, *Food Chemistry*, Vol. 73, No. 4, pp. 501-514 (2001).
- [7] Anderson, K. A. and Smith, B. W.: Chemical Profiling to Differentiate Geographic Growing Origins of Coffee, *J. Agric. Food Chem.*, Vol. 50, No. 7, pp. 2068-2075 (2002).
- [8] Fernández-Cáceres, P. L., Martín, M. J., Pablos, F. and González, A. G.: Differentiation of Tea (*Camellia sinensis*) Varieties and Their Geographical Origin According to their Metal Content, *J. Agric. Food Chem.*, Vol. 49, No. 10, pp. 4775-4779 (2001).
- [9] Simpkins, W. A., Louie, H., Wu, M., Harrison, M. and Goldberg, D.: Trace Elements in Australian Orange Juice and Other Products, *Food Chemistry*, Vol. 71, No. 4, pp. 423-433 (2000).
- [10] Anderson, K. A., Magnuson, B. A., Tschirgi, M. L. and Smith, B.: Determining the Geographical Origin of Potatoes with Trace Metal Analysis Using Statistical and Neural Network Classifiers, *J. Agric. Food Chem.*, Vol. 47, No. 4, pp. 1568-1575 (1999).
- [11] 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, Vol. 52, No. 11, pp. 969-978 (2003).
- [12] 安井明美, 進藤久美子: 玄米中の無機元素組成による産地判別, 分析化学, Vol. 49, No. 6, pp. 406-410 (2000).
- [13] Martin, G. J., Guillou, C., Martin, M. L., Cabanis, M.-T., Tep, Y. and Aerny, J.: Natural Factors of Isotope Fractionation and the Characterization of Wines, *J. Agric. Food Chem.*, Vol. 36, No. 2, pp. 316-322 (1988).
- [14] McDonald, M. S., Hughes, M., Burns, J., Lean, M. E., Matthews, D. and Crozier, A.: Survey of the Free and Conjugated Myricetin and Quercetin Content of Red Wines of Different Geographical Origin, *J. Agric. Food Chem.*, Vol. 46, No. 2, pp. 368-375 (1998).
- [15] Moses S. Charikar, "Similarity Estimation Techniques from Rounding Algorithms", in Proc. of 4th annual ACM Symposium on Theory of computing, pp.380-388. 2002.