

サンプル選別と半教師有り学習型 SOM を統合した たんぱく質発現量からの食品機能性推定法

山 森 一 人^{†1} 吉 原 郁 夫^{†1}
江 藤 望^{†2} 西 山 和 夫^{†2}

重みの次元数を拡張し、拡張した次元の重みを出力として扱う自己組織化マップの半教師有り学習法を用いた、たんぱく質発現量からの生理活性推定法を提案する。学習を通じた推定には適切な学習サンプルが必要である。そのため、複数の測定値の分散を求め、一定以上の分散となる対象物のデータは除外し、さらにスミルノフ-グラブス検定により外れ値と思われる測定値を除外する。たんぱく質発現量と生理活性値はそれぞれ別個に測定されているため、単回帰分析を応用した対応付けを行って学習サンプルを構成し、これらを半教師有り学習型自己組織化マップに学習させることで推定精度向上を図る。抗酸化ストレス活性など 6 種類の活性について実データに基づく検証を行い良好な結果を得た。

Physiological Activity Estimation from Protein Expression Levels using Semi-supervised Self-organizing Map with Reliability-reinforced Samples

KUNIHITO YAMAMORI,^{†1} IKUO YOSHIHARA,^{†1} NOZOMU ETO^{†2}
and KAZUO NISHIYAMA^{†2}

This paper proposes a method to estimate physiological activities of foods from protein expression levels by semi-supervised self-organizing map. To reinforce reliability of samples, some chemicals and constituents of foods that have large variance among measured protein expression levels or physiological activities are eliminated. In addition, outliers are removed by Smirnov-Grubbs test. The remaining protein expression levels are made correspondence with the physiological activities by simple-regression analysis to make training or validation samples. Experimental results shows that our proposed method drastically improves estimation accuracy.

1. はじめに

食品を通じた健康の維持や増進は、特段の生活習慣の改善を強いられることなく実行可能であり注目を浴びている。食品にはもともと、栄養としての第一次機能、味覚を刺激する第二次機能がある。食品が持つ体調調節機能は第三次機能と呼ばれ、生活習慣病の予防以外にも、例えばがん予防などへの応用が考えられている。例えば、村越ら¹⁾はカボチャやニンジンに含まれる α -カロテンにがん細胞増殖抑制機能があることを報告しており、健康増進法に定める特定保健用食品として、消費者庁は平成 23 年 3 月現在で 951 品目を挙げている²⁾。

食品がもつ第三次機能を利用するためには、どのような食品にどのような機能があるのかを調べる必要がある。しかし、食品や食品成分の種類は膨大であり、調理加工法によってもその活性は異なる。したがって、食品の生理活性を簡便に評価する方法が求められている。Buciński³⁾らはアブラナの新芽が持つ抗酸化機能を、アスコルビン酸量や総トコフェロールなど 6 種類の指標から階層型ニューラルネットワークにより推定する手法について報告している。本報告では、たんぱく質アレイ等により複数種の一括測定が可能なたんぱく質発現量を入力、推定したい生理活性値を出力とし、これらの関係を学習により獲得することを試みる。

本報告では、たんぱく質発現量からの生理活性推定の精度向上を目的とする。そのため、測定値の分散による対象物のスクリーニング、スミルノフ-グラブス検定による外れ値の除外、単回帰分析を応用した測定値間の対応付けを経て学習用、検証用サンプルの作成を行う。これらを半教師有り学習型自己組織化マップに学習させ、たんぱく質発現量の類似性を隣接する競合ノード間の近傍関係に対応づけると同時に、重みの一つとして出力である生理活性値を保持させる手法を提案する。提案手法により、抗酸化ストレス活性などを従来法と比べ高い精度で推定できることを示す。

2. 測定値の選別とサンプル作成

2.1 測定対象物の種類と項目

たんぱく質発現量と生理活性値は、医薬品や工業的に精製された食品成分 30 種類、農産

^{†1} 宮崎大学工学部
Faculty of Engineering, University of Miyazaki
^{†2} 宮崎大学農学部
Faculty of Agriculture, University of Miyazaki

表 1 主な測定対象物とその濃度

Table 1 Examples of constituents of foods or chemicals and their concentrations.

対象物名	濃度 (μM)		
	0.1	0.3	1.0
Lipoic acid	0.1	0.3	1.0
GABA	100	300	1000
Genistein	10	20	60
Quercetin	5	15	60
IFN	100	300	1000
玉葱葉抽出物	100	300	1000
ヘベス果皮抽出物	120	350	-

品抽出物 21 種類について、3 種類の濃度で測定されている。測定の対象物と使用した濃度の例を表 1 に示す。表 1 のヘベス果皮抽出物のように、一部の濃度で測定されていない場合もある。

51 種類の測定対象物において、13 種類のたんぱく質、および 6 種類の生理活性が測定されている。たんぱく質は Thioredoxin, Survivin, HSP70, XIAP, FADD, TXNRD1, HSP90, MxA, tNOX, NQO1, ERK2, p53, Bcl2 である。生理活性は以下の 6 種類について測定されている。

抗酸化ストレス活性： 酸化反応によって引き起こされる有害な酸化ストレスを抑制する活性。

細胞増殖抑制活性： がん細胞が増殖するのを抑制する活性。

抗炎症活性： アレルギー性炎症疾患の症状を低減する活性。

抗転移活性： がん細胞の転移の原因となる細胞の遊走を低減する活性。

血管新生抑制活性 (HRE)： がん細胞に酸素や栄養を供給する血管の形成を低酸素下で誘導する HRE を介した遺伝子発現を低減する活性。

血管新生抑制活性 (VEGF)： がん細胞に栄養を供給する血管を新生する際、血管内皮細胞を刺激するために生産する VEGF を低減する活性。

本報告では、上記のたんぱく質発現量を入力、各生理活性値を出力とし、これらの入出力間の関係を自己組織化マップにより自律的に獲得させた時の推定精度について議論する。

2.2 測定値の規格化

医薬品や農産品抽出物を HepG2 細胞に作用させ、たんぱく質発現量や生理活性値を測定する。測定された値の例を表 2 に示す。表 2 の濃度 $0\mu\text{M}$ は当該対象物を細胞に与えずに測定した値を表す。表 2 に示したように、同一対象物の各濃度において、測定が複数回行われ

表 2 抗酸化ストレス活性測定値と規格化の例

Table 2 Anti-oxidative stress activities and their normalized values.

濃度 (μM)	測定値					平均値
	0	362.0	230.0	232.0	226.0	
0.1	261.0	259.0	278.0	290.0	304.0	-
濃度 (μM)	規格化後の測定値					平均値
	0.1	1.034	1.026	1.101	1.149	

表 3 Lipoic acid を作用させたときの FADD の規格化発現量

Table 3 Normalized expression levels of protein FADD by the effect of lipoic acid.

濃度 (μM)	発現量						分散
	0.1	0.529	0.601	0.630	0.728	0.280	
0.3	0.350	0.263	0.947	0.896	0.123	0.104	0.144
1.0	0.518	0.560	1.177	1.061	0.142	0.106	0.201

ている。こうした測定値は細胞の状態等で値が変動するため、対象物を与えなかった場合の平均測定値（コントロール）で値を除すことで規格化を行う。表 2 の例では、濃度 $0\mu\text{M}$ 時の平均測定値 252.4 で、濃度 $0.1\mu\text{M}$ の測定値 5 つをそれぞれ除することになる。これにより、規格化後の値が 1.0 であることは、当該対象物が細胞に影響を与えないことを意味し、規格化後の測定値が 1.0 より大きい、あるいは小さいという判断基準で対象物の生理活性を判定できる。以降、測定値とは規格化後の値を指す。なお、6 種類の活性のうち、抗酸化ストレス活性については測定値が 1.0 より大きいほど活性が大きく、それ以外については測定値が 1.0 より小さいほど活性が大きいことを表す。

2.3 分散による測定対象物の選別

濃度の高い対象物を細胞に作用させた場合など細胞内部の状態が大きく変動したり、あるいは細胞が死滅するなどして、他の場合と比べて大きく異なる値が測定される場合がある。表 3 に、Lipoic acid を作用させたときのたんぱく質 FADD の測定値を示す。

表 3 に示したように、濃度 $0.3\mu\text{M}$ の Lipoic acid を作用させたときのたんぱく質 FADD の発現量の最小値は 0.104、最大値は 0.947、分散が 0.144 と大きく、細胞状態が大きく変化している可能性がある。濃度 $1.0\mu\text{M}$ のときは分散が 0.201 とさらに大きくなり、これらの測定値をそのまま利用した場合、正確な推定ができない恐れがある。そこで、測定値の分散がある一定以上の場合、細胞状態の変化により正しい測定が行われなかったものとして、学習サンプル等を構成する対象物から除外する。しきい値とした分散は 0.05, 0.10, 0.20 の

3種類とした。しきい値を 0.05 とした場合、たんぱく質 FADD の発現量は Lipoic acid の濃度 $0.1\mu\text{M}$ での測定値のみを用いることになる。また、しきい値を 0.20 とした場合、同じく濃度 $0.1\mu\text{M}$, $0.3\mu\text{M}$ の 2 つの場合の測定値を用いる。

2.4 スミルノフ-グラブス検定による測定値の選別

分散により対象物を選別したあとでも、個別の測定値を見ると一部に他とは値がはずれているものが存在することがある。例えば表 3 に示した濃度 $0.1\mu\text{M}$ の Lipoic acid を細胞に作用させたときのたんぱく質 FADD の発現量を見ると、4 個の測定値が $0.5 \sim 0.7$ 程度にあるのに対し、2 個の測定値が $0.2 \sim 0.3$ 程度とやや離れている。そこで、ある濃度の対象物での測定値に対し、スミルノフ-グラブス検定⁴⁾を行うことで外れ値の除外を行う。

まず、ある対象物での測定値の最大値(または最小値) x_i は「外れ値ではない」という仮説を立てる。一方、 x_i は「外れ値である」という対立仮説を立てる。次に、式 (1) で示した検定統計量 τ_i を求める。

$$\tau_i = \frac{|x_i - \mu|}{\sigma}. \quad (1)$$

ここで、 μ は測定値の平均、 σ は標準偏差である。

スミルノフ-グラブス検定では、検定統計量 τ_i と有意水準 α によって決定される有意点 t を比較することで仮説の採否を決定する。帰無仮説の採否を決定するための有意点は、 t 分布を利用した近似である式 (2) で求められる。

$$t = \frac{(n-1)u}{\sqrt{n(n-2+u^2)}}. \quad (2)$$

ここで、 n は標本数、 u は有意水準 α で自由度を $(n-2)$ とした t 分布の、 $\frac{\alpha}{n} \times 100$ のパーセントイルである。

式 (1) から求めた検定統計量 τ_i が、式 (2) で求めたスミルノフ-グラブス検定における有意点 t より大きい場合、測定値 x_i は棄却域に入ることから外れ値として除外する。この操作を、どの測定値も外れ値として除外されなくなるまで繰り返す。

2.5 単回帰分析を応用したサンプルの作成

入力信号に相当するたんぱく質発現量と教師信号に相当する生理活性値はそれぞれ別々に測定されているため、何らかの方法で適切に組み合わせ、学習や検証に用いるサンプルを構築する必要がある。そこで、分散による対象物の選定、スミルノフ-グラブス検定による外れ値の除外を経た測定値は、同一対象物、同一濃度では線型の関係にあるという前提に立ち、単回帰分析を応用した対応付けを行う。

表 4 たんぱく質発現量と生理活性の組み合わせ

Table 4 Combination of protein expression levels and physiological activities.

組み合わせ			背反確率
(x_1, y_1)	(x_2, y_2)	(x_3, y_3)	p_1
(x_1, y_1)	(x_2, y_3)	(x_3, y_2)	p_2
\vdots	\vdots	\vdots	\vdots
(x_1, y_3)	(x_2, y_2)	(x_3, y_1)	p_6

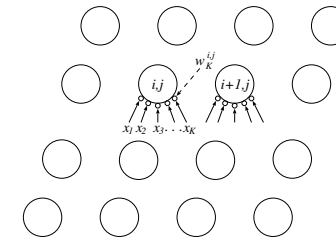


図 1 SOM の例
 Fig. 1 An example of SOM.

たんぱく質発現量 $\{x_1, x_2, x_3\}$ 、生理活性値 $\{y_1, y_2, y_3\}$ が検定後に残っている場合を考える。このとき、たんぱく質発現量と生理活性値の組み合わせ方は表 4 に示す 6 通りとなる。次に、表 4 に示した各組み合わせに対し単回帰分析を行い、たんぱく質発現量と生理活性値が互いに無関係である背反確率 p_i を求める。すべての組み合わせの中で p_i が最小となるものを、学習用、あるいは検証用サンプルとして採用する。

3. 自己組織化マップ

3.1 自己組織化マップの構造と学習アルゴリズム

自己組織化マップ (Self-organizing map, SOM) は、Kohonen⁵⁾ により提案されたニューラルネットワークの一種であり、教師なし学習アルゴリズムにより高次元の入力ベクトル間の類似関係を 2 次元格子状に配置された競合ノード間の隣接関係に対応づけることができる。図 1 に SOM の例を示す。時刻 t において、入力ベクトル $\mathbf{x}(t) \in R^k$ が二次元格子上に並べられた競合ノードすべてに提示された場合を考える。すべての競合ノードについて、入力ベクトルと競合ノード (i, j) が持つ重みベクトル $\mathbf{w}^{i,j}(t) \in R^k$ とのユークリッド距離 d_{ij} を式 (3) により計算し、 d_{ij} を最小とする勝者ノード c を決定する。

$$d_{ij} = \sum_{k=1}^K \sqrt{(x_k(t) - w_k^{ij}(t))^2}. \quad (3)$$

次に、勝者ノード c を中心としたある範囲（近傍領域）内にある競合ノードの重みを、入力ベクトルに近づくよう式 (4) に基づき更新する。

$$w^{ij}(t+1) = w^{ij}(t) + h_c(t)(\mathbf{x}(t) - w^{ij}(t)). \quad (4)$$

式 (4) の $h_c(t)$ は学習率であり、 t の増加に対して単調に減少する。近傍領域を学習が進むにつれて小さくしていくことで、競合ノード上にはそれぞれ入力ベクトルに類似した重みが獲得され、なおかつ隣接する競合ノードは類似した重みを持つ。

3.2 半教師有り学習型 SOM

3.1 節から分かる通り、SOM は出力に相当する値を持たないため、本研究のような用途には直接用いることができない。これに対し、半教師有り学習型 SOM^(6,7) では、入力ベクトルに教師信号 $\mathbf{y}(t) (\in R^L)$ を付加して $(\mathbf{x}(t), \mathbf{y}(t))$ とし、これに合わせて重みベクトルも $\mathbf{w} \in R^{K+L}$ に拡張する。勝者ノードの決定は式 (3) により教師信号を用いずに行い、重みの更新時には教師信号に対応する重みも含めてすべての重みを更新する。これにより、入力信号とは性質の異なる教師信号の影響を受けることなく、入力信号間の類似関係を競合ノード間の隣接関係に対応づけて、同時に重みの一部として出力値を各競合ノードに持たせることができる。

4. 食品機能性推定実験

4.1 実験条件

2 章で述べた操作を行って作成したサンプルを、学習用、検証用の 2 つに分けて推定実験を行う。学習用サンプルと検証用サンプルの選択はランダムに行い、5 セット分を準備する。対象とした活性は 2.1 節に示した 6 種類である。また、2.3 節で述べた分散による対象物選別の結果、たんぱく質 FADD は分散が大きく除外されたものが多かったため、FADD は入力から除いた。したがって、学習用、検証用のサンプルは入力信号であるたんぱく質発現量 12 種類、教師信号である生理活性値 1 種類の計 13 次元から成る。

自己組織化マップの競合層ノード数等のパラメータを表 5 に示す。学習率には、学習が進むにつれて線型に単調減少する式 (5) を用いる。

$$h_c(t) = \frac{\eta}{D_c^{i,j} + 1} \left(1 - \frac{t}{T}\right). \quad (5)$$

表 5 推定実験のパラメータ
Table 5 Parameters for experiments.

マップサイズ ($N \times N$)	15 × 15, 20 × 20, 25 × 25
しきい値とした分散	0.05, 0.10, 0.20
有意水準 (α)	0.10
最大学習回数 (T)	2,000
初期学習係数 (η)	0.5
初期近傍範囲係数 (r)	0.5
初期重み範囲	-2.0~2.0
試行回数	初期重みを変えて 20 回

表 6 学習用、検証用サンプル数 (学習用, 検証用)
Table 6 The number of samples for training and validation.

活性	σ^2	case1	case2	case3	case4	case5
抗酸化ストレス	0.05	(98,31)	(104,25)	(99,30)	(106,23)	(101,28)
	0.10	(245,67)	(262,50)	(248,64)	(248,64)	(254,58)
	0.20	(271,79)	(281,69)	(284,66)	(284,66)	(277,73)

ここで、 η は初期学習係数、 $D_c^{i,j}$ は勝者ノード c と近傍領域内の競合ノード (i, j) 間のユークリッド距離、 T は最大学習回数である。また、近傍領域の大きさ l は式 (6) に従い、学習が進むにつれて小さくする。

$$l = \left\lfloor Nr \left(1 - \frac{t}{T}\right) \right\rfloor. \quad (6)$$

ここで、 N は競合層の 1 辺をなす競合ノード数、 r は初期近傍範囲係数である。競合層端での不連続を避けるため、上下端、左右端をそれぞれ結合させたトラス型 SOM を用いる。

推定実験に用いる学習用サンプル、検証用サンプル数を表 6 に示す。なお、表 6 では、(学習用サンプル数, 検証用サンプル数) の形で記載している。推定実験は表 6 に示した各ケースについて初期重みを変えて 20 回づつ行う。

4.2 抗酸化ストレス活性推定実験結果

表 7 に、抗酸化ストレス活性を初期重みを変えて 20 回、自己組織化マップに学習させたときの、学習サンプルに対する平均二乗誤差 (RMSE) の最小値を示す。表 7 での太字は、5 ケースのうち誤差が最小のものを表す。表 7 から、しきい値とした分散の値が小さくサンプル数が少ない場合、マップサイズが大きいとサンプル 1 つに競合ノード 1 つが対応してしまい、学習サンプルに対する誤差が 0 となる場合が生じる。こうした場合はいわゆる過

表 7 抗酸化ストレス活性の各ケースでの学習サンプルに対する RMSE 最小値
Table 7 The minimum RMSE of anti-oxidative stress activity in each case.

σ^2	マップサイズ	case1	case2	case3	case4	case5
0.05	15 × 15	0.021	0.019	0.014	0.022	0.023
	20 × 20	0.000	0.001	0.000	0.002	0.000
0.10	15 × 15	0.093	0.109	0.097	0.099	0.103
	20 × 20	0.050	0.061	0.046	0.050	0.061
	25 × 25	0.021	0.025	0.021	0.020	0.026
0.20	20 × 20	0.078	0.082	0.079	0.079	0.077
	25 × 25	0.031	0.040	0.039	0.041	0.035

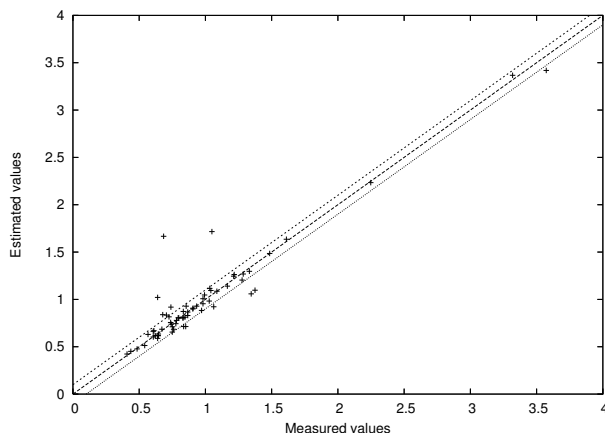


図 2 抗酸化ストレス活性推定結果 ($\sigma^2 = 0.10, 25 \times 25$)
Fig. 2 Estimation results of anti-oxidative stress activity ($\sigma^2 = 0.10, 25 \times 25$).

表 8 従来手法と提案手法での生理活性推定精度の比較

Table 8 Accuracy comparison of proposed method with previous works.

手法	抗酸化	細胞増殖抑制	抗炎症	抗転移	HRE	VEGF
提案手法	92.2%	98.4%	76.8%	86.8%	87.3%	83.8%
GP	41.4%	96.5%	67.7%	-	60.3%	-
MMNN	40.6%	92.2%	66.0%	43.1%	60.8%	39.2%

で推定できたものは 52 個, 81.3%であった。実際に活性があるサンプルを活性ありと推定できた割合を表す感度は 89.5%, 活性のないサンプルを活性なしと推定できた割合である特異度は 91.1%であり, 極めて高い精度で推定できている。

4.3 考 察

従来手法と比較しつつ, 実験結果について考察する。比較対象として, 上口ら⁸⁾によるサンプルの選別を行っていない遺伝的プログラミングによる推定結果, サンプル選別を行ったうえマルチモーダルニューラルネットワーク (MMNN) による推定を行った近藤ら⁹⁾による結果を用いる。なお, 文献 8), 9) では, 推定精度を検証用サンプルのうち絶対誤差 0.2 以下で推定できた割合で議論しているため, 本節では提案手法についても同様に絶対誤差 0.2 以下となった検証用サンプルの割合を改めて示す。

表 8 に, 従来手法および提案手法での推定精度の一覧を示す。表 8 より, すべての活性について提案手法の推定精度が上回っていることが分かる。

提案手法について, 特に推定精度の改善が著しい抗酸化ストレス活性を例に説明する。図 2 の例での, 競合ノードが持つ生理活性値に対応する重みの値を画像として図 3 に示す。図 3 では, 生理活性値に対応する重みが最大の競合ノードを白 (輝度値 255), 最小の競合ノードを黒 (輝度値 0) とし, その間を線型に補間している。また, 競合ノード間の境界に白線を引いている。競合ノードの座標は左上を (0,0) とし, 縦方向を y , 横方向を x として (y,x) の形で表す。

検証サンプル中最も生理活性値が大きいものは Chamomile-200 μ M であるが, この検証サンプルは競合ノード (11,21) に割り当てられている。これは図 3 の右から 4 列目の中央やや上にある白い競合ノードに対応する。

Chamomile-200 μ M の 6 サンプルはすべて 3.5 程度の大きな生理活性値を持っている。GP や MMNN では学習を通じてこれらのサンプルにも適合するよう個体や重みを変化させようとする。ところが, Chamomile-200 μ M 以外のほとんどのサンプルは大ききても 1.5 程度の生理活性値であるため, Chamomile-200 μ M のようなサンプルに適合させようとする変化は,

学習の状態に近く, 未知の検証サンプルに対する推定精度が低下する恐れがある。そのため, ある程度の学習サンプルと検証サンプル数が確保できている $\sigma^2 = 0.10$, マップサイズ 25×25 , case4 での検証用サンプルの推定結果を図 2 に示す。図 2 の横軸は抗酸化ストレス活性の測定値, 縦軸はその推定値であり, 推定精度が分かりやすい $y = x$ の直線とともに, その ± 0.1 となる $y = x + 0.1$, $y = x - 0.1$ の直線も合わせて描いている。

図 2 から, 他のサンプルとは値が離れた測定値 3.0 近辺のサンプルについても精度良く推定できていることが分かる。全検証サンプル 64 個のうち, 測定値との絶対誤差が 0.1 以下

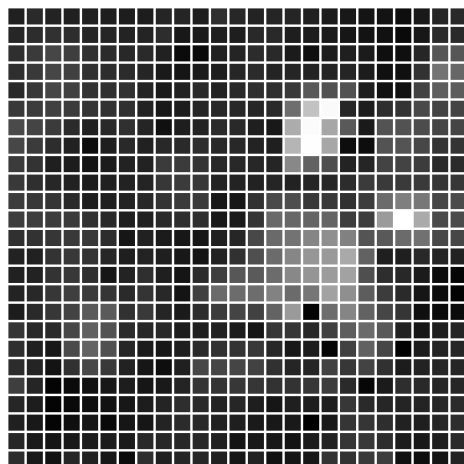


図3 抗酸化ストレス活性推定用マップ ($\sigma^2 = 0.10$, 25×25)

それ以外の大多数のサンプルへ対応しようとする変化で相殺されてしまう。一方、SOMでは学習が進むにつれて重みの更新範囲を縮小する。このため、Chamomile-200 μ Mのような他と異なる生理活性を持つサンプルについても、たんぱく質発現量が他と区別できる程度に異なっていれば、競合ノードのどれかが最終的に単独で割り当てられることにより、他のサンプルの影響を受けることなく生理活性値を高精度で近似できる。これにより、従来推定が困難であった高い、あるいは低い生理活性値を持つサンプルの推定精度が改善し、トータルでの推定精度が大幅に向上したと考えられる。

5. おわりに

本報告では、食品が持つ生理活性を測定が容易なたんぱく質発現量から高精度に推定する手法を提案した。対象物を細胞に作用させたときのたんぱく質発現量と生理活性値間の関係を学習により獲得するためには、適切なサンプルを用いることが重要である。そこで、あらかじめ対象物毎の測定値の分散を求め、分散が大きい対象物は正確な測定ができていないものとして除外した。さらに、同一対象物の複数の測定値に対しミルノフ-グラブス検定を行い、外れ値を取り除くことで測定値の妥当性を検証した。最後に、別々に測定されたたんぱく質発現量と生理活性値を対応づけてサンプルを構成するため、単回帰分析を応用して背反確率が最小となる組み合わせをもって学習、検証サンプルを作成した。

作成したサンプルを半教師有り学習型自己組織化マップに学習させ、たんぱく質発現量に基づいて自己組織化を行わせつつ、重みの一つとしてたんぱく質発現量に応じた生理活性値を獲得させ、6種類の生理活性について推定精度を評価した。その結果、測定値との絶対誤差0.1以下で推定できたサンプルは、最も低い抗炎症活性で約64%、最も高い細胞増殖抑制活性で89%となり、従来法に比べ大幅に推定精度を向上させることができた。

今後の課題として、脂質代謝機能などの他の活性への本手法の適用や、使用するたんぱく質を推定する活性に適したものみに精選することによる評価が挙げられる。

謝辞 本研究の一部は、日本科学技術振興機構 地域結集型共同研究「食の機能を中心としたがん予防基盤技術創出」により行われた。関係各位に感謝する。

参考文献

- 1) Murakoshi, M., Takayasu, J. et al.: Inhibitory Effects of α -Carotene on Proliferation of the Human Neuroblastoma Cell Line GOTO, *J. National Cancer Institute*, Vol.81(21), pp.1649–1652 (online), DOI:10.1093/jnci/81.21.1649 (1989).
- 2) 食品表示課：健康や栄養に関する表示の制度について、消費者庁（オンライン）、入手先(<http://www.caa.go.jp/foods/index4.html>)（参照 2011-3-24）。
- 3) Buciński, A., Zieliński, H. and Kozłowska, L.: Artificial Neural Networks for Prediction of Antioxidant Capacity of Cruciferous Sprouts, *Trends in Food Science & Technology*, Vol.15, pp.161–169 (2004).
- 4) 石村貞夫：入門はじめての統計解析，東京図書，東京（2006）。
- 5) Kohonen, T.: 自己組織化マップ，シュプリンガーフェアラーク東京，東京，3rd edition (2005). 監修：徳高平蔵他。
- 6) Heikkonen, J. and Oja, E.: Self-Organizing Maps for Visually Guided Collision-free Navigation, *Proc. 1993 International Joint Conference on Neural Networks*, pp.669–672 (1993).
- 7) de A.Barreto, G. and Araújo, A.F.: Temporal Associative Memory and Function Approximation with the Self-Organizing Map, *Proc. 2002 12th IEEE Workshop on Neural Networks for Signal Processing*, pp.109–118 (2002).
- 8) 上口真由美，山森一人，吉原郁夫，永濱清子：GPを用いたバイオマーカー発現量からの機能性食品の生理活性値推定，*情処研報（バイオ情報学）* 2009(25) (2009)。
- 9) Kondo, K., Yamamori, K. and Yoshihara, I.: Neural network with exponential output neuron for estimation of physiological activities from protein expression levels, *Proc. 16th International Symposium on Artificial Life and Robotics*, pp.363–366 (2011).