

GPを用いたバイオマーカ発現量からの 機能性食品の生理活性値推定

上口 真由美[†] 山森 一人^{††} 吉原 郁夫^{††} 永濱清子^{†††}

[†] 宮崎大学院工学研究科 ^{††} 宮崎大学工学部情報システム工学科

^{†††} 財団法人宮崎県産業支援財団

GPを用いたバイオマーカ発現量からの生理活性値推定法について報告する。HepG2細胞に機能性既知の精製された食品成分を作用させてバイオマーカ発現量を測定し、GPによりモデル式を自動生成した。食品抽出物を同細胞に作用させたときのバイオマーカ発現量はこのモデル式に入力し、抗酸化ストレス・抗血管新生活性等の生理活性値の推定を行った。

Estimation of Physiological Activity Values from Protein Expression Levels using GP

Mayumi KAMIGUCHI[†] Kunihito YAMAMORI^{††} Ikuo YOSHIHARA^{††}
Kiyoko NAGAHAMA^{†††}

[†]Graduate Course in Computer Science and Systems Engineering, Graduate School of Engineering,
University of Miyazaki

^{††}Department of Computer Science System Engineering, Faculty of Engineering, University of Miyazaki

^{†††}Miyazaki Prefectural Industrial Foundation

This research reports a new procedure to estimate physiological activity values of foods from protein expression levels using Genetic Programming(GP). Protein expression levels and physiological activities by refined constituents of foods are measured in advance and they are used to make a model by GP. Then physiological activity values of food extracts are estimated by the above model, from their protein expression levels only.

1 はじめに

食品が生命活動に与える影響である三次機能の評価指標として生理活性値がある。しかし、生化学的に生理活性値を測定するには多様な測定器を駆使して複雑な手順を踏む必要がある。

そこで、効能が既知である医薬品や精製された食品成分を細胞に与えて生理活性値と生体指標¹⁾であるバイオマーカ発現量(細胞のタンパク質増減量)を測定し、実際の食品抽出物から得られたバイオマーカ発現量とこれらを照合することで生理活性値を推定する方法が研究されている^{2,3,4)}。しかし従来の研究においては十分な推定精度を得られず、その原因の一つとしてバイオマーカ発現量と生理活性値間の関係に非線形性が含まれて

いるためであると考えられる。

本研究では、推定モデルに非線形性を取り入れるため、遺伝的プログラミング(GP:Genetic Programming)を用いてバイオマーカ発現量から非線形項を含む生理活性値推定モデルの自動構築を行った。このとき、バイオマーカ発現量と生理活性値はそれぞれ独立して複数回測定されているため、これらを単回帰分析により適切に組み合わせる事とした。生理活性値推定実験では、始めにGPを用いて、効能が既知である成分から測定されたバイオマーカ発現量を説明変数に持つ推定モデルを生成した。また、GPによって探索された個体の中で最もデータに対する適合度が高いと思われるモデルに対して実際の食品抽出物により測定されたバイオマーカ発現量を入力として推定を行い、その推

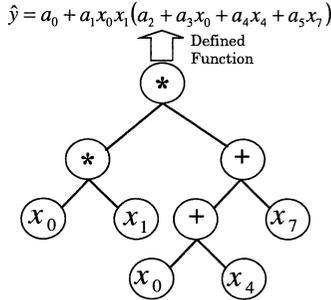


図1 木構造により推定関数 $f(x_1, x_2, \dots, x_n)$ を表現した例

定精度を評価した。

2 モデルの自動生成

2.1 遺伝的プログラミング (GP)

遺伝的アルゴリズム (GA: Genetic Algorithms) は、進化論的な考え方に基づいて解候補である個体を操作することで最適化の問題や学習、推論を扱う手法であり、近年盛んに研究されている。GP⁵⁾ は GA を拡張し、グラフ構造や図1のような木構造といった構造的表現を扱えるようにしたものである。

2.2 GP を用いた生理活性値推定モデルの生成

\hat{y} を n 個のバイオマーカ発現量 x_1, x_2, \dots, x_n から推定される生理活性推定値とすると、その推定モデルは (1) 式のように表せる。

$$\hat{y} = f(x_1, x_2, \dots, x_n). \quad (1)$$

この推定関数 $f(x_1, x_2, \dots, x_n)$ は例えば図1に示すような木構造によって表現することができる。この構造的表現において、説明変数であるバイオマーカ発現量は葉に対応し、演算子はノードに対応している。本研究ではノードに対応する演算子には四則演算を使用する。

木構造によって表わされた関数 $f(x_1, x_2, \dots, x_n)$ を一種の染色体と考え、図2のような処理を行うことで新しい個体を生成する。処理を行う以前の個体群を親世代、行った後の個体群を子世代とすると、世代を重ねるに従い構造的表現の染色体が少しずつ変化し、目的に合わせた関数が探索される。主な遺伝的操作を以下に示す。

交叉 親個体群から2つの個体を選択し、部分木を交換する。

突然変異 個体のノードが表している演算子を他の演算子に書き換える。

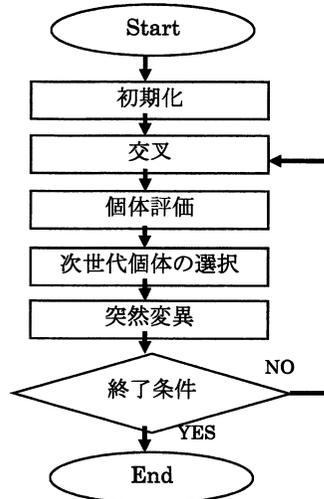


図2 GPの手順

個体評価 個体の評価は適合度 *Fitness* によって行う。 n 個のバイオマーカ発現量と1個の生理活性値をあわせて1個のデータセットとすると、 k 個のデータセットにおける適合度 *Fitness* は以下の式から計算される。ここで、 y_i 及び \hat{y}_i は i 番目のデータセットにおける生理活性値の実測値と推定値をそれぞれ表しており、 C は定数、 m はノード数を表している。

$$Fitness = \frac{1}{E}. \quad (2)$$

$$E = \frac{1}{2} \sum_{i=1}^k (\hat{y}_i - y_i)^2 + C \times m. \quad (3)$$

3 測定データの前処理

推定に用いるデータ群は、化合物を3種類の濃度でそれぞれ HepG2 細胞に与えた場合のバイオマーカ発現量と生理活性値を測定したものである。ここで化合物とは医薬品や精製された効能が既知の食品成分を指す。

3.1 測定対象化合物

推定に使用するバイオマーカ発現量と生理活性値の測定において、細胞に与えた化合物を以下に示す。

- ポリフェノール
 - フラボノイド系
 - * イソフラボン： Genistein, Daizein, Glycitein
 - * フラボノール： Kaempferol, Galangin, Quercetin

- * アントシアニン: Cyanidin, Delphinidin, Pelargonidin
 - * フラバノール: EGCG, EGC
 - * スチルベノイド: Resveratrol
 - クロロゲン酸系: Chlorogenic Acid
 - クルクミン系: Curcumin
 - その他: Rosmarinic Acid
- 脂肪酸
 - 共役リノール酸 (CLA), Arachidonic Acid, Linoleic Acid
 - 高血圧治療薬 (スタチン類)
 - Fluvastatin Na, Atorvastatin Ca, Simvastatin, Lovastatin, Pravastatin
 - 抗ウイルス薬
 - IFN, Ribavirin
 - その他
 - Capsaicin (Alkaloid), Lipoic Acid, GABA, BITC

3.2 測定項目

バイオマーカー: 化合物各 3 濃度に対して 13 種類のバイオマーカー Thioredoxin, Survivin, HSP70, XIAP, FADD, TXNRD1, HSP90, MxA, tNOX, NQO1, ERK2, p53, Bcl2 の発現量がそれぞれ 6 点ずつ測定されている。

生理活性: 本研究では抗酸化ストレス活性と抗血管新生活性、細胞増殖抑制活性、抗炎症活性の計 4 種類の生理活性について推定を行う。各生理活性ではバイオマーカー発現量と同様に、化合物各 3 濃度に対して 3~20 点ずつの生理活性値が測定されている。

3.3 正規化

推定に用いる生理活性値は、各化合物ごと当該化合物を与えなかった時の生理活性値の平均値によって正規化した値を使用する。従って、当該生理活性値が '1' であるということはその化合物が細胞に対し特に作用を与えないことを意味する。

3.4 データセットの作成

バイオマーカー発現量と各生理活性値のデータは、全て独立に測定されている。そのため、推定を行う前にバイオマーカー発現量と生理活性値間で対応付けを行う必要がある。本研究では各バイオマーカー発現量と生理活性値の間で単回帰分析を行い、二つの値がお互いに対して独立でないことへの背反確率 P を求め、 $P \leq 0.05$ かつ最小となる組み合わせに基づいてデータセットを作成した。また、これらの対応付けは 4 つの活性すべてにおいて行った。

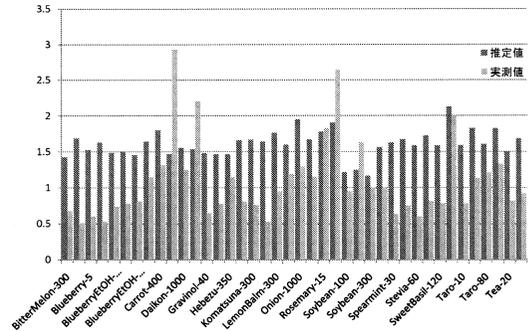


図3 抗酸化ストレス活性推定結果

4 生理活性値の推定実験

GP を用いて実際に抗酸化ストレス活性、抗血管新生活性、細胞増殖抑制活性、抗炎症活性について推定モデルを自動生成し、推定を行った。推定モデルの自動生成及びその各パラメータの決定には化合物による測定値を用いた。また、テストデータとしては実際の食品抽出物から測定されたバイオマーカー発現量を使用した。

4.1 抗酸化ストレス活性推定実験

測定対象化合物から測定された 600 個を用いて GP による推定モデル生成及びパラメータ最適化を行い、食品抽出物から得られたデータ 242 個をテストデータに用いて推定実験を行った結果を図 3 に示す。なお、以下のグラフでは食品抽出物と濃度毎に推定値を平均化して表示している。

抗酸化ストレス活性の推定実験では、図 3 から分かる通り良い推定精度を得られなかった。これは、テストデータに Carrot-400 のようにモデル構築用データに存在しない特出した活性値が存在しているためと考えられる。

4.2 抗血管新生活性推定実験

抗酸化ストレス活性と同様に測定対象化合物から測定された 600 個をモデル構築用データに、食品抽出物から得られたデータ 242 個をテストデータとして用いて推定実験を行った。推定結果を図 4 に示す。

抗血管新生活性推定実験においては、比較的良い推定精度を得た。Komatsuna-100 などについては、推定値と実測値の誤差は大きいですが、双方とも同じ 1 以下のクラスに属しているため問題にはならない。しかし、図 4 から Stevia-200 や Onion-300 のような急激に増大・減少するデータの推定精度が低いことが分かる。

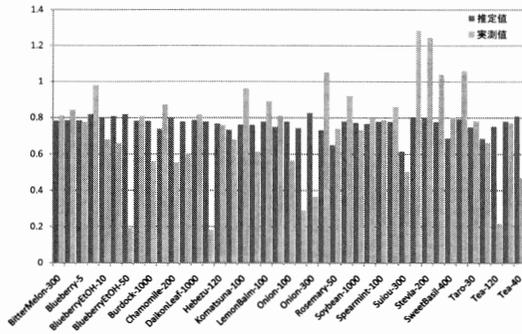


図4 抗血管新生活性推定結果

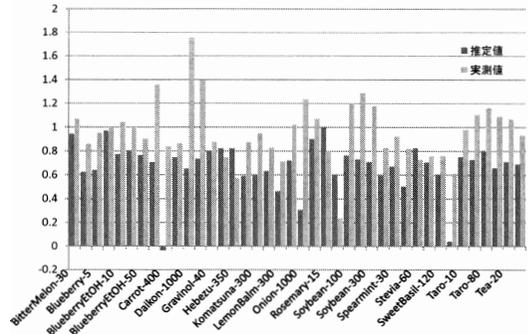


図6 抗炎症活性推定結果

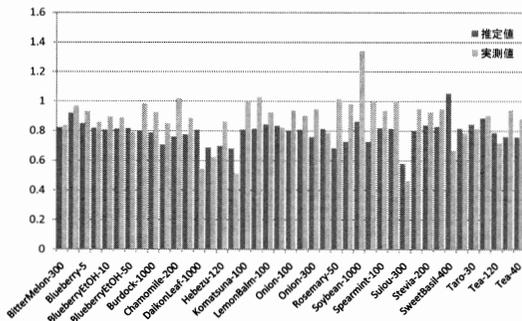


図5 細胞増殖抑制活性推定結果

4.3 細胞増殖抑制活性推定実験

他の活性と同様に、600個のモデル構築用データと242個のテストデータを用いて推定実験を行った。推定結果を図5に示す。

細胞増殖抑制活性推定実験においては、比較的良好な推定精度を得た。また、Hebezu-120などについても抗血管新生活性と同様に、双方とも同じ1以下のクラスに属しているため問題にはならない。しかし、図5からSoybean-1000のような急激に増大するデータの推定精度が低いことが分かる。

4.4 抗炎症活性推定実験

他の活性と同様に、600個のモデル構築用データと242個のテストデータを用いて推定実験を行った。推定結果を図6に示す。

抗炎症活性も、図6から分かる通り良好な推定精度を得られなかった。抗酸化ストレス活性と同様に、テストデータにGravinol-40のようなモデル構築用データに存在しない活性値が存在している。

5 おわりに

本研究では、効能が既知である化合物から測定されたバイオマーカー発現量と生理活性値を用いてGPによる非線形項を含む生理活性値推定モデルの自動構築を行い、得られたモデルにより実際の食

品抽出物から測定されたバイオマーカー発現量を入力としてその生理活性推定実験を行った。その結果、抗血管新生活性、細胞増殖抑制活性については比較的良好な推定結果を得られた。

また、推定実験を鑑みると、推定精度が低い急激に増大・減少するようなデータを表すことのできる新たな演算子の導入と、モデル構築用データにおけるデータ不足への対応が今後の課題として残されている。

謝辞

本研究は、独立行政法人科学技術振興機構・地域結集型共同研究事業「食の機能を中心としたがん予防基盤技術創出」の一部として行なわれ、バイオマーカー発現量や生理活性値は宮崎大学農学部及び宮崎県産業支援財団コア研究室にて測定されたものである。関係者各位に深く感謝する。

参考文献

- [1] 吉川敏一, 大澤俊彦: アンチエイジングと機能性食品-今なぜバイオマーカーか-, シーエムシー出版(2006).
- [2] 山森一人, 佃晋輔, 吉原郁夫: 揺らぎを含むタンパク質発現量からのニューラルネットワークによる食品の生理活性値の推定, *MEMOIRS OF THE FACULTY OF ENGINEERING UNIVERSITY OF MIYAZAKI*, No. 36, pp. 345-350 (2007).
- [3] Togo, S., Yamamori, K., Yoshihara, I. and Nagahama, K.: Estimating physiological activities of functional foods from protein expression levels using bayesian classifier, *The Thirteenth International Symposium on Artificial Life and Robotics 2008*, pp. OS12-3 (2008).
- [4] 上口真由美, 山森一人, 吉原郁夫, 古谷博史, 永濱清子: Estimation of physiological activities from protein expression levels with Multiple Regression Analysis., *APBC2008* (2008).
- [5] 伊庭斉志: 情報学セミナー 遺伝的プログラミング, 東京電機大学出版局(1996).