

ニューラルネットワークによる 生理活性値推定支援システムの開発

吉原郁夫¹⁾川西千尋²⁾山森一人³⁾

Development of Support System to Estimate Physiological Activity Values Using Neural Network

Ikuo YOSHIHARA

Chihiro KAWANISHI

Kunihito YAMAMORI

Abstract

Recently westernized daily life has gradually changed our living circumstances and eating habits. Today most of the people are deeply concerning to functionality of foods, especially to the 3rd functionality. To know that, we had developed a proto-type system to estimate physiology activity values from protein expression levels. However, the proto- system is a little tough to operate for the analysts inexperienced in computers. This paper attempts to extend the proto-system with graphical user interface which is expectedly to let the system easier to use.

Key Words:

neural network, physiological activity, estimation system

1 はじめに

近年、我々を取り巻く環境や食生活が変わり、食と健康に関する意識が高まってきている。食品には3つの機能があり、そのうち高次の生命活動に対する調節機能である第三次機能は、人々の健康への関心の高まりにともなって注目されてきている¹⁾。

本研究は、宮崎県産業支援財団による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」の一環として行われており、この中で我々は生理活性値を推定するシステムの構築の研究を行ってきた。生理活性値とは、食品に含まれる成分が生体に与える影響の大きさを数値で表したものである。食品中の成分を抽出し、細胞に作用させて生理活性値を調べるには、多くの労力が必要となる。ある化合物を細胞に加えると生理活性値とバイオマーカ発現量が変動する。この

二つの値のうち、バイオマーカ発現量は生理活性値よりも容易に測定することができることから、バイオマーカ発現量を測るだけで生理活性値が推定できるシステムの開発が望まれており、そのための方法として階層型ニューラルネットワークを用いる方法を提案し、宮崎大学農学部の共同研究者の方々の試用に提供してきた²⁾。

ところが、そのプログラムはUNIX上で動くものであり、キーボードでコマンドを入力しなければならないため、コンピュータの操作に不慣れな人には使いづらいものであった。

そこで我々は、コンピュータの操作に不慣れな人でも簡単に使えるような、生理活性値推定の対話型ユーザーインターフェースを開発することとした。

2 推定支援システムの構想

2.1 システムの具備機能

使いやすさ向上のため、まず世の中のOSの大半を占めるWindows上で動かせるようにする。

¹⁾情報システム工学科，教授

²⁾情報システム工学科学生

³⁾情報システム工学科，准教授

また、グラフィカルユーザーインターフェース (Graphical User Interface:GUI) を用いて容易に操作が出来るようにする。

操作はマウスオペレーションを基本とし、バイオマーカー発現量の値は決められた箇所に入力する“fill in the blank”方式とする。こうすることにより、プログラムを動かすためのコマンドをわざわざ調べたり覚えたりする手間が必要なくなる。

また、備えるべき機能として、複数の生理活性値を同時に推定できること、学習と推定が一環してできること、プログラムの設定が容易にできること、データをファイルから読み込めること、結果をファイルに書き込めること等がある。

2.2 システムの基本構成

本推定支援システムの構成を、図1に示す。本システムは、推定エンジン部、入力部、出力部、データ部から構成される。

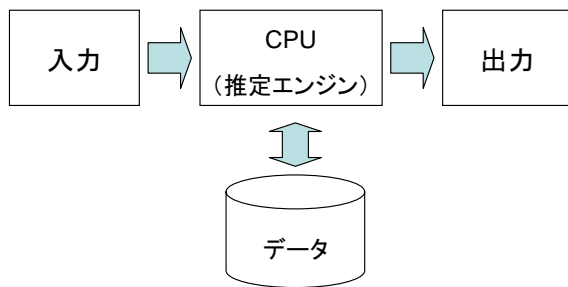


図.1 システム構成

推定エンジンには、階層型ニューラルネットワークを使用する。我々の研究室では、生理活性値の推定にニューラルネットワーク、SOM、ベイズ識別、統計を用いる方法を試みている³⁾⁴⁾⁵⁾。これらの方法の中で、今のところ最も推定の精度が高い方法がニューラルネットワークであることから、推定支援システムの推定方法にはニューラルネットワークを用いている。

ニューラルネットワークとは、生物の脳神経系を模した数学モデルで、学習、連想機能を有する。本研究では、3層の階層型ニューラルネットワークを使用し、その学習には、階層型ニューラルネットワークの代表的学習法である誤差逆伝搬法を使用する⁶⁾。

入力部は、GUIによって、わかりやすく容易に操作が出来るようにする。入力部は、バイオマーカー発現量の値を指定箇所に入力したり、データファイルを選択するだけにする。

出力部は、結果をわかりやすく表示し、ファイルに書き出せるようにする。また、出力は実際の測定値レベルに変換する必要がある。ニューラルネットワークは0~1までの出力しか出さないが、実際の生理活性値は1以上のものもある。そのため、ニューラルネットワークの出力を実際の測定値に合うよう線形変換している。

2.3 データ

本推定支援システムで推定する生理活性値は、抗酸化ストレス活性、細胞増殖抑制活性、抗炎症活性、血管新生活性の4つである。使用するバイオマーカーはThioredoxin、Survivin、HSP70、XIAP、FADD、TXNRD1、HSP90、MxA、tNOX、NQO1、ERK2、p53、Bcl2の13種類で、化合物は30種類(濃度3種類ずつ)である。

また、データ部には実験データや学習済みのニューラルネットワークの重みファイルがある。

3 プロトタイプシステムの作成

3.1 試作システムの構成

プロトタイプは、CPU：Pentium 4、クロック：2.60GHz、主記憶：1GBのハードウェア上で作り、ソフトウェアはOS：Windows XP、Visual Studio 2005を使用した。処理の流れを図2に示す。

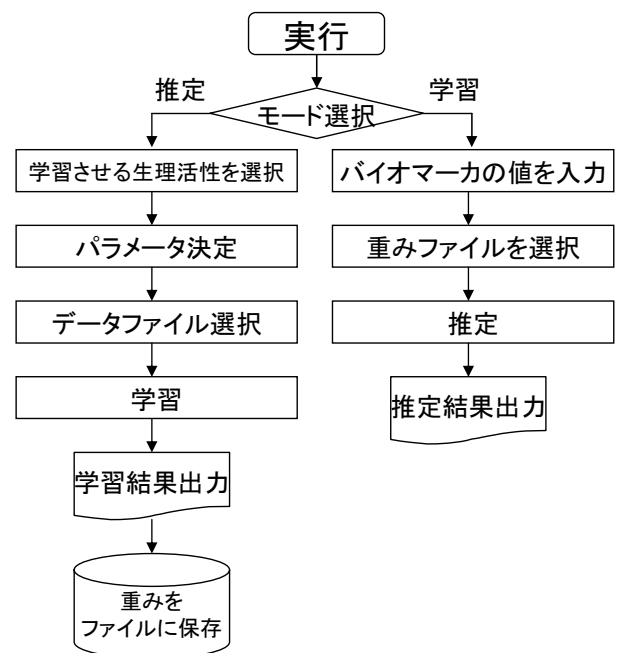


図.2 処理の流れ

表. 1 実験データの例

化合物名 (濃度)	バイオマーカ発現量 (13個)					教師信号
	Thioredoxin	Survivin	Bcl2	生理活性値
BITC (1.5 μ M)	0.759	0.842	1.214	1.380
CLA12C (10 μ M)	0.892	0.798	1.156	0.662
EGCG (20 μ M)	0.992	0.651	0.556	0.679
...

また、推定支援システムの開発言語には、Visual Basic(VB)を使用した⁷⁾⁸⁾。VBはインターフェースの開発が容易であり、GUIアプリケーションを作成することに適している。また、インターフェイス部と計算処理部が同時に書けることから、VBを採用した。

3.2 実験データ

測定された生理活性値とバイオマーカ発現量は、それぞれ測定されたデータの数が異なるため、本実験では少ないほうの数にあわせて実験データを作る。バイオマーカ13個と生理活性値1つの14次元ベクトルが1つの実験データを構成する。表1に実験データの例を示す。

今回、抗酸化ストレス活性、細胞増殖抑制活性、抗炎症活性は生理活性値が10個、バイオマーカ発現量が6個だったため、同じ化合物の同じ濃度に対して6個ずつ実験データを作る。全部で化合物30種類×3濃度×6で540個である。血管新生は生理活性値が3個、バイオマーカ発現量が6個だったため、同じ化合物の同じ濃度に対して3個ずつ実験データを作る。全部で化合物30種類×3濃度×3で270個である。

実験に用いるデータは、まず基準のタンパク質であるGAPDHの発現量で正規化し、各化合物ごとにその化合物を加えなかったときの値で割ったものを使用する。また、生理活性値についても各化合物ごとに濃度0 μ Mの時の生理活性値の平均値によって割り正規化して使用する。

また、生理活性値とバイオマーカ発現量はそれぞれ個々独立に測定されているため、バイオマーカ発現量の測定値と生理活性の測定値を対応付ける必要がある。そこで、本研究では各生理活性値と各バイオマーカ発現量の関係を単回帰分析で

調べ、 p 値を求め、 $p \leq 0.05$ かつ最小の組み合わせを学習用データとして選定した。

3.3 実験

推定支援システムには学習モードと推定モードがある。

学習モードは、推定モードに必要なニューラルネットワークの重みを作るために使う。推定モードは、バイオマーカ発現量から生理活性値を推定するために使う。

学習モードと推定モードは、タブを選択するだけで簡単に切り替えられるようになっている。

また、出力の線形変換は、各生理活性の測定値の最大値と最小値がニューラルネットワークの出力範囲(0,1)に収まるように、線形変換の係数を定める。各生理活性値の係数を表2に示す。

表. 2 各生理活性値の線形変換の係数

生理活性値	係数
抗酸化ストレス活性	$a = 5.9, b = 0.1$
細胞増殖抑制活性	$a = 1.75, b = -0.075$
抗炎症活性	$a = 1.95, b = 0.05$
血管新生活性	$a = 2.79, b = 0.01$

学習は、実験データを学習用データと検証用データに分けて行う。抗酸化ストレス活性、細胞増殖抑制活性、抗炎症活性のデータは、540個中450個を学習用に、90個を検証用に使用する。血管新生のデータは、270個中225個を学習用に、45個を検証用に使用する。

以下に、学習モードの実行の仕方を説明する。

実験を行う際の各パラメータの設定値を表3に示す。

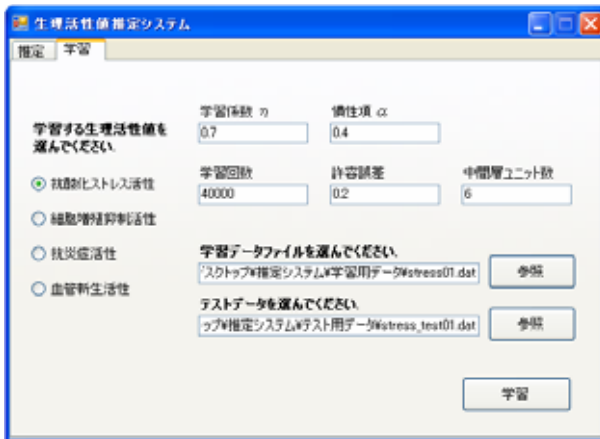


図.3 学習モードの入力画面

図3は学習モードの入力画面である。学習させたい生理活性値(今回は抗酸化ストレス活性)を選び、パラメータの設定を変更する場合は値を書き換え、学習用データと検証用データを選択し、学習ボタンを押す。

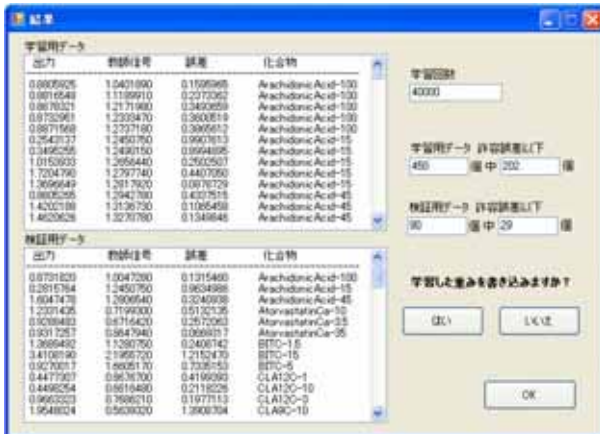


図.4 学習結果の表示画面

図4は学習結果の表示画面であり、学習が終了すると表示される。ここで得られた重みを使う場合は「はい」を押し、書き込み先とファイル名を指定してファイルに保存する。この保存した重みファイルを、生理活性値の推定に使用する。

次に、検証実験について報告する。ニューラルネットワークは毎回学習結果にばらつきがあるため、検証の精度は何回か実験をして平均で評価する。今回、学習用データと検証用データを6つずつ作ったため、1つにつき5回ずつ、各生理活性値ごとに計30回ずつ実験を行い、平均を調べた。

表.3 各パラメータの設定

学習係数	0.7
慣性項	0.4
最大学習回数	40000
許容誤差	0.2
中間層ユニット数	6
初期重み	(-2.0,2.0)の乱数

図5は、検証の成功率である。検証用データに対する推定結果が許容誤差以下になった時、検証は成功したとする。

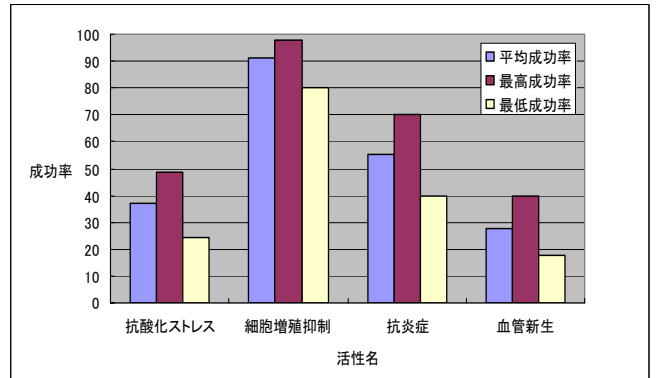


図.5 検証の成功率

実験の結果、推定精度は活性の種類によってかなり差があることがわかった。図5からわかるように、細胞増殖抑制活性の精度が一番高く、血管新生活活性の精度が一番低かった。

以下に、推定モードの実行の仕方を説明する。



図.6 推定モードの入力画面

図6は推定モードの入力画面である。バイオマーカー発現量の値を入力し、使用する重みファイルを選択した後、推定ボタンを押す。

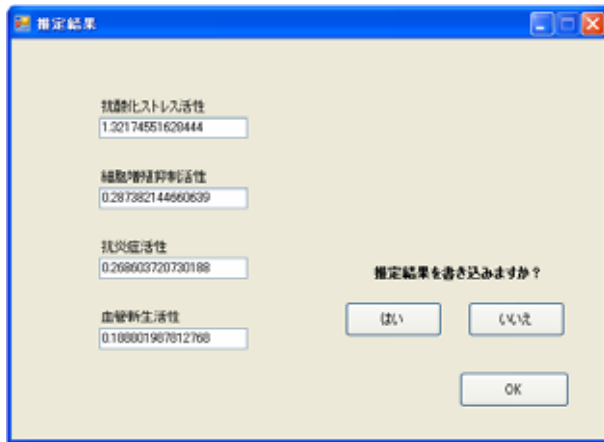


図. 7 推定結果の表示画面

図7は推定結果の表示画面である。推定結果はファイルにテキスト形式で書き込めるようになっている。結果をファイルに書き込むときは「はい」を押し、書き込み先とファイル名を指定して保存する。

3.4 システムの試用評価

この推定支援システムを共同研究者の農学部の方たちに実際に使ってもらい、直感的に操作できたか、インターフェースは使いやすかったか等の質問に5段階評価(5:とても良い、4:良い、3:普通、2:悪い、1:とても悪い)で答えてもらったところ、全体的に良い評価が得られた。その評価を以下の図に示す。

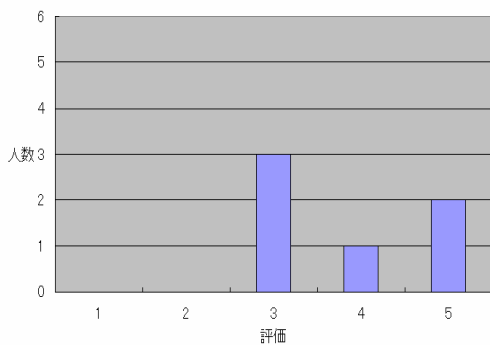


図. 8 学習モードは直感的に操作できたか

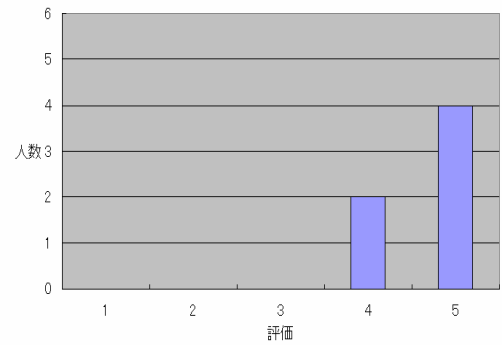


図. 9 推定モードは直感的に操作できたか

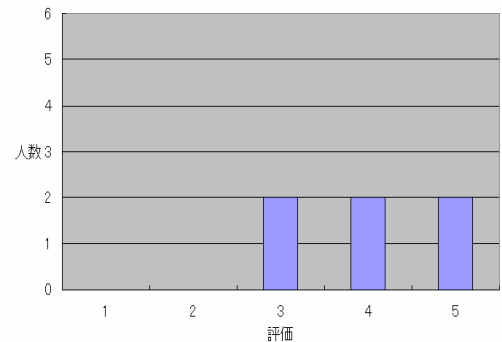


図. 10 学習モードのインターフェースは使いやすかったか

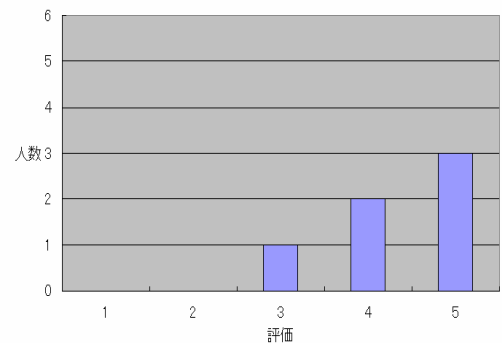


図. 11 推定モードのインターフェースは使いやすかったか

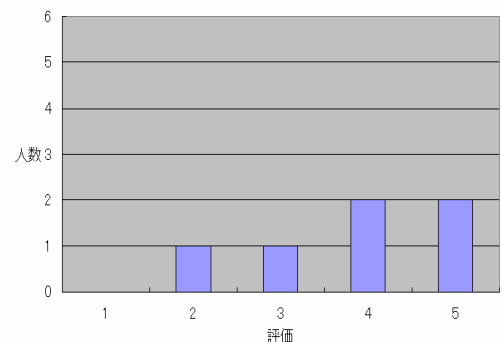


図. 12 学習結果の表示はわかりやすかったか

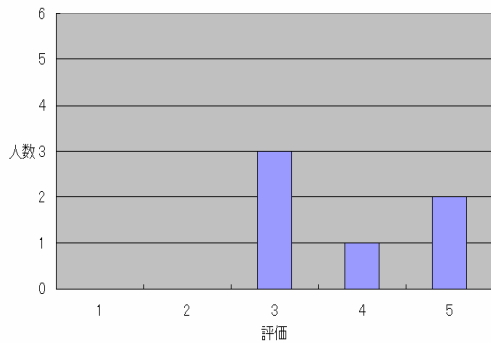


図. 13 推定結果の表示はわかりやすかったか

また、自由記述では以下のような意見があった。

- 選択したり入力したりするだけなので、Windowsしか使ったことの無い人にとっては簡単に操作しやすい
- 簡単で非常にわかりやすい
- バイオマーカの数を変更できるようにしてほしい

4 おわりに

当初生理活性値の推定用に我々が提供していたプログラムは、コマンドを入力しなければならぬもので、コンピュータの操作に不慣れな人には使いづらいものであった。本研究ではこの点を改善すべく、対話型のユーザーインターフェースを備えた生理活性値推定支援システムを試作し、評価した。

この推定支援システムを共同研究者の宮崎大学農学部の方々に実際に使ってもらい、評価してもらったところ、全体的に良い評価を得られた。その結果、プログラムを動かすためのコマンドをわざわざ調べて覚えたりしなくてすむようになり、被験者全員がすぐに使えるようになった。

今後の課題として、農学部から要望があったバイオマーカの数を変更できるようにすることが挙げられる。他にも、推定の出力を図でわかりやすく表示させることやニューラルネットワーク以外の推定方法が使えるようにすること等が挙げられる。また、使いやすさとは別に、知的財産保護の観点から重み係数の値を暗号化してストアしておくべきである。

また、余りにも大きすぎる生理活性値があると精度が悪くなるため、それらへの対応策を作る必要がある。

謝辞

本研究の一部は、財団法人宮崎県産業支援財団・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」の研究として行なわれ、バイオマーカ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から提供を受けた。関係各位に感謝する。

参考文献

- [1] 須見洋行. 食品機能学への招待 - 機能性食品とその効能 - . 三共出版, 1995.
- [2] Patrik Henry Winston. *ARTIFICIAL INTELLIGENCE Third Edition*, pp. 443-469. ADDISON WESLEY, 1992.
- [3] 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).
- [4] I. Yoshihara, T. Fukushima, and K. Yamamori. Estimating Physiology Activity of Functional Foods with Self-Organizing Map. *Memoirs of the Faculty of Engineering, University of Miyazaki*, No. 35, pp. 243-246, (2006).
- [5] I. Yoshihara, S. Tougou, and K. Yamamori. A Physiological Activity Value Prediction of Functional Foods using Bayse Classifier. *Memoirs of the Faculty of Engineering, University of Miyazaki*, No. 35, pp. 247-250, (2006).
- [6] 吉富康成. ニューラルネットワーク. 朝倉書店, 2002.
- [7] 林晴比古. 明快入門 Visual Basic 2005 ビギナー編 . SoftBank Creative, 2006.
- [8] ミューテック. Visual Basic テクニックブック. C&R 研究所, 2007.