

## リカレントニューラルネットワークによる細胞内反応経路の脆弱・頑健部位の推定手法

北風 裕教<sup>†</sup>      神田 全啓<sup>†</sup>      中司 弘樹<sup>††</sup>      池田 信彦<sup>†††</sup>  
 松野 浩嗣<sup>††††</sup>      宮野 悟<sup>†††††</sup>

Prediction of Fragile Points for Robustness Checking of Cell Systems  
 with Recurrent Neural Networks

Hironori KITAKAZE<sup>†</sup>, Masahiro KANDA<sup>†</sup>, Hiroki NAKATSUKA<sup>††</sup>,  
 Nobuhiko IKEDA<sup>†††</sup>, Hiroshi MATSUNO<sup>††††</sup>, and Satoru MIYANO<sup>†††††</sup>

あらまし 細胞システムは外部環境の変化に対して頑健な構造をもち、内部異常に対しては分子機構が互いに補償することで恒常性を保っている。しかし何かかぎとなる要因により、状態が崩壊してしまう脆弱性も同時に持ち合わせており、バイオシミュレーションツールを用いた脆弱部位の推定手法が求められている。ハイブリッド関数ベトリネット (HFPN) は、多次元の微分方程式の解法を必要とせず、モデルの改変を容易に実現できるが、ノックアウトの実行を繰り返す必要があり多くの処理時間を要する問題が指摘されている。本論文では、HFPN で表現した細胞内反応経路を維持する形でリカレントニューラルネットワーク (RNN) に再形成し、自動で連続的にノックアウトを実現することにより、処理時間の問題を改善する手法を提案する。また、HFPN の速度情報を RNN の結合荷重として与えることで、脆弱部位・頑健部位の推定率の向上も試みる。システムの有効性の評価を上皮成長因子のシグナル伝達モデルを用いて検証し、評価結果に対する考察を述べる。

キーワード ハイブリッド関数ベトリネット, 生命パスウェイ, 脆弱部位, リカレントニューラルネットワーク

### 1. ま え が き

細胞システムは、外部環境の変化に対して頑健な構

造をもち、内部異常に対しては分子機構が互いに補償することで恒常性を保つメカニズムを有している。しかし、様々な要因によりかぎとなる分子やシステムに異常が生じた場合、途端に生体の恒常性が崩れるといった脆弱性も同時に持ち合わせている [1], [2]。ここで、引き金の要因と考えられる部位 (脆弱部位) と恒常性が強い部位 (頑健部位) を予測し判定できれば、細胞システムの脆弱性を補い頑健な構造へと改善する適切な薬剤の開発などができるようになる。また、事前に細胞システムについて研究対象の焦点を絞り込むことができることから、しらみつぶしに行っていた生物実験による膨大な時間や費用そして実験用動物の犠牲を削減でき、生物学の発展に貢献できる。

近年、細胞システムにおける代謝系、遺伝子制御ネットワーク、シグナル伝達のカスケードなどの生体内におけるイベント (パスウェイ情報) に対してバイオシミュレーションツールを用いて分析を行う研究が進められている。バイオシミュレーションツールを用いてシステムの挙動を求める手法に、生命現象の伝達

<sup>†</sup> 大島商船高等専門学校, 山口県  
 Shipping Technology Department, Oshima National  
 College of Maritime Technology, 1091-1 Komatsu,  
 SuouOshima-cho, Oshima-gun, Yamaguchi-ken, 742-2193  
 Japan

<sup>††</sup> 九州工業大学大学院生命体工学研究科, 北九州市  
 Graduate School of Life Science and Systems Engi-  
 neering, Kyushu Institute of Technology, 2-4 Hibikino,  
 Wakamatsu-ku, Kitakyushu-shi, 808-0196 Japan

<sup>†††</sup> 徳山工業高等専門学校情報電子工学科, 周南市  
 Dept. of Computer Science & Electronic Engineering,  
 Tokuyama College of Technology, Gakuen-dai, Shunan-shi,  
 745-8585 Japan

<sup>††††</sup> 山口大学大学院理工学研究科, 山口市  
 Graduate School of Science and Engineering, Yamaguchi  
 University, 1677-1 Yoshida, Yamaguchi-shi, 753-8511  
 Japan

<sup>†††††</sup> 東京大学医科学研究所ヒトゲノム解析センター, 東京都  
 Human Genome Center, Institute of Medical Science, Uni-  
 versity of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo,  
 108-8639 Japan

経路（生命パスウェイ）を多次元の微分方程式を用いて表現する手法がある [3], [4]。この手法は、経路情報や分子作用を変更する際のモデルの改変が容易ではなく、そのつど微分方程式を立て直しその解を求める必要があるため、分析において膨大な時間を要する。また、生命パスウェイが複雑となれば連立方程式の数も増加する問題がある。更に、ある局面の頑健性を向上させれば他の部分において脆弱性が増すトレードオフがあり [5]、微分方程式をそのつど変更して確認することは困難である。これらの要因から、生命パスウェイ全体の脆弱部位と頑健部位を自動で推定する手法は現在までに確立されておらず、生命パスウェイの一部を対象にし、どの経路がターゲットの濃度変化に強く影響を及ぼすかについて感度解析などの結果報告がされているにすぎない [6]。

松野ら [7] は、生命パスウェイのモデル化手法としてペトリネット [8] 理論を基本手法として使い、ハイブリッドペトリネット [9]、関数ペトリネット [10] 及びハイブリッドオブジェクトネット [11] の機能を包含するように拡張したハイブリッド関数ペトリネット (Hybrid Functional Petri Net, 以下 HFPN と略記) を提案した。この方法は、連続的なイベント・離散的なイベントに対して関数を適用できるため多種の反応系を柔軟にモデル化でき、その分子作用の相互関係を数式に変換することなく相互関係を維持したままモデル化できる。また、HFPN を基本アーキテクチャとして開発したソフトウェア Cell Illustrator [12], [13] を用いれば、作成したモデルのシミュレーションを行うことができ、これまでに多くの生命モデルを実現している [14], [15]。

HFPN を用いて脆弱部位の推定を行うためには、生物実験で反応経路の停止（ノックアウト）を行う作業と同様の作業が必要であり、生命パスウェイモデル上で経路の連結を削除してシミュレーション分析を行う必要がある。すべての経路において繰り返し操作を行うことで、ネットワーク全体の脆弱性の度合を確認することが可能となる。しかし、この一連の作業に要するシミュレーション時間と、脆弱部位の有無の確認作業に要する時間は生命パスウェイが拡充されるに伴い増加するため、逐次実行を繰り返す必要のない脆弱部位の自動推定手法が求められる。

そこで本研究は、多くの処理時間を要する HFPN から、システム形成後にはそれほどの処理時間を要さないニューラルネットワークを用いて脆弱部位を推定

するシステムの開発を試みる。ニューラルネットワークには時間記憶が可能なりカレントニューラルネットワーク (Recurrent Neural Networks, 以下 RNN と略記) [16] を用いる。HFPN から RNN へネットワークの再形成を行い、通時的逆伝搬 (Back Propagation Through Time, 以下 BPTT と略記) 法 [16] の学習により自動で連続的に脆弱部位を推定するシステムを提案する。この手法はネットワーク移行のために RNN の学習時間を要するが、連続的に経路のノックアウトが行えるため推定結果を得るまでの時間を改善できる。

一方で、次のような問題もある。提案手法は処理時間の短縮が可能な反面、RNN 学習の非線形性により HFPN の完全な写像を形成することができないことから、推定率の低下につながる問題が残る。また、HFPN においてプレース間のトークン量の変化は速度として表現されるが、提案手法では入力ユニットへのトークン量の変化のみで速度情報を定義したことが推定率の低下につながる。この問題を解決するため、HFPN の速度情報を RNN の結合荷重として反映させる手法を提案し、推定率の向上を試みる。

最後に生命パスウェイモデルによるシミュレーション実験を行い、システムの有効性を検討する。このモデルには、生物実験により内部構成の研究が比較的進んでいる上皮成長因子のシグナル伝達モデル [17] を用いる。上皮成長因子のシグナル伝達モデルの経路は、生命現象を表す基本的な構造の“結合・分裂・崩壊・翻訳・フィードバックループ”などをもつ構造をしているため、提案手法の脆弱部位推定率を評価するモデルに適している。これにより、提案手法が実システムとして有効なシステムであるかを考察する。

## 2. HFPN の基本構成

ペトリネット (Petri Net, 以下 PN と略記) はグラフ表現の一種であり、大別してプレース、トランジション、アーク、トークンの四つの要素から構成される。物質の個数変化をトークン値として離散値で表現できる離散 PN と、物質の濃度変化をトークン値として連続値で表現できる連続 PN を同時に取り扱うハイブリッド PN (HPN) を利用すれば、細胞システムの多くの現象変化を、経路（生命パスウェイ）を用いてモデル化でき、シミュレーションを実行できる。HPN の基本構成要素を図 1 に示す。

HPN による生命パスウェイを用いたモデル化では、蛋白質や mRNA などの遺伝子産物の濃度は連続プ

|                         | 離散要素<br>(discrete elements) | 連続要素<br>(continuous elements)       |
|-------------------------|-----------------------------|-------------------------------------|
| プレース<br>(place)         | ○                           | ◎                                   |
| トランジション<br>(transition) | ■                           | □                                   |
| トークン<br>(token)         | 整数値<br>(integer)            | 実数値<br>(real number)                |
| アーク<br>(arc)            | ノーマル<br>(normal)            | 促進<br>(test)      抑制<br>(inhibitor) |

図 1 HPN の基本構成要素  
Fig.1 Basic component of HPN.

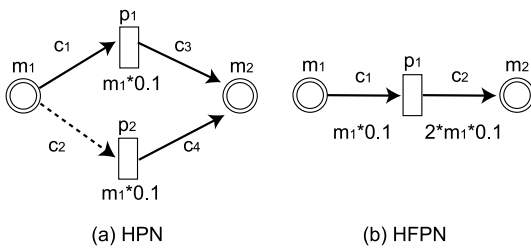


図 2 1 分子 2 量体から 2 分子 1 量体への分解 (a) HPN, (b) HFPN  
Fig.2 Decomposition from dimer to monomer. (a) HPN, (b) HFPN.

レースを用いて表現し、細胞の形状、状態、条件は離散プレースを用いる [7]。また、リン酸化、結合、崩壊などの現象には連続トランジションを用い、その反応速度を連続トランジションの発火速度で表す。一方シグナルや外部刺激、形態変化のタイミングや条件の判断に要する時間は離散トランジションを用いて表現し、各信号が影響を与えるまでの時間を離散トランジションの遅延時間に対応させる。反応の方向や状態の移行、活性化、抑制化などの影響の種類についてはアークを用い、プレース・トランジション間を連結する。これらの組合せにより、HPN を用いれば生物事象を模式図に沿った生命表記で実現できる。

しかし、HPN は連続トランジションの発火速度から消費されるトークン数と出力されるトークン数の比が一定であるため、速度変化のある自然界の生命パスウェイ表記においてモデル化が複雑となる。例えば、2 分子 1 量体が 1 分子 2 量体へ結合する事象や、1 分子 2 量体が 2 分子 1 量体へ分解する事象である。図 2 に示す分解の事象を考えた場合、HPN の記述法では、二つのトランジションを準備し、一方はノーマルアーク

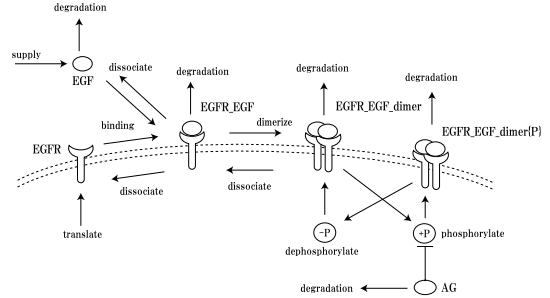


図 3 上皮成長因子におけるシグナル伝達機構の模式図  
Fig.3 Schematic diagram of the structure of signal transmission in epidermal growth factor.

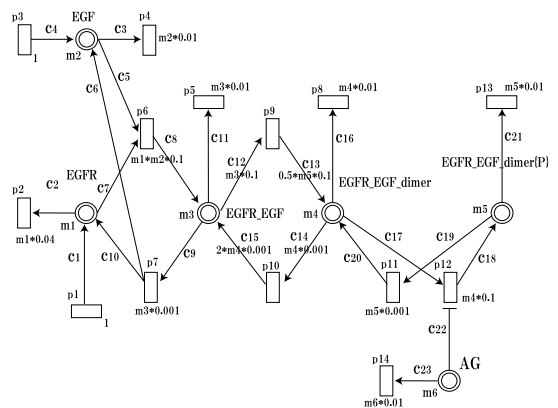


図 4 HFPN 表記を用いた上皮成長因子のシグナル伝達機構  
Fig.4 The mechanism of signal transmission in epidermal growth factor using HFPN.

クでプレースからトランジション間を連結し、もう一方はテストアークを用いて連結を行う必要がある (図 2 (a))。ノーマルアークは接続先のトランジションが実行可能になれば、接続元のプレースからトークンの移動が生じるが、テストアークは接続先のトランジションが実行可能になっても、接続元のプレースにトークンの変化が生じない。そのため、目的とする分解の事象を実現できるが、直観的なイメージからかけ離れた生命パスウェイの表現となる。そこで、トランジションの発火速度をアークごとに定義付けできるように HPN を拡張したハイブリッド関数ペトリネット (HFPN) を生命パスウェイの表記に用いる。HFPN を用いて 1 分子 2 量体から 2 分子 1 量体への分解を表した結果を図 2 (b) に示す。

図 3 と図 4 はそれぞれ文献 [17] に掲載されている上皮成長因子におけるシグナル伝達系の模式図と、そ

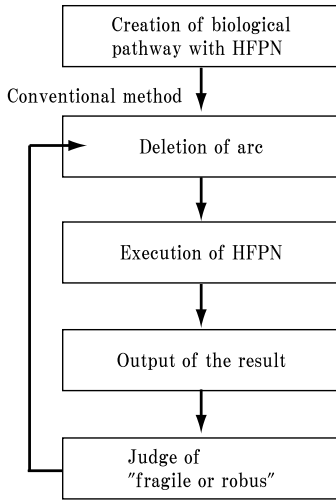


図 5 脆弱部位推定法の従来手法の流れ図  
Fig. 5 Flowchart of the conventional technique.

れを HFPN を用いて生命パスウェイを表記したものである。このように、模式図に対して HFPN 表記は非常に似た状態で描くことが可能であることからパスウェイ作成が容易な表記法といえる。

### 3. 脆弱部位・頑健部位推定の従来手法

HFPN を用いて脆弱部位・頑健部位を推定する従来手法 [17], [18] について図 5 にフローチャートを示す。従来手法では、基本的にユーザがシミュレーションツールを利用して HFPN 表記により生命パスウェイを作成した後、ノックアウトを手動で連続的に実行し、脆弱部位か頑健部位かの判断をする。ここでノックアウトは、ノックアウトの対象とする部位に連結される経路(アーク)を削除することで実現できる。図 4 の HFPN 表記を用いた上皮成長因子におけるシグナル伝達機構において、ノックアウト前の EGF, EGFR, EGFR\_EGF, EGFR\_EGF\_dimer, EGFR\_EGF\_dimer{P} のトークン値の時間変化を図 6 に示す。ここでノックアウトにより重要なアークを削除した場合、以降の連結されたプレースの値(トークン値)は発散またはある値へ収束するが、これは破綻状態を意味する。このとき、削除したアークを脆弱部位とする。一方、重要ではない部位をノックアウトした場合は、ノックアウトをしていない場合と比べてもトークン値にほとんど変化が見られない。このとき、削除したアークを頑健部位とする。連続トランジション  $p_1$  と EGFR 間のアーク  $c_1$  をノックアウトした際のそれらの変化を図 7 に示す。

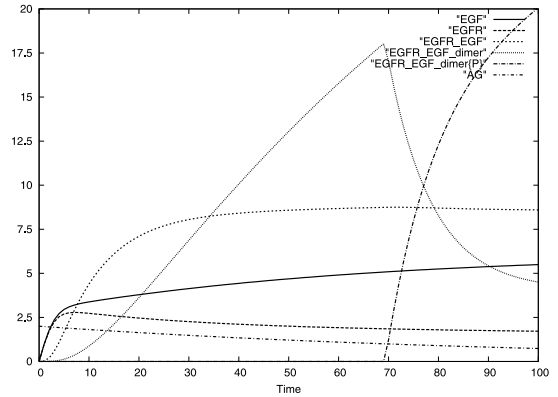


図 6 ノックアウト前の結果  
Fig. 6 The result before knockout.

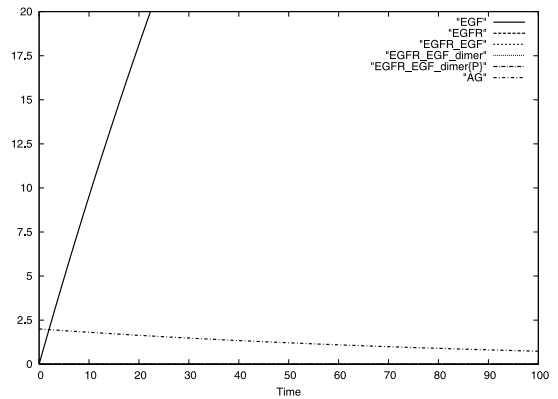


図 7 ノックアウト後の結果  
Fig. 7 The result after knockout.

図 6 のノックアウト前の結果に対して図 7 のノックアウト後の結果は、AG はノックアウトの影響をほとんど受けていないが、EGF は発散し、EGFR\_EGF, EGFR\_EGF\_dimer, EGFR\_EGF\_dimer{P} は 0 へとトークン値が収束しているため、脆弱部位である重要な経路を削除したことになる。

この作業をすべてのアークにおいて繰り返し行うことにより、生命パスウェイ全体における頑健部位と脆弱部位を確認することができる。しかし、この推定法によるユーザの作業はアークの数だけ必要となるため、推定にばく大な時間を要する。

### 4. 脆弱部位・頑健部位推定の提案手法

図 8 に脆弱部位・頑健部位推定の提案手法を示す。提案手法は HFPN によりモデル化した生命パスウェイから 1st stage: HFPN から RNN への再形成, 2nd

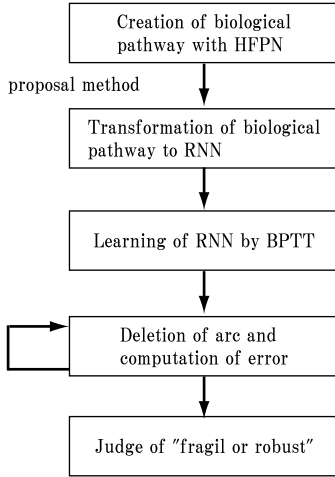


図 8 脆弱部位推定法の提案手法の流れ図  
Fig. 8 Flowchart of the proposed technique.

stage : RNN の学習, 3rd stage : 連続ノックアウトによる誤差の推定, そして 4th stage : 脆弱部位の断定の 4 段階の工程を, すべて短時間で自動的に行うことを可能にしている. 提案手法について順に詳しく説明する.

4.1 1st stage : HFPN から RNN への再形成  
HFPN によりモデル化した生命パスウェイを RNN に再形成する際, ネットワークの連結情報を保持する必要がある.

HFPN のプレースの集合を  $M = \{m_i : i = 1, \dots, K\}$ , トランジションの集合を  $P = \{p_i : i = 1, \dots, L\}$ , プレースから得られたトークンの集合を  $T = \{token_i : i = 1, \dots, K\}$  とする. ここで,  $j$  から  $i$  への結合を  $(i, j)$  と表す. これらの一部はアークを形成する. これにより, アークの集合を  $C = \{(i_k, j_k) \in (M \cup P) \times (M \cup P) | k = 1, 2, \dots, N\}$  とする. ここで, HFPN から RNN への再形成の定義を次のように与える.

[定義 1] HFPN で抽出したプレースから得られる時系列のトークンのデータは, RNN の入力ユニットへの入力信号とする.

$$x_i(t), \quad i \in T \quad (1)$$

[定義 2] HFPN のプレースとトランジションを RNN の隠れユニットとする.

$$y_i(t), \quad i \in M \cup P \quad (2)$$

ただし,

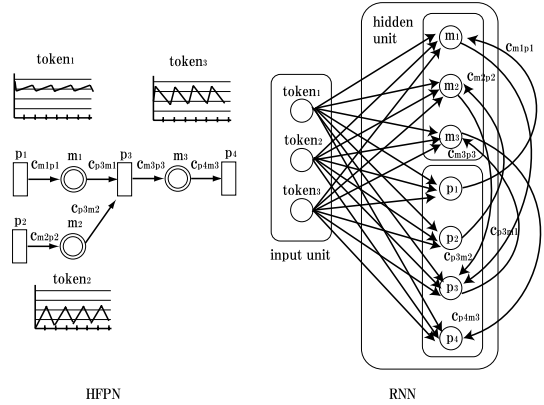


図 9 1st stage : HFPN から RNN への再形成  
Fig. 9 1st stage : re-formation from HFPN to RNN.

$$y_i(0) = \begin{cases} 0.5, & M \cup P \text{ を順に並べて奇数番目} \\ -0.5, & M \cup P \text{ を順に並べて偶数番目} \end{cases} \quad (3)$$

隠れユニットと出力ユニットは区別を行わない. ここで, 出力信号  $z_i(t)$  とすると,

$$z_i(t) = \begin{cases} x_i(t), & i \in T \\ y_i(t), & i \in M \cup P \end{cases} \quad (4)$$

となる. また RNN のユニット数は, HFPN のプレース数とトランジション数によって増減する.

[定義 3] HFPN のプレースからトランジションまたはトランジションからプレースへ結合するアークを, RNN の出力ユニット間 (隠れユニット) のネットワーク結合とする.

$$c_{ij} = \begin{cases} 1, & (i, j) \in C \\ 1, & j \in T \\ 0, & \text{その他} \end{cases} \quad (5)$$

この再形成法により, 図 9 左の HFPN は, 図 9 右の RNN の形で表現される.

#### 4.2 2nd stage : RNN の学習

4.1 では, HFPN の連結情報を RNN の連結情報として再定義した. しかし, この段階では HFPN におけるトークン値の時系列データを RNN には形成できてはいない. 4.2 では, トークン値情報を RNN に導入すると同時に脆弱部位と頑健部位を自動推定できるシステムを構築する. 具体的には, 4.3 で示すノックアウトを行った際, RNN のネットワーク全体

が非常に頑健でノックアウトの影響を受けない場合は RNN の各出力ユニットは“0”，逆にノックアウトで RNN のネットワークが影響を受けてしまう場合は脆弱とし，“0.3”，“0.5”，“0.8”，“1.0”など (0,1) 区間の数値によりその割合を出力するように学習する．ここで，ノックアウトによる破綻状態の例として，過剰反応による発散現象・振幅増幅，反応停止による収束・減衰，位相変化，生命パスウェイの一部分のみの破綻や全体破綻など様々なパターンが存在する．しかし，これらのすべての時系列パターンを RNN に再形成することは非常に複雑であり，学習時間の増大にもつながる．そこで HFPN を用いた生命モデルをノックアウトして破綻させた場合，図 7 の上皮成長因子モデルの例のようにトークン値は“0”へ収束する場面が多いという事実から，学習は正常な場合と入力信号をすべて“0”とした破綻状態の極端な場合について行う．脆弱か頑健かの判断は，学習後の RNN にノックアウトによるトークン値の時系列パターンを入力し，それに対する出力の誤差量を求めることによって行う．誤差量の算出は正常な場合を基準としても，破綻状態の極端な場合を基準としてもよいが，ここでは正常な場合を基準として用いる．RNN の学習の詳細を以下に示す．RNN の学習には BPTT 法を用いる．

BPTT で学習するために，RNN を時間展開したものを図 10 に示す．出力ユニット間の連結は時刻が変化した場合も基本となる RNN の連結には変化はないため同様の連結が続く．RNN における階層数は HFPN のトークン値の時系列データ数と一致する．したがって，HFPN においてトークン値のサンプリング周波数が高い場合，再形成において RNN の階層数が増加する．それに伴い，BPTT 学習に要する計算オーダも増加するため，HFPN のサンプリング間隔を広げるか，若しくはサンプリング後のデータを間引くことにより RNN の学習オーダを減少させることが可能となる．また，結合荷重  $w_{ij}$  の初期設定を乱数とする．

各ユニットの入出力関係は次式で与えられる．

$$\begin{aligned}
 y_i(t+1) &= f_i(s_i(t+1)) \quad i \in M \cup P \\
 s_i(t+1) &= \sum_{j \in T} c_{ij} w_{ij} x_j(t) + \sum_{j \in M \cup P} c_{ij} w_{ij} y_j(t) \\
 &= \sum_{j \in T \cup M \cup P} c_{ij} w_{ij} z_j(t) \quad (6)
 \end{aligned}$$

ここで， $w_{ij}$  はユニット  $j$  からユニット  $i$  への結合荷重， $s_i(t+1)$  はユニット  $i$  の内部状態を表し，入出

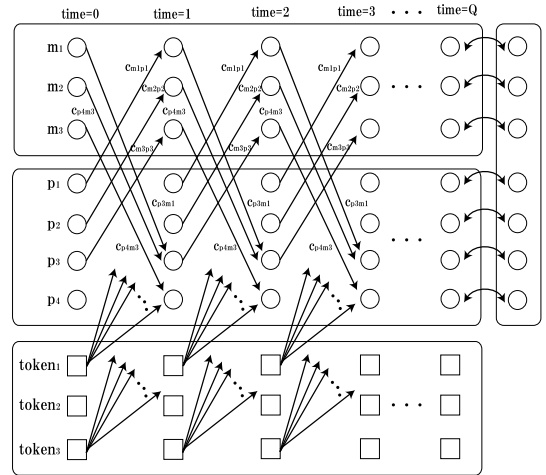


図 10 2nd stage : 時間展開した RNN を BPTT 法により学習

Fig.10 2nd stage : The study of RNN, expressed with time course by BPTT method.

力関数  $f_i$  は微分可能な単調増加関数のシグモイド関数である．

$$f_i(s) = \frac{1}{1 + e^{-s}} \quad (7)$$

RNN [16] はその構造上，入力ユニットはすべての隠れユニットと結合した形になっており，HFPN の構造と異なる．しかし，HFPN の連結情報が RNN における  $c_{ij}$  により表現されていることから，実際の学習においては HFPN において連結がないプレースやトランジションへの入力（トークン値）の影響はない．すなわち，RNN は実質的に HFPN と同じ連結状態を実現する．

入力信号と教師信号については学習に用いるパターンを 2 種類とし，次式のように定義する．

[パターン 1] HFPN の入出力関係を RNN でも同様の関係が維持できるように行うための標準学習用の入力信号  $z_i$  と教師信号  $d_i$  である．入力信号はプレース内の濃度変化であるトークンの内容で，それに対応する教師信号は 0 とする．

$$z_i = \begin{cases} x_i, & i \in T \\ y_i, & i \in M \cup P \end{cases} \quad (8)$$

$$d_i = 0, \quad i \in M \cup P \quad (9)$$

[パターン 2] HFPN 生命パスウェイが破綻したと想定してすべての反応が停止した状態を記憶させるため

の学習用の入力信号と教師信号である．入力信号は破綻した際の状態としてすべて 0 で，教師信号は 1 とする．

$$z_i = \begin{cases} 0, & i \in T \\ y_i, & i \in M \cup P \end{cases} \quad (10)$$

$$d_i = 1, \quad i \in M \cup P \quad (11)$$

このように定義した 2 種類のパターンを学習することで，RNN のネットワークにおいて異常がなければ，RNN の出力信号は 0 に近い値が出力され，RNN のネットワークにおいて異常が生じた場合には，RNN の出力信号は 1 に近い値が出力されるようになる．これにより，ネットワークが頑健であるか脆弱であるかを判断できるようになる．

また，生命パスウェイが複雑化すればするほど，連結するアークの数が増加するため安定状態からの誤差量の変動を大きく観測することが可能となり，後述する脆弱部位推定精度を向上させることが可能となる．HFPN でアークを削除しロックアウトを行う処理は，ある  $(i, j) \in C$  に対し，ネットワーク結合  $c_{ij} = 0$  にすることである．この処理を学習後の RNN において行い，ロックアウト処理を施していない RNN の出力信号とロックアウト処理を施した RNN の出力信号の差から，生命パスウェイにおけるロックアウトした経路の重要性を判定する．

図 9 左の HFPN のパスウェイにおいて，トランジション  $p_3$  からプレース  $m_3$  へのアーク  $c_{m_3p_3}$  をロックアウトにより削除した図 11 左の場合を考える．これは，展開した RNN では図 11 右のように対応した RNN において隠れユニット  $p_3$  から隠れユニット  $m_3$  への連結  $c_{m_3p_3}$  の削除に相当する．HFPN におけるロックアウトは再形成した RNN では再帰的構造に時間展開した各時刻における連結の削除に相当する．展開した時間分だけユニット間の連結が同時に削除される．したがって，BPTT 法によって時間展開した RNN は，図 10 において， $time = 0$  から  $time = Q$  までのすべての  $p_3$  から  $m_3$  への連結  $c_{m_3p_3}$  が削除されたネットワークとなる．このロックアウトは HFPN の  $c_{m_3p_3}$  のアークを例に挙げた場合であるが，この処理をすべてのアークにおいて行えば生命パスウェイ全体のロックアウトを実現できる．今，ロックアウト処理を施した RNN の出力信号を  $\tilde{z}_i(Q)$  とする．そのとき，重要性を判断するための誤差量を次式で定める．

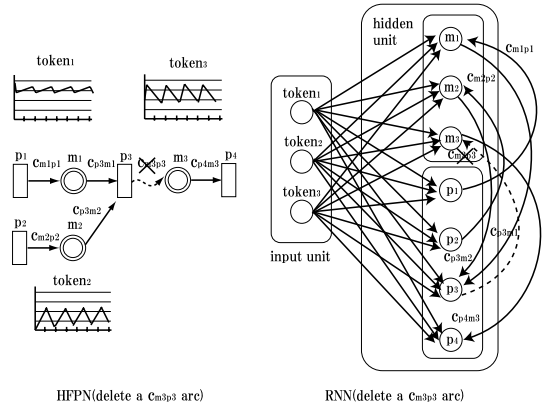


図 11 3rd stage (1) : アーク削除 (a5) を行った HFPN と RNN の関係  
 Fig. 11 3rd stage (1) : The relation between HFPN and RNN which performed arc elimination (a5).

$$D = \sum_{i \in M \cup P} |\tilde{z}_i(Q)| \quad (12)$$

ここで， $D$  の値が小さい場合は，ロックアウトによりアークを削除しても誤差量が少ないことから，対象としたアークは生命パスウェイにおいて重要な経路ではなく，他の経路でこの経路の信号を補うことができるような頑健な経路であり，それと連結されるプレースは頑健部位であると考えられる．それに対し， $D$  の値が大きい場合は，ロックアウトによりアークを削除することによって誤差量が大きくなることから，対象としたアークは生命パスウェイにおいて重要な経路であるといえる．したがって，他の経路でこの経路の信号を補うことができない脆弱な経路において，それと連結されるプレースは脆弱部位であると考えられる．この誤差量の大きさにより頑健部位であるか脆弱部位であるかを判定できるようになる．この処理を，すべてのアークにおいて順に誤差量を求め判定を行うことで，生命パスウェイ全体の脆弱部位を判定することが可能となる．

4.3 4th stage : 脆弱部位の断定

すべてのアークにおいてロックアウトを行い誤差量を求めた結果を次式とする．

$$D_i, \quad i \in C \quad (13)$$

誤差量が大きい場合は破綻の可能性が高く，誤差量が小さい場合は頑健である可能性が高いことから，誤差量の大きい順にソート処理を行い，その結果を  $\tilde{D}_i$  と

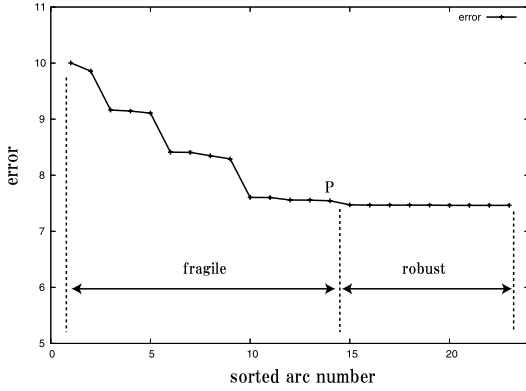


図 12 脆弱部位と頑健部位のしきい値決定

Fig. 12 Determination of the threshold of the fragile part and the robust part.

する。ここで、

$$O_i = \{j | \tilde{D}_i \leq \tilde{D}_j, j \geq i\} \quad (14)$$

とし、 $O_i$  の個数が最大となる  $\tilde{D}_p$  の値を脆弱部位と頑健部位のしきい値とする。ソート処理により誤差の大きい順に並べ換えられた結果を図 12 に示す。横軸は、ソート後に定義し直したアークの番号であり、縦軸は誤差量である。ここで、1 から P 点までは、ノックアウトの影響で誤差の変動が存在するのに対して、P 点以降は、ノックアウトの影響があるにもかかわらず、誤差が小さい状態で一定値となる。この“P”点を脆弱部位と頑健部位のしきい値として定義する。これにより、各々のアークが脆弱部位か頑健部位かを判定することができる。 $\tilde{D}_p$  の妥当性については、付録 1. で議論する。

## 5. 速度情報導入による脆弱部位推定法の改善

前章では、HFPN を RNN に再形成する手法を示したが、この手法は HFPN で定義される速度情報については考慮されていない。後述の生命モデルを用いたシミュレーション実験の結果 (表 3) から明らかであるが、頑健部位の推定において、RNN 手法の的中率が良くない結果となる。これは、HFPN で記述した生命パスウェイにおいて、プレース中を移動するトークン量の変化は、各々のアークに設定される速度情報の影響を非常に大きく受けることにある。そこで RNN の学習を行う前処理として HFPN の速度情報を RNN の各ユニット間の結合荷重に与える結合係数として定義することで、速度情報を反映させ BPTT 法

により学習を行う。

離散時刻  $t$  で、ユニット  $j$  からユニット  $i$  における速度情報を  $c_{ij}(t)$  とすると、式 (5) は

$$c_{ij}(t) = \begin{cases} c_{ij}(t), & (i, j) \in C \\ 1, & j \in T \\ 0, & \text{その他} \end{cases} \quad (15)$$

となり、式 (6) は、次のように再定義できる。

$$\begin{aligned} y_i(t+1) &= f_i(s_i(t+1)) \quad i \in M \cup P \\ s_i(t+1) &= \sum_{j \in T} c_{ij}(t) w_{ij} x_j(t) + \sum_{j \in M \cup P} c_{ij}(t) w_{ij} y_j(t) \\ &= \sum_{j \in T \cup M \cup P} c_{ij}(t) w_{ij} z_j(t) \end{aligned} \quad (16)$$

## 6. 細胞内反応モデルを用いたシミュレーション実験

### 6.1 上皮成長因子の基本モデル

細胞内反応モデルによるシミュレーション実験を行い、本手法の脆弱部位推定における有効性の検討を行う。このモデルには、図 3 で示した上皮成長因子受容体 (EGFR) に増殖のかぎとなる上皮成長因子 (EGF) が結合することで、受容体が活性化するためのシグナル伝達モデルを用いた。このシグナル伝達が、細胞の分化・増殖へとつながる。図 3 では、EGFR 阻害剤として知られる AG 化合物を加え、阻害剤の代謝とリン酸化反応の抑制関係を導くことができる。このように、導入された AG 化合物が代謝されることで、AG 化合物の量が抑止アークのしきい値以下になるとき、リン酸化反応の抑制が外れ、活性化した EGF と EGFR の複合体 EGFR\_EGFR\_dimer{P} の 2 量体が増えていく現象を確認できる [17]。しかし、上皮成長因子のモデルは頑健部位・脆弱部位について生物学的実験では明らかにされていない。

ここで脆弱部位を推定するために、上皮成長因子のシグナル伝達モデルを HFPN を用いてモデル化する (図 4)。ここで、トークンの初期値やトランジションパラメータの調節は、実際にシミュレーションを行い、その結果を踏まえてパラメータの変更を逐次修正することによって構築されている。

### 6.2 HFPN を用いた手作業によるノックアウト指標結果

あらかじめ HFPN を用いて従来の手作業によるノッ



表 1 HFPN を用いた手作業による指標判定

Table 1 Creating indicators by manual procedures using HFPN.

| 始点                | 終点                | 破綻数 | 指標 |
|-------------------|-------------------|-----|----|
| AG                | p14               | 3   | 脆弱 |
| AG                | p12               | 1.5 | 脆弱 |
| EGFR_EGF_dimer{P} | p13               | 1   | 頑健 |
| EGFR_EGF_dimer{P} | p11               | 0   | 頑健 |
| EGFR_EGF_dimer    | p12               | 2.5 | 脆弱 |
| EGFR_EGF_dimer    | p10               | 0   | 頑健 |
| EGFR_EGF_dimer    | p8                | 0.5 | 頑健 |
| EGFR_EGF          | p9                | 3   | 脆弱 |
| EGFR_EGF          | p7                | 0   | 頑健 |
| EGFR_EGF          | p5                | 0   | 頑健 |
| EGF               | p6                | 3   | 脆弱 |
| EGF               | p4                | 1.5 | 脆弱 |
| EGFR              | p6                | 2   | 脆弱 |
| EGFR              | p2                | 2   | 脆弱 |
| p12               | EGFR_EGF_dimer{P} | 1   | 頑健 |
| p11               | EGFR_EGF_dimer    | 0   | 頑健 |
| p10               | EGFR_EGF          | 0   | 頑健 |
| p9                | EGFR_EGF_dimer    | 2   | 脆弱 |
| p7                | EGF               | 0   | 頑健 |
| p7                | EGFR              | 0   | 頑健 |
| p6                | EGFR_EGF          | 3   | 脆弱 |
| p3                | EGF               | 5   | 脆弱 |
| p1                | EGFR              | 5   | 脆弱 |

クアウトを行うことで指標判定を行った結果を表 1 に示す。始点と終点は、ロックアウトを行い削除したアークに連結するプレースまたはトランジションを表す。ロックアウトを行っていない正常動作の場合とロックアウトを行った場合を目視で比較し、次のように破綻数を求める。始点から終点につながるアークをロックアウトした場合に、各プレース中のトークン値を出力し、通常動作の場合に比べてトークン値の時間変化に差が見られないものを 0 とする。また、トークン値の時間変化が 0 に収束するまたは無限大に発散する場合は 1 とし、それ以外をすべて 0.5 とする。これを各プレースで確認していき、すべてのプレースの合計を求めることで指標判定用の破綻数とする。したがって、上皮成長因子の基本モデルにおいてすべてのプレースで破綻した場合は 6 となり、すべてのプレースで破綻が起こらなかった場合は 0 となる。

ここで、指標判定を行うために、しきい値  $T_h$  を定める。ロックアウトによりアークを一つ抜くと、そのアークの連結先以降のトランジションやプレースは連鎖的に影響を受ける。しかし、生命現象において崩壊のように次に伝わる伝達経路がないときのように、HFPN 表現した場合、アークの連結先がトランジションで終わりその後の影響を受けない場合や、他のアーク

クがロックアウトの影響を補う経路をもつ頑健な場合でない限り、必ず一つのプレースは破綻が生じる。このことから、一つのアークを抜いて一つのプレースしか影響しないシステムは頑健なシステムといえる。逆に一つのアークを抜いて二つ以上のプレースに連鎖的に影響がある場合は、脆弱なシステムであることがいえるが、少し条件を厳しくして頑健と脆弱のしきい値を 1.5 に設定する。よって、アークのロックアウトにより破綻数が 0 から 1 であった場合は、システム全体のバランスは保つことができる頑健部位であると判断して指標判定を“頑健”とした。また、1.5 以上の破綻数の場合は、システム全体のバランスが崩れてしまう脆弱部位であると判断して指標判定を“脆弱”とした。

### 6.3 速度情報を含まない場合の RNN による脆弱部位推定の結果

速度情報を含まない場合における上皮成長因子のシグナル伝達モデルにおいて RNN 手法による推定を行った。BPTT 法の学習パラメータは、学習係数を  $\alpha = 0.8$  とし、学習回数を 10000 回で行った。入力ユニットへの信号は HFPN のトークン値をそのまま利用するのではなく、サンプリング間隔を 4 ポイント間隔とし時間展開する階層数の減少を図った。

学習された RNN において自動で連続的にロックアウトを行い、ソート処理で脆弱部位の可能性が高い順番に並べ換え、しきい値により脆弱部位か頑健部位かの判定を行った結果の例を表 2 に示す。表 2 において脆弱部位と判定されたのは、EGFR\_EGF\_dimer{P} から p11 へのアークまでであり、EGFR\_EGF\_dimer から p8 へのアーク以降は頑健部位であると判定された。実験は結合荷重の初期値を変えて 30 回行ったが、すべて同じ推定結果が得られた。表 1 で指標判定した結果と、RNN により推定した結果が一致し正しく推定された場合を正当性の項目でとし、正しく推定できなかった場合を  $\times$  とした。

各々の予測率を示した結果を表 3 に示す。指標判定の脆弱部位 12 箇所、頑健部位 11 箇所に対し、速度情報を含まない RNN 手法による脆弱部位の予測結果は、脆弱部位であると正しく予測できた結果が 100.0% (12 箇所) であり、頑健であると誤った率が 0.0% (0 箇所) となった。しかし、正しく頑健部位であると予測した結果が 36.4% (4 箇所) で、脆弱部位であると誤った率が 63.6% (7 箇所) となった。全体の結果を総合的に判断した場合、23 のアークに対し、16 のアークが正しく予測でき、正当率は 69.6% であり、誤り率

表 2 RNN を用いた自動推定結果 (速度情報なし)

Table 2 The result of presumption automatically by using RNN. (There is no information of the speed)

| 始点                | 終点                | 指標 | 推定 | 評価 |
|-------------------|-------------------|----|----|----|
| EGFR_EGF_dimer{P} | p13               | 頑健 | 脆弱 | ×  |
| p1                | EGFR              | 脆弱 | 脆弱 |    |
| p3                | EGF               | 脆弱 | 脆弱 |    |
| p11               | EGFR_EGF_dimer    | 頑健 | 脆弱 | ×  |
| p9                | EGFR_EGF_dimer    | 脆弱 | 脆弱 |    |
| p12               | EGF_REG_Fdimer{P} | 頑健 | 脆弱 | ×  |
| EGF               | p4                | 脆弱 | 脆弱 |    |
| EGFR              | p2                | 脆弱 | 脆弱 |    |
| EGFR              | p6                | 脆弱 | 脆弱 |    |
| EGF               | p6                | 脆弱 | 脆弱 |    |
| p6                | EGFR_EGF          | 脆弱 | 脆弱 |    |
| p10               | EGFR_EGF          | 頑健 | 脆弱 | ×  |
| EGFR_EGF          | p7                | 頑健 | 脆弱 | ×  |
| EGFR_EGF          | p5                | 頑健 | 脆弱 | ×  |
| EGFR_EGF          | p9                | 脆弱 | 脆弱 |    |
| AG                | p14               | 脆弱 | 脆弱 |    |
| EGFR_EGF_dimer    | p12               | 脆弱 | 脆弱 |    |
| AG                | p12               | 脆弱 | 脆弱 |    |
| EGFR_EGF_dimer{P} | p11               | 頑健 | 脆弱 | ×  |
| EGFR_EGF_dimer    | p8                | 頑健 | 頑健 |    |
| EGFR_EGF_dimer    | p10               | 頑健 | 頑健 |    |
| p7                | EGF               | 頑健 | 頑健 |    |
| p7                | EGFR              | 頑健 | 頑健 |    |

表 3 速度情報を含まない場合の RNN による脆弱・頑健部位推定の結果

Table 3 The result of presumption of the fragile part and the robust part by RNN under condition in which there is no information of the speed.

|           | 脆弱部位        | 頑健部位      | 合計         |
|-----------|-------------|-----------|------------|
| HFPN の指標  | 12          | 11        | 23         |
| RNN 手法的中率 | 100.0% (12) | 36.4% (4) | 69.6% (16) |
| RNN 手法誤り率 | 0.0% (0)    | 63.6% (7) | 30.4% (7)  |

は 30.4% であった。

#### 6.4 速度情報を導入した場合の RNN による脆弱・頑健部位推定の結果

速度情報を含む場合における上皮成長因子において RNN 手法による推定を行った。BPTT 法の学習パラメータは、速度情報を含まない場合と同様に学習係数を  $\alpha = 0.8$  とした。学習回数は、速度情報を導入した場合、誤差の推移が速く 10000 回の学習を必要としないため、5000 回とした。サンプリング間隔は 4 ポイント間隔とし時間展開する階層数の減少を図った。

BPTT 法により学習された RNN において自動で連続的にノックアウトを行い、ソート処理で脆弱部位の可能性が高い順番に表の上から並べ換え、しきい値を決定し脆弱部位と頑健部位を分けた結果を表 4 に

表 4 RNN を用いた自動推定結果 (速度情報あり)

Table 4 The result of presumption automatically by using RNN. (There is information of the speed)

| 始点                | 終点               | 指標 | 推定 | 評価 |
|-------------------|------------------|----|----|----|
| p12               | EGF_REG_dimer{P} | 頑健 | 脆弱 | ×  |
| p6                | EGFR_EGF         | 脆弱 | 脆弱 |    |
| EGFR_EGF          | p9               | 脆弱 | 脆弱 |    |
| p9                | EGFR_EGF_dimer   | 脆弱 | 脆弱 |    |
| EGFR_EGF_dimer    | p12              | 脆弱 | 脆弱 |    |
| EGF               | p6               | 脆弱 | 脆弱 |    |
| EGFR              | p6               | 脆弱 | 脆弱 |    |
| EGFR_EGF_dimer{P} | p13              | 頑健 | 脆弱 | ×  |
| AG                | p12              | 脆弱 | 脆弱 |    |
| p3                | EGF              | 脆弱 | 脆弱 |    |
| p1                | EGFR             | 脆弱 | 脆弱 |    |
| EGF               | p4               | 脆弱 | 脆弱 |    |
| EGFR              | p2               | 脆弱 | 脆弱 |    |
| p11               | EGFR_EGF_dimer   | 頑健 | 脆弱 | ×  |
| p10               | EGFR_EGF         | 頑健 | 頑健 |    |
| p7                | EGFR             | 頑健 | 頑健 |    |
| p7                | EGF              | 頑健 | 頑健 |    |
| EGFR_EGF          | p7               | 頑健 | 頑健 |    |
| EGFR_EGF          | p5               | 頑健 | 頑健 |    |
| EGFR_EGF_dimer    | p10              | 頑健 | 頑健 |    |
| EGFR_EGF_dimer{P} | p11              | 頑健 | 頑健 |    |
| EGFR_EGF_dimer    | p8               | 頑健 | 頑健 |    |
| AG                | p14              | 脆弱 | 頑健 | ×  |

表 5 速度情報を含む場合の RNN による脆弱・頑健部位推定結果

Table 5 The result of presumption of the fragile part and the robust part by RNN under condition in which there is information of the speed.

|           | 脆弱部位       | 頑健部位      | 合計         |
|-----------|------------|-----------|------------|
| HFPN の指標  | 12         | 11        | 23         |
| RNN 手法的中率 | 91.7% (11) | 72.7% (8) | 82.6% (19) |
| RNN 手法誤り率 | 8.3% (1)   | 27.3% (3) | 17.4% (4)  |

示す。表 4 において脆弱部位と判定されたのは、p11 から EGFR\_EGF\_dimer へのアークまでであり、p10 から EGFR\_EGF へのアーク以降は頑健部位であると自動推定された。

各々の予測率を示した結果を表 5 に示す。指標判定の脆弱部位 12 箇所、頑健部位 11 箇所に対し、速度情報を含む RNN 手法による脆弱部位の予測結果は、脆弱部位であると正しく予測できた結果が 91.7% (11 箇所) であり、頑健であると誤った率が 8.3% (1 箇所) であった。一方、正しく頑健部位であると予測した結果が 72.7% (8 箇所) で、脆弱部位であると誤った率が 27.3% (3 箇所) と、頑健部位に対しては大幅に推定率が向上した。全体の結果を総合的に判断した場合、23 のアークに対し 19 のアークが正しく予測できたこ

となり、正当率は 82.6% であり、誤り率は 17.4% であった。

速度情報を導入した場合、脆弱部位の推定率は多少低下するものの頑健部位の推定に対しては、推定率の向上につながる事が明らかになった。また全体としても推定率が向上することが分かった。しかしまだ十分な推定率とはいえ今後の改善が求められる。

## 7. 脆弱・頑健部位推定システムの評価

速度情報を含まない場合の RNN による脆弱部位推定の結果、そのほとんどが脆弱部位であると誤った推定をしているために、頑健部位を 4 箇所しか推定ができなかった。これは、脆弱部位をあらかじめ予測し、生物実験での焦点を絞り込むという目的を果たすことができないため、脆弱部位の推定率が高くても決して良い結果であるとはいえない。それに対し、速度情報として、HFPN の速度パラメータを、RNN のユニット間の結合荷重へかかる重みとして組み込むことで脆弱部位、頑健部位の推定率に 10% 以上の向上が認められた。また、学習回数が少ない段階で結果を得ることも可能になった。この結果は、HFPN において速度パラメータが非常に重要な要素であることを示していると同時に、RNN への再形成で HFPN の速度パラメータをうまく付加できたことを意味する。

しかし、推定率の面からは脆弱部位・頑健部位全体として 82.6% であり、いまだ十分とはいえない。今回、脆弱部位と頑健部位を分類するしきい値は式 (14) により決定したが、より適切なしきい値を選ぶことによって推定率の更なる向上が期待される。今後、更なる有効なしきい値決定方法の検討が必要である。

## 8. む す び

コンピュータを用いて脆弱部位・頑健部位の推定を行うためには、生物実験と同様にモデル上でノックアウトを行う必要がある。HFPN はノックアウトによる確認作業を、対象とするアークを削除し実行処理を施す一連の作業を逐次すべてのアークで行う必要があるため、ばく大な処理時間がかかり、更にネットワークが大きいものは作業途中で人的なミスが生じる可能性があった。また、すべてをコンピュータを用いて連続的に HFPN でノックアウトを行わずに自動化できたとしても、ペトリネットを用いている以上、アルゴリズム的な変更がない限り、計算処理時間を削減することはできない。

本研究ではこの問題を改善するために、脆弱部位と頑健部位の自動推定システムの開発において、HFPN で描かれた生命モデルから、ネットワークの連結情報と時間的な特性、速度情報などの特徴抽出を行い、これを RNN へ記憶保持させることで、想起処理により短時間で結果を得る手法を提案した。

生命現象を表す基本的な構造の“結合・分裂・崩壊・翻訳・フィードバックループ”をもつ上皮成長因子のシグナル伝達モデルの経路を利用し、脆弱・頑健部位推定システムの開発を行い、有効性の検討を行った。HFPN をアーキテクチャにもつソフトウェア Cell Illustrator の熟練者が手作業で行った場合、すべての推定結果を表にまとめるまでに 1.5 時間程度の時間を要した。内訳は、1 本のアークの計算処理時間が入出力設定の処理時間を含めて約 1.5 分で 23 本のアークに対して約 35 分、グラフ確認による各トークン値の状態確認とシステム全体の脆弱性・頑健性の判断に約 55 分であった。これに対し RNN を用いて分類を行った結果、速度情報がない場合、CPU のクロック数が 3 GHz の計算機において 10000 回の学習で約 16 分の学習時間と 2 秒程度の想起時間で推定結果を得ることができた。また、速度情報がある場合、5000 回まで学習回数を削減でき、5000 回の学習で約 10 分の学習時間と 2 秒程度の想起時間で推定結果を得ることができた。計算機の処理能力の向上により更なる推定時間の削減が期待できる。また、推定率においては HFPN の速度情報を組み込んだ RNN の再形成を行った結果、脆弱部位で 90% を超える結果を得ることができた。頑健部位に関しては、脆弱部位ほどの結果を得ることができず 72.7% の推定率であったが、全体的には、82.6% の推定率となった。上皮成長因子のモデルでは、アークの数が 23 本、プレースの数が 6 個と小さなモデルの例であるが、HFPN を用いてこれまでに実現されてきた生命モデルのようにアークの数が 200 本以上、プレースの数が 100 個以上の分裂酵母モデル [18] などにおいては、従来の手法で数週間かけて脆弱部位・頑健部位の推定を行っていた作業が、数時間程度の学習時間となり、有効的なシステムにつながったことがいえる。

薬剤開発のため、薬品投与のタイミングや量、そしてノックアウトの場所などらみつぶしに行っていた生物実験において、あらかじめ対象範囲の絞り込みを行えるほか、予測実験をシミュレーションすることで膨大な費用・時間・生命の犠牲を抑えることが期待できる。

謝辞 本研究の一部は文部科学省科学研究費補助金特定領域研究「生命システム情報」の援助を受けて行われた。ここに記して感謝の意を表する。

## 文 献

- [1] 山下富義, “生命をシステムとしてみる薬剤開発,” 月刊薬事, vol.46, no.7, pp.15–17, 2004.
- [2] H. Kitano, “Biological robustness,” Nature Review Genetics, vol.5, pp.826–837, 2004.
- [3] M. Tomita, K. Hashimoto, K. Takahashi, T. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. Venter, and C.A. Hutchison, “E-CELL: Software environment for whole cell simulation,” Bioinformatics, vol.15, pp.72–84, 1999.
- [4] M. Tomita, “Whole-cell simulation: A grand challenge of the 21st century,” Trends Biotechnology, vol.19, pp.205–210, 2001.
- [5] J.M. Carlson and J. Doyle, “Highly optimized tolerance: A mechanism for power laws in designed systems,” Phys. Rev., vol.60, no.2, pp.1412–1427, 1999.
- [6] H. Shinto, Y. Tashiro, M. Yamashita, G. Kobayashi, T. Sekiguchi, T. Hanai, Y. Kuriya, M. Okamoto, and K. Sonomoto, “Kinetic modeling and sensitivity analysis of acetone-butanol-ethanol production,” Biotechnology, no.131, pp.45–56, 2007.
- [7] H. Matsuno, Y. Tanaka, H. Aoshima, A. Doi, and S. Miyano, “Biopathways representation and simulation on hybrid functional Petri net,” Silico Biology, vol.3, no.3, pp.389–404, 2003.
- [8] W. Reisig and G. Rozenberg, eds., Lecture on Petri Nets I: Basic Models, Lecture Notes in Computer-Science 1941, Springer-Verlag, 1998.
- [9] H. Alla and R. David, “Continuous and hybrid Petri nets,” J. Circuits Syst. Comput., vol.8, pp.159–188, 1998.
- [10] R. Hofstadt and S. Thelen, “Quantitative modeling of biochemical networks,” Silico Biology, vol.1, pp.39–57, 1998.
- [11] R. Drath, “Hybrid object nets: An object oriented concept for modeling complex hybrid systems,” Proc. Hybrid Dynamical Systems, 3rd Internatiuonal Conference on Automation of Mixed Process, pp.437–442, 1998.
- [12] M. Nagasaki, A. Doi, H. Matsuno, and S. Miyano, “Genomic object net : I, A platform for modeling and simulating biopathways,” Applied Bioinformatics, vol.2, pp.181–184, 2004.
- [13] <http://www.cellillustrator.com> (株)ジーエヌアイ公式 Cell Illustrator 用 URL トップページ
- [14] H. Matsuno, C. Li, and S. Miyano, “Petri net based descriptions for the systematic understanding of biological pathways,” IEICE Trans. Fundamentals, vol.E89-A, no.11, pp.3166–3174, Nov. 2006.
- [15] A. Doi, M. Nagasaki, H. Matsuno, and S. Miyano, “Simulation-based validation of the p53 transcriptional activity with hybrid functional Petri net,” Silico Biology, vol.6, pp.1–13, 2006.
- [16] R. Williams and D. Zipser, “A learning algorithm for continually running fully recurrent neural networks,” Neural Comput., vol.1, pp.270–280, 1989.
- [17] 土井 淳, 長崎正朗, 斉藤あゆむ, 松野浩嗣, 宮野 悟, システム生物学がわかる!, 共立出版, 2007.
- [18] S. Fujita, M. Matsui, H. Matsuno, and S. Miyano, “Modeling and simulation of fission yeast cell cycle on hybrid functional Petri net,” IEICE Trans. Fundamentals, vol.E87-A, no.11, pp.2919–2928, Nov. 2004.
- [19] J.H. Holland, Adaptation in natural and artificial system, University of Michigan Press, 1975.

## 付 録

### 1. しきい値の妥当性の検討

提案する脆弱部位推定の一連の操作によって得られたしきい値の有効性を確認するために, 上皮成長因子の基本モデルを例にして検討を行う。ここで, 提案手法で定義した式 (14) から得られた  $O_i$  の個数が最大となるしきい値  $\tilde{D}_p$  が, 従来の HFPN を用いた手法と比べてどの程度距離があるかを比較検討する。しかし, 従来の方法では表 1 を作成した後, 破綻数から脆弱か頑健かを導くしきい値  $T_h$  を主観的に決定するため, 提案手法のしきい値  $\tilde{D}_p$  と直接比較することができない。そこで従来手法のしきい値  $T_h$  をメタヒューリスティックな最適化アルゴリズムである遺伝的アルゴリズム (GA) [19] を用いて, 提案手法のしきい値  $\tilde{D}_p$  と直接比較できるしきい値  $\tilde{T}_h$  に変換する。

#### 1.1 しきい値 $\tilde{T}_h$ の変換

しきい値  $\tilde{T}_h$  の変換を次の七つの過程で行う。

##### step 1: 遺伝子型の決定

コーディングを行うため, 遺伝子型の設定を行う。GA では遺伝子型に 2 進数 (0,1) でコード化した染色体を用いる場合が一般的であるが, ここでは RNN を BPTT により学習して得られた結果  $D_i$  のような実数値を遺伝子型とする。

##### step 2: 初期母集団の生成

$D_i$  に対して,  $[\min\{D_i\}, \max\{D_i\}]$  の範囲で乱数を用いて初期遺伝子を決定する。初期遺伝子集団  $B = \{B_u\}$  とし, その適応度を  $F_u$  とする。各々の集団サイズを 20 とする。

##### step 3: 染色体の評価

表 A.1 は上皮成長因子の基本モデルに対して, 本手法を用いて得られた結果であり, これを評価用のデータとする。RNN の学習回数は 10000 回とした。

表 A.1 評価用のデータ  
Table A.1 Data for evaluation.

| 誤差量 ( $D_i$ ) | 破綻数 $\times 10(h_i)$ | 指標 ( $d_i$ ) |
|---------------|----------------------|--------------|
| 7.462710433   | 30                   | 1            |
| 8.290739479   | 15                   | 1            |
| 8.347663434   | 10                   | 0            |
| 7.462839644   | 0                    | 0            |
| 9.107557604   | 25                   | 1            |
| 7.462849603   | 0                    | 0            |
| 7.462839633   | 5                    | 0            |
| 9.163007920   | 30                   | 1            |
| 7.465870956   | 0                    | 0            |
| 7.465776776   | 0                    | 0            |
| 8.411056466   | 30                   | 1            |
| 7.557467721   | 15                   | 1            |
| 8.407344026   | 20                   | 1            |
| 7.555774098   | 20                   | 1            |
| 10.00383740   | 10                   | 0            |
| 7.544107645   | 0                    | 0            |
| 7.471113021   | 0                    | 0            |
| 9.144240393   | 20                   | 1            |
| 7.466306727   | 0                    | 0            |
| 7.466349238   | 0                    | 0            |
| 9.857082812   | 30                   | 1            |
| 7.604164569   | 50                   | 1            |
| 7.601894861   | 50                   | 1            |

表 A.2  $\delta_{ui}$  の値  
Table A.2 Value of  $\delta_{ui}$ .

| $B_u$ と $D_i$ の関係 | $d_i$ |       |
|-------------------|-------|-------|
|                   | 1(脆弱) | 0(頑健) |
| $B_u \leq D_i$    | 1     | -1    |
| $B_u > D_i$       | -1    | 1     |

この評価用データを用いて適応度  $F_u$  を次のように定義する。

$$F_u = \sum_i \delta_{ui} \times h_i \tag{A.1}$$

ただし、 $\delta_{ui}$  の値は表 A.2 のように定義する。

ここで  $B_u$  は仮のしきい値であり、破綻数は影響を及ぼす度合（脆弱度）として考えられる。仮のしきい値  $B_u$  と RNN の誤差量  $D_i$  を比較して、予想どおり正しく評価できた場合は脆弱度をプラスし、予想に反して正しく評価できない場合は、脆弱度をマイナスにする。この評価方法を用いて GA を通すことによりしきい値  $B_u$  が最良の値に近づき、 $B_u$  の指標となる適応度  $F_u$  を求めることができる。

適応度  $F_u$  を降順ソートした結果を  $\tilde{F}_u$  とし、それに対応した順に遺伝子の要素  $B_u$  を並べ換える。並べ換え後の遺伝子集団を  $\{\tilde{B}_u\}$  とする。

step 4：適応度による選択

淘汰・増殖の選択方法は一般的に用いられる適応

度比例方式を用い、遺伝子交叉を行うための確率を設定する。 $\tilde{B}_u$  に対応する確率を 50%, 25%, 12.5%, 6.25%, 3.125%, ... と 1/2 ずつ減少したものを定義する。

$$p\tilde{r}o_u = \left(\frac{1}{2}\right)^u \tag{A.2}$$

遺伝子交叉のために、 $p\tilde{r}o_u$  の確率で適応度比例方式を用いて二つの遺伝子を  $\tilde{B}_u$  から選択し、親 1, 親 2 とする。

step 5：交叉処理

step 4 で得られた  $\tilde{B}_u$  の数値を 10 ビットの文字列へ変換して次世代の個体群の中から 2 組対の個体に対してランダムに遺伝子型の交叉を行う。交叉は 1 点交叉を用いる。

step 6：突然変異処理

局所安定へ状態が陥らないようにするために、ランダムに選ばれた個体に対して遺伝子座の遺伝子をあらかじめ決めておいた突然変異率の生起確率に従って、他の対立遺伝子に置き換える。ここでは、5% の生起確率を与えた。

step 7：繰返し

step 3 へ戻る。遺伝的アルゴリズムの処理を繰り返すことで、最終的な遺伝子型が最適なしきい値となる。本確認では、10 万回の処理を繰り返した。

step 1 から step 7 を繰り返す。GA の結果  $\tilde{B}_1$  を 20 回分得る。その平均値をとることで確率的にしきい値  $\tilde{T}_h$  を決定する。

### 1.2 しきい値 $\tilde{T}_h$ としきい値 $\tilde{D}_p$ との比較検討

以上のことから、HFPN を用いて破綻確認をあらかじめ行い結果が分かっている場合の最良のしきい値  $\tilde{T}_h$  が求められる。これに対し、あらかじめ HFPN の破綻確認を行わず結果が分かっていない場合における本手法で提案するしきい値  $\tilde{D}_p$  を比較した。その結果、 $\tilde{T}_h = 7.555371372$  に対して RNN を 5000 回学習した場合の結果が  $\tilde{D}_p = 7.544107645$  となり、その差が 0.011263727 であり、それほど差がないことが明らかとなった。この結果から、4.3 に記した方法で選択した  $\tilde{D}_p$  は有効なしきい値であるといえる。

(平成 20 年 1 月 17 日受付, 5 月 7 日再受付)



北風 裕教 (正員)

平 12 北陸先端科学技術大学院大学情報科学研究科・修士課程了。同年(株)興学社入社。平 13 大島商船高等専門学校商船学科助手。現在、大島商船高等専門学校商船学科助教。山口大学大学院自然共生科学専攻博士後期課程在学中。平 18 本会中国支部奨励賞。日本バイオインフォマティクス学会会員。



神田 全啓

昭 47 山口大・文理・理学科数学専攻卒。昭 48 大島商船高等専門学校機関学科助手。現在、大島商船高等専門学校一般教科教授。数学教育の研究に従事。



中司 弘樹 (学生員)

平 16 (有)新設計コンサルタント入社。平 20 大島商船高等専門学校電子・情報システム工学専攻科修了。現在、九州工業大学大学院生命体工学研究科・修士課程在学中。



池田 信彦 (正員)

昭 51 山口大学大学院修士課程了。同年徳山工業高等専門学校助手。同校講師，助教授を経て，平 5 教授，現在に至る。この間，昭 58 内地留学，平 8 外地留学(米国)。工博。主に連想記憶，自己組織理論，BCI などニューラルネットワークの研究に従事。日本神経回路学会会員。



松野 浩嗣 (正員)

昭 57 山口大・工・電子卒。昭 59 同大学院修士課程了。昭 59~62 山口短期大，昭 62~平 6 大島商船高専勤務。平 7 山口大・理・助教授。平 17 同教授。計算機ネットワーク構築技術と生命のシステムの理解に関する研究に従事。理博。IEEE，情報処理学会各会員。



宮野 悟 (正員)

昭 52 九大・理・数学卒。昭 54 同大学院修士課程了。理博。同年九大理学部助手。昭 60 より昭 62 フンボルト財団研究員。同年九大理附属基礎情報学研究施設助教授。平 5 同教授。平 8 東大医科学研究所ヒトゲノム解析センター教授。システムバイオロジー，バイオインフォマティクスなどの研究に従事。日本 IBM 科学賞，坂井記念特別賞などを受賞。日本バイオインフォマティクス学会，International Society for Computational Biology 等各会員。PLoS Computational Biology, Bioinformatics, J. Computational Biology and Bioinformatics などのエディター。