

ニューラルネットワークを用いたがん細胞の認識検証

Batnasan Baasanchuluun[†] 古屋 智子^{††} 松野 浩嗣^{†††} 北風 裕教[†]

† 大島商船高専情報工学科

†† 山口大学大学院医学系研究科

††† 山口大学大学院理工学研究科

1. はじめに

病理医師は、がん患者から細胞を抽出し、個人に適した治療方針や治療方法、薬剤の選別を行う。しかし、がん細胞は進行速度や治療薬の効果など、目視では分類することが困難なタイプも知られており、その形態学的な特徴では分類が困難である。

本研究グループでは、レーザースキャニングサイトメーター (LSC) を用いてがん細胞のタンパク質情報を取得し、サポートベクターマシン (SVM) を用いてがん細胞ごとの SVM を知識 DB へ構築する仕組みを提案してきた。未知のがん細胞を分類する際は、知識 DB へ LSC を介してデータを受け渡し、DB から過去に作成した最適な SVM を特定することで実現できる。一方で、この手法はニューラルネットワークの誤差逆伝播法 (BP 法) で置き換えることが可能であり、SVM の手法と適合率の評価から精度比較を行うことが求められてきた。

そこで本研究では、SVM の学習アルゴリズムを BP 法で置き換えたシステムを開発し、未知データによる適合実験を行ったので報告する。

2. システム概要

本研究では、BP 法を用いて認識実験を行う。以下に、利用するがん細胞データと、学習の前に行う事前準備、BP 法で学習を行う学習用ツール、未知のデータがどのがん細胞と似ているか認識する想起用ツールについて記述する。

2.1 利用するデータの特徴抽出

実験では、LSC から得られた 10 種類のがん細胞のタンパク質情報のデータを 200 個ずつ準備する。タンパク質情報には、Area (核の反応面積)、P_Integral (タンパク質量)、P_MaxPixel (タンパク質の凝集度)、G_Integral (DNA 量)、G_MaxPixel (DNA の凝集度) がある。しかし、これらのデータを学習させる前に特徴抽出しなければならない。特徴抽出では、200 個のデータから 100 個のデータを乱数で選び、それらの平均、標準偏差を求め、すべての値を図 1 に示す方法を使って 4 ビットの 2 進数に変換させ、1 列に並べる。これが 1 つの学習パターンとなる。この方法を利用して、50 個分のパターンデータを作成した。予備実験において、このデータだけで学習を行ったが、特徴を分離できず誤差が収束に繋がらなかった。そこで、面積とタンパク質量、面積とタンパク質の凝集度の相関係数を事前に求め、データに加えて学習を行う事で、十分なデータを抽出できるようにした。

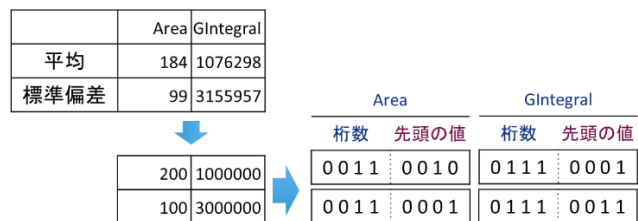


図 1. 2 進数変換方法

2.2 BP 学習のパラメータ

学習アルゴリズムには、BP 法を利用した。1 つの学習パターンには合計 88 ビットの 2 進数と教師信号が含まれる。1 つのがん細胞から特徴抽出して取り出した 50 個のデータ、10 種類のがん細胞から合計 500 個のデータを学習した。ネットワークの構成は、入力層 88 ユニット、中間層 100 ユニット、出力層 1 ユニットとする。学習をすべてのパターンが収束まで行った。

($\eta=0.5$, $\alpha=0.8$, 誤差 0.1)

2.3 実験結果

シミュレーション実験では 10 種類のがん細胞の学習に利用されていないデータを 200 個ずつ準備し、学習前に特徴抽出を行った。図 2 にシミュレーション結果を示す。全てのがんで適合率が最も高い結果となった。

	A431	CCK-81	G-EMC-SS	HeLaS5	HMC-1-8	LK-2	MDA-MB-435	PC-3	SW-13	TIG-7
A431	42 (84%)	0	0	0	0	0	0	8 (16%)	0	0
CCK-81	0	49 (98%)	0	0	0	0	1 (2%)	0	0	0
G-EMC-SS	0	0	50 (100%)	0	0	0	0	0	0	0
HeLaS5	0	0	0	49 (98%)	0	0	0	1 (2%)	0	0
HMC-1-8	0	1 (2%)	0	0	30 (60%)	0	18 (36%)	0	1 (2%)	0
LK-2	0	0	0	0	1 (2%)	32 (64%)	0	17 (34%)	0	0
MDA-MB-435	0	0	0	0	0	14 (28%)	36 (72%)	0	0	0
PC-3	0	0	0	0	0	18 (36%)	0	32 (64%)	0	0
SW-13	1 (2%)	0	0	0	0	0	0	2 (4%)	47 (94%)	0
TIG-7	0	0	0	0	0	0	0	0	0	50 (100%)

図 2. シミュレーション結果

3. おわりに

SVM の学習アルゴリズムを BP 法で置き換えたシステムを開発し、未知データによる適合実験を行った結果、従来の SVM の結果に対し、高い適合率を得ることができた。今後は最適なパラメータの自動設定を行いたい。

謝辞

本研究は、科学研究費補助金・基盤研究 (B) 研究課題番号: 23300110 の補助により行われた。ここに、謝意を表す。