

# ガウシアンカーネルのパラメータの自動化取得を導入した SVM によるがん細胞の分類手法の検証

原田 志織<sup>†</sup>古屋 智子<sup>††</sup>松野 浩嗣<sup>†††</sup>北風 裕教<sup>†</sup><sup>†</sup> 大島商船高専情報工学科<sup>††</sup> 山口大学大学院医学系研究科<sup>†††</sup> 山口大学大学院理工学研究科

## 1. はじめに

現在、がん患者の検体を病理医師は顕微鏡によって目視で判断を行い、知識と経験によって病理診断を行っている。我々の研究グループでは、がん患者から抽出した未知のがん細胞情報と、事前学習した多種のがん細胞の知識データベース(認識 DB)とを比較することで、病理医師が最適な診断結果を導き出すことのできる補助システムについて提案してきた。このシステムでは、知識 DB はサポートベクターマシン(SVM)の学習により特徴化されており、細胞数の増加に対しても、再学習を必要とせず知識 DB の最適な候補を導き出すことを可能にしている。しかしながら、本システムで用いる SVM のカーネル関数は Gaussian カーネルであり、そのパラメータ値はユーザによって最適値を決定する必要があり、全自動化の必要性が求められた。

そこで本研究では、Gaussian カーネルの最適値を自動で取得できる手法を導入し、その性能を評価したので報告する。

## 2. 分類手法

### 2.1 SVM のシステム概要

本システムは2つの経路からなり、1つは、事前に取得しているがん細胞を、レーザースキャニングサイトメーター(LSC)を介してがん細胞のタンパク質情報を抽出し(がん細胞 DB へ保存)、SVM 学習用システムによって特徴化したがん細胞の SVM を形成して、SVM を認識 DB に保存する経路である。もう1つは、がん患者から抽出したがん細胞を、LSC を用いてタンパク質情報を取得し、認識用システムを用いて、事前に作成した知識 DB へアクセスする経路である。SVM 認識用システムでは、認識 DB から可能性の高い候補順に結果を表示する。

### 2.2 SVM による認識

SVM 学習用システムにおいて、各がん細胞の SVM 学習を行う。使用する Gaussian カーネルを式(1)に示す。

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1)$$

この $\sigma$ 値の最適なパラメータ設定により、汎化能力の向上が期待できる<sup>[1]</sup>。

### 2.3 パラメータの自動取得

パラメータ $\sigma$ 値によって正と負の境界(最大マージン化)を決定する(図 1)。最大マージン化では、正と負の領域において、最も近い正と負の点の半分の距離に相当す

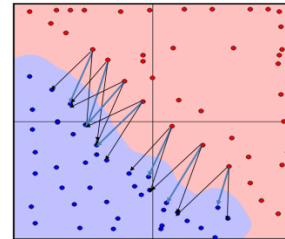


図1 マージン最大化の作成

る。それを全ての点で繰り返すことで、最大マージンの境目を作る<sup>[2]</sup>。

## 3. 分析結果の比較

未知の10種類のがん細胞における適合実験を行った結果を図 2(従来法)と図 3(自動化手法)に示す。従来手法においても提案手法においても統計的に正しく識別が可能となっているが、自動化手法の方がより他のがん細胞と比べて識別率の向上を可能にした。

	A431	CCK-81	G-EMC-SS	HeLaS3	HMC-1-8	LK-2	MDA-MB-435	PC-3	SW-13	TIG-7
A431	41	4	8	9	5	2	2	9	6	9
CCK-81	16	29	7	13	3	4	9	16	5	13
G-EMC-SS	6	3	50	5	4	3	1	16	13	11
HeLaS3	15	8	5	28	4	3	5	6	3	11
HMC-1-8	14	5	3	8	28	9	5	11	14	11
LK-2	5	3	4	6	17	44	3	9	9	6
MDA-MB-435	7	6	6	5	3	2	57	21	5	2
PC-3	14	8	2	5	5	1	12	21	5	7
SW-13	6	28	14	15	20	30	7	9	36	8
TIG-7	17	8	7	8	8	4	1	4	8	24

図 2 従来手法(手作業)の結果

	A431	CCK-81	G-EMC-SS	HeLaS3	HMC-1-8	LK-2	MDA-MB-435	PC-3	SW-13	TIG-7
A431	56	4	4	3	7	6	0	9	10	3
CCK-81	13	20	5	16	14	7	0	13	8	6
G-EMC-SS	5	3	47	4	7	5	2	3	8	18
HeLaS3	5	2	4	33	7	3	14	7	8	29
HMC-1-8	15	7	6	9	32	10	1	6	9	7
LK-2	3	4	2	2	8	62	1	18	1	2
MDA-MB-435	5	3	5	26	5	1	46	1	3	9
PC-3	11	8	6	4	5	20	1	31	13	5
SW-13	12	6	14	4	11	2	3	8	34	8
TIG-7	6	4	13	23	3	1	6	3	7	35

図 3 自動取得した結果

## 5. 今後の課題

Gaussian カーネルの最適値を自動で取得できる手法を導入し、その性能を評価した。その結果、手作業によるパラメータ設定に比べてより適合率を向上させ、作業効率を向上させることに繋げることを可能にした。

### 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。謝意を表す。

### 参考文献

- [1] Nello Cristianini, John Shawe-Taylor サポートベクターマシン入門
- [2] Dynamically Adapting Kernels in Support Vector Machines Nello Cristianini, Colin Campbell, John Shawe-Taylor