

SVM を用いた病理診断システムにおける がん細胞の種類増加に伴う認識検証

重本 昌也[†]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

近年、学習アルゴリズムを用いてがんの分類を行い、治療方針を決定するシステムの開発が進められている。本研究グループでは、がん細胞のタンパク質情報に着目し、SVMを用いたがん細胞分類システムを実現し、病理医師による診断の補助を実現するシステムの開発に努めてきた。また、適合実験を行うことで、システムの有効性を検討してきた。しかし、これまでの実験では、がん細胞のデータ数に制限のある小規模な場合に限られた実験しか実現してはいなかった。

そこで本研究では、適合実験の規模を従来の2倍のがん細胞の種類まで拡張して実験を行い、従来の結果と比較検証を行ったので報告する。

2. SVMを用いた病理診断システム

本システムの構成を図1に示す。がん細胞を蛍光免疫染色してレーザースキャニングサイトメーター(LSC)でデータ化し、データベースへ取り込む。ここでデータを統合してCSVファイル形式に変換して数値データとして取り扱えるようにする。

次に、がん細胞データベースから目的のがん細胞データを抽出し、SVM学習用システムで学習を行い、各がん細胞に対応するSVM識別器を構築する。学習では、カーネル関数を用いて非線形学習が行われ、結果は認識データベースへ保存される。最後に、未知のがん細胞を、LSCを介して情報化し、認識データベースを参照して各認識率を取得する過程である。この過程では、認識検証も行え、認識率の根拠となる結果を分析することができる。

3. 適合実験による検証

3.1 適合実験に用いるデータ

今回、実験に用いるがん細胞は A431：皮膚癌、CCK-81：ヒト大腸癌、G-EMC-SS：軟骨肉腫、HeLaS3：ヒト子宮頸癌、HMC-1-8：乳癌、LK-2：肺癌、MDA-MB-435：乳癌、PC-3：前立腺癌、SW-13：副腎癌、TIG-7：ヒト正常線維芽細胞の10種類である。認識したいがん細胞のデータを200、異なるがん細胞のデータを200、合計400のデータを学習に用い、識別機を構築した。未知のがん細胞データには想定して学習には用いなかったデータを250個準備して実験を行った。

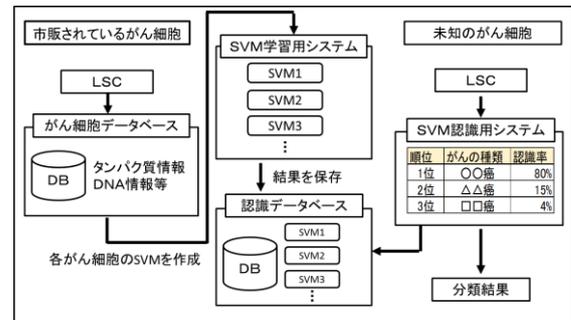


図1. システム概要

3.2 適合実験の結果

従来の実験であるA431, CCK-81, G-EMC-SS, MDA-MB-435, SW-13の5種類の識別機を用意し、A431を未知データとして適合実験を行った結果を表1(左)に示す。最も高い可能性を示したのはA431の識別機で、73%と高い適合率を示した。また、2位のSW-13は15%と大きい差が現れた。

これに対し10種類の識別器を用意して同様の適合実験を行った結果を表1(右)に示す。種類が10種類に増加したにも関わらず、適合率は若干の低下しか見られず、A431の順位は1位で識別された。適合率は60%で、2位と3位の8%、7%に比べても高い結果を得ることができた。

表1. SVM 識別機5種類(左)と10種類(右)の結果

順位	細胞名	適合率	順位	細胞名	適合率
1位	A431	73%	1位	A431	60%
2位	SW-13	15%	2位	SW-13	8%
3位	CCK-81	6%	3位	PC-3	7%

4. おわりに

がん細胞の種類を増加して認識実験を行った結果、大きな認識率の低下はみられなかった。また、SVM 認識用システムに入力し分析するサンプル数を増やすことで、より正確な分別結果が本システムでは期待できると考えられる。

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。