

サポートベクターマシンを用いたがん細胞の認識検証

重本 昌也[†] 北風 裕教[†] 松野 浩嗣^{††}
[†] 大島商船高等専門学校 ^{††} 山口大学大学院理工学研究科

1. 序文

がんは、同一部位から抽出した同じ組織型の細胞であっても、進行速度や薬の効果に違いがあり、病理医師の顕微鏡による目視の分類を困難にしている。我々の研究グループでは、レーザースキャニングサイトメータ(LSC)を用いてがん細胞のタンパク質情報の特徴抽出を実現し、SOM など生体情報処理アルゴリズムを用いて学習システムを構築することで分類実験を行ってきた。その結果、数種類のがん細胞においては高い認識率を示すことが確認できた。しかし、これらの手法では、新しいタンパク質情報を追加するたびに、全体のデータで再学習を行う必要があり、細胞数の増加に応じて認識率が低下するという問題点があった。

そこで本研究では、余計な追加学習を行う必要がなく、多次元データでも高い認識率を獲得可能なシステムを SVM により実現し、病理医師のがん細胞の分析を援助する統合システムを開発した。認識実験を行ったので報告する。

2. システムの概要

本システム構成を図 1 に示す。

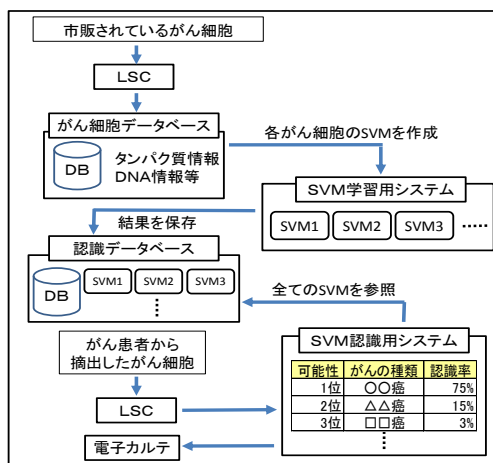


図1. がん細胞分類の統合システムの構成図

LSC を用いて、市販されるがん細胞のタンパク質情報(タンパク質の反応面積、量、凝集度、DNA 量、DNA 凝集度)を抽出し、がん細胞データベースに保存する。SVM(カーネル関数はガウシアンカーネル)学習用システムを利用してデータベース上にある全てのタンパク質情報を学習し、認識データベースに蓄積する。がん細胞

データベースに新しい情報が追加された場合は、がん細胞の SVM のみ再学習を行う。

がん患者から抽出したがん細胞は一旦 LSC に通し、タンパク質情報を抽出する。その情報を認識用システムで参照し、認識データベースに保存してある全ての SVM と自動的に比較を行う。そして、どのがん細胞に近似しているかを認識率という値で求める。比較が終了したら、可能性の高いがん細胞とその認識率を、病理医師が用いる電子カルテに表示する。

3. 認識実験

統合システムを用いてがん細胞の分類が可能か認識実験を行った。認識実験に用いたデータと実験結果について以下に記述する。

3.1 実験に用いたデータ

本実験で用いたがん細胞の種類は皮膚癌、軟骨肉腫、乳癌、副腎癌、ヒト大腸癌の 5 種類である。認識したいがん細胞のデータを 200 個、異なるがん細胞のデータを 200 個、合計 400 個のデータを学習に利用し、認識データベースに保存した。また、未知のデータとして、学習には用いなかったがん細胞を 200 個用意した。

3.2 実験結果

未知データとして皮膚癌を認識用システムに読み込んだ結果を図 2 に示す。一番認識率が高かったのは皮膚癌の SVM であり、認識率は 78%であった。また、他の種類のがん細胞でも同様の実験を行った結果、同じがん細胞の SVM において一番高い認識率が得られたことを確認した。

可能性	がんの種類	認識率
1位	皮膚癌	78%
2位	副腎癌	10%
3位	ヒト大腸癌	6%
4位	軟骨肉腫	6%
5位	乳癌	1%

図 2. 皮膚癌の認識結果

4. 結文

未知のがん細胞の特徴抽出化を行い、データベースに存在するがん細胞情報と比較して認識率を得る統合システムを実現した。今後は、がん細胞の種類を増やして認識実験を行い、正しく識別を行えるか確認したい。