



# 展 TENBO 望

## 生物医学画像を自動分類する ソフトウェア“カルタ”の開発



松永 幸大  
Matsunaga Sachihiko  
(東京理科大学)

### 1 はじめに

画像技術の急速な発達により、診断や研究の現場では必ず画像データが活用されるようになった。遺伝子・タンパク質などの分子レベルから、細胞内外の微細構造・細胞・組織・器官・個体・地球レベルまで、生命現象の解析に画像化（イメージング）技術は必要不可欠である。イメージング技術の発達にともない、近年、生物医学画像データの多様化と大規模化が急速に進んでいる。これは、文字や記号だけのデータよりも画像の方が多くの情報を含むことと、通信システムやデータ保存技術の発達により、巨大容量のデータを簡単に伝達、保存できるようになったからである。しかしその一方で、多様なイメージング機器から得られる画像データの判定や分類については、実務経験の豊富な医師や研究者の目視で行うことがいまだに主流である。これは画像を使用した研究推進や医学的診断の律速段階になっている。自動化された装置で多量の画像データを取得し保存しても、分類や解析ができないために放置されるといったことが研究現場で起き始めているのだ。

実際に、今日の医療現場では、顕微鏡・X線撮像法・コンピューター断層撮影法（CT）・核

磁気共鳴画像装置（MRI）・陽電子放出断層撮影（PET）などによる画像診断が広く用いられている。しかし、限られた画像診断医による読影診断を受け、診断結果が確定するまで、多くの時間を要している。筆者の共同開発機関である国立がん研究センター東病院画像診断部では、1人の画像診断医が1日6,000枚以上、年間100万～150万枚の画像を判定しているという。画像診断医の免許は、6年間所定の研究を行い、試験に合格して初めて取得できるため、画像診断医の数を急増させることは難しい。高齢化社会による患者数の増加と画像診断機器の発達による相乗効果により、膨大な数の画像を判定する画像診断医の労力は増加の一途を辿っている。このような画像診断医の負担軽減や画像判定の効率化を図りたいとの思いも、生物医学画像の自動分類ソフトウェアの開発を目指した動機の1つである。

### 2 今までの画像分類の問題点

これまでも多量の生物医学画像を分類するソフトウェアが開発されてきた。ただ、その多くが、画像の種別や分類目的ごとに、評価を行う目的で個別に開発されてきている。人間は複数

の絵画から画風を元に作者を知ることでもできるし、特に判定基準を意識しなくても多くのサンプルから異常なものを選定することが可能である。このような人間に近いパターン認識をソフトウェアに実行させようとするとき、教師データ（入力 A の場合、出力 B になると教えたデータ）をコンピューターに与えなくてはならない。画像分類の場合の教師データとは、アノテーション情報（画像が何を意味するかの情報）を事前に付与した見本データや手本データのことである。そのデータを元にして機械学習を行う手順が教師付機械学習アルゴリズムである。このような学習アルゴリズムを基盤にした自動分類は、教師データを与えれば与えるほど分類精度が高まる。しかしそのためには、専門家が事前に教師データを作成する必要があり、労力と時間がかかる。

また、目的や対象、更には撮影装置や撮影手法が異なる画像を分類するためにソフトウェアを作成しようとするならば、そのたびに教師データを用意してアルゴリズムを作成し直さなければ

ならない。例えば、同じ細胞の明視野像と蛍光像であっても、それぞれの像を分類するためには異なる教師データを準備し、異なるアルゴリズムを開発して初めて自動分類が可能となる。そのため、個別のソフトウェア開発に加え、それぞれの分類精度の比較などに伴って煩雑なデータ処理が生まれるなど、多くの時間と労力が必要となる（図 1 上）。膨大で多種多様な画像を、1つのソフトウェアを用いて自動的かつ高精度で分類できる方法を開発すれば、これらの問題を解決できると期待されてきた。

### 3 カルタの技術的背景

生物医学画像の自動分類法確立の道を切り開く1つの手段として、2012年8月にカルタを作成し、*Nature Communications* 誌（2012年8月28日号）に論文発表した<sup>1)</sup>。CARTA（カルタ）は Clustering-Aided Rapid Training Agent の頭文字をとっており、分類結果表示が日本伝統文化のカルタを並べた様子に似ていることから

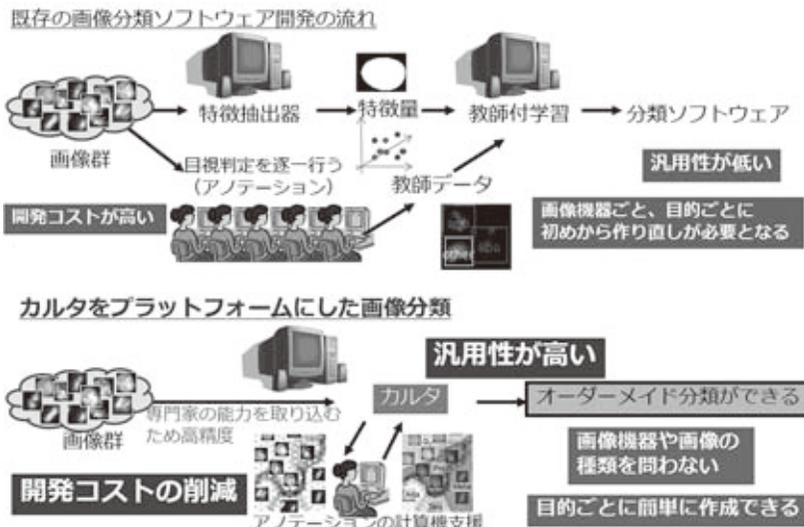


図 1 従来の画像分類ソフトウェアとカルタの違い  
従来の画像分類ソフトウェア（上）は、開発コストが高く、汎用性の低い画像分類しかできなかった。カルタ（下）では、開発コストを削減でき、精度と汎用性の高い画像分類ができる

日本発の技術として親しみを込めて命名した。

カルタは、ユーザーや専門家が入力する目的やアドバイスを繰り返し学習しながら画像分類を行うことが可能になった能動学習型ソフトウェアである。このように人間の意見を取り入れたり、不明な点を質問しながら学習するソフトウェアは能動学習型ソフトウェアと呼ばれている<sup>2)</sup>。カルタを用いれば、専門家や診断士の知力と経験を自ら取り込むことで、最小限の教師情報により精度の高い分類器を作成することが可能となる<sup>3)</sup>。

また、もう1つの特徴としてカルタは半教師付機械学習アルゴリズムを利用している<sup>4)</sup>。人間が与えた少数の教師データに加え、多数の教師なしデータ（出力が与えられていないデータ）からも学習することで、予測精度の高い分類を実現するアルゴリズムとなった。これは教師付機械学習の弱点である個別かつ多数の教師データを準備する難点を解決したと言えよう（図1下）。

教師なしデータから学習するために、カルタは自己組織化マップ（self organizing map：SOM）を利用している。自己組織化マップは1982年にフィンランドのTeuvo Kohonen博士が開発したニュートラルネットワークの教師なし機械学習法である<sup>5)</sup>。入力データの類似度を2次元平面のマップ場に距離で示すことで、視覚的に相関関係やクラスタリング（群れや集団を形成すること）を示すことができる<sup>6,7)</sup>。しかし、クラスタリングが常に目的に合うように類似度を示すわけではない。そこでカルタでは、初期の分類結果を専門家が見て、間違っ て分類された画像を指摘したり、新たにアノテーション情報（教師データ）を指示することができる。

このように SOM による画像のク

ラスタリングを介して、専門家の意見を繰り返し学習すること（iterative clustering）で、研究や検査目的にあった的確な分類基準を自動的に検討する（図2）。クラスタリングに用いる特徴量は自由に入れ替えることが可能であり、遺伝的アルゴリズムを使用して最適な分類を達成した段階で自動的に検討作業を止め、その結果をコンピューター画面上に表示する。

#### 4 カルタの性能—MR 画像を例にして—

実際に各種バイオ画像を分類・判定することで、カルタの性能を実証した。基礎研究分野で重要なタンパク質の細胞内局在解析や特定タンパク質の発現を低下させた細胞表現型の評価、蛍光画像と明視野画像の細胞生死判定などについて、従来法（1枚ずつ目視で判断した場合）と分類・判定精度を比較した。その結果、カルタを使用した場合の判定精度は、目視による従来法を上回り、判定速度についても2倍以上に

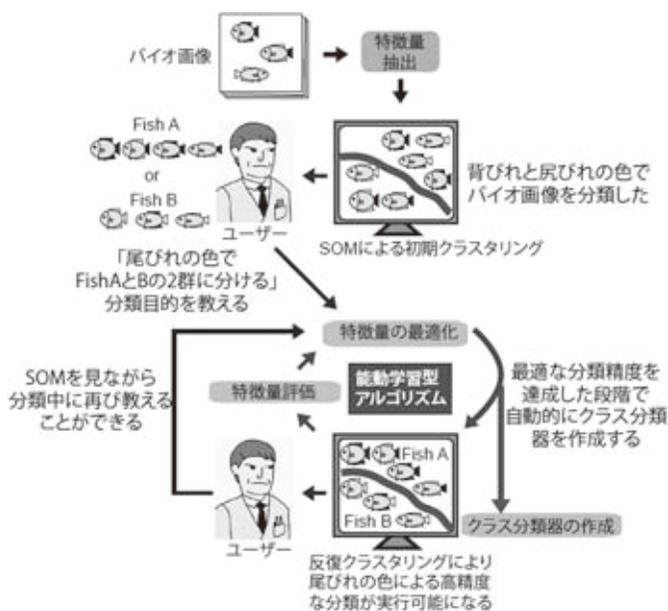


図2 カルタの概略図と分類操作例  
ユーザーが教えた分類目的に最適な特徴量の組み合わせを、クラスタリングの繰り返しによって能動的に学習する

スピードアップすることができた<sup>1)</sup>。

次に、判別が難しい2種類のがんについて核磁気共鳴画像装置 (MRI) で画像を取得し、カルタを用いて分類した結果を紹介する (図3)。がんの腫瘍画像からがんの種類を見分けることは、そのがんが原発性腫瘍なのか転移して形成された腫瘍なのかを判断し、治療方針を決める上でも最も重要な診断の1つである。肉腫由来のがん細胞 S180 と乳がん由来のがん細胞 FM3A をマウスに皮下注射し、形成された腫瘍を動物用 MRI で撮像した 286 枚の画像を分類に用いた。画像診断医はこれらの MR 画像から2種類の腫瘍を見分けることができたが、筆者には困難であった。

カルタ分類結果のコンピューター画面への表示方法は、タイルマップと円グラフマップの2種類がある。いずれも 10×10 の格子状のマップの格子1点1点に類似度の高い画像群が分類されている。タイルマップは格子1点に集まった画像群の中で代表的な画像を表示している。円グラフマップは集まった画像の判定情報を円グラフで表示したマップである。円の大きさは分類された画像数に比例して大きくなる。

カルタを用いて分類した結果をタイルマップで解析すると、左側に FM3A、右側に S180 として分類された画像群がクラスタリングされていた。さらに、円グラフマップで解析すると遠く離れた位置の格子には、一方のがん腫瘍の種類のみ分類されていることが分かる。このように、カルタを用いて、画像診断医にしか判別できない、がんの MR 画像も分類することができた。しかし、円グラフチャートの境界部分に分類された画像群には FM3A と

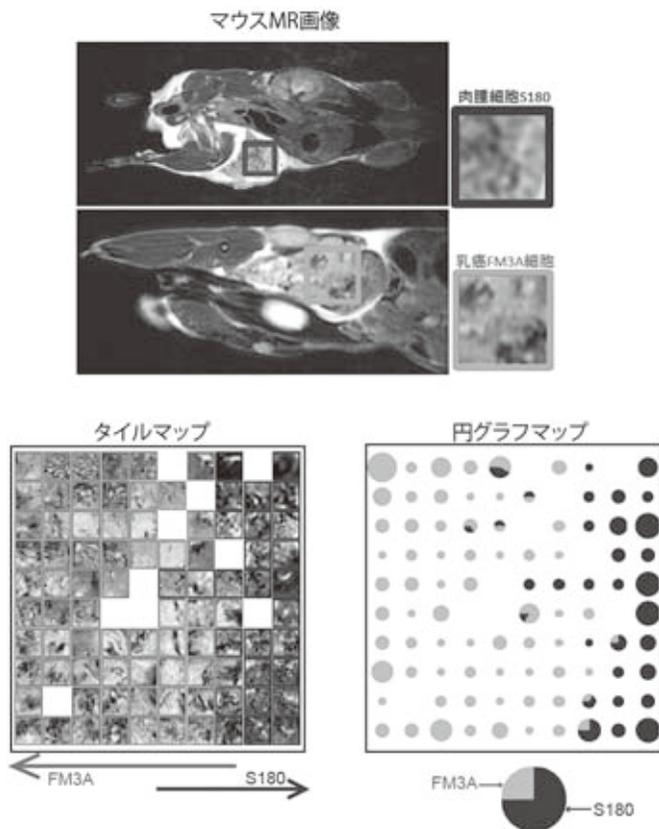


図3 MRIにより取得された腫瘍画像のカルタによる分類  
肉腫由来のがん細胞 S180 と乳がん由来のがん細胞 FM3A を 19 個体のマウスに移植して形成された腫瘍のマウス MR 画像 286 枚 (上: MR 画像例) を、カルタによって分類した。カルタは画像分類結果をタイルマップ (左) と円グラフマップ (右) で表示する。その結果、カルタは左側に FM3A、右側に S180 の画像群を分類した。マップの格子点には数枚から数十枚の画像が分類されている。タイルマップは、分類された画像群の中で、代表的な画像を示している。円グラフマップは分類群の中で、FM3A と S180 の割合を示すが表示している

S180 が混在していた。この場合、境界部分に分類された画像群だけを抽出して画像診断医に判定してもらえばよいと考えている。このように、カルタを画像診断のプレ診断ソフトウェアとして活用することで、明らかに分類が明確な画像群を省き、分類の境界線上にある判断が難しい画像群だけを抽出して、画像診断医に判定用画像群として供与できる。カルタが画像診断の補助ソフトウェアとして活用されることで、

画像診断医は膨大な画像を端から1枚1枚目視する必要がなくなり、画像診断の労力軽減とスピード向上が期待される。

## 5 今後の展望

MRIとは別の、放射線診断法であるレントゲン・CT・SPECT・PETなどの診断画像へのカルタの適用は今後の課題である。原理的に適用は可能であり、更にカルタのアルゴリズムを改良する試みが進行中である。また、画像機器に搭載する際には、カルタをそのまま搭載するのではなく、目的にあった形でカルタを用いて作成した分類プログラムをシステムに組み込むことが実用的には優れていると考えられる。今後も、放射線診断に関わる専門医と連携して、カルタの医療現場への実用化を検討していきたい。

### 【謝辞】

カルタは、科学技術振興機構の研究成果展開事業 先端計測分析技術・機器開発プログラム

「生物画像のオーダーメイド分類ソフトウェアの開発」プロジェクトの支援を受け、東京大学大学院新領域創成科学研究科の朽名夏磨助教、桧垣匠特任助教、馳澤盛一郎教授、国立がん研究センター 東病院の山口雅之ユニット長、藤井博史分野長との共同研究により開発されました。関係各位に深謝いたします。

### 参考文献

- 1) Kutsuna, N., *et al.*, *Nature Commun.*, **3**, 1032 (2012)
- 2) Wang, M. and Hua, X.S., *ACM Trans. Intell. Syst. Technol.*, **2**, 1-21 (2011)
- 3) 朽名夏磨, 桧垣匠, 馳澤盛一郎, 松永幸大, *細胞工学*, **31**, 1372-1373 (2012)
- 4) Chapelle, O., Schölkopf, B. and Zien, A., *Semi-Supervised Learning.*, Vol. 2, MIT press Cambridge (2006)
- 5) Kohonen, T., *Biol. Cybern.*, **43**, 59-69 (1982)
- 6) Vesanto, J. and Alhoniemi, E., *IEEE Trans. Neural Netw.*, **11**, 586-600 (2000)
- 7) Wu, S.T. and Chow, T.W.S., *Pattern Recog.*, **37**, 175-188 (2004)