

一細胞ラマン計測と情報科学の融合による細胞診断: 計測迅速化に向けて

Cell diagnosis by combining single cell Raman imaging and information science: toward accelerating measurements

小松崎 民樹*¹
Tamiki Komatsuzaki

*¹ 北海道大学電子科学研究所附属社会創造数学研究センター

Research Center of Mathematics for Social Creativity, Research Institute for Electronic Science, Hokkaido University

Single cell Raman measurements have a great advantage to enable us to monitor the cell states such as cancer/noncancer free from disturbing the system, compared to dye molecule insertion. Recent advance in information science such as Bandits problem and sparse modelling can provide us a new information technology to design speedy and highly accurate measurement. We aim for the creation of a molecular measurement technique in the fields of life science, medicine, and drug-discovery by enabling early diagnosis of diseases that are difficult to diagnose. To begin, we analyze spectral imaging data of cells in the diagnosis of cancer, which currently relies on morphometric information of cells and tissues.

1. はじめに

ラマン分光法は分子振動を検出する強力な分析法である。通常のイメージング技術では、細胞に(下村脩先生のノーベル化学賞 2008 年に代表される GFP タンパク質など) 光る分子(色素分子)を特定のタンパク質に標識することで、そのタンパク質の含有量や動態を観察する。1 細胞ラマン分光イメージング [Palonpon 2013]は、細胞や生体組織を無標識に生きたまま 1 細胞レベルで観察することができ、かつ(色素分子で標識できないような小分子も含めて) 試料内分子を網羅的に分析することができる。また、それらのラマン分光イメージング画像から、細胞分裂前後で有意な変化がみられること[市村 2014]、および、(間葉系幹細胞、胚性幹細胞などの) 正常な細胞と(乳癌、子宮頸癌などの) 癌細胞が(波数の数の次元をもつ) 膨大なスペクトル空間のうちのごく少数の次元で識別できることが分かってきた。

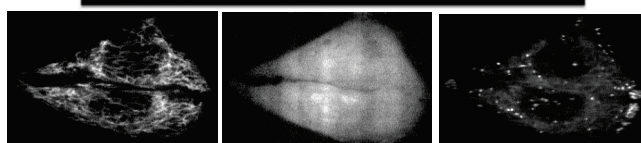
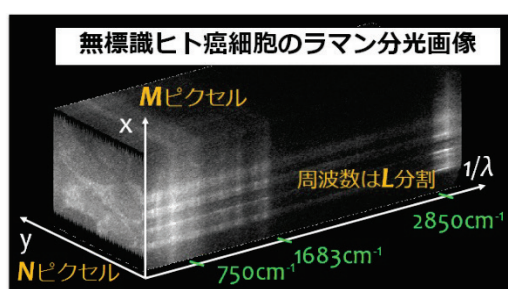


図 1. 無標識ヒト癌細胞のラマン分光画像のデータ構造: 2次元空間($M \times N$)に対して、 L 個(周波数方向を L 分割)の画像が存在。下段(左)シクロロム c (ピロール環)に相当する 750cm^{-1} 、(中央)たんぱく質の B シート(アミド I)に相当する 1683cm^{-1} (右)脂質 (CH_2 振動)に相当する 2850cm^{-1} 。

このことは、従来、細胞や細胞核の形の異常性などに基づいて行われてきた細胞病理診断の信頼性の質的向上を意味する。しかしながら、1 細胞のラマン分光イメージングを取得するだけで、市販の装置ではおよそ 1 分掛かるため、計測の迅速化が医療応用においては喫緊の課題となっている。

イメージング画像に対するデータ科学的な特徴としては、大別して 3 つ挙げられる。第一に、データが膨大であること、空間方向($M \times N$)、波数方向(L)、時間方向を離散化すると 10^7 以上のピクセルに実数が付与されている、第二に、ラマンシフトの信号が微弱であるため、信号のゆらぎやノイズの取り扱いが極めて重要であること、そして第三に、(画像に対して)多様なスパース性をもつことである。たとえば、スパースモデリングを検討する場合、図1下段(右)の画像は実空間にスパース性を仮定できる反面、(左)(中央)の画像は風景や人の写真のような自然画像により近い画像のため、画像の滑らかさ、連続性が重要で、実空間ではスパースではない。そこでは、隣接ピクセルの強度差や Wavelet 変換による波数空間にスパース性を見出す必要がある。このようにラマン分光イメージング画像のスパースモデリングによる画像復元の精度は(周波数方向に対して)均一ではない。最適なスパース尺度がデータを構成する次元(ここでは周波数)に依存して変化するデータ構造は磁気共鳴分光イメージング(MRSI)等にも当てはまる。

2. 研究戦略

ラマン分光イメージングで「どのような情報を抽出したいか」が重要となる。例えば、細胞病理の診断応用において、少数の主成分空間で細胞種が分類できるという経験的事実は、なんらかの少数の特徴量が存在し、すべての周波数成分を計測する必要がない可能性を示唆している。そこで、我々はすべての周波数成分のイメージング画像を均一に復元するのではなく、細胞状態(癌/非癌、病変など)の判定精度を復元する戦略を取る。そのためには、第一には、シグナル・ノイズ比が低いラマンスペクトルデータからの識別規則の構築、少数の特徴量の同定が、第二には、その知識を応用し、ラマン計測による細胞病理診断を迅速化する情報科学的な戦略が必要となる。本発表では、第一の

連絡先: 小松崎民樹, 北海道大学, 札幌市北区北 20 条西 10 丁目, TEL/FAX 011(706)9434, tamiki@es.hokudai.ac.jp

一步としてシグナル・ノイズ比の低いスペクトルからの癌／非癌検出に関する研究成果に関して報告する。

3. ラマン分光画像が張る特徴空間におけるファジー分類:

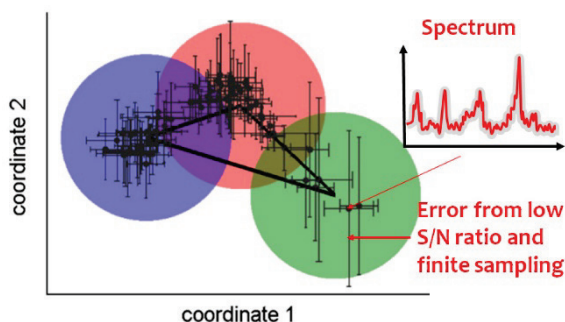
分布関数 i, j 間の距離に相当する Kantorovich 測度 $d(i, j)$ を用いて、主成分解析などで低次元線形空間に情報を限定することなく、スペクトルの全情報を反映したスペクトル分布関数間の距離空間を設定する。さらに、計測誤差、有限サンプルに由来する数揺らぎ誤差を評価しつつ、(1 分子解析で成功を収めた) 誤差を考慮するファジークラスタリング[Taylor2015, Li2015]に基づいて状態数(スペクトル形状が有意に異なるクラスターの総数 N_c)を同定し、各スペクトル g_i がどの細胞状態(クラスター) C_k に属するかの確からしさ(確率) $p(C_k|g_i)$ を評価する。

具体的には、スペクトルとクラスターのあいだの平均相互情報量

$$I(C; g) = \sum_{k=1}^{N_c} \sum_{i=1}^N p(C_k|g_i)p(g_i) \log \left[\frac{p(C_k|g_i)}{p(C_k)} \right]$$

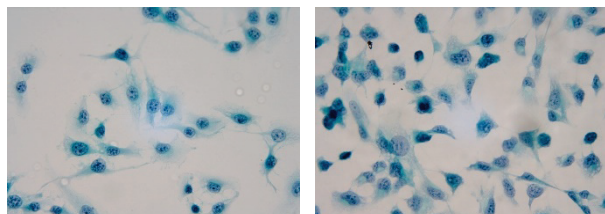
を最小化することでデータ圧縮すると同時に、クラスター内のスペクトルができるだけ類似するように Kantorovich 測度の期待値

$$\langle d \rangle = \sum_{k=1}^{N_c} \sum_{i,j=1}^{N,N} p(C_k)p(g_i|C_k)p(g_j|C_k)d(i, j)$$



を最小化する。クラスター内のスペクトル類似度の制約の強さは、スペクトルデータの計測誤差、有限サイズサンプリングに由来する数揺らぎによって、与えられたデータ毎に自動的に決定する。直感的に言えば、クラスター内に帰属されるスペクトル類似度の制約の強さが大きいほど、クラスターの数は多く、類似度も高くなるが、計測誤差、数揺らぎを越えてクラスター内のスペクトル間の測度が小さくなると、データが保証する以上に分類することに相当する。

形態情報だけでは診断が不可能とされる甲状腺濾胞癌のラマン分光画像による細胞診断の有用性を検証した。図 2 は濾胞癌細胞株、および正常な濾胞上皮細胞株を各々培養したも



濾胞癌細胞株(FTC133) 濾胞上皮細胞株(Nthy-ori)

図 2. 濾胞癌細胞株、および正常な濾胞上皮細胞株を各々培養し、パパニコロウ染色したもの。

ので、細胞標本染色法の 1 つであるパパニコロウ染色したものである。この癌は形態が正常なものほとんど変わらず、癌診断が難しいとされているものの代表例である。

これらのラマン分光画像を細胞核、ミトコンドリア領域などに分類して、それらのラマン分光画像の集合をファジークラスタリングした結果、形態情報だけでは診断ができない癌もラマン分光イメージングの情報を用いることで高い確率で診断できること、ならびにオルガネラ単位でその予測性能が有意に異なることなどが判明した。

4. おわりに・展望

近年、パーシスタ細胞[Balaban 2004, 若本 2013]のように、細胞集団のなかには遺伝子上の違いではなく表現型の違いによって獲得された(抗生物質耐性などの)機能をもつ細胞、リーダー・フォロアー細胞のように集団の中から役割分担[Reffay2014]をする細胞が存在することが明らかとなってきた。このような細胞における「集団と個」の問いは、酵素分子が複数の形を取り得、その形に依存して分子認識のし易さが変化する分子個性(Englishら Nature Chem. Biol. 2006)の細胞版に当たるが、表現型の違う細胞群を如何に分類し得るかという挑戦的な問いに応え得るデータ科学は未だ確立されていない。シグナル・ノイズ比が低いラマン分光データからの細胞病理、計測迅速化のほか、細胞個性の学術的な問いは本研究課題のカバーし得る重要問題である。

参考文献

- [Palonpon2013] Palonpon A. F., et al., **Nat. Protocols**, **8**, 677 (2013)
- [市村 2014] Ichimura T., Chiu L.-d., Fujita K., Kawata S., Watanabe T. M., Yanagida T., Fujita H., **PLOS ONE**, **9**, e84478 (2014)
- [Taylor2015] J. N. Taylor, C.-B. Li, D. Cooper, C. F. Landes, T. Komatsuzaki, **Scientific Reports** **5**, 9174 (2015)
- [Li2015] C.B. Li, H. Ueno, R. Watanabe, H. Noji, T. Komatsuzaki, **Nature Communications** **6**(10223), (2015)
- [Balaban2004] N.Q. Balaban, J. Merrin, R. Chait, L. Kowalik, S. Leibler, **Science**, **305**, 1622-1625 (2004)
- [Wakamoto2013] Y. Wakamoto, N. Dha, R. Chait, K. Schneider, F. Signorino-Gelo, S. Leibler, J.D. McKinney, **Science**, **339**, 91-95 (2013)
- [Reffay2014] M. Reffay, M. C. Parrini, O. Cochet-Escartin, B. Ladoux, A. Buguin, S. Coscoy, F. Amblard, J. Camonis & P. Silberzan, **Nature Cell Biology**, **16**, 217-223 (2014)

1.