

仮説検定法を利用した MS Imagingデータ解析ソフトの開発

(株式会社島津製作所 基盤技術研究所)

○梶原茂樹、森永浩子、小河潔

Development of the Software Using Hypothesis Testing for MS Imaging Data

○S. Kajihara, H. Morinaga, K. Ogawa
(Technology Research Laboratory, Shimadzu Corporation)



Short Abstract

The preliminary studies are conducted to apply the hypothesis testing for MS imaging. This result suggests that t-test is useful to find the common or different peaks in two regions.

はじめに

質量分析 (MS) イメージングを取得した場合、測定点ごとにMSスペクトルを持つことになり、極めて膨大なデータとなる。この膨大なデータから人間が目で見ても有意な情報を引き出すことは多大な時間と労力を必要とする。このため、多変量解析などの統計的手法を利用して、自動的に膨大なデータから有意な情報を得るためのソフトウェアが開発されている。

今回、MSイメージングにおいて、ROI (Region Of Interest) で指定した2つの領域内で同じ強度、あるいは異なる強度のピークを見つけるため、t検定を利用する方法を検討したので報告する。

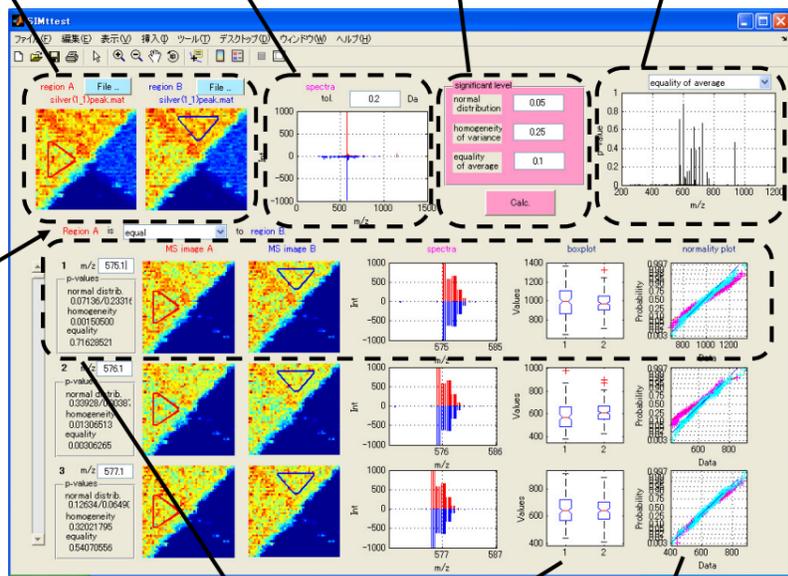
方法

あらかじめ撮影した光学画像、あるいはTIC (total ion current) によるMS画像上で、比較対象の2つの関心領域 (ROI) を指定し、各ROIに含まれるピクセル群において、同一のm/zのピーク強度に違いがあるかを調べる。このため、「領域A内のピクセル群と領域B内のピクセル群において、同じm/zで観測されるピーク強度の平均値の差はゼロである」という帰無仮説に対する検定を行う。ここでは、MSスペクトルのピーク強度は正規分布に従うと仮定した場合の検定方法を説明する。

まず、各領域で特定のm/zにおけるピーク強度の母集団が正規分布に従うかを検定する。このため、Lilliefors検定を利用し、有意水準5%で両側検定を行う。次に、2つの領域の両方で、「正規分布している」という帰無仮説が棄却されなかったm/zにおいて、等分散性の検定を行う。すなわち、「2つの領域の母分散は同じである」という帰無仮説に対し、検定統計量F値を使い、有意水準0.25で両側検定を行う。最後に、平均値の差について、等分散性が棄却されない場合はt検定、等分散性が棄却された場合はWelchのt検定を、有意水準0.1の両側検定で行う。

開発ソフトウェアの概観

- ① 2つのデータファイルを選択する。TICによる画像が表示される。各画像上で、マウスにより検定対象領域 (ROI) を指定する。
- ② ROI内部のピクセルの平均スペクトルが表示される。縦軸は、ROI A (赤) がプラス、ROI B (青) がマイナスで表示される。
- ③ 正規分布の検定、等分散性の検定、平均値の差の検定について、有意水準を入力する。Calc. ボタンを押すと、検定計算を実行する。
- ④ 各m/zにおけるp値の計算結果が表示される。リストメニュー選択により、正規分布、等分散性、平均値の差に対する計算結果を切り替えて表示する。



- ⑤ リストメニューにより「ROI AとROI Bの平均値が等しい」、「ROI Aの平均値はROI Bより大きい」、「ROI Aの平均値はROI Bより小さい」のいずれかを選択する。
- ⑥ 有意な結果を示すm/zでのMS画像、平均スペクトル拡大図、箱髷図、正規確率プロット図が、セットで表示される。
- 箱髷図：中央値、中央値のばらつき、第1四分位点と第3四分位点、異常値などを表し、検定とは違う角度で、統計値の視覚化に役立つ。2つのボックスプロットのノッチ (へこんだ部分) が重ならない場合、5%の有意水準で異なる中央値をもつと言える。
- 正規確率プロット図：データの分布が正規分布しているかどうかを目視で判断するためのグラフであり、検定とは違う角度で、正規分布性を立証したり警告を与える。すべてのデータ点が直線の近くにある場合、正規分布しているという仮定が正しいことを示す。

結果

インクジェットプリンター (Calario™ PX-G5000@EPSON) でカラー図形を印刷し、MSイメージングを測定した。図1Aに光学画像、図1BにTICによるMS画像を示す。カラー図形は、シアンインクで左三角形 (▽)、マゼンタで右三角形 (△) を描画したものであり、上部の領域 (▽) では2色が混合されている。図1Bに検定対象として設定した2つのROIを示す。ROI Aはシアン単色領域、ROI Bはシアンとマゼンタの混合領域である。ROI A、ROI B内部のピクセル群の平均MSスペクトルを図1Cに示す。

正規性が棄却されないm/zにおいて、ピーク強度平均値が等しいことを否定できないものとしてm/z575の同位体ピーク群が見つかった (図2A)。また、ピーク強度平均値が等しいとは言い難いピークとしてm/z1152が見つかった (図2B)。この結果は、m/z575とm/z1152はシアンインクの成分であるとともに、m/z1152はマゼンタインクとの混合により、イオンサブレーションが起きたことを示唆している。

さらに、マゼンタ単色領域にROI Cを設定し、ROI Bとの比較を行った。その結果、ピーク強度平均値が等しいとは言い難いピークの1つとしてm/z647が見つかった (図3A)。なお、PCAで見つかったm/z387のピークについて調べると、強度が正規分布していないことがわかった (図3B)。

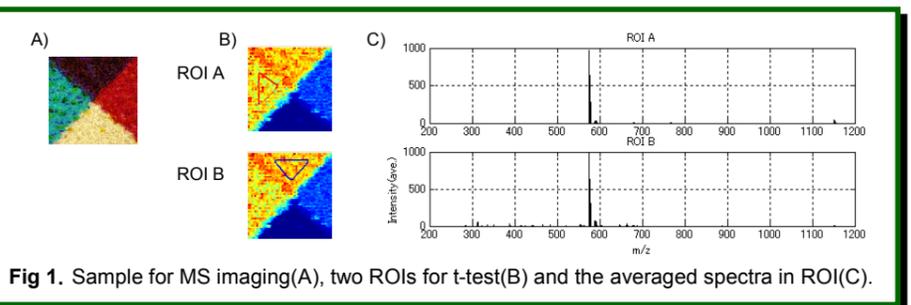


Fig 1. Sample for MS imaging(A), two ROIs for t-test(B) and the averaged spectra in ROI(C).

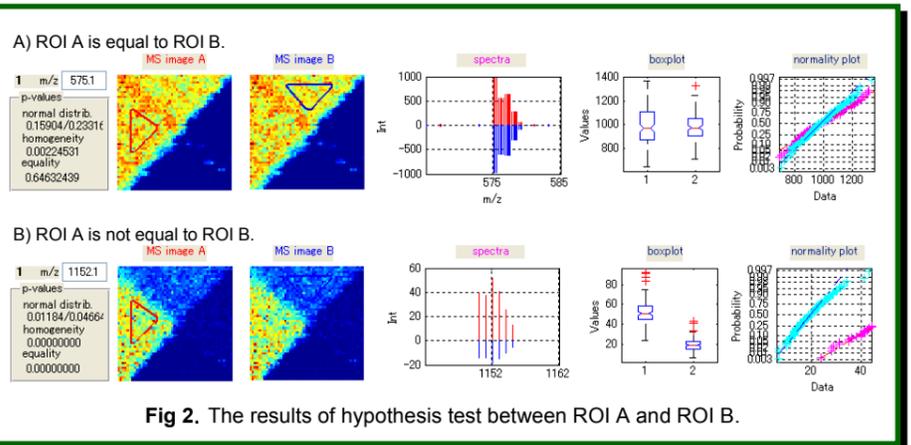


Fig 2. The results of hypothesis test between ROI A and ROI B.

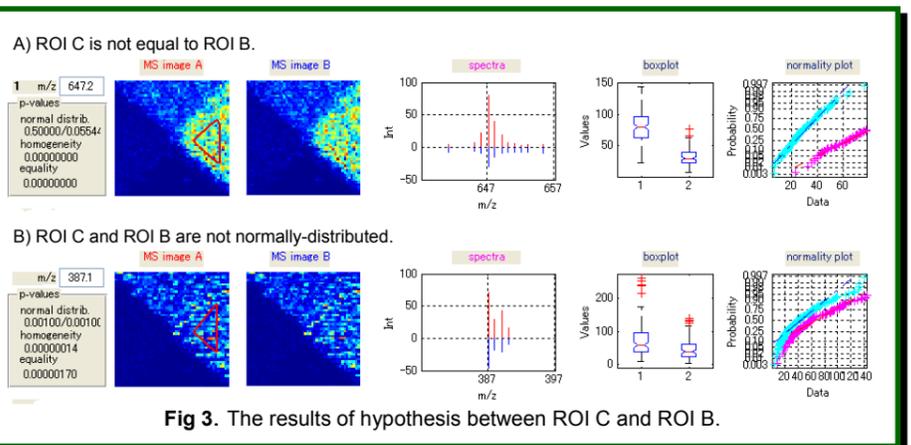


Fig 3. The results of hypothesis between ROI C and ROI B.

結論

MSイメージングにおいて、t検定を利用するソフトウェアを開発し、その有効性を確認した。このような方法は、従来、主観的に議論されていたMSイメージングの結果を、統計量という客観的指標により説明できるという点でも意義深いと考える。

今後、正規分布を仮定しないU検定の利用法も検討する予定である。