SHIMADZU
Excellence in Science

Application News

No. M307

GC-MS　GCMS-QP™2020 NX, GCMS-TQ™ NX シリーズ

脂肪酸による判別分析モデルの構築

中川 裕貴

ユーザーベネフィット

- ◆ LabSolutions Insight™から機械学習ソフトへ簡便にデータを移動し判別モデルを作成可能
- ◆ 機械学習ソフトはスタンドアローンなためWebにあげる必要がなくデータセキュリティーが高い
- ◆ 新たなサンプル群を発見・分類したり分類の決め手となる化合物を判断

■はじめに

国立研究開発法人農業・食品産業技術総合研究機構と島津製作所は「食」の機能性成分解析を目的とした共同研究により農産物や食品に含まれる機能性成分の簡便で迅速かつ正確な分析手法の開発を行っています。脂肪酸は栄養成分であるとともに機能性成分として高度不飽和脂肪酸や中鎖脂肪酸などが知られており、穀類、豆類などの農産物の機能性成分の一斉分析を行う際には脂肪酸にも注目する必要があります。脂肪酸についての多変量解析は食用油の識別などで活用されているが、ここでは、農産物1品目の品種・系統の違いを形質以外でグルーピングできるのではないかと考えて解析を実施しました。

Application News M306「GC/MSによる高速脂肪酸分析」で紹介した方法により、大麦48サンプルを用いてグリセリド（トリグリセリド、ジグリセリド、モノグリセリド、レシチンなど）、遊離脂肪酸、ステロールエステルなどの脂肪酸高速分析を実施し既知3群のいずれかに未知サンプルを判別するモデルを作成しました。

判別モデル作成にはLabSolutions InsightとOrange Data Mining（University of Ljubljana）を用いました。

■LabSolutions Insight

判別モデル作成のために、まずLabSolutions Insightを使用して大麦48サンプルを解析しました。

飽和しているピークは面積比較できないので波形処理せず、後で欠損値として消すために面積値0としました。LabSolutions Insightではバッチ内のサンプルを一度に波形処理できることから48サンプルと多検体にも関わらず約30分で解析は完了しました。

サンプル間のマトリックスの差は非常に少ないと予想されることから、検量線で定量された精密な濃度値ではなく面積値を最終結果としてcsvの形で出力しました。

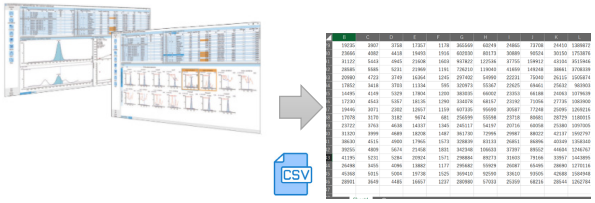

表 1　測定化合物　FAME 37（Merk Millipore、P/N：CRM48775）

Abbreviation	Common Name (Methyl Derivative)
C4:0	Methyl butyrate
C6:0	Methyl hexanoate
C8:0	Methyl octanoate
C10:0	Methyl decanoate
C11:0	Methyl undecanoate
C12:0	Methyl laurate
C13:0	Methyl tridecanoate
C14:0	Methyl myristate
C14:1(9c)	Methyl myristoleate
C15:0	Methyl pentadecanoate
C15:1(10c)	Methyl cis-10-pentadecenoate
C16:0	Methyl palmitate
C16:1(9c)	Methyl palmitoleate
C17:0	Methyl heptadecanoate
C17:1(10c)	cis-10-Heptadecanoic acid methyl ester
C18:0	Methyl stearate
C18:1(9t)	trans-9-Elaidic acid methyl ester
C18:1(9c)	cis-9-Oleic acid methyl ester
C18:2(9t,12t)	Methyl linolelaidate
C18:2(9c,12c)	Methyl linoleate
C18:3(6c,9c,12c)	Methyl γ-linolenate
C18:3(9c,12c,15c)	Methyl linolenate
C20:0	Methyl arachidate
C20:1(11c)	Methyl cis-11-eicosenoate
C20:2(11c,14c)	cis-11,14-Eicosadienoic acid methyl ester
C21:0	Methyl heneicosanoate
C20:3(8c,11c,14c)	cis-8,11,14-Eicosatrienoic acid methyl ester
C20:4(5c,8c,11c,14c)	cis-5,8,11,14-Eicosatetraenoic acid methyl ester
C20:3(11c,14c,17c)	cis-11,14,17-Eicosatrienoic acid methyl ester
C22:0	Methyl behenate
C20:5(5c,8c,11c,14c,17c)	cis-5,8,11,14,17-Eicosapentaenoic acid methyl ester
C22:1(13c)	Methyl erucate
C22:2(13c,16c)	cis-13,16-Docosadienoic acid methyl ester
C23:0	Methyl tricosanoate
C24:0	Methyl lignocerate
C24:1(15c)	Methyl nervonate
C22:6(4c,7c,10c,13c,16c,19c)	cis-4,7,10,13,16,19-Docosahexaenoic acid methyl ester

■ データ前処理

Orange Data Miningは多変量解析や機械学習などが行えるスタンドアローンの統計ソフトです。

csvデータを機械学習ソフトOrange Data Miningに読み込ませた後、主成分解析などの多変量解析や機械学習のモデルを作成する前にデータの前処理をする必要があります。

この実験では、以下のような前処理を行いました。

① 飽和化合物の消去
② 欠損値（面積値0）を1サンプルでも含む化合物の消去
③ 正規分布になっていない化合物の消去（図1）
④ 面積値が正か負の相関関係のある2つの化合物の面積値を平均して1つの新たな化合物作成（図2）

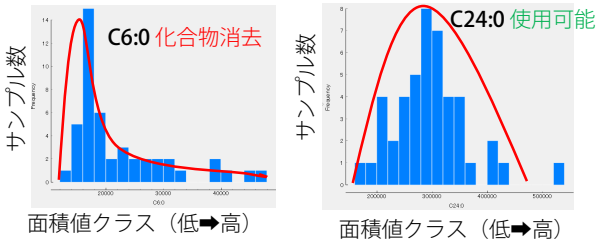

図 1　正規分布の確認

化合物の数が37と限られていたので、p値（>0.05以上の化合物）による化合物除去、化合物の種類（TCA回路の化合物のみ、高沸点化合物のみ、など）による化合物除去などは行いませんでした。外れ値を生むサンプルや飽和を含むサンプルは除去しました。

また最終の正規化はソフトが自動で行いました。

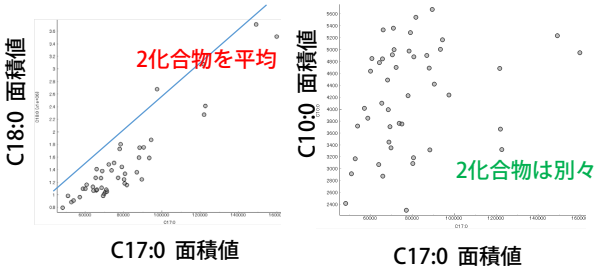

図 2　相関関係にある2つの化合物検索

■ 主成分分析

分析された48サンプルのデータを前処理し、主成分分析で可視化しました。PC1とPC2の累積寄与率は約80 %でした。

PC1は中鎖と長鎖脂肪酸のあり（右）なし（左）で左右に分かれることから「中・長鎖脂肪酸」と名付けました。PC2は短鎖脂肪酸が多い（上）と少ない（下）で上下に分かれることから、「短鎖脂肪酸」と名付けました。

マイナスに引く化合物が少なかったことと、変数が脂肪酸で絞られていたことで左右上下を「多い」「少ない」で示すことができました。

図3の通り3つのサンプル群に分かれることが確認できました。

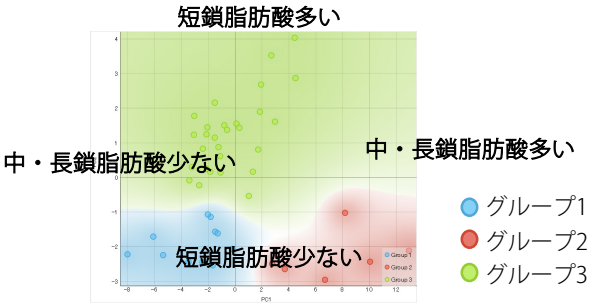

図 3　PCA

■ 判別モデル

csvデータを機械学習ソフトOrange Data Miningに読み込ませデータの前処理を完了し主成分分析などの多変量解析によって群を確認した後、判別モデルを作成しました。

Orange Data Miningに「Tree」というボタンがあるのでクリックすると判別モデルが自動的に作成されます（図4）。AUCが0.810の判別精度の高いモデルが作成されました。

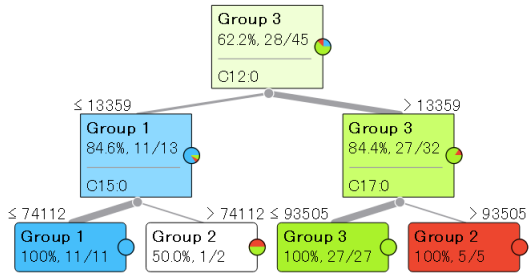

図 4　Decision Tree

■ 未知サンプルの判別

3つの未知サンプル（実験のため実際にどの群に属するか実験書では既知）を分析し、判別モデルで群を判断したところ、サンプル1と3が第2群、サンプル2が第3群と正しい結果になりました。

表 2　未知サンプルの判別結果

	classification	データファイル名	C6:0	C8:0	C10:0	
1	Group 2	Unknown 1	26175	4767	4681	1649
2	Group 3	Unknown 2	28901	3649	4485	1665
3	Group 2	Unknown 3	46260	5417	5162	2317

■ まとめ

GCMS-QP2020 NXを用いて脂肪酸の高速分析法を開発しました。

開発した高速分析方法によって大麦サンプルの脂肪酸分析を行いLabSolutions Insightと機械学習ソフトを用いて判別モデルを作成しました。高い精度のモデルが作成され、未知サンプル3点も正しく判別できました。

本実験を進めるにあたり国立研究開発法人農業・食品産業技術総合研究機構の山本万里先生、十一浩典研究員、市来弥生研究員から多大な助言を賜りました。厚く感謝を申し上げます。

株式会社 島津製作所 分析計測事業部 グローバルアプリケーション開発センター

初版発行：2020 年 12 月

島津コールセンター 0120-131691