

Agilent GC/MSD、Mass Profiler Professional、サンプルクラス予測モデルを用いた焼酎の汚染物質の検出

アプリケーションノート

食品分析・農芸化学

著書

Takeshi Serino
Agilent Technologies, Inc.
Santa Clara, CA
USA

概要

サンプルクラス予測 (SCP) モデルは、きわめて複雑なサンプルの質量分析データを用いて、汚染物質などの分類を予測することもできる強力なツールです。このアプリケーションノートでは、SCP を用いて、焼酎サンプル中の汚染物質を検出および分類するメソッドを開発しました。このメソッドは、焼酎およびその他の飲料製品などの製造プロセスの品質保証 (QA) および品質管理 (QC) に適用することができます。

はじめに

焼酎は、16 世紀から日本でつくられている蒸留酒です。一般に、アルコール濃度は 25 % ほどで、米、芋、黒糖などを 1 回または複数回蒸留して製造されます。ワインや一般的な酒類とは異なり、焼酎では酵母ではなく麴による発酵が用いられます。日本では 2000 年代はじめに焼酎ブームが起き、女性を中心とする若者に人気の酒となっています。焼酎には、血栓症、心臓発作、糖尿病の防止などの健康上の効果があるとされています。そのため、日本における焼酎の消費量は、国税庁統計年報書によると、2010 年の段階で日本酒を上回っています。

こうした商業上の価値を備えた焼酎の品質管理は、消費者の満足度を保つうえで重要となります。焼酎の製造過程では、機械油やゴム手袋などに起因する汚染が生じるおそれがあると予想されます。こうした汚染は、味や香りに影響を与えるため、メーカーにとっては売上を脅かすものになります。ブランドの評判を守るためには、製品が瓶詰めされ、出荷されるまえに、迅速に汚染を見極めることが不可欠です。



Agilent Technologies

このアプリケーションノートでは、発酵および瓶詰めプロセスにおける焼酎の汚染を検出できる判別予測モデルの可能性について検討しました。この判別予測モデルでは、ワインの分類 [1] やエキストラバージンオリーブ油の感覚テストの合否判定 [2] に用いられたアプリケーションノートと同様のノンターゲット化合物分析手法が使われています。LTM (Low Thermal Mass : 低熱質量) カラムを搭載した Agilent 7890 GC システムと、シングル四重極 Agilent 5975 GC/MS システムを使用し、サンプルを測定しました。GC/MS のデータを比較したところ、トータルイオンクロマトグラム (TIC) 上では、大部分のサンプル条件間で包括的に違いを見つけることは容易ではありませんでした。こうした違いを明らかにするため、NIST の自動質量スペクトルデコンボリューション 同定ソフトウェア (AMDIS) を用いて GC/MS のデータをデコンボリューションして、多変量解析や判別予測モデル作成機能を搭載した Mass Profiler Professional (MPP) を用いて解析しました。

汚染されていない焼酎サンプルと、ゴムおよび機械油により意図的に汚染させたサンプルを分析しました。測定した GC/MS データを 3 種類の方法でフィルタリングし、エンティティと呼ばれる化合物情報 (化合物のリテンションタイム、質量スペクトル、各サンプルのピーク強度を含む情報) を作成しました。その後、各フィルタリングによりノイズとなるエンティティを除去して、サンプル間で有意差のあるエンティティを用いて、サンプルクラス予測 (SCP) モデルを作成しました。このモデルを別のバッチで意図的に汚染したデータに適用して、判別分析モデルの汚染物質の検出能力について検証しました。今回検討を行ったデータセットにおいては、3 種類いずれかの方法でフィルタリングして得られたデータに決定木 (Decision Tree) モデルを適用したところ、モデルの作成に用いなかったサンプル中の汚染物質の存在を 100 % の精度で判別予測することができました。サポートベクターマシン (SVM) モデルでは、分散分析 (ANOVA) と 45 % 変動係数 (CV) によりデータをフィルタリングした場合のみ 100 % の精度で判別予測することができました。

実験手法

サンプル

汚染のない焼酎サンプル、および既知量の洗剤、殺虫剤、ゴム手袋、機械油を添加した 8 種類のサンプルを作成しました。サンプル GA と GB については、種類の異なる市販のゴム手袋の破片 20~23 mg を焼酎 1 g に添加しました。サンプル DA については、塩素系洗剤 10 mg を焼酎 1 g に添加し、サンプル DC については、焼酎 1 g に殺虫剤を噴霧 (スプレー 2 回) しました。サンプル OB については、機械油 10~15 mg を焼酎 1 g に添加しました。サンプル名と供給元を表 1 にまとめています。

表 1. 分析したサンプルの種類

サンプル名	由来	汚染物質
IOSK	大阪	なし
ITKO	東京	なし
IUSA	サンノゼ	なし
DA	大阪	塩素系洗剤
DC	大阪	殺虫剤
GA	大阪	ゴム手袋 A
GB	大阪	ゴム手袋 B
OB	大阪	機械油 B

装置

この研究では、低熱質量 (LTM) モジュールを搭載した 7890 GC システムと、5975 GC/MS システムを使用しました。機器の設定条件を表 2 に示しています。Gerstel MPS2 オートサンブラを用いて、固相マイクロ抽出 (SPME) サンプル前処理をおこないました。

表 2. GC および質量分析条件

GC 分析条件	
プレカラム	なし
分析カラム	10 m × 0.18 mm、0.18 μm DB-1ms LTM カラムモジュール (p/n 100-2000LTM)
注入メソッド	SPME (Supelco 57341-U)、1 cm 注入
注入口温度	240 °C
注入モード	1.52 分スプリットレス、72 kPa
オープン温度	GC オープン: 200 °C で 11.83 分維持 (等温) LTM モジュール: 35 °C で 120 秒維持 30 °C/min で 35 °C~240 °C 240 °C で 3 分維持
カラム流量	1.1 mL/min 定流量
キャリアガス	ヘリウム
トランスファーライン温度	240 °C
GC 分析時間	11.83 分
MS 条件	
イオン化モード	EI
イオン源温度	230 °C
取り込みモード	スキャン (35~450 amu)
微量イオン検出	オン
チューニング	atune.u

サンプル前処理

各サンプルの揮発性成分を、SPME によりサンプリングしました。各焼酎サンプルを 20-mL ヘッドスペースバイアルに入れました。100 μm \times 1 cm ポリジメチルシロキサン (PDMS) SPME ファイバー (Supelco 57341-U) を 40 $^{\circ}\text{C}$ で 40 分間、攪拌なしでサンプルのヘッドスペースに曝露しました。SPME ファイバーに吸着した揮発性化合物を、240 $^{\circ}\text{C}$ で 1.5 分間、注入口に熱脱離させました。各注入後、ベイクステーションでファイバーを 260 $^{\circ}\text{C}$ で 5 分間ベイクアウトしました。

データ処理と統計解析

GC/MS データからの化合物抽出には、Agilent MSD Productivity Chemstation (E.02.02)、および NIST の AMDIS を使用しました。AMDIS の ELU ファイルを MPP にインポートし、サンプル間で有意差のある化合物を見つける操作をおこないました。

データのフィルタリングと統計解析には Mass Profiler Professional (B.02.02) を、サンプルクラス予測 (SCP) モデルの作成には Agilent Sample Class Predictor を使用しました。データ処理のステップを図 1 に示しています。

- サンプル全体の化合物ピークのフィルタリングおよびアライメント
- エンティティのフィルタリング
- 主成分分析 (PCA)
- 階層的クラスター分析 (HCA)
- 予測モデルの作成

結果と考察

データ採取

焼酎サンプルを分析し、GC/MSD で検出可能な化合物を調べました (図 2)。AMDIS を用いて、GC/MS データから成分を抽出しました。得られたデータの内訳は、ゴム、洗剤、機械油を添加した焼酎の各 4 回の繰り返し測定結果、大阪で販売されていた汚染のない焼酎の 23 回の繰り返し測定結果、東京で販売されていた焼酎の 18 回の繰り返し測定結果、カリフォルニア州サンノゼで販売されていた焼酎の 13 回の繰り返し測定結果です。全体として、クロマトグラフィデコンボリューションにより、330~380 のピークが検出されました。トータルイオンクロマトグラム (TIC) では、大部分のサンプル条件において、ほとんど違いが見られませんでした。サンプル間の違いを明らかにするためには、さらなるデータ処理が必要でした。



図 1. GC/MSD をもとにした焼酎中汚染物質の予測モデル作成の統計解析ワークフロー

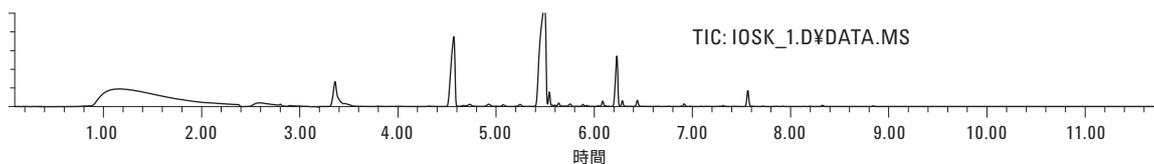


図 2. 汚染されていない焼酎 (この図は IOSK サンプル) の典型的なトータルイオンクロマトグラム (TIC)

データのフィルタリング

データセットを Mass Profiler Professional (MPP) ソフトウェアにインポートし、スペクトルの類似性およびリテンションタイムのウィンドウ幅をもとに、各 TIC に含まれるデコンボリューションされた化合物ピークのアラインメントを実施し、2,376 化合物のエンティティが作成されました。これらのエンティティに対して、各種サンプルの判別予測モデルで検証する 3 種類のエンティティリストを作成し、それぞれ検証しました (図 3)。

Frequency によるフィルタ (フィルタ 1) では、多くの化合物が 1 つのサンプルに固有のものであることがわかりました。実際には、226 エンティティがこのフィルタを通過しました。確率 p 値を .05 (当該エンティティが有意である確率が 95 %) として一元配置分散分析 (ANOVA) フィルタ (フィルタ 2) を用いたところ、1,080 のエンティティが得られました。第 3 のフィルタでは、一元配置 ANOVA フィルタを通過したエンティティを選択し、すべてのサンプルについて変動係数 (CV) フィルタを 45 % 未満に設定しました。ANOVA と $CV < 45 %$ を用いたのは、意図的に強力なフィルタを作成し、その結果得られる SCP モデルの精度における影響を調べるためです。

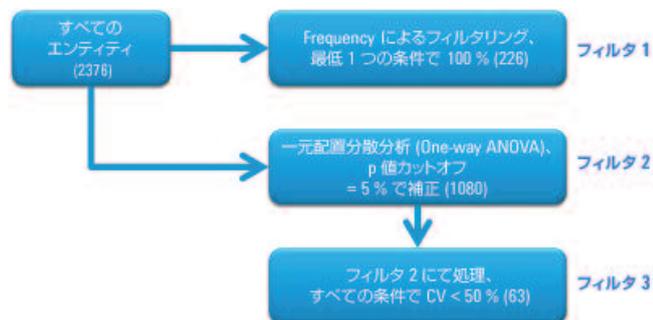


図 3. エンティティのスクリーニングと SCP モデルの作成に用いた 3 つのデータ処理フィルタ

主成分分析

3 つのフィルタで得られたエンティティリストを用いて、主成分分析 (PCA) をおこないました。PCA スコアプロットを図 4 に示します。すべてのサンプルをもっとも良く分離でき、もっとも正確な SCP モデルを作成できるのは、ANOVA + 45 % CV フィルタ (フィルタ 3) であることがわかります。

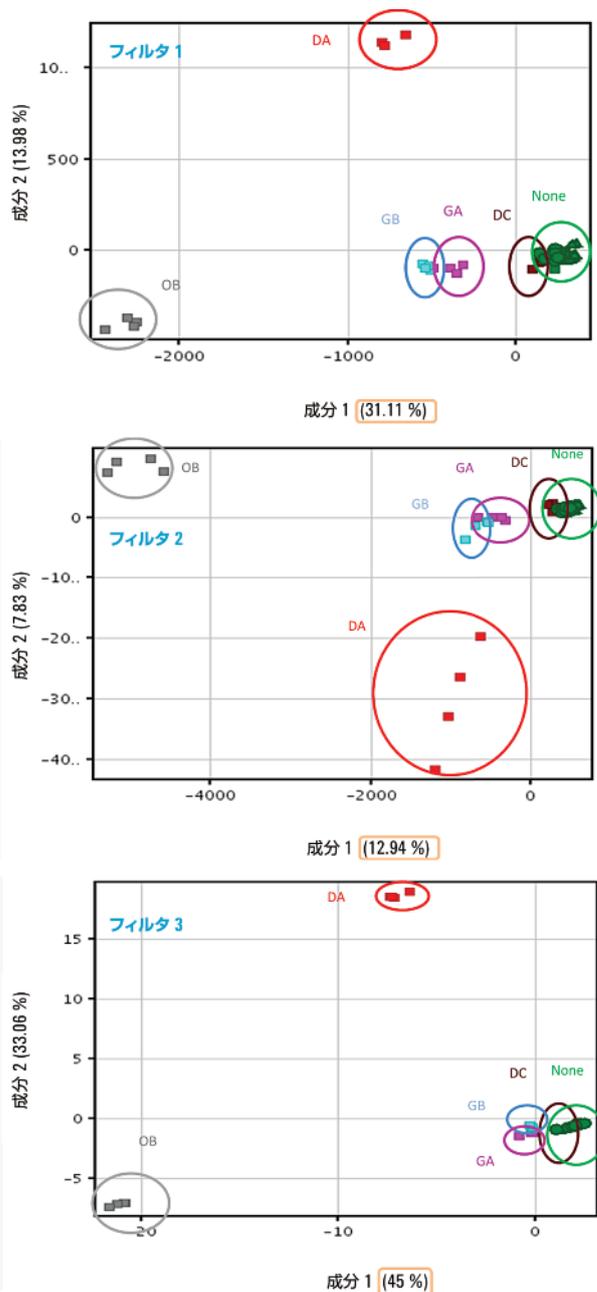


図 4. 主成分分析 (PCA) スコアにより、各サンプルをもっとも良好に分離できるのは、フィルタ 3 (ANOVA + 45 % CV) であることがわかります。

階層的クラスター分析 (HCA)

クラスター分析は、各化合物のピーク強度変動の類似性をもとにエンティティとサンプルをクラスターで分類する手法です。階層的クラスタリング (HCA) は、もっともシンプルなクラスター分析の1つで、GC/MS のクラスター分析などに広く用いられています。HCA は、もっとも類似性の高いアバンドンスプロフィールを組み合わせ、1つのグループを作成します。それらをさらに組み

合わせてピーク強度変動の類似度のツリー構造を作成し、最終的にすべてのデータが1つのグループを形成するようにします。フィルタ3で生成されたエンティティのHCAでは、汚染されていないサンプルと汚染されているサンプルを区別することができた33の化合物が特定されました(図5)。Wiley 9th/Nist08 MS ライブラリを用いて、これらの汚染物質のライブラリ検索をおこなったところ、そのうちの8つの化合物が同定されました(表3)。

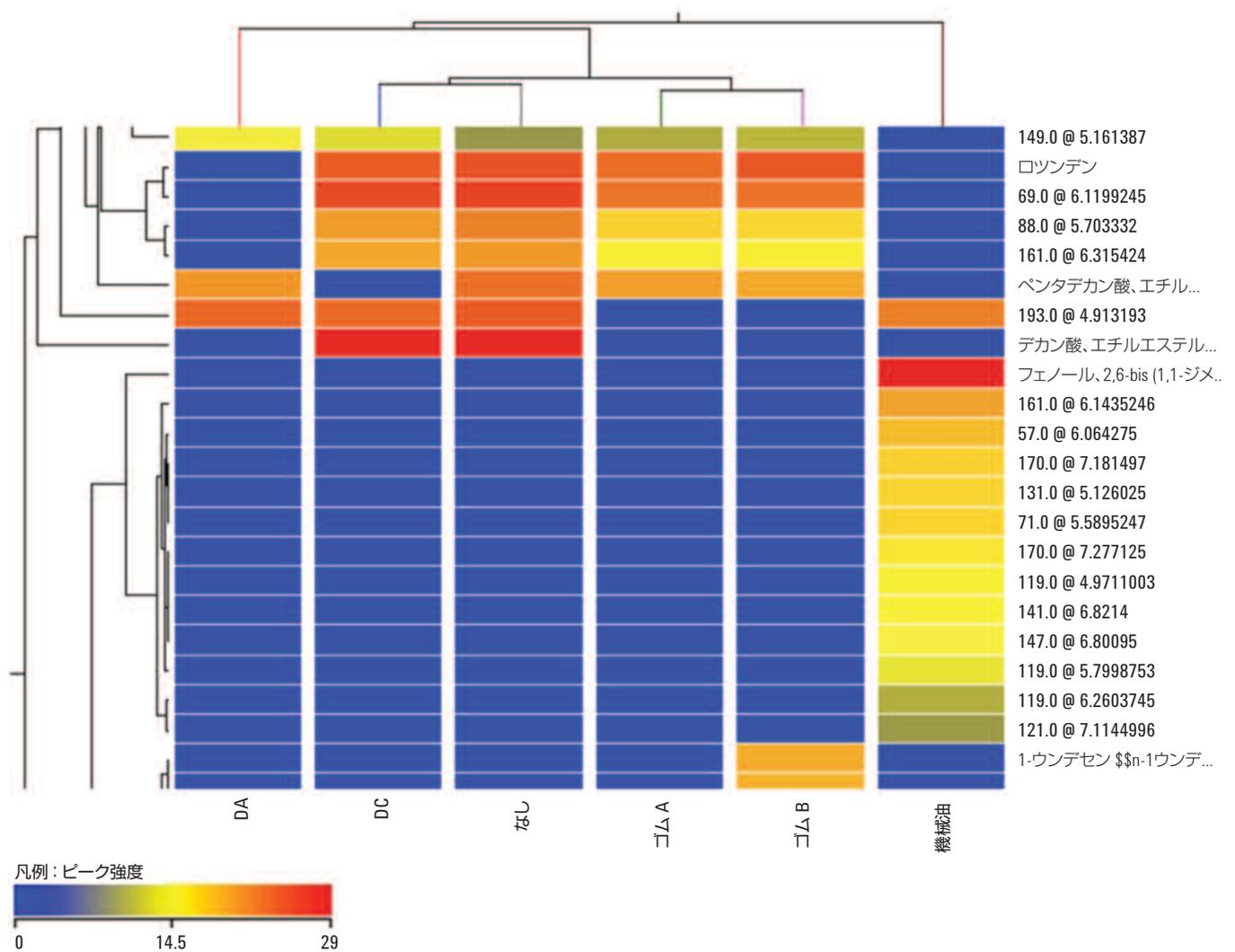


図5. 各種サンプルで検出された関連化合物に関する階層的クラスター分析 (HCA) ヒートマップの拡大図。このマップでは、汚染されていないサンプルと汚染されているサンプルを区別する33化合物のうち、23化合物を示しています。マップの色は、各条件における各化合物のピーク強度を示しています: DA、なし (汚染されていない機械油)、DC、その他。ピーク強度の低い化合物は青で表示、中程度の強度の化合物は黄色、強度の高い化合物は赤で表示して、各化合物の相対強度を示しています。

表 3. 同定された汚染物質

サンプルクラス	化合物
機械油	フェノール、 2,6-bis(1,1-ジメチルエチル)-4-メチル- (CAS)
ゴム B	2-イソプロピル-5-メチル-1-ヘプタノール 1-ウンデセン
洗剤 A	ヘプタデカン酸、エチルエステル 1-ノナデセン 1-デカノール (CAS)
洗剤 C	ベンズアミド、N,N-ジエチル-3-メチル- 1,1'-ビフェニル、3-(1-メチルエチル)-

判別予測モデル

クラス分類の目的は、測定によって得られた結果をもとに、一般的な仮説を立てることにあります。これらの結果は、複数の変数で説明され、クラス情報に対応する既知のラベルで特定されます。クラス分類では、複数の変数から結果をマッピングする方法を確立する必要があります。そのための手法として、統計的手法や人工知能を用いた多くの手法が開発されています [3]。このアプリケーションノートにおける目標は、汚染と関連することが示された 33 の化合物のピーク強度をもとに、焼酎サンプルの汚染の有無を判断することです。

5 種類のアルゴリズム (決定木 (DT)、サポートベクターマシン (SVM)、ナイーブベイズ (NB)、ニューラルネットワーク (NN)、PLS 判別分析 (PLS-DA)) を検証し、汚染物質のスクリーニングにもっとも適したアルゴリズムを検討しました。3 つのフィルタにより得られたデータセットを用いて、各アルゴリズムをテストしました。

DT アルゴリズムでは、木のように配置された条件選択シーケンスが用いられます。決定木の適切な経路をたどっていくことで、サンプルを分類します。SVM アルゴリズムは、サンプルを空間中の点と考え、2 クラスの点を隔てる (分類する) 平面を決定することで、サンプルを各クラスに分類するものです。NB アルゴリズムでは、任意のクラスの特性による影響が、他の特性の価値とは独立しているものと仮定されます。この仮定は、条件付き独立と呼ばれます。ニューラルネットワークは、生物学的な神経ネットワークの構造や機能にヒントを得た並列システムで、直線的な演算では不可能なパラダイムを解くことができます。PLS 判別分析は、測定された変数よりも観測値が少ない場合に特に適用されます。クラス間の最大距離を得る場合など、観測値グループ間の区分を際立たせるために用いられます。

分類モデルの構築の最初のステップは、データによりモデルを調整することです。これには、5 つのモデルアルゴリズムと 3 つのデータフィルタを使用、同じデータで各モデルの有効性を検証しました。これらのデータについて、3 つのデータフィルタを用いた場合の 5 つのモデルの予測精度を表 4 に示しています。SVM および DT アルゴリズムで、3 つのフィルタすべてについて 100 % の精度が得られました。

表 4. 調整後の SCP モデル精度 (%)

アルゴリズム	フィルタ 1	フィルタ 2	フィルタ 3
決定木	100%	100%	100%
単純ベイズ	99	100	99
ニューラルネットワーク	69	73	69
部分最小二乗判別分析	90	90	90
サポートベクターモデル	100	100	100

第 2 のステップは、未知サンプルのデータで各モデルをテストします。このテストではモデル作成に用いなかった別の 12 サンプルを使用しました (表 5)。これらのサンプルを用いた未知汚染物質の予測では、DT モデルの堅牢性が高く、3 つのフィルタすべてで 100 % の精度が得られました (表 5)。それに対して、SVM モデルでは、100 % の精度が得られたのは ANOVA + 45 % CV フィルタ (フィルタ 3) のみでした。この結果は、一部のエンティティが予測モデルに大きな影響を与えることを示唆しています。PLSDA モデルでは、汚染サンプルの同定の効率がきわめて悪くなりました。各モデルアルゴリズムで判定精度が異なることから、任意のデータセットに適用する場合は、適切なフィルタおよびアルゴリズムを検討する必要があることを示しています。

表 5. 未知サンプルの SCP モデル精度 (%)

アルゴリズム	フィルタ 1	フィルタ 2	フィルタ 3
決定木	100%	100%	100%
単純ベイズ	92	100	92
ニューラルネットワーク	50	50	50
部分最小二乗判別分析	50	42	33
サポートベクターモデル	83	83	100

結論

サンプルクラス予測 (SCP) は、焼酎の品質を確認するための有効な手法となります。GC/MSD データの多変量解析を用いれば、サンプル間の小さな違いを明確に見極めることができます。

もっとも予測精度の高い SCP モデルを作成するためには、高いデータ品質が不可欠です。データの品質が高ければ、サンプルに適したフィルタリングおよび精度の高い予測モデルを構築することができます。SCP でもっとも良い結果が得られるのは、サンプルデータが適切にフィルタリングされ、適切な予測アルゴリズムを用いた場合です。複数の予測モデルを使えば、任意の分析における各予測モデルの検証やカスタマイズが可能になります。質の高いエンティティリストがあれば、より優れた SCP 予測モデルを構築し、食品分析の品質保証 (QA) および品質管理 (QC) のワークフローを向上させることができます。

参考文献

1. L. Vaclavik, O. Lacina, J. Hajslova, J. Zweigenbaum. "The use of high performance liquid chromatography-quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety." *Anal Chim Acta*.**685**, 45-51 (2011).
2. S. Baumann and S. Aronova. 「Agilent GC/Q-TOF MS および Mass Profiler Professional ソフトウェアによるオリーブオイルのキャラクタリゼーション」、アジレント・テクノロジー、アプリケーションノート、5991-0106JAJP.
3. J. Boccard, J. L. Veuthey, S. Rudaz. "Knowledge discovery in metabolomics: an overview of MS data handling." *J Sep Sci*.**33**, 290-304 (2010).

詳細情報

アジレント製品とサービスの詳細については、アジレントのウェブサイト www.agilent.com/chem/jp をご覧ください。

www.agilent.com/chem/jp

アジレントは、本文書に誤りが発見された場合、また、本文書の使用により付随的または間接的に生じる損害について一切免責とさせていただきます。

本資料に記載の情報、説明、製品仕様等は予告なしに変更されることがあります。

アジレント・テクノロジー株式会社

© Agilent Technologies, Inc. 2012

Printed in Japan

August 2, 2012

5991-0975JAJP



Agilent Technologies