SHIMADZU
Excellence in Science

Application News

No. M307

GC-MS GCMS-QP™2020 NX, GCMS-TQ™ NX series

Classification of Unknown Samples by Fatty Acids

Y. Nakagawa

User Benefits

- Simple workflow from LabSolutions Insight™ to multivariate analysis/machine learning software
- R&D by identifying new sample groups with a list of compounds critical in separating those groups
- Predict/Classify an unknown sample into the known sample groups

■ Introduction

Fatty acids, particularly those with a high degree of unsaturation and a carbon backbone of a middle chain length, are known as functional nutrients.

Multivariate analysis is often used in profiling oils of food origin, but it can also be applied in profiling brans to facilitate their class characterization aside from mere physical traits.

Shimadzu Corporation Japan has entered the second year of the three year collaborative research with National Agriculture and Food Research Organization (NARO) on the amount of functional nutrients (e.g. fatty acids) in foods such as brans, tea leaves and rice.

As a part of the collaboration, fatty acids (Table 1) were quantitated in 48 bran samples. While details of the results are in non-disclosure agreement, analysis workflow will be discussed in this article.

Orange Data Mining (University of Ljubljana) was employed to identify three distinct groups out of the 48 bran samples and create a list of compounds that were important in classifying those clusters.

Fig. 1 GCMS-QP™2020 NX

■ LabSolutions Insight

Two features of LabSolutions Insight in particular facilitated data analysis: the automatic integration and the flagging.

Automatic integration allowed any manual integration that needed to be performed to be applied to all samples. The flagging highlighted compounds with detector saturations which would be unsuitable for subsequent area comparison.

48 samples were analyzed for FAME 37 in less than 20 mins. A csv file that tabulated sample and compound in terms of peak area was exported from LabSolutions for further analysis.

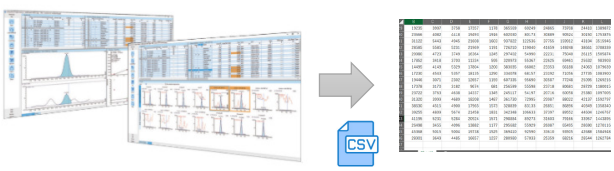

Table 1 A List of Target Compounds

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

■ CSV File Loading

It is a common error to load a csv file without "preparing" data first. The preparation involves removal of any compound from the Excel column that lacks peak area value (i.e. 0, not integrated) for any of the 48 samples. Compounds with saturation were not integrated in Insight and deleted in this step. 37 compounds were now down to 25 compounds.

Orange Data Mining further reduced the number of compounds by identifying and deleting compounds that do not have a normal distribution (Fig. 2). This reduced the number of compounds from 25 to 16.

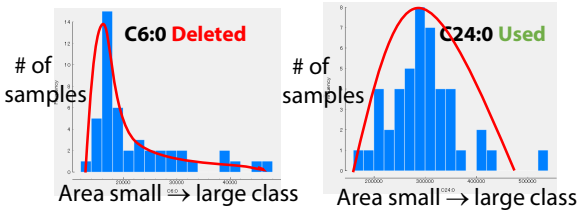

Fig. 2 Distribution Examination

Finally, positive or negative correlation between any 2 of the 16 compounds were examined. Highly correlated compounds were combined into one by averaging (Fig. 3).

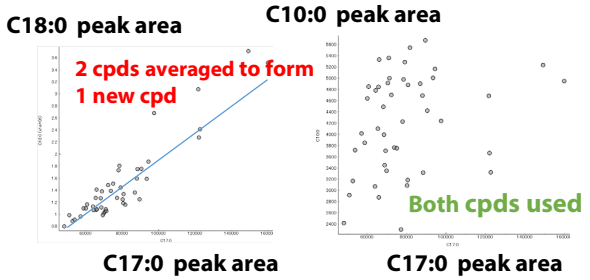

Fig. 3 C17:0 and C18:0

■ Principle Component Analysis

PCA was performed to identify clusters within the 48 samples. The 3 clusters were identified with the PC1 and 2 cumulative of ca. 80 %.

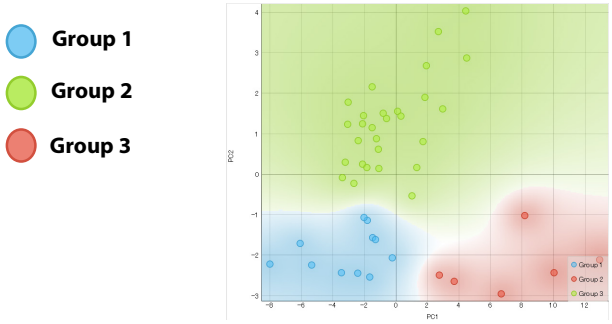

Fig. 4 PCA Plot

■ Predict and Classify Unknown

A few clicks gave the compound lists that were important in separating the groups. For instance, C12:0 was decisive in separating the group 1 and 3 (Fig. 5)

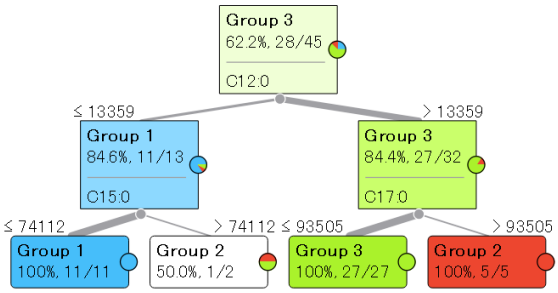

Fig. 5 Classification Tree

Three unknown sample were run , analyzed and subsequently classified into one of the three groups (Fig. 6)

	classification	Sample	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

Fig. 6 Classification of Unknown Sample

■ Summary

48 bran samples were analyzed for 37 fatty acids, using GCMS-QP™2020 NX. The objective was to group similar samples into a cluster, identify a list of compounds characteristic to a given group and create a model that classifies an unknown sample.

The objectives were met with Orange Data Mining (University of Ljubljana). LabSolutions Insight exported a csv file that was loaded onto the software. With the software being stand alone type, there was no need to upload data onto internet.

Three clusters were found and important fatty acids in the classification process were identified. The model was created and tested with unknown samples, giving the correct results.

Shimadzu Corporation would like to thank Hironori Juichi, Yayoi Ichiki and Mari Yamamoto from NARO for their guidance in leading the project.

SHIMADZU
Shimadzu Corporation
Analytical & Measuring Instruments Division
Global Application Development Center

www.shimadzu.com/an/

First Edition: Dec. 2020