

Clasificación multivariante de rones añejos cubanos

Estrella Patricia Zayas Ruiz

Correo electrónico: estrella@quimica.cujae.edu.cu

Instituto Superior Politécnico José Antonio Echeverría, La Habana, Cuba

Artículo Original

Magdalena Lorenzo Izquierdo

Correo electrónico: magdalena.lorenzo@icidca.edu.cu

Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar, La Habana, Cuba

Dianet Hernández Saíenz

Correo electrónico: dhernandez@quimica.cujae.edu.cu

Instituto Superior Politécnico José Antonio Echeverría, Cujae, La Habana, Cuba

Idania Blanco Carvajal

Correo electrónico: idania.blanco@icidca.edu.cu

Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar, La Habana, Cuba

Resumen

En el análisis de alimentos se emplean numerosas técnicas analíticas instrumentales que proporcionan gran cantidad de datos. El empleo del análisis multivariante permite extraer la información más importante de estos datos de forma eficiente. La clasificación supervisada se utiliza principalmente para construir reglas de clasificación para un número de subgrupos conocidos. Las nuevas muestras desconocidas son entonces asignadas al subgrupo más probable basado en estas reglas. El presente trabajo tuvo como objetivo la aplicación de dos técnicas de clasificación supervisada, análisis de clases por modelación suave independiente (SIMCA) y el análisis discriminante por mínimos cuadrados parciales (PLS-DA) a datos de cromatografía de gases y acidez para 52 muestras de rones añejos cubanos, para lograr la clasificación de los rones añejos oscuros entre varios añejos analizados en el Centro de Referencia de Alkoholes y Bebidas (CERALBE) del Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar (ICIDCA). Se utilizó el programa UNSCRAMBLER v. 8.0 (CAMO ÅS, N-7401). El resultado más relevante obtenido es que el 100% de las muestras de añejos oscuros fue bien clasificado por el SIMCA y por el PLS-DA. De igual forma, se analizan las ventajas y desventajas de ambos métodos, así como la utilización futura de estos modelos en la industria.

Palabras claves: Clasificación multivariada, quimiometría cromatografía, ron añejo

Recibido: 30 de abril del 2013

Aprobado: 1 de julio del 2014

INTRODUCCIÓN

En el análisis de alimentos el empleo de la quimiometría permite extraer la mayor cantidad de información de los datos de forma eficiente, especialmente cuando se emplean técnicas analíticas que producen grandes cantidades de datos, como la cromatografía de gases y la espectroscopia [1] [2]. En la autenticación de alimentos se hace uso de estas técnicas con el objetivo de reconocer marcas, origen de los alimentos se variedad botánica o especie animal [3].

Entre los estudios realizados relacionados con la clasificación multivariante se pueden citar las investigaciones

de Giovanetti [4] y colaboradores que diferenciaron los jugos de dos variedades de manzanas, Fuji y Gala, que presentan un patrón químico que puede ser utilizado para esta clasificación. Ellos emplearon análisis de componentes principales (PCA) y modelación suave independiente de análisis de clases (SIMCA) para el análisis exploratorio y clasificatorio, obteniendo un 88 % y 64 % respectivamente como pertenecientes ambas a un dominio mixto.

Para determinar la composición de la fracción volátil de muestras de vinos blancos, de varias marcas españolas con certificados de origen (Rueda, Ribeiro, Penedés y Condado

de Huelva), Jurado y colaboradores [5] emplearon la microextracción en fase sólida de espacio de superior de cromatografía de gases. Se aplicó el análisis de componentes principales (PCA) y análisis discriminante lineal (LDA), y redes neuronales como herramientas quimiométricas para realizar los estudios de reconocimiento de patrones. Se obtuvo el reconocimiento de las diferencias de los vinos certificados antes mencionados.

Con relación a los destilados de la caña de azúcar, Patterson P. de Souza y colaboradores, estudiaron muestras de cachaza y ron, destilados de la caña de azúcar de composición similar y cuya diferenciación está sujeta a disputas comerciales y retos analíticos, mediante masas con ionización negativa por electroatomización. Se detectaron diferencias marcadas en los espectros y es posible encontrar juegos de iones para cada tipo de bebida, mediante la aplicación de PCA y análisis de clúster jerárquico [6]. Después de estudiar el aguardiente de Cocuy de penca que es una bebida espirituosa muy popular en la región nordeste de Venezuela, Hernández Caraballo y colaboradores [7] determinaron concentraciones de cobre, cinc y hierro mediante espectroscopia de absorción atómica, que son marcadores de la producción ilegal de esta bebida. Se realizaron análisis discriminantes lineales y cuadráticos, así como trabajos con redes neuronales artificiales donde estas últimas dieron el mejor intervalo de predicción para la clasificación de las distintas bebidas de acuerdo con las concentraciones de dichos elementos. Posteriormente [8] utilizaron una red neuronal artificial de base radial para la discriminación de especímenes del aguardiente de Cocuy de penca y se logra obtener una red con muy buenas cualidades predictivas del 90 % para discriminar entre la producción legal y la ilegal de esta bebida.

El análisis de componentes principales (PCA) es un método estadístico multivariante de simplificación o reducción de la dimensión de una tabla de casos variables con datos cuantitativos, para obtener otra de menor número de variables, combinación lineal de las primitivas, que se denominan componentes principales o factores, cuya posterior interpretación permitirá un análisis más simple del problema estudiado. El PCA se considera una técnica exploratoria y de pretratamiento de los datos (reducción de dimensiones), como paso previo a la obtención del modelo de calibración o clasificación.

La clasificación o análisis clasificatorio puede catalogarse como un método predictivo y utilizarse para determinar el origen geográfico de una materia prima, a partir de niveles de alguna impureza o para aceptar o rechazar un producto dependiendo de su calidad, en la misma se construyen modelos capaces de pronosticar la pertenencia de un objeto a una categoría sobre la base de las características del objeto. La matriz de datos contiene al menos una variable de categoría, que indica la categoría a la que pertenece cada objeto y que constituye la respuesta o variable que se quiere

predecir, y una o más variables de escala que describen otras tantas características de los objetos y que se utilizan como variables predictoras.

El método SIMCA se basa en hacer modelos de componentes principales para cada clase en el conjunto de entrenamiento; posteriormente las muestras desconocidas se comparan con los modelos de clases desarrollados y son asignadas a las clases de acuerdo con su analogía con las muestras de entrenamiento.

El PLS-DA es un método de clasificación basado en modelar las diferencias entre varias clases con mínimos cuadrados parciales, a diferencia del SIMCA. Si solo hay dos clases a separar, el modelo de mínimos cuadrados parciales usa una variable respuesta (Y), la cual codifica como miembro de clase, de la siguiente forma: -1 para miembros de una clase y +1 para miembros de la otra. Se utiliza entonces el algoritmo PLS1 (Mínimos cuadrados parciales).

Otros autores que han trabajado en bebidas destiladas son Lachenmeier y colaboradores, quienes han probado la detección de vodkas adulterados a partir de mediciones de conductividad [15], en tequilas y su control de calidad [16], la adulteración en vodka y rones [17], [18]. Estos trabajos representan un indicio fundamentado de que la investigación presentada en el artículo tiene realmente un futuro de aplicabilidad en la industria del ron en Cuba. En Cuba el doctor Pino [19] ha trabajado en la caracterización de rones añejos por cromatografía de gases y microextracción en fase sólida planteando que el uso de solamente 15 compuestos volátiles permite la separación entre rones de 3 y 7 años de añejamiento lo que también señala la posibilidad real de llegar a establecer un modelo que permita la clasificación multivariante de rones cubanos.

Con los rones cubanos se han hecho algunos intentos para lograr la determinación de su autenticidad o al menos algún tipo de clasificación entre marcas o tipos [1], [2], pero no se tiene un análisis de métodos de clasificación aplicables a añejos cubanos. Es por eso que el presente trabajo tiene como objetivo lograr la discriminación de los añejos oscuros de otros tipos de añejos mediante la aplicación de dos técnicas de reconocimiento supervisado de patrones ampliamente discutidas en la literatura [9], [10], SIMCA y el PLS-DA a datos de cromatografía de gases y acidez.

MATERIALES Y MÉTODOS

Las muestras utilizadas fueron de dos tipos: rones añejos y añejos extra. Estas son muestras comerciales colectadas en tiendas durante el período 2006-2008 y proceden de diferentes productores del país.

Se determinaron las concentraciones de acetaldehído, acetato de etilo, metanol, propanol, isobutanol, alcohol isoamílico, acetal, 1-butanol, 2-metil-1-butanol determinados por cromatografía de gases [11], acidez [12], y porcentaje alcohólico v/v [13], las dos primeras expresadas en g/100 L de alcohol absoluto, de 52 muestras confidenciales obtenidos

en CERALBE (Centro de Referencia de Alcoholes y Bebidas). Se empleó el programa Unscrambler v. 8.0 (CAMO ÅS, N-7401, Trondheim, Noruega) [9].

Se aplicó además el análisis de componentes principales (PCA) como técnica exploratoria y de pretratamiento de los datos y paso previo a la obtención del modelo de clasificación.

El PLS-DA es un método de clasificación que permite modelar las diferencias entre varias clases con mínimos cuadrados parciales, a diferencia del SIMCA. Si solo hay dos clases a separar, el modelo de mínimos cuadrados parciales usa una variable respuesta (Y), la cual codifica como miembro de clase, de la siguiente forma: -1 para miembros de una clase y +1 para miembros de la otra. Se utiliza entonces el algoritmo PLS1 (Mínimos cuadrados parciales).

Para la clasificación multivariada en el trabajo se emplearon los métodos SIMCA y PLS-DA. Las variables utilizadas para ambos métodos fueron las concentraciones de los componentes mayoritarios, grado alcohólico y acidez, y se analizaron 44 muestras en total, 22 muestras de cada clase (rones añejos y añejos extra), en la construcción de los modelos y 8 para la validación o clasificación. No es posible ofrecer los datos analíticos debido a la confidencialidad de la información. La leyenda de los datos tendrá la letra de la clase y después el número correspondiente (Ej.: O20 o A1). Los conjuntos de datos se muestra en la tabla 1:

$Y_{estimada} < 0,5$ que significa "cerca de 0" puede decirse que es "no miembro de la clase añejos extra o lo que es igual miembro de la clase añejos [9].

RESULTADOS Y DISCUSIÓN

Análisis exploratorio

Del PCA realizado a los 44 datos del conjunto de entrenamiento, que se muestra en la figura 1, se observa cómo hay dos grupos perfectamente separados donde se encuentran agrupadas las muestras de cada clase, o sea, rones añejos y añejos extra; el componente 1 explica el 87 % de la varianza y el componente 2 el 9 %, siendo el componente 1 el que indica el tipo de añejo. Este resultado es muy útil pues anticipa que la clasificación tiene grandes posibilidades de éxito. Con respecto a las variables o cargas (*loadings*) puede verse en la figura 2 como todas las variables son significativas para el análisis y la más cercana al origen es el grado alcohólico que si se desea puede extraerse del análisis, esto puede afirmarse ya que la elipse más externa muestra un 100 % de varianza explicada y la interna solo el 50 %, o sea, si está por debajo de la segunda elipse y cercano al origen, apenas tiene significación ya que explicaría muy poco de la varianza. El grado alcohólico de los añejos oscila entre 39 y 40 (% v/v), es decir, varía muy poco a diferencia de las otras variables utilizadas en el estudio. El resto de las variables tienen un alto nivel de significación en el estudio realizado.

Tabla 1
Conjuntos muestras de entrenamiento y validación

Conjunto muestras de entrenamiento		Conjunto muestras de validación	
Rones añejos (A)	Añejos extra (O)	Añejos	Añejos extra
1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 25, 26.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24, 26.	3, 8, 16, 24.	11, 19, 23, 25.
Total 22	Total 22	Total 4	Total 4

Clasificación SIMCA: Se crearon los modelos PCA con el conjunto de entrenamiento para cada clase empleando la validación cruzada y pesos unitarios para las variables pues se obtuvo una mejor clasificación con estas condiciones.

Clasificación PLS-DA: Se corrió un modelo de regresión, PLS 1, con una variable indicadora (Y) que tiene valor 1 para la clase oscuro y valor 0 para los que no son oscuros, empleando la validación cruzada y pesos unitarios para las variables. Posteriormente se realiza la predicción del conjunto de prueba con el modelo creado. Estimada la Y se puede predecir la pertenencia a las clases a partir de las variables X que describen las muestras. Este modelo se interpreta observando el grafico "estimados" contra "medidos" para la variable Y indicadora de clase:

$Y_{estimada} > 0,5$ que significa "cercano a 1" puede decirse que es "miembro de la clase añejos extra".

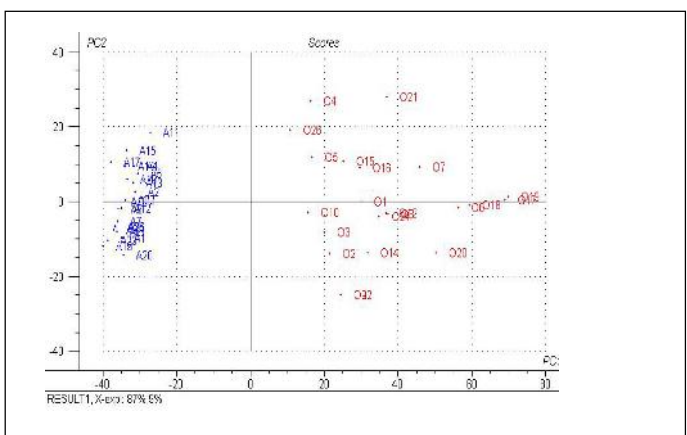


Fig. 1 Análisis de componentes principales para los 44 datos utilizados en el conjunto de entrenamiento

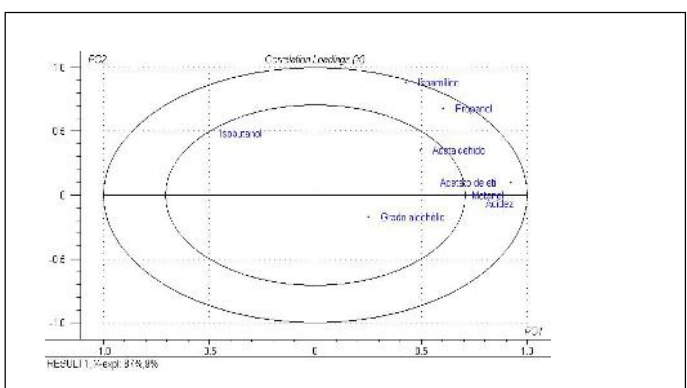


Fig. 2 Cargas o loadings correlacionados

Como el grado alcohólico no resultó significativo para el análisis en el conjunto de entrenamiento, se eliminó en el procesamiento y los modelos se crearon con solo 7 variables tanto para rones añejos como para añejos extra.

Clasificación por el método SIMCA

Los modelos de clasificación tuvieron las características siguientes:

Para los rones añejos

Método de calibración: PCA

Método de validación:

Validación cruzada

Tamaño del modelo: Completo

Número de PC calculados: 4

Número de PC sugeridos: 3

Para los añejos extra

Método de calibración: PCA

Método de validación:

Validación cruzada

Tamaño del modelo: Completo

Número de PC calculados: 4

Número de PC sugeridos: 4

En los rones añejos extra el modelo PCA se explica al 93 % de varianza con dos componentes, mientras que para los añejos con dos componentes se explica al 91 %, estos resultados avalan la calidad de los modelos creados.

En la validación o predicción se obtuvo una clasificación correcta en el 100 % de las muestras del conjunto de clasificación como se muestra en la tabla 2:

Estos resultados son muy satisfactorios, pues no hay ninguna muestra dudosa en el conjunto de validación, no obstante, debe ampliarse el tamaño de los conjuntos tanto para la creación del modelo como el conjunto de validación.

Tabla 2 Clasificación por el método SIMCA		
Muestras	Clasificadas como añejos	Clasificadas como añejos extra
O11		*
O19		*
O23		*
O25		*
A3	*	
A8	*	
A16	*	
A24	*	

Clasificación por el método PLS DA

El modelo de regresión PLS1 obtenido presenta buen ajuste como puede apreciarse en la figura 3, donde la línea diagonal y la de tendencia son prácticamente iguales. El modelo explica en las X el 89 % de la varianza con dos componentes y para la Y se explica el 96 % lo cual ratifica la buena calidad del ajuste del modelo creado.

En la tabla 3 se observa que todas las muestras del conjunto de prueba fueron clasificadas correctamente por el modelo, ya que si el valor estimado es mayor que 0,5 y la desviación no cruza la línea de 0,5 son considerados como miembros, mientras que si el valor estimado es menor que 0,5 y la desviación no cruza la línea de 0,5 son considerados como no miembros.

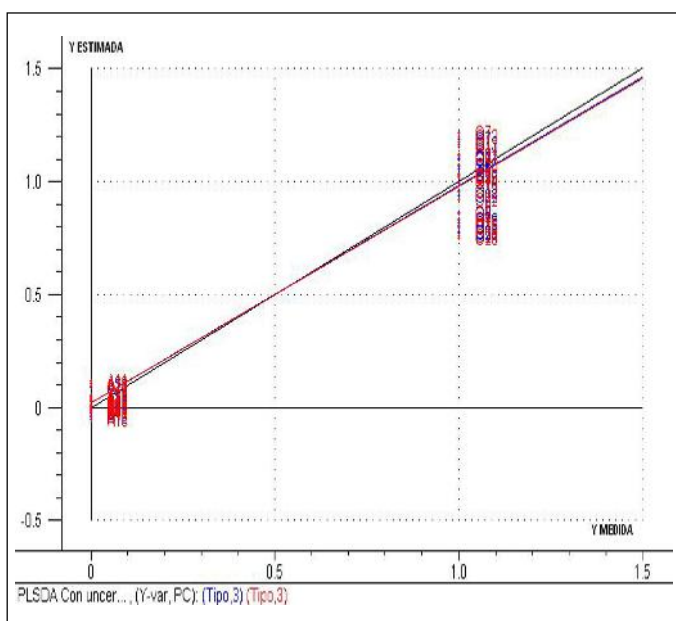


Fig. 3. Gráfico de $Y_{estimada}$ contra Y_{medida} para el modelo PLS1

Como se observa en la figura 4 las variables con mayor poder de discriminación fueron acidez, acetato de etilo e isobutanol. En el caso de la acidez pudiera estar relacionada con las transformaciones que sufren los compuestos en el añejamiento (alcohol ácidos ésteres). El acetato de etilo, compuesto mayoritario dentro de los ésteres en los rones, es característicos de las bebidas destiladas y generalmente se incrementa durante el proceso de añejamiento. Por otra parte el isobutanol es uno de los compuestos del grupo de los alcoholes superiores que está relacionado con una de las materias primas que le da origen al ron, el aguardiente, se encuentra en pequeñas concentraciones, siendo el mayoritario el alcohol amílico.

Tabla 3 Valores estimados con desviación para el método PLS-SA		
Muestras	Valores Estimados	Desviación
O11	1,041	0,058 62
O19	1,063	0,079 46
O23	0,910	0,068 83
O25	0,730	0,04612
A3	-0,037 11	0,095 53
A8	0,031 68	0,056 79
A16	-0,120	0,243
A24	-0,037 74	0,062 67

Con estos resultados pudo establecerse que la composición de los añejos, estudiados, permite diferenciar claramente los añejos extras de los demás añejos y se puede establecer un sistema de autenticación de este tipo de bebida alcohólica destilada, pues sus sistemas de añejamientos son diferentes. Con relación a las muestras es necesario aumentar la cantidad de las mismas en los conjuntos empleados en el análisis para dar mayor robustez a los modelos obtenidos. Alonso-Salces y colaborador [14], cuando analizaron la composición polifenólica de las cidras vascas y francesas utilizándolas para caracterizar las diferencias entre estas dos bebidas estableciendo un sistema de clasificación capaz de confirmar la autenticidad de ambos tipos de cidras, utilizaron un total de 165 muestras de cidra y 27 variables medidas, lo que ratifica la necesidad de aumentar el tamaño de muestra.

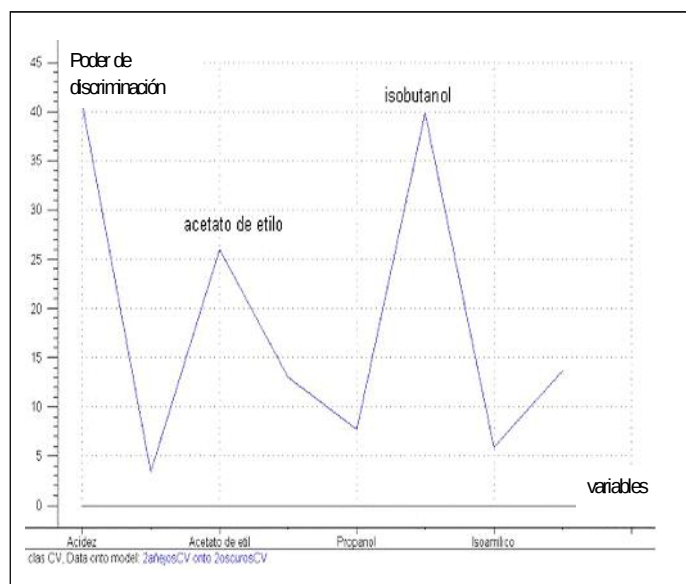


Fig. 4. Gráfico de poder de discriminación

CONCLUSIONES

El tratamiento quimiométrico a datos de: acidez, acetaldehído, acetato de etilo, metanol, propanol, isobutanol, isoamílico, acetal, 1-butanol y 2-metil-1-butanol en añejos cubanos, ha permitido mostrar las diferencias existentes entre los añejos extras y otros añejos que fueron estudiados en el trabajo.

En la aplicación del PCA se observaron dos grupos perfectamente separados donde se encuentran agrupados los rones de cada tipo, añejos extras y añejos, lo cual es muy conveniente para la clasificación.

La aplicación de las técnicas de clasificación SIMCA y PLS-DA permiten disponer adecuadamente las muestras de añejos extra y diferenciarlas de los otros añejos analizados. Ambos métodos presentaron un 100 % de clasificación correcta de las muestras. Las variables con mayor poder de discriminación fueron la acidez, las concentraciones de acetato de etilo e isobutanol, mientras que el grado alcohólico no resultó significativo para el análisis y se eliminó en la creación de los modelos y la validación.

RECONOCIMIENTOS

Los autores desean reconocer el apoyo brindado por el Centro de Referencia de Alcoholes y Bebidas (CERALBE) del Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar (ICIDCA) en la caracterización de las muestras empleadas para el trabajo, así como la colaboración de sus investigadores.

REFERENCIAS

- ZAYAS RUIZ, Estrella Patricia; O'DONNELL, Colm.** "Autenticidad de bebidas alcohólicas. Un estudio preliminar". Memorias del Tercer Simposio Internacional de Química, 5-8 de junio, 2007. ISBN 978959250373-3.
- ZAYAS RUIZ, Estrella Patricia; O'DONNELL, Colm; LORENZO IZQUIERDO, Magdalena.** "Aplicación de la quimiometría en la determinación de autenticidad de bebidas espirituosas. *Revista Cubana de Ingeniería*, vol. IV, núm. 1, enero - abril, 2013, pp. 33 - 40. ISSN 2223 -1781.
- SINGHAL, R.; KULKARNI, P.; REGE, D.** *Handbook of Indices of Food Quality and Authenticity*. England. Woodhead Publishing Limited. 1997, pp. 9-31. ISBN 1855732998.
- GIOVANETTI VIEIRA, Renato; PRESTES, Rosilene Aparecida, et al.** "Chemical pattern of brazilian apples. A chemometric approach based on the Fuji and Gala varieties". *Ciência e Tecnologia de Alimentos*, vol.31, núm.2, Apr./June 2011.
- JURADO, J. M.; BALLESTEROS, O. et al.** "Differentiation of certified brands of origins of Spanish white wines by HS-SPME-GC and chemometric". *Analytical Bioanalytical Chemistry*, vol. 390, 2008, pp. 961-970.
- DE SOUZA PATTERSON, P.; AUGUSTI, Daniella et al.** "Differentiation of rum and brazilian artisan cachaca via electrospray ionization mass spectrometry fingerprinting". *Journal of mass spectrometry*. 2007, vol. 42, pp. 1294-1299.
- HERNÁNDEZ CARABALLO, E.; ÁVILA GÓMEZ, Rita M. et al.** "Discrimination of Venezuelan spirituous beverages by a trace element- radial basis neural network approach". *Talanta*. 2007, vol. 60, pp.1259-1267.
- HERNÁNDEZ CARABALLO, Edwin; ÁVILA DE HERNÁNDEZ, Rita M.; RIVAS ECHEVERRÍA, Francklin; CAPOTE LUNA, Tarcisio.** "Discrimination of venezuelan spirituous beverages by a trace element-radial basis neural network approach". *Talanta*, vol. 74, 2008, pp. 871-878.
- NAES, T.; ISAKSSON, T.; FAERN, T.; DAVIES, T.** *A User-Friendly Guide to Multivariate Calibration and Classification*. Chichester, UK. Ed by NIR Publications. 2002. ISBN 0952866625.
- BERRUETA, Luis A.; ALONSO SALCES, Rosa M.; HÉRBERGER, Károly.** "Supervised pattern recognition in food analysis". *Journal of Chromatography A*. 2007, vol. 1158, pp. 196-214.

11. NORMAS CUBANAS. "Determinación de componentes volátiles mayoritarios en bebidas alcohólicas destiladas, aguardientes y alcohol etílico por cromatografía gas líquido". NC- 508, Cuba, 2007.
12. NORMAS CUBANAS. "Determinación de acidez en rones y aguardientes". NC- 291, Cuba, 2003.
13. NORMAS CUBANAS. "Determinación de grado alcohólico en rones y aguardientes". NC- 290, Cuba, 2003.
14. ALONSO SALCES, R. M.; GUYOT, S. *et al.* "Chemometric characterisation of Basque and French ciders according to their polyphenolic profiles". *Analytical Bioanalytical Chemistry*. 2004, vol. 379, pp. 464-475.
15. LACHENMEIER, Dirk W.; SCHMIDT, Beke; BRETSCHNEIDER, Thorsten. "Rapid and mobile brand authentication of vodka using conductivity measurement". *Microchimica Acta*. 2008, vol. 160, pp. 283-289.
16. LACHENMEIER, D. W.; RICHLING, E. *et al.* "Multivariate Analysis of FTIR and Ion Chromatographic Data for the Quality Control of Tequila". *Journal of Agricultural. Food Chemistry*. 2005, vol. 53, pp. 2151-2157.
17. LACHENMEIER, D. W.; ATTIG, R.; FRANK, W.; ATHANASAKIS, C. "The use of ion chromatography to detect adulteration of vodka and rum". *European Food Research and Technology*. 2004, vol. 218, núm. 1, pp. 105-110.
18. LACHENMEIER, D. "Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra". *Food Chemistry*. 2007, vol. 101, pp. 825-832.

19. PINO, Jorge A. "Characterization of rum using solid-phase microextraction with gas chromatography-mass spectrometry". *Food Chemistry*. 2007, vol. 104, pp. 421-42.

AUTORES

Estrella Patricia Zayas Ruiz

Ingeniera Química, Máster en Análisis de Procesos, Profesora Asistente, Facultad de Ingeniería Química, Instituto Superior Politécnico José Antonio Echeverría, Cujae, La Habana, Cuba

Magdalena Lorenzo Izquierdo

Licenciada en Bioquímica Analítica, Máster en Química Analítica, Centro de Referencia de Alcoholes y Bebidas (CERALBE), Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar, La Habana, Cuba.

Dianet Hernández Saíñz

Ingeniera Química, Instructora, Facultad de Ingeniería Química, Instituto Superior Politécnico José Antonio Echeverría, Cujae, La Habana, Cuba

Idania Blanco Carvajal

Ingeniera Química, Máster en Ingeniería Alimentaria, CERALBE, Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar, La Habana, Cuba

Multivariate Classification of Cuban Aged Rums

Abstract

Several instrumental analytical techniques that are employed in food analysis produce a large amount of data. The employment of the multivariate analysis allows extracting the most important information from these data efficiently. Supervised classification is primarily used to build classification rules for a number of known subgroups. New samples are then assigned to the most likely subgroup based on these rules. The present work had as objective the application of two techniques of supervised classification, modeling independent class analysis (SIMCA) and partial least square discriminant analysis (PLS-DA) to data of gas chromatography and acidity for 52 samples of aged Cuban rum, to achieve the dark aged rum classification among several aged rum analyzed in the Reference center of drinks and alcohols (CERALBE) of the Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar (ICIDCA). The software UNSCRAMBLER v. 8.0 (CAMO ÅS, N-7401) was used. The more important result obtained is that 100% of the samples of aged dark rum was very well classified by the SIMCA and by PLS-DA. The advantages and disadvantages of both methods are analyzed as well as the future use of these models in the industry.

Key words: multivariate classification, chemometrics, gas chromatography, aged rum