

IDENTIFICACIÓN DE ORIGEN EN MIELES A TRAVÉS DE UNA METODOLOGÍA DE EXPLOTACIÓN DE INFORMACIÓN

Lic. Gabriela Anahí CAYÚ gcayu@unrn.edu.ar

Laboratorio de Informática Aplicada – Sede Atlántica UNRN

RESUMEN

La valoración territorial de los productos alimenticios está estrechamente ligada a los atributos de calidad y son la base actual de los procesos de tipificación. En la construcción de identidad de la miel proveniente de las abejas obreras (*Apis mellifera* L), el conjunto de estudios integrados por análisis melisopolinológicos¹, sensoriales y parámetros físico-químicos, tienen un peso significativo en la definición de grupos pertenecientes a un área de procedencia. Las metodologías de análisis e identificación abordadas por la estadística clásica se basan en técnicas multivariadas², que encuentran limitaciones asociadas al gran número de atributos con conjuntos de muestras acotadas. Como alternativa de trabajo, en esta tesis se propone definir un procedimiento que permita, a partir de la explotación de información con sistemas inteligentes (Britos, 2008; Britos et al., 2008), determinar grupos de pertenencia y sus características descriptivas. Para el análisis se utilizaron 47 muestras de mieles provenientes de la región nordeste del distrito oriental del parque chaqueño (Formosa) sobre las que se utilizaron un conjunto de 7 parámetros fisicoquímicos, 20 descriptores sensoriales y 128 especies botánica (taxones). En primer lugar, se identificaron grupos afines, a través del algoritmo Kohonen (Kohonen, Hynninen, Kangas & Laaksonen, 1996) y (Kohonen, 2001), posteriormente se utilizaron reglas de inducción (Quinlan, 1997) para determinar las características de los grupos. Por otra parte, se identificó a través de reglas de inducción las características de los grupos de mieles Monoflorales (tipo de polen dominante) y Multiflorales (en la que intervienen más de un tipo de polen). Las técnicas utilizadas contribuyen a la simplificación metodológica de la identificación y en los procesos de valoración de la miel.

1 Es una rama de la palinología que estudia la presencia de grano de polen de diferentes orígenes florales presentes como trazas en la miel de abejas.

2 Es un conjunto de técnicas estadísticas que analizan simultáneamente más de dos variables en una muestra de observaciones, esta técnica estudia, interpreta y elabora el material estadístico sobre la base de un conjunto de $n > 1$ variables, las cuales pueden ser de tipo cuantitativo, cualitativo o una combinación de ambas (Carranza et al., 2011).

PROBLEMA A RESOLVER

El conjunto de datos habituales para la caracterización de las mieles de un territorio contempla en primera instancia la identificación de las familias botánicas de las cuales las abejas se aprovisionan de néctar de flores para la elaboración de miel. Esta metodología se denomina análisis melisopalinológicos y consiste en el conteo y cálculo de proporcionalidad de presencia de granos de polen como residuos en la miel. Si bien esta técnica, en primera instancia, permite clasificarlas o tipificarlas por su origen botánico y asociarlo a una procedencia (Tellería, 2010), no resulta definitorio, dado que una especie, aún con alto porcentaje de participación en el espectro polínico, puede estar sub o sobrerrepresentada en las muestras a causa de menor o mayor exposición de las anteras (estructura floral productora del polen) en los momentos que la abeja accede a los nectarios. Otra situación se encuentra asociada al aporte sensorial de cada especie. Aún con muy poca participación porcentual, determinadas familias botánicas pueden afectar los descriptores sensoriales y los parámetros físico-químicos.

En este sentido, el abordaje integral en la caracterización aplica como ventaja al momento de definir su identidad. Así, las metodologías de análisis e identificación tradicionales, abordadas por la estadística clásica, se basan en técnicas multivariadas, que encuentran limitaciones asociadas al gran número de atributos con conjuntos de muestras acotadas.

Propuesta de Solución

En esta tesis, para el análisis se utilizaron 47 muestras de mieles provenientes de la región nordeste del distrito oriental del parque chaqueño (Formosa) sobre las que se utilizaron un conjunto de 7 parámetros físico-químicos, 20 descriptores sensoriales y 128 especies botánica (taxones).

El mayor desafío de este proyecto es buscar una relación en los conjuntos de datos físico-químicos, melisopalinológicos y sensoriales, que permita caracterizar los grupos florales establecidos *a priori* en función del estudio individual de cada muestra y el encuadre que le corresponde a partir de la interpretación de la totalidad de los atributos y sus descriptores. Estos grupos se corresponden con mieles multiflorales o monoflorales de: *Aspidosperma quebracho blanco* (Ast), *Copernicia alba* (Cop), *Eugenia uniflora* (Eug),

Geoffroea decorticans (Geo), Prosopis (Pro), Schinopsis balansae (SchB), Ziziphus mistol (Ziz), Tipo Baccharis-Eupatorium (BacE), Eichhornia crassipes (Eic), Parkinsonia aculeata (Par), Senecio grisebachii (Sen).

Por ello, se define un procedimiento que permite, a partir de la explotación de información (Garcia-Martinez et. al., 2016) con sistemas inteligentes, determinar grupos de pertenencia y sus características descriptivas, estableciendo reglas de agrupamiento.

Objetivos Generales

Se definen como objetivos generales de esta tesis:

- Implementar una metodología de explotación de información para la identificación de origen en mieles y
- Utilizar procesos de explotación de información con sistemas inteligentes para obtener resultados.

Objetivos Específicos

A continuación, se detallan una serie de objetivos específicos que permiten en conjunto establecer los pasos a seguir para lograr cumplir con los objetivos generales:

- Utilizar la metodología propia de la investigación documental y de aplicación metodológica a un caso de estudio, con base en que:
 - Se producirá un relevamiento en primera instancia sobre metodologías para la Explotación de Información y a continuación la caracterización de origen en mieles.
 - Formalización de la metodología a través de un caso de estudio que permita determinar origen en mieles.
 - Para construir el conocimiento asociado a la presente tesis, se seguirá un enfoque de investigación clásico (Riveros y Rosas, 1985; Creswell, 2002) con énfasis en la producción de tecnologías (Sábato y Mackenzie, 1982); identificando procesos, materiales y abordaje metodológico necesarios para desarrollar la tesina:

CASO DE VALIDACIÓN: IDENTIFICACIÓN DE ORIGEN EN MIELES

Dada la extensa distribución fitogeográfica del Parque Chaqueño, sería importante definir «tipos regionales de miel» cuyas características sensoriales y físico-químicas, en estrecha relación con el origen floral puedan ser valoradas en los mercados nacionales y extranjeros. De allí, la importancia de obtener tipos de mieles monoflorales y poliflorales. A partir de este trabajo de investigación realizado con 47 muestras, se tipificaron mieles monoflorales de *Copernicia alba*, *Schinopsis balansae*, *Eugenia uniflora*, *Prosopis* sp., Tipo *Baccharis-Eupatorium* y mieles poliflorales, teniendo en cuenta el origen botánico, físico-químico y sensorial.

Para el análisis se trabajó con 7 parámetros físico-químicos (PH, AT (Acidez Total), AL (Acidez Lactónica), Cond (Conductividad), Humedad, Color, Cantidad); 20 atributos descriptores sensoriales (NarizFloralFrutado, NarizFresco, NarizQuímico, NarizAmaderado, NarizVegetal, NarizDegradado, NarizCálido, BocaFloralFrutado, BocaFresco, BocaQuímico, BocaAmaderado, BocaVegetal, BocaDegradado, BocaCálido, GustoPicante, GustoAstringente, GustoSalado, GustoAmargo, GustoÁcido, GustoDulce), divididos en tres grupos: aromas en nariz, aromas en boca y gustos; y 128 atributos de especies botánicas (AcaA, AcaB, AcaC, AcaP, Aci, Ade, Aga, All, Alo, Alt, Ama, Amb, Ani, Art, Asp, Ast, Bah, Bid, Bra, Bul, Cae, Cal, CapT, Car, Cas, Cec, Cel, Ces, ChA, Cir, Cis, Cit, CleM, CleS, Con, Cop, CorS, CorA, Cro, Cry, Cuc, Cup, Cyp, Dau, Des, Dip, Echg, Echp, Eic, Ent, Eri, EryE, EryD, Euc, Eug, Fos, Geo, Gle, Gly, Gom, Han, Hei, Hels, Helm, HypN, HypB, HypF, Jus, Lig, Lud, Lyc, Mas, May, Mel, Mim, MOR, Mut, Myr, Nic, Nym, Oco, Opu, Oxa, Par, Pas, Pau, Pel, Pet, PhyN, PhyL, Pis, Pla, PolS, PolH, Pon, Pou, Pro, Psi, Sag, Sal, SalO, SalG, Sap, SchB, Sch, Sco, Scu, Sen, SidR, SidO, SolS, SolC, Son, Ste, Sya, Tab, Ter, Tes, Teu, BacE, TriC, TriS, TriD, Verb, Vern, Vic, Zan, Ziz).

A continuación, se detalla el tratamiento que se les dio a los datos:

Limpieza de Datos

El proceso de limpieza de datos se realizó de la siguiente manera:

1. Separación de datos: los conjuntos de datos fueron ubicados en archivos separados en formato XML.
2. Recorte de datos: fueron seleccionados los datos que corresponden a las mismas muestras de los conjuntos de datos, dejando un resultante de 59 muestras a analizar.
3. Conversión del formato de los archivos de datos: conversión al formato CVS (comma separated values - valores separados por coma).

4. Merge de datos: la información en los distintos archivos se cruzó mediante el número identificador de muestra para unificar todos los atributos de las muestras en un único archivo.

Luego de realizar este proceso se obtuvo lo siguiente:

Especies botánicas de 128 atributos, pasó a tener 75, se eliminaron 53 atributos que contenían trazas con valor no mayor a 3 en las muestras, de acuerdo a información brindada por el experto.

Descriptores sensoriales de 20 atributos, pasó a tener 3, sensorial_nariz: 7 atributos; sensorial_boca: 7 atributos y sensorial_gustos: 6 atributos. Se agruparon de esta manera, porque el experto lo considero más representativo.

Procesos utilizados

Los procesos utilizados para el análisis de los datos son los que se detallan en la Figura 1.1 y 1.2.

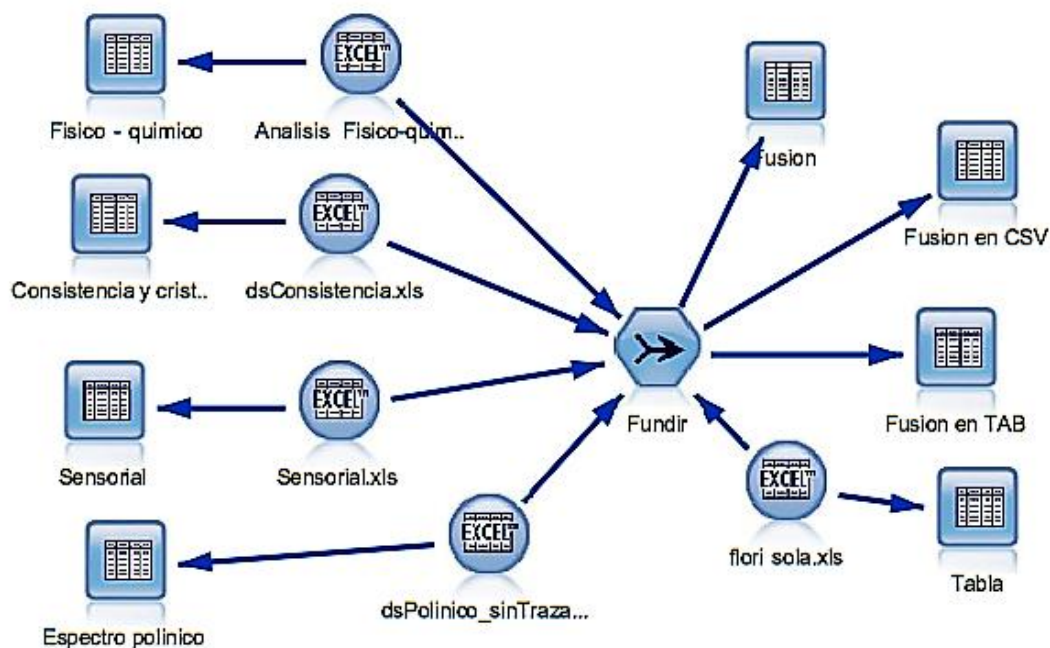


Figura 1.1 Fusión/Integración de datos

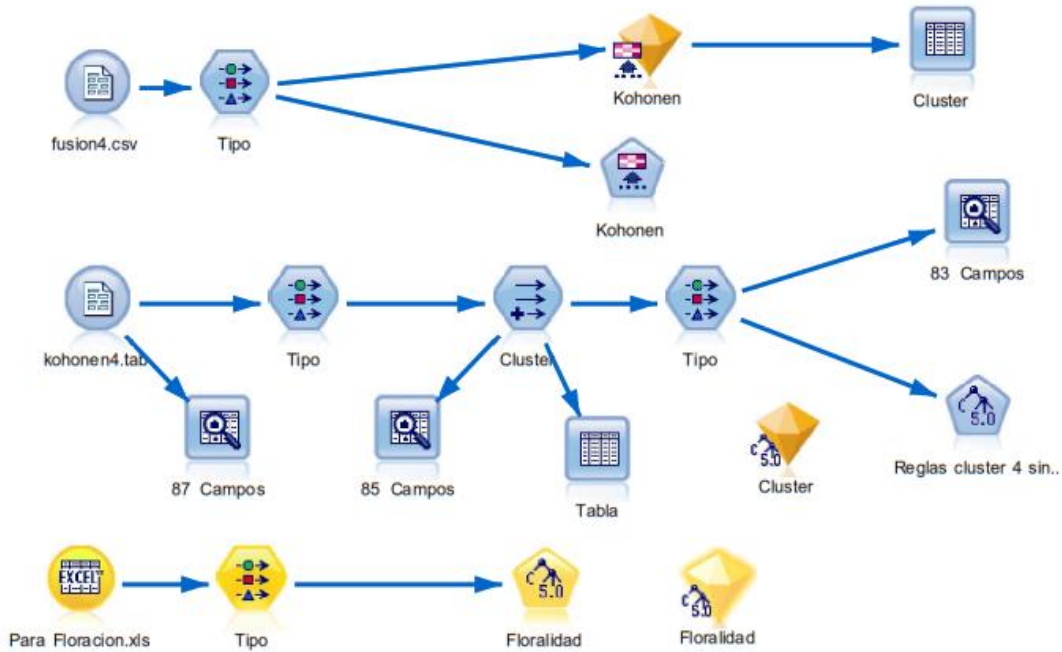


Figura 1.2 Procedimiento de agrupamiento con descubrimientos de reglas y agrupamiento

RESULTADOS

Se destaca la formación de reglas para los grupos preestablecidos, con caracterizaciones más ricas en descriptores de diferenciación que los utilizados en las metodologías clásicas utilizadas en esta temática.

Una de las características de las muestras es la numerosa cantidad de atributos descriptivos de las mieles en relación a las muestras (3 a 1 aproximadamente), muy típico en este campo de estudio. Es por esto que, para llevar adelante la solución al problema planteado, se utilizaron los procesos de descubrimiento de reglas para grupos florales establecidos a priori. Estos grupos se conformaban de muestras con presencia de polen mayor a 45%, denominado frecuencia dominante, mientras que el grupo multiflora se estableció con muestras en las cuales las frecuencias de polen no resultaban mayores a 25 %.

Como resultados podemos destacar que las reglas de inducción descubierta permiten:

1. Identificar grupos de pertenencia en aquellas mieles que siendo multiflorales presentan porcentajes elevados de más de un taxón.
2. Asociar los taxones a variables fisicoquímicas que aportan en la identificación del grupo de pertenencia floral, como es el caso de pH, CE y color en relación al grupo conformado número 3 que agrupó las mieles

de Copernicia alba (Mieles de Palma), las cuales presentan conflictividad en su definición por la variabilidad de sus parámetros físico-químicos.

3. La conjunción de especies florales asociadas en la definición de un grupo monofloral.

Por otra parte, se utilizó el descubrimiento de reglas de pertenencia a grupos sin clasificación de floralidad por parte del experto, utilizando las mismas características anteriormente mencionadas.

Estos resultados fueron publicados en (Cayú G, et. al., 2016).

CONCLUSIÓN Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Definición de grupos de pertenencia

A partir del análisis realizado es posible aportar reglas para la conformación de grupos florales, cumpliendo así la primera pauta de trabajo en el camino hacia la construcción de identidad terri-torial, las mismas aportaron mayor precisión en la definición de los mismos de acuerdo a los comentarios realizados por el experto.

Se puede observar que a través de los procesos empleados y sus algoritmos se ponderó el peso conferido por la estructura de datos proveniente de los taxones, y permitió la participación más activa de los parámetros físico-químicos. Se destaca la aparición de reglas de exclusión de taxones como definitorias en la conformación de grupos, lo cual confiere potencia a la metodología general de trabajo, no solo definiendo grupos locales, sino con proyección en la definición de grupos de regiones diferentes con floración parcialmente coincidente. Por otro lado, el agrupa-miento de los descriptores sensoriales en la conformación de un único valor denominado inten-sidad, no aplicó bien a la conformación de grupos, posiblemente en concordancia a lo descrito respecto de valores muy variables pero intensos en la totalidad de las mieles, razón que conduce a la clasificación y utilización de los descriptores individuales.

Por otra parte, la metodología aplicada mejora las posibilidades de tipificación de mieles a la vez que abre nuevas alternativas en otros alimentos con conjuntos de atributos de diferente origen analítico.

Finalmente, se considera que las reglas definidas deben ser confirmadas en nuevas temporadas y con mayor número de muestras para establecer generalidades.

Análisis de datos

Para el análisis de los datos, se tomó como parámetros los procesos indicados en (Britos, P; 2008). Por lo cual, en su comienzo se utilizó Descubrimiento de Reglas de Comportamiento, lo que permitió identificar grupos afines, a través del algoritmo Kohönen, posteriormente se utilizaron reglas de inducción para determinar las características de los grupos. Por otra parte, se identificó a través de reglas de inducción las características de los grupos de mieles, las cuales fueron predefinidas por el Experto Mg. Germán Balbarrey.

Una vez obtenidos los resultados se validó con el experto en explotación de información, evaluando la calidad de los procesos.

De esta manera, podemos concluir con que se validó la hipótesis propuesta en el Plan de Proyecto presentado como anteproyecto de la tesis, el cual fue aprobado:

“La metodología de explotación de información sería aplicable a la determinación de origen en mieles de distintos puntos geográfico.”

Para futuros trabajos se considera la posibilidad de la identificación de los productores de los cuales provienen las muestras de mieles, y analizar conjuntos de datos obtenidos en otra época del año.

BIBLIOGRAFÍA

Britos, P. (2008). Procesos de Explotación de Información basados en Sistemas Inteligentes. Tesis Doctoral. Universidad Nacional de La PlataLa Plata, Facultad de Informática. La Plata, Argentina. Recuperado en Agosto de 2018, de <https://goo.gl/WuNj78>.

Britos, P., Dieste, O., & García-Martínez, R. (2008). Requirements Elicitation in Data Mining for Business Intelligence Projects. Advances in Information Systems Research, Education and Practice IFIP – The International Federation for Information Processing, 139-150. doi:10.1007/978-0-387-09682-7-9_12.

Kohönen, T., Hynninen, J., Kangas, J., & Laaksonen, J. (1996). SOM_PAK: The self-organizing map program package. Report A31. Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland. 1996. Recuperado en Agosto de 2018, de http://www.cis.hut.fi/research/som_lvq_pak.shtml

Kohonen, T. (2001). Self-Organizing Maps. Series in Information Sciences, Vol. 30. Springer, Heidelberg. Three ed. ISBN 978-3-642-56927-2. Recuperado en Agosto de 2018, de <https://goo.gl/z3S7GQ>

Quinlan, J.R. See5. (1997). Recuperado en Agosto de 2018, de <http://www.rulequest.com/see5-info.html>.

Tellería, M. C. (2010, Diciembre diciembre 02). Avances y perspectivas en la tipificación de mieles argentinas. Recuperado en Agosto de 2018, de <http://sedici.unlp.edu.ar/handle/10915/16706>

García-Martínez, R., Britos, P., Martins, S., & Baldizzoni, E. (2016). Explotación de Información. Ingeniería de Proyectos. Editorial: Nueva Librería. ISBN 978-987-1871-34-6.

Riveros, H., & Rosas, L. (1985). El Método Científico Aplicado a las Ciencias Experimentales. Editorial Trillas. México. ISBN 96-8243-893-4.

Creswell, J. (2002). Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. Prentice Hall. ISBN 10: 01-3613-550-1.

Sábato, J., & Mackenzie, M. (1982). La Producción de Tecnología. Editorial Nueva Imagen. México. ISBN 968- 429-348-8.

Cayú, G. A., Agüero, G. A., Balbarrey, G. P., Cabrera, M. M., Carrera, C., Britos, P., & Vivas, H. L. (2016). Building honey-based territorial identity for the Formosa Monte through information exploitation using intelligent systems. IEEE CACIDI 2016 - IEEE Conference on Computer Sciences, doi: 10.1109/CACIDI.2016.7785980