

Agrupamiento espectral multiclase basado en particiones normalizadas

Diego Hernán Peluffo Ordóñez
Doctor en Ingeniería
Docente catedrático
Universidad Cooperativa de Colombia – Sede Pasto
diego.peluffo@ucc.edu.co



Recibido: 12 de febrero 2013
Aceptado: 5 de marzo 2013

Resumen

Se presenta una metodología completa de análisis no supervisado orientada a la clasificación, empleando técnicas espectrales. Esta metodología incluye una etapa de agrupamiento no supervisado desarrollada con el criterio multiclase de particiones normalizadas. Se prefiere esta técnica porque no requiere un algoritmo de agrupamiento adicional. Además, genera una partición resultante adecuada ya que considera información obtenida de la solución propia. Adicionalmente, se aplican medidas de afinidad adecuadas y se realiza la estimación automática del número de grupos con el fin de reducir el tiempo de procesado y mejorar la convergencia del algoritmo. Los resultados experimentales son obtenidos sobre una base de datos de imágenes. La calidad del agrupamiento se mide en los resultados de la segmentación. Se introduce también una medida de desempeño no supervisada.

Palabras clave: Agrupamiento espectral, medida de afinidad, solución propia, SVD.

Multi-way spectral clustering based on normalized partitions

Abstract

A comprehensive methodology for classification-oriented non-supervised analysis through the use of spectral techniques is presented here. This methodology includes a non-supervised clustering stage developed under the multiway criterion of normalized partitions. This technique is preferred because it does not require an additional clustering algorithm. Additionally, it results in a suitable partition, since it considers the information obtained from a solution of one's own. Besides, adequate affinity measurements are applied, and the number of clusters is automatically estimated in order to reduce processing time and improve algorithm convergence. Experimental results are obtained on an image database. The quality of the cluster is measured by means of segmentation results. Also, a non-supervised performance measurement was introduced.

Keywords: Spectral clustering, affinity measurement, proprietary solution, SVD.

1. Introducción

En el área de reconocimiento y clasificación de patrones, los métodos de agrupamiento basados en grafos y análisis espectral son relativamente nuevos, aunque ya han sido aplicados exitosamente en diversos contextos como el diseño de circuitos (Chan et al., 1994), balance eficiente de carga computacional orientado a programación en paralelo (Hendrickson & Leland, 1995), segmentación de imágenes (Ishikawa & Geiger, 1998; Shi & Malik, 2003), entre otros. Estos métodos son discriminativos, por tanto no requieren información a priori (suposiciones sobre la estructura de los datos), y las particiones son generadas a partir de información obtenida de la misma heurística del método, en este caso un criterio de decisión global que considera el valor estimado de la probabilidad de que dos elementos pertenezcan a un mismo grupo. Por esta razón, este tipo de análisis puede ser fácilmente entendido a partir de la teoría de grafos, donde dicha probabilidad puede asociarse a una medida de afinidad entre los nodos. Comúnmente, este criterio se aplica en un espacio de representación de dimensión menor que la del conjunto de datos, lo que implica una reducción de la dimensionalidad en la representación, conservando tan fiel como sea posible las relaciones entre los elementos, y por lo tanto converge en un análisis de vectores y valores propios.

Los métodos de análisis espectral son de gran interés en tareas de clasificación, debido a que la estimación de los óptimos globales en un dominio continuo no restrictivo se obtienen de la descomposición en valores y vectores propios (descomposición propia), con base en el teorema de Perron–Frobenius en el que se establece que los mayores valores propios, estrictamente reales, de una matriz positiva definida e irreducible definen el radio espectral de la misma (Anh & Thanh, 2007). En otras palabras, el espacio generado por los vectores propios está directamente relacionado con la calidad del agrupamiento. Considerando este principio y la posibilidad de obtener una solución discreta a través de los vectores propios, surgió un enfoque, relativamente nuevo, del agrupamiento no supervisado, denominado agrupamiento espectral.

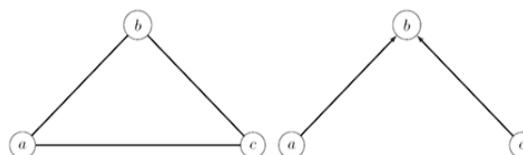
En el desarrollo de este trabajo, se explica detalladamente un método de agrupamiento no supervisado

denominado agrupamiento espectral normalizado. Para tal propósito, se comienza por mencionar algunos conceptos preliminares sobre teoría de grafos. Luego, se desarrolla el método de agrupamiento, deduciendo un criterio de particionado multi-clase a partir de un análisis topológico y, por último, se explica el algoritmo de agrupamiento. Adicionalmente, se incluyen dos etapas, en las que se establecen algunas medidas de afinidad orientadas al agrupamiento y se describen métodos para la estimación del número de grupos, respectivamente. Los resultados experimentales se obtienen con una base de datos de imágenes correspondiente a fotografías de granos de café. Las imágenes se caracterizan por píxeles, empleando espacios o modelos de color (HSV, YCC, LAB y LUV). El algoritmo de agrupamiento es evaluado en un sistema que incluye una etapa de estimación del número de grupos. Adicionalmente, se compara el desempeño de diferentes medidas de afinidad.

2. Marco teórico

2.1 Preliminares sobre grafos

Los grafos son conjuntos, no vacíos, de objetos denominados nodos o vértices, y de líneas denominadas aristas que unen por pares los vértices. Las aristas indican la relación entre los vértices, de forma que si dos vértices están relacionados se traza una línea entre ellos, en caso contrario no debe existir ningún trazo. En general, un grafo propiamente dicho puede ser descrito por el par ordenado $\mathbb{G} = (\mathbb{V}, \mathbb{E})$, donde \mathbb{V} es el conjunto de vértices y \mathbb{E} es el conjunto de aristas. Los grafos pueden clasificarse en dirigidos y no dirigidos. Los grafos no dirigidos no tienen definido un sentido en la relación entre sus elementos, es decir, si $a b \in \mathbb{V}$ entonces las aristas $a b$ y $b a$ y representan el mismo elemento (ver Figura 1, a). En los grafos dirigidos, las aristas tienen orientación definida, por tanto $ab \neq ba$ (ver Figura 1, b).



(a) Grafo conexo no dirigido (b) grafo no conexo dirigido

Figura 1. Ejemplos de grafos

Fuente: Stella & Shi (2003)

Una forma generalizada de representar un grafo corresponde a los grafos ponderados en los que, además del orden y el sentido (en el caso de los grafos dirigidos), importa el grado de relación o afinidad entre los nodos, definido a través de una matriz de pesos, donde es un valor no negativo que representa el peso de la arista ubicada entre los nodos i y j , como se muestra en la Figura 2. Además, en un grafo no dirigido, es evidente que $w_{ij} = w_{ji}$. Por tanto, un grafo ponderado puede ser descrito de la forma $G = (V, E, W)$, donde W es una matriz simétrica y semi-definida positiva.

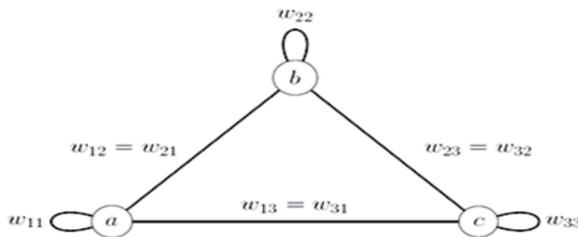


Figura 2. Grafo ponderado
Fuente: Stella & Shi (2003)

Existen diversas medidas, aplicables sobre los grafos, que resultan útiles para clasificarlos y realizar operaciones entre ellos. En este trabajo se tienen en cuenta dos medidas fundamentales, que serán de uso posterior: peso total de un subconjunto de vértices con respecto a otro y grado de un subconjunto.

Dado el grafo $G = (V, E, W)$ y los elementos $A, B \subset V$, el peso total de las conexiones entre A y B se puede calcular con

$$\text{lazos}(A, B) = \sum_{i \in A, j \in B} w_{ij} \quad (1)$$

El grado de un grafo no ponderado con respecto a un vértice equivale a la cantidad de aristas adyacentes al mismo. En los grafos ponderados, que es el caso de interés, el grado representa el peso total de las conexiones de un subconjunto con respecto al conjunto de todos los vértices, por tanto el grado del elemento A podría ser calculado empleando:

$$\text{grado}(A) = \text{lazos}(A, V) \quad (2)$$

El grado se emplea comúnmente en la normalización de los pesos de las conexiones. Dicha normalización

permite estimar la proporción de un subconjunto de conexiones con respecto a otro, por ejemplo, la proporción de conexiones ponderadas entre A y B con respecto al total de conexiones de A es:

$$\text{lazosnor}(A, B) = \frac{\text{lazos}(A, B)}{\text{grado}(A)} \quad (3)$$

2.2 Agrupamiento espectral normalizado

El análisis espectral orientado al agrupamiento de datos se basa en grafos ponderados no dirigidos y conexos $G = (V, E, W)$, en donde $V = \{1, \dots, n\}$ representa los índices del conjunto de datos a ser agrupado y W es la matriz de afinidad que, de acuerdo a lo discutido en la sección 1.2.1, se asume semidefinida positiva y simétrica. El objetivo del agrupamiento espectral es descomponer V y k en subconjuntos disjuntos, por tanto, $V = \cup_{l=1}^k V_l$ y $V_l \cap V_m = \emptyset, \forall l \neq m$. Dicha descomposición comúnmente se realiza empleando información espectral y transformaciones ortonormales. En el agrupamiento espectral, dos medidas son de particular interés: $\text{lazosnor}(A, A)$, que define la proporción de conexiones de A agrupadas correctamente, en otras palabras, indica el grado de conexión intra-particiones, y $\text{lazosnor}(A, A \setminus V)$, una medida complementaria a la primera que indica el grado de conexión entre particiones. Entonces, un buen agrupamiento se obtiene maximizando el grado de asociación o conexión intra particiones y minimizando el grado de desviación o conexión entre particiones. Estos objetivos se pueden lograr aplicando un criterio de particionado multi-clase que, por lo general, implica medir el total de las conexiones intra y entre particiones. Estas medidas pueden calcularse a través del total de asociaciones normalizadas (kasocn) y el total de desviaciones normalizadas (kdesvn), que corresponden, respectivamente, a:

$$\text{kasocn}(\Gamma_v^k) = \frac{1}{k} \sum_{l=1}^k \text{lazosnor}(V_l, V_l) \quad (4)$$

$$\text{kdesvn}(\Gamma_v^k) = \frac{1}{k} \sum_{l=1}^k \text{lazosnor}(V_l, V_l \setminus V) \quad (5)$$

donde es $\Gamma_v^k = \{V_1, \dots, V_k\}$ el conjunto de todas las particiones.

Debido a la normalización con respecto al grado, aplicada sobre las medidas de asociación y desviación, se puede verificar que:

$$\text{kasocn}(\Gamma_{\mathbf{V}}^k) + \text{kdesvn}(\Gamma_{\mathbf{V}}^k) = 1 \quad (6)$$

2.2.1 Criterio de particionado multi-clase

De la ecuación (6) puede concluirse que la maximización de la medida de asociaciones y la minimización de la medida de desviaciones ocurren simultáneamente, por tanto no es necesario distinguir entre estas medidas. Con esto, la función objetivo a maximizar podría plantearse como:

$$\varepsilon(\Gamma_{\mathbf{V}}^k) = \text{kasocn}(\Gamma_{\mathbf{V}}^k) \quad (7)$$

Entre los diversos criterios que existen para la optimización de las asociaciones, únicamente los criterios de desviaciones mínimas (Ishikawa & Geiger, 1998) y normalizadas (Shi & Malik, 2000) tienen esta propiedad de dualidad. Sin embargo, se prefiere el criterio que emplea desviaciones normalizadas porque es menos sensible al ruido y a la inicialización de la matriz de pesos que el método de desviaciones mínimas.

2.2.2 Representación matricial

En adelante, el conjunto $\Gamma_{\mathbf{V}}^k$ se representará con la matriz de particiones que indica la pertenencia de un elemento a un grupo, de manera que:

$$m_{il} = \langle i \in \mathbf{V}_l \rangle, i \in \mathbf{V}, l = 1, \dots, k \quad (8)$$

Donde m_{il} es el elemento il de la matriz \mathbf{M} , $\langle \cdot \rangle$ es un indicador binario: toma el valor de 1 si su argumento es verdadero y 0 en caso contrario.

Dado que un vértice es asignado a una única partición, se debe garantizar que $\mathbf{M}\mathbf{1}_k = \mathbf{1}_n$, donde $\mathbf{1}_d$ es un vector d-dimensional en el que todos sus valores son iguales a 1.

Sea la matriz \mathbf{D} el grado de la matriz de pesos definido como:

$$\mathbf{D} = \text{Diag}(\mathbf{W}\mathbf{1}_n) \quad (9)$$

Donde $\text{Diag}(\cdot)$ denota una matriz diagonal formada por el vector argumento, entonces las medidas representadas en las ecuaciones (1) y (3) pueden replantearse como sigue:

$$\text{lazos}(\mathbf{V}_l, \mathbf{V}_l) = \mathbf{M}_l^T \mathbf{W} \mathbf{M}_l \quad (10)$$

$$\text{grado}(\mathbf{V}_l) = \mathbf{M}_l^T \mathbf{D} \mathbf{M}_l \quad (11)$$

Con lo anterior, el criterio de particionado multi-clase puede expresarse como:

$$\max \varepsilon(\mathbf{M}) = \frac{1}{k} \sum_{l=1}^k \frac{\mathbf{M}_l^T \mathbf{W} \mathbf{M}_l}{\mathbf{M}_l^T \mathbf{D} \mathbf{M}_l} \quad (12)$$

$$\text{s.t.} \quad \mathbf{M} \in \{0,1\}^{n \times k}, \mathbf{M}_k = \mathbf{1}_n \quad (13)$$

Esta formulación se denotará con NCPM (Normalized cuts problem), la letra “M” se usa para expresar que el problema de optimización está planteado en términos de la matriz \mathbf{M} .

2.2.3 Algoritmo de agrupamiento

El algoritmo de agrupamiento espectral que se estudia en este trabajo se basa en el método desarrollado en Shi & Malik (2003), y consiste en resolver el problema de optimización planteado en la ecuación (12), en un dominio continuo y sin restricciones, aplicando descomposición en vectores propios, transformaciones ortonormales y un proceso de discretización de la solución, partiendo del hecho de que los vectores propios generan todas las soluciones más aproximadas a las soluciones óptimas continuas. A este algoritmo se le denomina método de desviaciones normalizadas. Este método es robusto con relación a la inicialización y converge más rápido que las técnicas clásicas, sin embargo podría generar un costo computacional más elevado por cada iteración, debido al cálculo de los valores singulares. Las transformaciones ortonormales se aplican sobre la descomposición propia para generar una familia completa de soluciones óptimas globales, que posteriormente se normalizan en longitud para hacer que cada óptimo encontrado corresponda a una partición solución en el dominio continuo. A dicha solución se aplica un proceso iterativo de discretización

que es llevado a cabo empleando descomposición en valores singulares (SVD) y seleccionando vectores propios de acuerdo con los valores propios máximos y la cantidad de grupos establecida. El diagrama que se muestra en la Figura 3 explica de forma gráfica la idea general del algoritmo de agrupamiento.

2.2.4 Idea general del algoritmo

El primer paso en este método es establecer el número de grupos y la matriz de afinidad W . En las secciones C y D se muestran algunas alternativas para estimar el número de grupos y seleccionar la medida de afinidad. Luego, se descompone en vectores propios Z^* la matriz W normalizada con respecto a su grado. El conjunto Z^* genera todas las soluciones globales óptimas en un dominio continuo a través de la transformación ortonormal R . Después de una normalización en longitud, cada óptimo encontrado corresponde a una partición solución \tilde{M}^* . Por último, en un proceso iterativo se encuentra la solución discreta M^* más aproximada a \tilde{M}^* . El algoritmo termina con el par (\tilde{M}^*, M^*) , donde $\varepsilon(\tilde{M}^*) \approx \varepsilon(M^*)$. En el ejemplo del diagrama explicativo de la Figura 3, la convergencia ocurre en la segunda iteración con $(\tilde{M}^{*(2)}, M^{*(2)})$. La optimalidad de \tilde{M}^* garantiza

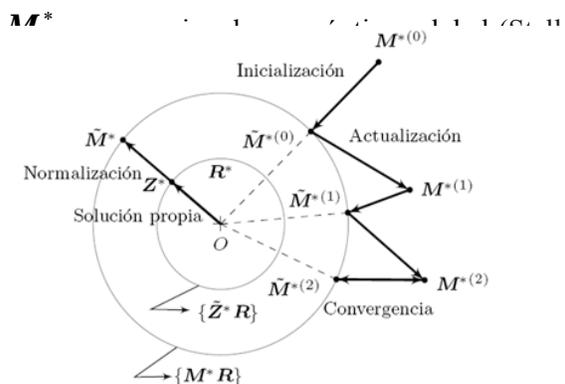


Figura 3. Diagrama explicativo del algoritmo de agrupamiento espectral
Fuente: Stella & Shi (2003)

2.2.5 Solución del problema de optimización

En esta sección se estudia más a fondo la solución del problema de optimización del agrupamiento espectral. Como se comentó antes, este problema puede resolverse en dos etapas: en la primera se obtiene un óptimo global continuo, y en la segunda se realiza una discretización aproximada de la solución obtenida en la primera etapa.

La convergencia a un óptimo global continuo se garantiza a través de la transformación ortonormal de la representación en vectores propios de la matriz W normalizada, por tanto resulta conveniente replantear el problema de optimización en términos de Z^* . Para esto, se define Z como la matriz escalada de particiones (Pak et al., 1993), que corresponde a un escalamiento de M de la forma

$$Z = M(M^T D M)^{-\frac{1}{2}} \tag{14}$$

Dado que $M^T D M$ es una matriz diagonal, las columnas de Z corresponden a las columnas de M escaladas por el inverso de la raíz cuadrada del grado de W . Con lo anterior, puede establecerse que la condición que deben cumplir las soluciones factibles del problema de optimización es $Z^T D Z = I_k$, donde I_k representa una matriz identidad de $k \times k$. Omitiendo las restricciones iniciales, se puede plantear un nuevo problema de optimización empleando la matriz Z , así:

$$\max \varepsilon(Z) = \frac{1}{k} \text{tr}(Z^T W Z) \tag{15}$$

$$\text{s.a. } Z^T D Z = I_k \tag{16}$$

El hecho de emplear la variable Z en un dominio continuo y omitir las restricciones iniciales, hace que el problema discreto de optimización se convierta en un problema continuo relativamente sencillo. Resolver este problema implica tener en cuenta la propiedad de invariancia ortonormal y el criterio de la solución óptima basada en vectores propios (solución propia).

En adelante, el problema de optimización continuo del método de desviaciones normalizadas se denotará como *NCPZ*.

Proposición 1.1. (Invariancia ortonormal) Sea \mathbf{R} una matriz de $k \times k$ y \mathbf{Z} una posible solución a *NCP*, entonces $\mathbf{ZR} = \mathbf{R}^T \mathbf{R} = \mathbf{I}_k$ y $\varepsilon(\mathbf{ZR}) = \varepsilon(\mathbf{Z})$.

Por tanto, una solución factible de *NCPZ*, pese a la rotación arbitraria debida a la transformación aplicada, conserva las mismas propiedades, es decir, sigue siendo igualmente probable a ser considerada como un óptimo global. En la Proposición 1.2 se plantea un nuevo enfoque del *NCPZ* orientado a la representación en términos de \mathbf{Z}^* , y se comprueba que entre los óptimos globales se encuentran los vectores propios de la matriz de afinidad normalizada \mathbf{P} definida como:

$$\mathbf{P} = \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}} \quad (17)$$

La matriz \mathbf{P} es estocástica (Maila & Shi, 2001), por tanto puede verificarse que $\mathbf{1}_n$ es un vector propio trivial de \mathbf{P} asociado al mayor valor propio que, en este caso, es 1 debido a la normalización.

Proposición 1.2. (Solución propia óptima) Sea (\mathbf{V}, \mathbf{S}) la descomposición propia de \mathbf{P} :

$\mathbf{PV} = \mathbf{VS}$, $\mathbf{V} = [v_1, \dots, v_n]$ y $\mathbf{S} = \text{Diag}(s)$ con valores propios ordenados de forma decreciente

$$s_1 \geq \dots \geq s_n,$$

obtenida a partir de la solución propia ortonormal

$(\tilde{\mathbf{V}}, \mathbf{S})$ de la matriz simétrica $\mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$ donde

$$\mathbf{V} = \mathbf{D}^{-\frac{1}{2}} \tilde{\mathbf{V}} \quad (18)$$

$$\mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}} \tilde{\mathbf{V}} = \tilde{\mathbf{V}} \mathbf{S}, \quad \tilde{\mathbf{V}}^T \tilde{\mathbf{V}} = \mathbf{I}_n \quad (19)$$

entonces cualquier conjunto de k vectores propios es un candidato a ser un óptimo local de *NCPZ* y los k primeros vectores conforman el óptimo global.

El valor de la función objetivo de los conjuntos candidatos a ser solución de *NCPZ*, se puede escribir como

$$\varepsilon[v_{\pi_1}, \dots, v_{\pi_k}] = \frac{1}{k} \sum_{l=1}^k s_{\pi_l} \quad (20)$$

donde $\boldsymbol{\pi}$ es un vector de índices compuesto por k enteros distintos de $\{1, \dots, n\}$. Entonces, de acuerdo a la Proposición 1.2, los óptimos globales son obtenidos con $\boldsymbol{\pi} = \{1, \dots, k\}$ y pueden escribirse como:

$$\mathbf{Z}^* = [v_1, \dots, v_k] \quad (21)$$

cuya matriz de valores propios y función objetivo son, respectivamente:

$$\mathbf{A}^* = \text{Diag}([s_1, \dots, s_k]) \quad (22)$$

$$\varepsilon(\mathbf{Z}^*) = \frac{1}{k} \text{tr}(\mathbf{A}^*) = \max_{\mathbf{Z}^T \mathbf{D} \mathbf{Z} = \mathbf{I}_k} \varepsilon(\mathbf{Z}) \quad (23)$$

En conclusión, el óptimo global de *NCPZ* es un subespacio generado por los vectores propios asociados a los k mayores valores propios de \mathbf{P} a través de matrices ortonormales:

$$\{\mathbf{Z}^T \mathbf{R} : \mathbf{R}^T \mathbf{R} = \mathbf{I}_k, \mathbf{P} \mathbf{Z}^* = \mathbf{Z} \mathbf{A}^*\} \quad (24)$$

El valor de función objetivo de las soluciones de la forma $\mathbf{Z}^* \mathbf{R}$, representa una cota superior al valor de las soluciones de *NCPM*. Además, esta representación satisface las propiedades de acotado superior por medio de los valores propios y monotonicidad decreciente con respecto al incremento del número.

Después de obtener la solución en términos de los vectores propios, es necesario aplicar una transformación de \mathbf{Z} para volver al problema inicial, por tanto si \mathbf{T} es la función de mapeo de \mathbf{M} a \mathbf{Z} , entonces \mathbf{T}^{-1} es la normalización de \mathbf{M} , es decir:

$$\mathbf{Z} = \mathbf{T}(\mathbf{M}) = \mathbf{M}(\mathbf{M}^T \mathbf{D} \mathbf{M})^{-\frac{1}{2}} \quad (25)$$

$$\mathbf{M} = \mathbf{T}^{-1}(\mathbf{Z}) = \text{Diag}(\text{diag}(\frac{1}{2}(\mathbf{Z} \mathbf{Z}^T))) \mathbf{Z} \quad (26)$$

Donde $\text{diag}(\cdot)$ representa la diagonal de la matriz

de su argumento. Desde una perspectiva geométrica, considerando las filas de \mathbf{Z} como coordenadas de un espacio k -dimensional, se puede decir que T^{-1} es una normalización en longitud que ubica los puntos dentro de una hiper-esfera unitaria. Con esta normalización, se transforma un óptimo continuo $\mathbf{Z}^* \mathbf{R}$ del espacio de \mathbf{Z} al espacio de \mathbf{M} . Dado que \mathbf{R} es ortonormal se tiene que:

$$T^{-1}(\mathbf{Z}^* \mathbf{R}) = T^{-1}(\mathbf{Z}^*) \mathbf{R} \quad (27)$$

Esta simplificación permite caracterizar directamente los óptimos continuos con $T^{-1}(\mathbf{Z}^*)$ en el espacio de \mathbf{M} , así:

$$\{\tilde{\mathbf{M}}^* \mathbf{R} : \tilde{\mathbf{M}} = T^{-1}(\mathbf{Z}^*), \mathbf{R}^T \mathbf{R} = \mathbf{I}_k\} \quad (28)$$

La segunda etapa para resolver NCPMes obtener la solución discreta óptima (OPD: Optimal discretization problem). Los óptimos de NCPZ, en general, no son soluciones óptimas de NCPM, sin embargo resultan útiles para obtener la solución discreta. Dicha solución no puede ser el absoluto maximal de NCPM, pero sí es aproximado al óptimo global debido a la naturaleza continua (continuidad) de la función objetivo. Con esto, el objetivo de la discretización es encontrar una solución discreta que satisfaga las condiciones binarias del problema original, mostradas en (8) y (13), de forma que las soluciones encontradas se aproximen al óptimo continuo, como se describió en (28).

Resolver OPD implica llevar a cabo dos etapas alternantes: primero, encontrar el óptimo discreto aproximado $\tilde{\mathbf{M}}^* \mathbf{R}$ (OPDR), luego determinar la transformación ortonormal más apropiada (OPDM), e iterar hasta lo que establezca un determinado criterio de convergencia. Este método iterativo solamente garantiza la convergencia a un óptimo local, pero este puede aproximarse a un óptimo global empleando una buena estimación inicial. Este método es ampliamente explicado en Stella & Shi (2003).

2.2.6 Medida de afinidad

El grado de relación o conexión entre nodos, también denominado grado de afinidad, se puede representar con una matriz simétrica y semidefinida positiva. En-

tonces, una forma aleatoria de establecer la afinidad es $\mathbf{W}_r = \mathbf{X}_r \mathbf{X}_r^T$, donde \mathbf{X}_r es una matriz aleatoria de $n \times d$. La forma trivial corresponde a una matriz que captura los productos internos entre todas las filas de la matriz de datos, es decir, $\mathbf{W}_t = \mathbf{X} \mathbf{X}^T$, donde \mathbf{X} corresponde a la matriz de datos. Esta es la medida de afinidad de uso más frecuente.

Dado que la afinidad es una medida de similitud, es decir, tiene un sentido opuesto a la distancia o disimilitud, intuitivamente podrían plantearse alternativas como $w_{ij} = 1/d(x_i, x_j)$ para $(i \neq j)$ con, donde $d(\cdot, \cdot)$ es alguna medida de distancia. Esta medida de afinidad presenta una relación inversamente proporcional a la distancia, y esto no es conveniente en todos los casos, porque podría generar cambios bruscos en la matriz \mathbf{W} , por esta razón, se han propuesto versiones suaves empleando una función exponencial, así: $w_{ij} = e^{-d^2(x_i, x_j)}$ con $w_{ii} = cte$. El valor de la constante w_{ii} se establece de acuerdo con las condiciones de análisis, en caso de que no sea de interés la afinidad de un elemento con el mismo, $w_{ii} = 0$.

Una mejora a la anterior medida, propuesta en Zelnik & Perona (2004), corresponde a la afinidad exponencial escalada de forma que $w_{ij} = e^{-d^2(x_i, x_j)/(\sigma_i \sigma_j)}$, con $w_{ii} = 0$, donde $\sigma_i = d(x_i, x_N)$ y x_N es el N -ésimo vecino más cercano. La selección de N es independiente de la escala y se da en función de la dimensión de los datos (Chan et al., 1994). En la Tabla 1 se muestran algunas medidas de afinidad.

Tabla 1. Medidas de afinidad

Medida de afinidad	Expresión
Aleatoria	$\mathbf{W}_r = \mathbf{X}_r \mathbf{X}_r^T$, donde \mathbf{X}_r es una matriz aleatoria de $n \times d$
Trivial	$\mathbf{W}_t = \mathbf{X} \mathbf{X}^T$, donde \mathbf{X} corresponde a la matriz de datos
Exponencial	$w_{ij} = e^{-d^2(x_i, x_j)}$ con $w_{ii} = cte$

Exponencial escalada	$w_{ij} = e^{-d^2(x_i, x_j) / (\sigma_i \sigma_j)}$, con $w_{ii} = 0$, donde $\sigma_i = d(x_i, x_N)$ y x_N es el N -ésimo vecino más cercano (Zelnik & Perona, 2004)
----------------------	--

Fuente: Elaboración Propia (2013)

2.2.7 Estimación automática del número de grupos

La estimación del número de grupos puede lograrse teniendo en cuenta los vectores propios de la matriz de afinidad (Zelnik & Perona, 2004). Para esto se divide la matriz \mathbf{P} en c submatrices $\{\mathbf{P}^1, \dots, \mathbf{P}^c\}$, y luego se genera una matriz diagonal por bloques de $n \times c$. El valor de $c (c \leq n)$ representa la cantidad tentativa de grupos con la que se inicia la heurística de búsqueda, este parámetro se fija manualmente. La matriz diagonal por bloques $\hat{\mathbf{V}}$ agrupa los vectores propios en la diagonal de la matriz y los demás elementos son 0, así:

$$\hat{\mathbf{V}} = \begin{bmatrix} \mathbf{v}^1 & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \dots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n & \mathbf{v}^c \end{bmatrix}_{n \times c} \quad (29)$$

Donde \mathbf{v}^i es de $1 \times n$ y representa el i -ésimo vector propio de la submatriz \mathbf{P}^i , y $\mathbf{0}_n$ es un vector columna n -dimensional en el que todos sus elementos son 0.

El valor propio igual a 1 se repite con multiplicidad igual a la cantidad de grupos, por lo mismo, la solución propia pudo haber sido generada por cualquier otro conjunto de vectores ortogonales que abarcan el subespacio de las mismas columnas de $\hat{\mathbf{V}}$. Por tanto, $\hat{\mathbf{V}}$ puede ser reemplazado por $\mathbf{V} = \hat{\mathbf{V}}\mathbf{R}$, donde \mathbf{R} es cualquier matriz de rotación ortonormal de $c \times c$. Para esto se debe garantizar que exista una rotación $\hat{\mathbf{R}}$, tal que cada una de sus filas tenga un único elemento diferente de 0.

Sea $\mathbf{Z} = \mathbf{V}\mathbf{R}$ y $\beta_i = \max_j z_{ij}$, entonces para encontrar un espacio en donde los grupos queden bien representados, en términos de separabilidad, se puede plantear la siguiente función de costo:

$$J = \sum_{i=1}^n \sum_{j=1}^c \frac{z_{ij}^2}{\beta_i^2} \quad (30)$$

Minimizar esta función de costo sobre todas las posibles rotaciones genera la mejor rotación de los vectores propios, por tanto el número de grupos se escoge como el valor comprendido entre 1 y c que genere el valor mínimo de J . En caso de que más de un valor cumpla esta condición, se escoge el mayor de ellos. El proceso de optimización se lleva a cabo empleando el esquema del gradiente descendiente estocástico descrito en Zelnik & Perona (2004).

3. Resultados y discusión

La base de datos está conformada por fotografías de alta resolución de ramas de café con iluminación controlada. El contenido de cada imagen corresponde a granos de café verdes y maduros, y algunos tallos. En la Figura 4 se muestran algunos ejemplos de la base de datos.

Las imágenes fueron caracterizadas usando los modelos de color: HSV, YCC, LAB y LUV. Estas características conforman el conjunto de datos de 45288 píxeles (filas) por 12 características (columnas). Los píxeles fueron agrupados empleando el algoritmo de agrupamiento espectral basado en el criterio de particiones normalizadas, aplicando diferentes medidas de afinidad y estimando el número de grupos. Los resultados se aprecian en las Figuras 5 y 6, y en la Tabla 2. La calidad del agrupamiento se mide en términos del porcentaje de acierto CP y la medida no supervisada generada por valor de la función objetivo del algoritmo de agrupamiento (ver ecuación 12).



Figura 4. Base de datos de imágenes empleada

Fuente: Elaboración Propia (2013)

En la Figura 5 se presenta una de las imágenes de la base de datos (a), una máscara que representa las etiquetas de la misma imagen, considerando los elementos de interés a segmentar (b), y los resultados del agrupamiento con diferentes medidas de afinidad en las tres imágenes de la derecha (c), (d) y (e). Se aprecia que la afinidad exponencial escalada presentó el mejor desempeño (CP = 98.1 %), entre tanto las medidas de afinidad trivial y exponencial generaron, respectivamente, CP = 96 % y CP = 91.2 %. Esto se debe al escalamiento local de la matriz de afinidad exponencial.

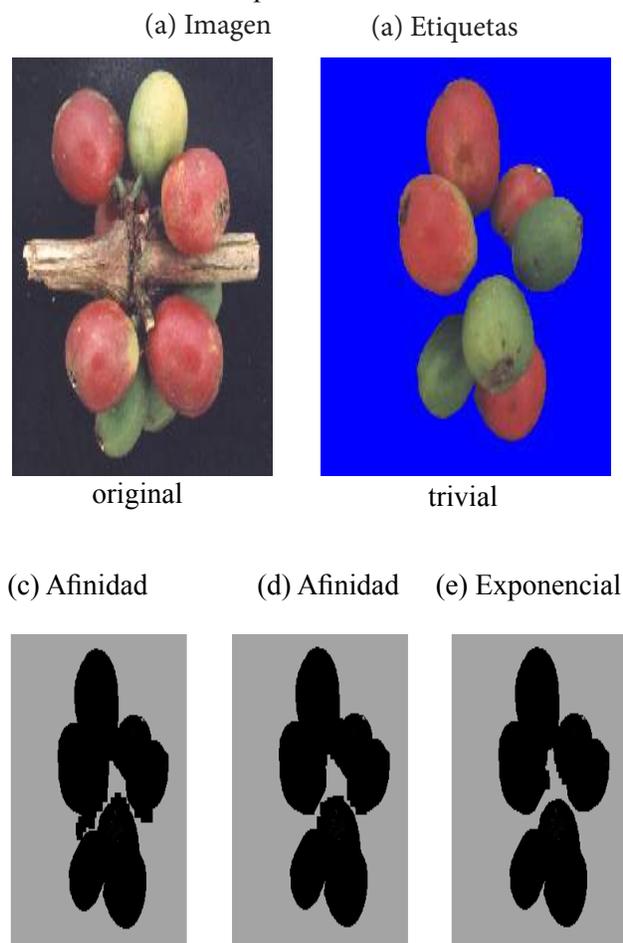


Figura 5. Base de datos de imágenes empleado

Fuente: Elaboración Propia (2013)

En general, se aprecia que las afinidades de tipo exponencial presentan mejor desempeño, esto puede ser atribuido a su naturaleza suave. La afinidad trivial tan solo calcula el producto interno de las observaciones, en cambio la matriz de exponencial es más versátil, puede usarse con diversas medidas de distancia y escalarse con diferentes valores de ϵ .

En la Tabla 2 se muestran los resultados de la tasa de acierto, el valor de ϵ y el número de grupos estimado obtenidos para 5 imágenes de la base de datos. Se registran los valores de la media y desviación estándar (σ). Aquí también se comprueba que la afinidad exponencial es la que mejor funciona, no solo genera una mejor tasa de acierto y valor de convergencia, sino que mejora la estimación del número de grupos, debido al análisis exploratorio local de los datos.

Tabla 2. Resultados obtenidos sobre la base de datos aplicando métodos de agrupamiento no supervisado

Medida de afinidad	CP (%)	ϵ_M	k
Trivial	90 – 3.71	0.77 – 2.1	8 – 3.12
Exponencial	95.3 – 2.2	0.85 – 1.1	6 – 1.7
Exponencial escalada	97.4 – 1.16	0.89 – 0.9	5 – 0.9

Fuente: Elaboración Propia (2013)

En la Figura 6 se relacionan los valores de la función objetivo y el número de grupos asignado para realizar el agrupamiento, obtenidos con 5 imágenes de la base de datos. Se muestra la media y la desviación estándar para cada valor de k.

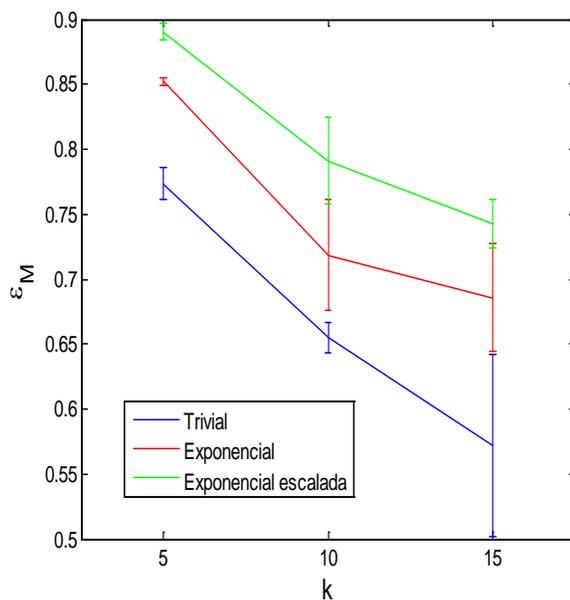


Figura 6. Base de datos de imágenes empleado
Fuente: Elaboración Propia (2013)

Los resultados comprueban la monotonía decreciente, es decir, que el valor de la función objetivo disminuye conforme aumenta la cantidad de grupos. Además, también se aprecia que la afinidad exponencial presenta mejores resultados cuando el número de grupos es el apropiado. En este caso 5 es un número adecuado porque las imágenes tienen 5 clases: café verde, café maduro, tallo, fondo y sombra.

4. Conclusiones y trabajos futuros

El agrupamiento espectral es un método de agrupamiento que presenta buen desempeño, aun cuando los grupos no son linealmente separables, debido a que emplea información de los vectores propios y transformaciones de los datos, aunque es de costo computacional elevado. Por tanto, es necesario diseñar un protocolo de inicialización que mejore la velocidad de convergencia sin disminuir su efectividad. Las matrices de afinidad de tipo exponencial representan una buena alternativa, además se puede incluir una etapa de estimación del número basada en análisis espectral también, con el fin de inicializar apropiadamente el algoritmo de agrupamiento.

La medida no supervisada asociada a la coherencia de los grupos, obtenida a través de un criterio de particionado espectral, es un buen índice del desempeño del

proceso de agrupamiento, porque penaliza el número de grupos y, con una matriz de afinidad apropiada, genera un valor cercano a 1 cuando el agrupamiento es correcto.

Como trabajo futuro, se estudiarán propiedades de algoritmos particionales y espectrales con el fin de desarrollar un sistema de análisis no supervisado que permita superar problemas como costo computacional y sensibilidad a la inicialización, incluyendo etapas de inicialización, estimación del número de grupos y selección de características, y agrupamiento.

5. Agradecimientos

El autor agradece al programa de beca para estudiantes sobresalientes de posgrado de la Universidad Nacional de Colombia y al programa de financiación para Doctorados Nacionales de Colciencias. Igualmente, al grupo de trabajo académico Grupo de Control y Procesamiento Digital de Señales (GC&PDS) de la Universidad Nacional de Colombia – sede Manizales.

Referencias

- Chan, P. K., Schlag, M. D. F. & Zien, J.T. (1994). K-way ratio cut partitioning and clustering. *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, N° 13, pp. 1088-1096.
- Hendrickson, B. & Leland, R. (1995). An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM Journal on Scientific Computing*, Vol. 16, N° 2, pp. 452-469.
- Ishikawa, H. & Geiger, D. (1998). Segmentation by grouping junctions. In *CVPR'98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (p. 125). Washington: IEEE Computer Society.
- Shi, J. & Malik, J. (2003). Normalized cuts and image segmentation. *Proceedings of the Ninth IEEE International Conference on Computer Vision* (p. 313). Washington: IEEE Computer Society.
- Anh, B.T. & Thanh, D. D. (2007). A perron-frobenius theorem for positive quasi polynomial matrices associated with homogeneous difference equations. *Journal of applied mathematics*. Doi:

10.1155/2007/26075

Maila, M. & Shi, J. (2001). Learning segmentation with random walk. In *Neural Information Processing Systems (NIPS)*.

Zelnik, L. & Perona, P. (2004). Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems* (pp. 160-168). N° 17. MIT Press.