

Selección sub-óptima del espectro asociado a la matriz de afinidad

Luciano Lorenti, Lucía Violini, Javier Giacomantone

Instituto de Investigación en Informática (III-LIDI),
Facultad de Informática - Universidad Nacional de La Plata - Argentina.
La Plata, Buenos Aires, Argentina.
{llorenti,lviolini,jog}@lidi.info.unlp.edu.ar

Resumen. En este artículo se presenta un método de agrupamiento espectral que incorpora un etapa de selección sub-óptima de características. Los métodos de agrupamiento espectral tienden a determinar la estructura subyacente en un conjunto de patrones, donde otros métodos convencionales por la disposición y características particulares de los agrupamiento, no obtienen los resultados esperados. En este trabajo se propone utilizar un método particular de selección de características que tiene como objetivo determinar el mejor conjunto de autovectores de la matriz de afinidad normalizada. La determinación correcta del subconjunto de autovectores más relevantes, puede ser utilizada para mejorar las características de las particiones generadas. El método es evaluado con datos sintéticos simulando estructuras de datos específicas, y en datos reales obtenidos con una cámara de tiempo de vuelo.

Palabras clave: Clustering espectral, Selección de características, Visión por Computadora, Robótica.

1 Introducción

Una etapa general de un sistema de reconocimiento automático de patrones es la de reducción de dimensión [1][2]. La reducción del número de características puede realizarse en modo supervisado o no supervisado y las técnicas empleadas pueden ser clasificadas como de selección o de extracción de características [3]. En selección de características se definen criterios para elegir el mejor subconjunto de características de un conjunto original [4][5], y en extracción de características se generan nuevas características mediante transformaciones lineales o no lineales del mismo. El problema particular para el cual el sistema es diseñado y el tipo de datos que define los objetos o fenómenos tratados, determinará si se incluirán las dos etapas, una de extracción y una de selección, o solamente una etapa. El diseño de un módulo de selección de características involucra múltiples consideraciones sobre el conjunto de patrones disponibles, como valores atípicos, imputación de valores, tipo de normalización y compromisos en el número óptimo de características, dadas por el fenómeno de pico [6]. Sistemas con restricciones

temporales o donde el número de características es mucho mayor que el número de muestras son alguno de los casos en que se plantea utilizar sistemas de reducción de dimensión. En este trabajo se propone un método en el cual se seleccionan autovectores en el contexto de un método de agrupamiento (clustering) y no como una etapa clásica en el diseño de un sistema de clasificación. Cuando la estructura de los datos no corresponde a regiones convexas, no es lineal o cuando los métodos clásicos de agrupamiento, jerárquicos o particionales, no obtienen resultados satisfactorios, una alternativa son los métodos de agrupamiento espectral (spectral clustering) [7][8][9]. Las técnicas de agrupamiento espectral utilizan los autovectores de la matriz de afinidad, o de matrices derivadas, para generar una partición del conjunto de muestras en agrupamientos disjuntos, que presenten valores altos de la medida de semejanza adoptada para patrones en un mismo conjunto, y bajos para patrones de conjuntos diferentes. En general el valor de los correspondientes autovalores determina un criterio que permite establecer la prioridad de los autovectores, esta no necesariamente genera la mejor partición del espacio muestral. Es posible entonces aplicar técnicas de selección de características para determinar qué combinación de autovectores genera la mejor partición [10][11][12]. En este artículo se propone un método que combina el potencial de los métodos de agrupamiento espectral con un método específico de selección sub-óptima de características [13]. Se presentan resultados experimentales utilizando datos artificiales e imágenes de rango reales. En particular se utiliza una cámara de tiempo de vuelo [14], utilizada en aplicaciones generales de visión por computadora [15] y robótica [16][17].

En la sección 2 se describe el método de selección sub-óptima de búsqueda secuencial flotante hacia adelante adoptado. En la sección 3 se describe método de cortes normalizados y en la sección 4 se presenta el método propuesto. En la sección 5 se muestran resultados experimentales. Finalmente, en la sección 6 se presentan las conclusiones.

2 Selección Sub-óptima de características

2.1 Selección de Características

El problema de selección de características consiste en, dado un conjunto Y de D características determinar cuál es el subconjunto X de tamaño $d < D$ que genera la mayor contribución a la discriminación entre clases. Se puede plantear el problema en términos de la optimización de una función criterio J para un subconjunto de tamaño d .

$$J(X) = \max_{Z \subset Y, |Z|=d} J(Z)$$

El método de selección determinará una función criterio y el algoritmo de búsqueda adecuado para el problema analizado. Una vez determinada la función criterio es necesario definir un algoritmo de búsqueda que tenga como objetivo un compromiso entre optimización y complejidad que resulte viable. El método sub-óptimo

adoptado en este artículo es el método de búsqueda secuencial flotante hacia adelante (SFFS - Sequential Forward Floating Selection) [13][18] y la función criterio adoptada basada en matrices dispersas. El método SFFS fue propuesto para evitar el problema de anidamiento que genera el método de búsqueda secuencial hacia adelante (SFS - Sequential Forward Selection) [19].

2.2 Búsqueda Secuencial Flotante hacia Adelante

Dado el conjunto completo de D mediciones $Y = \{y_j \mid j = 1, \dots, D\}$, seleccionar k características, para formar el conjunto $X_k = \{x_j \mid j = 1, \dots, k, x_j \in Y\}$, que optimice la función criterio $J(X_k)$ correspondiente.

Inicialización:

$$X_0 := \emptyset; \quad k := 0$$

en la práctica se puede comenzar con $k = 2$ por aplicar SFS dos veces.

Paso 1 (*Inclusión*):

$$x^+ := \arg \max_{x \in Y - X_k} J(X_k + x)$$

x^+ es la característica más significativa con respecto a X_k .

$$X_{k+1} := X_k + x^+; \quad k := k + 1$$

Paso 2 (*Exclusión Condicional*):

$$x^- := \arg \max_{x \in X_k} J(X_k - x)$$

x^- es la característica menos significativa en X_k .

```

if  $J(X_k - \{x^-\}) > J(X_{k-1})$  then
     $X_{k-1} := X_k - x^-$ ;  $k := k - 1$ 
    ir a Paso 2
else
    ir a Paso 1

```

Terminación:

Parar cuando k es igual al número de características requerido

2.3 Medida de Semejanza entre clases

Debido al uso recursivo de la función criterio, se propone utilizar la siguiente medida de separación entre clases basada en matrices dispersas.

$$J = \text{tr}\{S_w^{-1} S_m\} \quad \text{donde} \quad S_w = \sum_{i=1}^M P_i \Sigma_i \quad \text{y} \quad S_m = E[(x - \mu_0)(x - \mu_0)^T]$$

siendo M el número de clases y P_i la probabilidad a priori de cada clase.

3 Agrupamiento Espectral

El algoritmo de cortes normalizados propuesto por Shi y Malik [20] modela la segmentación de imágenes como un problema de partición de un grafo. Un grafo $G=(V,E)$ está formado por un conjunto de vértices V y un conjunto de aristas E que relacionan elementos de V . Con el objetivo de construir grafos a partir de imágenes, los vértices son generados a partir de los pixeles que la constituyen. El conjunto de las aristas E está formado por elementos que denotan la semejanza entre los pixeles. Dado un grafo pesado no dirigido $G = (V, E)$, donde V son los nodos y E son las aristas. Sea A,B una partición de un grafo: $A \cup B = V, A \cap B = \emptyset$. La semejanza entre dos grupos es llamada *cut*

$$cut(A, B) = \sum_{i \in A, j \in B} w(i, j)$$

donde $w(i, j)$ es el peso de la arista que conecta el vértice i con el vértice j . El criterio propuesto por Shi y Malik utiliza un criterio de semejanza normalizado para evaluar la partición:

$$Ncut(A, B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)}$$

Una de las ventajas más importantes para usar el criterio de cortes normalizados es que se puede obtener una buena aproximación de la partición óptima de forma muy eficiente. Sea $W_{ij} = w(v_i, v_j)$ la matriz de pesos del grafo y sea D la matriz diagonal de forma que $D_{ii} = grado(v_i) = \sum_{v_j \in V} w(v_i, v_j)$

Shi y Malik demostraron que una partición óptima se puede obtener calculando:

$$y = \arg \min_y Ncut = \arg \min_y \frac{y^T (D - W) y}{y^T D y}$$

donde y es un vector indicador binario que especifica a qué grupo pertenece cada pixel. Relajando el carácter discreto de y , la ecuación anterior puede ser aproximada resolviendo el sistema de autovalores generalizado:

$$(D - W)y = \lambda D y$$

El segundo autovector de este sistema es la solución real del problema discreto de cortes normalizados.

3.1 Ncut simultáneo de k-vías

Se define el corte normalizado simultáneo de k-vías que da como resultado k segmentos en una sola iteración de la siguiente forma:

$$Ncut(A_1, A_2, \dots, A_k) = \frac{cut(A_1, A_1)}{assoc(A_1, V)} + \frac{cut(A_2, A_2)}{assoc(A_2, V)} + \dots + \frac{cut(A_n, A_n)}{assoc(A_n, V)}$$

Sea $L = (D - V)$, dado un vector indicador v de un sub-grafo A_j tal que

$$v_i = \begin{cases} \frac{1}{\sqrt{\text{assoc}(A_n, V)}} & \text{si } i \in A_j \\ 0 & \text{si } i \notin A_j \end{cases}$$

resulta que

$$v_i^T L v_i = \frac{\text{cut}(A_j, A_j)}{\text{assoc}(A_j, V)}$$

Sea H la matriz formada por k vectores indicadores puestos en columnas, minimizar $Ncut(A_1, A_2, \dots, A_k)$ es equivalente a minimizar:

$$\min_{A_1, A_2, \dots, A_k} \text{Tr}(H^T L H) \text{ sujeto a } H^T D H = I$$

Se puede hallar eficientemente una aproximación continua de los autovectores discretos que minimizan la traza. Estos son los autovectores correspondientes a los k autovalores más pequeños del sistema de autovalores $(D - W)u = \lambda D u$.

4 Método Propuesto

Sea $M \in R^{n \times m}$ una matriz que contiene n patrones de m características.

1. Se construye la matriz de afinidad W de forma que

$$W(i, j) = w(v_i, v_j)$$

Para imágenes se utiliza:

$$w(v_i, v_j) = e^{\frac{-\|F(i) - F(j)\|_2}{\alpha_I}} * \begin{cases} e^{\frac{-\|X(i) - X(j)\|_2}{\alpha_X}} & \text{si } \|X(i) - X(j)\|_2 < r \\ 0 & \text{c. c.} \end{cases}$$

donde $X(i)$ es la locación espacial (x, y) del pixel i y $F(i)$ es el nivel de intensidad del pixel i .

2. Se calcula la matriz diagonal D tal que $D(i, i) = \text{grado}(v_i) = \sum_{v_j \in V} w(v_i, v_j)$
3. Sea H la matriz formada por los k autovectores correspondientes a los autovalores más pequeños del sistema de autovalores $(D - W)u = \lambda D u$.
4. Considerando las columnas de H como características y las filas como patrones se aplica el método de selección sub-óptima descrito en la sección 2.
5. Se aplica el algoritmo k -medias sobre el conjunto de patrones utilizando las mejores características.

5 Resultados Experimentales

En esta sección se presentan resultados experimentales preliminares del método propuesto aplicado a imágenes simuladas y a imágenes reales. Las evaluaciones comparativas del método han sido realizadas con distintos conjuntos de imágenes de rango, y con diferentes parámetros en la adquisición de las mismas. Las capturas reales fueron obtenidas utilizando la cámara de tiempo de vuelo MESA SwissRanger SR4000 [14] y las imágenes simuladas utilizando el software Blensor [21].

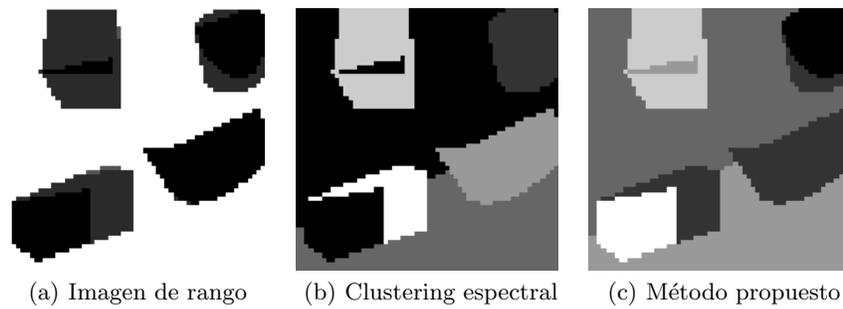


Fig. 1. Resultados sobre imagen de rango simulada

En la figura 1(a) se muestra una imagen de rango simulada compuesta por cuatro objetos equidistantes del plano focal. La figura 1(b) presenta la segmentación obtenida utilizando las tres primeras columnas de H . En la figura 1(c) se muestra la segmentación obtenida utilizando las columnas de H especificadas en la tabla 1 seleccionadas por el método propuesto.

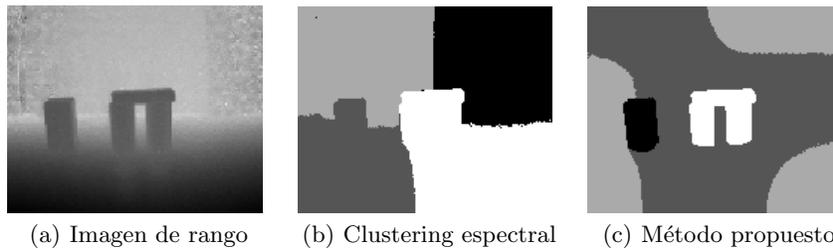


Fig. 2. Resultados sobre imagen de rango real

La figura 2 muestra imágenes del experimento realizado sobre una imagen real obtenida utilizando la cámara de tiempo de vuelo MESA SR4000. En la figura

2(a) se muestra una imagen de rango correspondiente a dos objetos. La figura 2(b) presenta la segmentación obtenida utilizando clustering espectral. En la figura 2(c) se muestra la segmentación obtenida utilizando el método propuesto. En la tabla 1 se presentan los parámetros utilizados y las columnas de H seleccionadas.

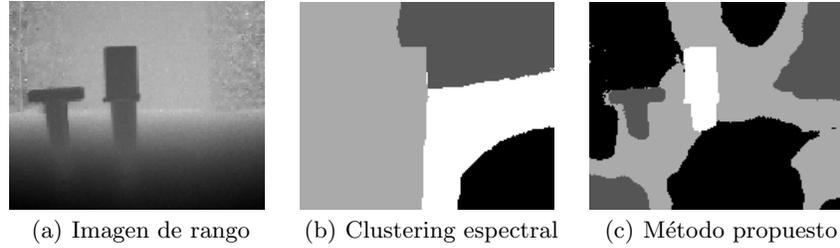


Fig. 3. Resultados sobre imagen de rango real

La figura 3 muestra imágenes del experimento realizado sobre otra imagen real. En la figura 3(a) se muestra una imagen de rango correspondiente a dos objetos con niveles de profundidad similares. La figura 3(b) presenta la segmentación obtenida utilizando clustering espectral. En la figura 3(c) se muestra la segmentación obtenida utilizando el método propuesto. En la tabla 1 se presentan los parámetros utilizados y las columnas de H seleccionadas.

	Método	Columnas de H	r	α_I	α_X
Prueba 1	Espectral	1,2,3	4	8	2
	Propuesto	1,2,7	4	8	2
Prueba 2	Espectral	1,2,3	4	0.02	4
	Propuesto	5,7,10	4	0.02	4
Prueba 3	Espectral	1,2	4	0.045	4
	Propuesto	4,5	4	0.045	4

Tabla 1.

6 Conclusiones

En este artículo se propone utilizar un método de selección de características sub-óptimo embebido en un método de clustering espectral. Los resultados experimentales preliminares sobre datos simulados y reales permiten concluir que la selección adecuada del conjunto de autovectores de la matriz de afinidad mejora la determinación de la estructura subyacente en los patrones de entrada ya sea en datos simulados como reales. El método propuesto muestra potencial para su aplicación en imágenes de rango y de intensidad para problemas que presenten restricciones temporales, si se utilizan cámaras con características similares a la SR4000 de bajo tiempo de adquisición. El método es sensible al ajuste de los principales parámetros del mismo, por lo tanto como parte del trabajo futuro es necesario analizar el comportamiento de los mismos con respecto a los agrupamientos generados. Un segundo aspecto involucra evaluar el mecanismo de selección en otros métodos espectrales y realizar una evaluación comparativa de los resultados para estructuras de agrupamientos complejas.

Referencias

1. M. A. Carreira-Perpinan. A review of dimension reduction techniques. Technical report CS-96-09, Department of Computer Science, University of Sheffield, 1997.
2. I. K. Fodor. A survey of dimension reduction techniques. Technical report URL-ID-148494, Center for Applied Scientific Computing, Lawrence Livermore Laboratory, 2002.
3. P. Narendra, K. Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, vol. C-26, no. 9, pp.917-922, 1977.
4. A. Blum, P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, vol. 97(1-2), pp. 245-271, 1997.
5. M. Kudo, J. Sklansky. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, vol. 33, pp. 25-41, 2000.
6. R. E. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.
7. J. Shi, J. Malik, Normalized cuts and image segmentation. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 731-737. 1997
8. P. Perona, W. T. Freeman. A factorization approach to grouping. *ECCV*, pp. 655-670. 1999.
9. A. Y. Ng, M. I. Jordan, Y. Weiss. On Spectral Clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, vol. 14, 2002.
10. T. Xiang, S. Gong. Spectral clustering with eigenvector selection. *Pattern Recognition Vol. 41*, 1012-1029, 2008.
11. F. Zhao, L. Jiao, H. Liu, X. Gao, M. Gong. Spectral clustering with eigenvector selection based on entropy ranking. *Neurocomputing Vol. 73*, pp. 1704-1717, 2010.
12. S. A. Toussi, H. S. Yazdi. Feature Selection in Spectral Clustering. *International Journal of Signal Processing, Image Processing and Pattern Recognition* vol. 4, No. 3, pp. 179-194, 2011.
13. P. Pudil, J. Novovicova, J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters* vol. 15, pp. 1119-1125, 1994.

14. M. Cazorla, D. Viejo, C.Pomares. Study of the SR 4000 camera. XI Workshop de Agentes Físicos, Valencia, 2010.
15. A. A. Dorrington, C. D. Kelly, S. H. McClure, A. D. Payne, M. J. Cree. Advantages of 3D Time of Flight Range Imaging Cameras in Machine Vision Applications. 16th Electronics New Zealand. Dunedin, New Zealand, 95-99, 2009.
16. S. May, B. Werner, H. Surmann, K. Pervolz. 3D time-of-flight cameras for mobile robotics. IEEE International Conference on Intelligent Robots and Systems, pp.790-795, 2006.
17. A. Prusak, O. Melnychuk, H. Roth, I. Schiller, R. Koch. Pose estimation and map building with a Time of Flight camera for robot navigation. International Journal of Intelligence Systems Applications, vol. 5, pp. 355-364, 2008.
18. P. Pudil, F.J. Ferri, J. Novovicova, J. Kittler. Floating Search Methods for Feature Selection with Nonmonotonic Criterion Functions. Proceedings of 12th IAPR International Conference vol. 2, pp. 278-283, 1994.
19. A. W. Whitney. A direct method of nonparametric measurement selection. IEEE Transactions on Computers, vol.20, pp. 1100-1103, 1971.
20. J. Shi, J. Malik, Normalized Cuts and Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Inteligence, Vol. 22, No. 8, pp. 888-905, 2000.
21. M. Gschwandtner; R. Kwitt; A. Uhl, BlenSor: Blender Sensor Simulation Toolbox. Proceedings of 7th International Symposium In Advances in Visual Computing. ISVC 2011, 2011.