FORTHCOMING PAPER ICOR2020-91C20-01

OVERVIEW ON KERNELS FOR LEAST-SQUARES SUPPORT-VECTOR-MACHINE-BASED CLUSTERING: EXPLAINING KERNEL SPECTRAL CLUSTERING

Y. Fernández*[1], I. Marrufo**[2], M. A. Paez**, A. C. Umaquinga-Criollo**, P. D. Rosero***[3], D. H. Peluffo-Ordóñez**,****,*****[2,4,5]

*Universidad Politécnica Estatal del Carchi, Tulcán - Ecuador

**SDAS Research Group (www.sdas-group.com)

***Universidad Técnica del Norte, Ibarra Ecuador

****Yachay Tech University - Urcuquí, Ecuador

*****Corporación Universitaria Autónoma de Nariño – Pasto, Nariño, Colombia.

ABSTRACT

This letter presents an overview on some remarkable basics on kernels as well as the formulation of a clustering approach based on least-squares support vector machines. Specifically, the method known as kernel spectral clustering (KSC) is of interest. We explore the links between KSC and a weighted version of kernel principal component analysis (WKPCA). Also, we study the solution of the KSC problem by means of a primal-dual scheme. All mathematical developments are carried out following an entirely matrix formulation. As a result, in addition to the elegant KSC formulation, important insights and hints about the use and design of kernel-based approaches for clustering are provided.

KEYWORDS: Support vector machine (SVM), Clustering, Kernel spectral clustering KSC, kernel principal component analysis

MSC: 91C20

RESUMEN

En esta investigación, se presenta una descripción general de algunos conceptos básicos e importantes sobre kernels, así como la formulación de un enfoque de agrupación basado en máquinas de vectores de soporte usando mínimos cuadrados. Específicamente, es de interés el método conocido como kernel spectral clustering (KSC). Se explora los enlaces entre KSC y una versión ponderada del análisis de componentes principales con kernels (WKPCA). Además, estudiamos la solución del problema de KSC, por medio de un esquema primal-dual. Todos los desarrollos matemáticos presentados se llevan a cabo siguiendo una formulación completamente matricial. Como resultado, además de una formulación de KSC, se proporcionan nociones y aspectos importantes sobre el uso y el diseño de enfoques basados en kernels para agrupación.

PALABRAS CLAVE: Support vector machine SVM, Clustering, Kernel spectral clustering KSC, kernel principal component analysis WKPC.

1. INTRODUCTION

Clustering is a typical task in many real-life scenarios, and its automated approaches have therefore become a widely-used, powerful tool [9]. Broadly speaking, clustering can be defined as the procedure aiming at splitting a set of objects into subsets -known as clusters. More technically, clustering approaches lie on the field of unsupervised learning wherein the exploratory analysis of data is carried with no prior knowledge nor any supervised information

Among the remarkable applications where clustering has taken place, it is worth mentioning: Medical applications such as cardiac arrhythmia identification from electrocardiographic signals through heartbeat clustering [23, 20, 3, 21], and sleep stages classification via clustering of

[1] *yasmany.fernandez@upec.edu.ec

electroencephalogram derived patterns [22], and patient stratification [10]. Transportation optimization by grouping areas [7]. Moving objects using dynamic data clustering [31, 15, 19]. And very recently, clustering is one of the benchmark methods for Big Data applications, especially, for those involving representation and pre-processing stages for high-dimensional data [14, 13, 1].

In spite of being widely used, highly recommended for exploratory analysis problems, and greatly versatile [9], clustering algorithms -even the most recent ones- still lack a definite solution to local minima convergence issue. In the pursue of such a solution, the density-based [5], deep learning [8], and spectral [29] approaches have shown to be the most suitable ones. Indeed, some works have studied how to make a decision on what is the optimal clustering alternative for a certain application [17].

In this article, we focus on spectral clustering (SC), which estimates the cluster indicators from the spectrum of the graph Laplacian related to a weighting matrix, also called kernel matrix. Literature reports that SC is able to dealing with complex-structure data (holding highly non-linearly separable classes) [18]. Nonetheless, when the calculation of the eigenvectors is prohibitive due to complexity or computational cost issues, there are alternatives that solve the same graph-based clustering formulation with no spectral calculations. For instance, by means of either a heuristic search [3] or using quadratic forms formulations [11].

This letter is intended to overview some concepts on kernels to devise a clustering approach followed from support vector machines (SVM) within a least-squares framework. Particularly, we outline the so-named kernel spectral clustering (KSC) proposed in [2], which is based on a weighted kernel principal component analysis (WKPCA) interpretation of spectral clustering with primal-dual least squares SVM formulations. The contribution of this work is that not only a fully matrix, and elegant formulation for KSC is provided but also some previous works have been extended [28] to provide more insights and hints to study and use kernel-based approaches for clustering.

The scope of this manuscript encompasses the description of the kernel spectral clustering method in Section 3. To do so, it starts by explaining some definitions and basics about kernels in Section 2. Some experimental results are presented in Section 4. Finally, some concluding remarks are drawn in Section 5.

2. KERNELS FOR CLUSTERING

This section gathers new perspective to define the slippery concept of "kernel" within the context of data exploratory analysis. From a geometric point of view, the term kernel can be understood as a function quantifying somehow the similarity among given input elements -which relies on graph theory.

Let us define the data matrix to be clustered as $X \in R^{N\times D} = \left(x_1,\ldots,\ x_N\right)^T = \left(x^{(1)},\ldots,x^{(D)}\right)$, where $x_i \in R^D$ is the i-th data sample, $x^{(l)} \in R^N$ is the l-th variable, $i \in \{1,\ldots,N\}$, and $l \in \{1,\ldots,D\}$. Also, let us consider a function to map from the D-dimensional space to that D_h dimensional one is in the form $\phi(\cdot)$, such that:

$$\begin{aligned} \phi(\cdot) &: \mathrm{R}^d \rightarrow R^{d_h} \\ x_i &\rightarrow \phi(x_i) \end{aligned} \tag{2.1}$$

Then, the matrix $\Phi = \left(\phi(x_1)^T,\ldots,\ \phi(x_N)^T\right)^T, \Phi \in R^{N\times D_h}$ becomes a high dimensional representation of the original data X.

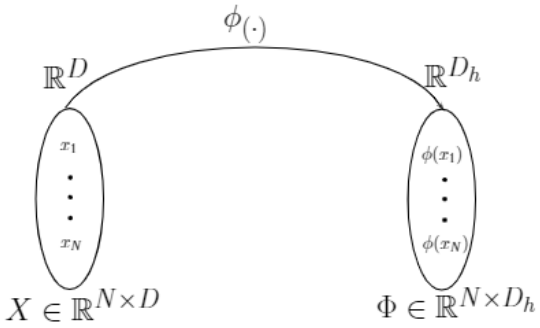

Figure 1: High dimensional mapping

That said, we may define kernels as functions allowing for mapping from a D-dimensional input space representing a data set to a significantly higher dimension D_h space, where $D_h \ggg D$. In terms of clustering, the advantage of mapping the original data space onto a higher one lies in the fact that the latter space may provide more cluster separability as seen in Figure 2. Furthermore, it must be taken into account that the mapping is done before carrying out any clustering process. Then, the success of the clustering task can be partly attributed to the kernel-matrix-building function when grouping algorithms are directly associated with the chosen kernel.

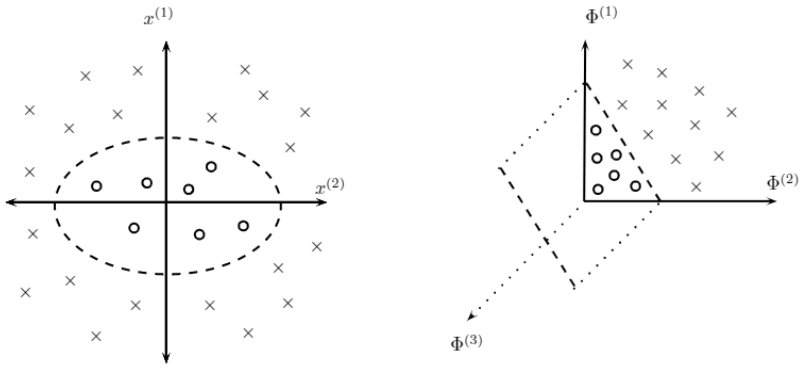

Figure 2: Feature space to high dimension.

Some kernels with special structure aimed to attend particular interests have been proposed. For instance, in [26], a structural cluster kernel (SCK) is introduced incorporating similarities induced by a structural clustering algorithm to improve graph kernels recommended by literature. Also, Mercer kernels have been employed for unsupervised partitioning with automatic estimation of the inherent number of groups [24], as well as for solving multi-cluster problems [4].
This section describes some basics and fundamental aspects regarding kernels, in particular, for clustering purposes. The remaining of this section is as follows: Section 2.1. presents in general terms the definition of kernel and some properties and related concepts. Finally, in Section 2.2., the most common kernel functions coming from positive definite matrix are described.

2.1. Kernel function

In terms of human learning theory [16], one of the fundamental problems is the discrimination among elements or objects. Take the following instance: We have a set of objects formed by two different classes; then, when a new object appears the task is to determine to which class such an object belongs. This is usually done by taking into account the object's properties as well as similarities and differences with regards to the two previously known classes. According to the above, and regarding kernel theory, we need to create or choose a similarity or affinity measure to compare the data. Kernels considered in this work are positive semi-definite, and then their corresponding Gram Matrix (or kernel matrix) is to be positive semi-definite as well. In other words, we can define a kernel function in the form

$$\begin{aligned} \kappa(\cdot,\cdot) &: \mathrm{K}^d \times K^d \rightarrow K \\ x_i, x_j &\rightarrow \kappa(x_i, x_j) \end{aligned} \tag{2.2}$$

where $K = C$ or R. Note that in this case we have assumed elements x_i to be real and D-dimensional. Then, if we have a total of N elements or data points, a $N \times N$ matrix K with entries $k_{ij} = \kappa(x_i, x_j)$ is called Gram matrix or kernel matrix as well. Therefore, the kernel matrix must be a positive semi-definite matrix, i.e., a $N \times N$ complex matrix K satisfying

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i \bar{c}_j k_{ij} \geq 0 \tag{2.3}$$

for all $c_i \in C$ is called positive definite, being $\bar{c}_i$ the complex conjugate of c_i. Similarly, a real symmetric $N \times N$ matrix K satisfying 2.3 for all $c_i \in$ is also called positive definite [25] Note that a symmetric matrix is positive if and only if all its eigenvalues are non-negative. In the literature, a number of different terms are used for positive definite kernels, such as reproducing kernel, Mercer kernel, admissible kernel, support vector kernel, non-negative definite kernel and covariance function [25].

An interesting and strongly useful property is the so-called *kernel trick*. This property gains importance in kernel theory, since it permits to replace a positive definite kernel with another kernel that is finite and approximately positive definite. For instance, from a given algorithm formulated in terms of a positive definite kernel $\kappa(\cdot,\cdot)$, one can construct an alternative algorithm by replacing it by another positive semi-definite kernel $\tilde{\kappa}(\cdot,\cdot)$ [25], in such a manner that $\Phi\Phi^T = K$. Then, in this case, kernel $\Phi\Phi^T$ has been estimated as K. The fact to use K as an alternative estimation of $\Phi\Phi^T$ is known as kernel trick.

2.2. Types of kernel functions

Radial basis function (RBF) kernels are those that can be written in terms of similarity or dissimilarity measure, in the form:

$$\kappa(x_i, x_j) = f(\mathrm{d}(x_i, x_j)), \tag{2.4}$$

where $\mathrm{d}(\cdot,\cdot)$ is a measure on the domain of X, in this case R^d, so:

$$\begin{aligned} \mathrm{d}(\cdot,\cdot) &: \mathrm{R}^D \times R^D \rightarrow R^+ \\ x_i, x_j &\rightarrow \mathrm{d}(x_i, x_j) \end{aligned} \tag{2.5}$$

and $f(\cdot)$ is a function defined on R^+. Usually, such measure arises from the inner product;

$$\mathrm{d}(x_i, x_j) = \left\| x_i - x_j \right\| = \sqrt{\left\langle x_i - x_j, x_i - x_j \right\rangle}$$

In Table 1, some conventional kernels are mentioned -all of them are defined over the domain R^D.

Kernel name	Definition
Linear	$\left\langle x_i, x_j \right\rangle$
Polynomial	$\left\langle x_i, x_j \right\rangle^D$

Rational quadratic	$1-\frac{\left\|x_i-x_j\right\|^2}{\left\|x_i-x_j\right\|^2+\sigma}, \sigma \in R^+$
Exponential	$\exp\left(-\frac{\left\|x_i-x_j\right\|}{2\sigma^2}\right), \sigma \in R^+$
Gaussian	$\exp\left(-\frac{\left\|x_i-x_j\right\|^2}{2\sigma^2}\right), \sigma \in R^+$

Table 1: Some conventional kernel functions.

3. LEAST-SQUARES SUPPORT-VECTOR-MACHINE-BASED FORMULATION FOR KERNEL SPECTRAL CLUSTERING

This section outlines the formulation and solution of a clustering problem based on least-squares SVMs, so-named kernel spectral clustering (KSC). KSC -just as any other clustering approach- is aiming to split the input data matrix X into K disjoint subsets. In the following, the clustering model is described. Let $e^{(l)} \in R^N$ be the l-th projection vector, which is assumed in the following latent variable form:

$$e^{(l)} = \Phi w^{(l)} + b_l 1_N, \tag{3.1}$$

Where $w^{(l)} \in R^{D_h}$ is the l-th weighting vector, b_l is a bias term, n is the number of considered latent variables, notation 1_N stands for a N dimensional all-ones vector, and the matrix Φ is a high dimensional representation of the input data. Therefore, $e^{(l)}$ represents the latent variables from a set of n_e binary cluster indicators obtained with $sign(e^{(l)})$, which are to be further encoded to obtain the K resultant groups. Grounded on the least-squares SVM formulation of equation (3.1), the following optimization problem can be stated:

$$max_{e^{(l)},\, w^{(l)},\, b^{(l)}} \frac{1}{2N}\sum_{l=1}^{n_e}\gamma_l\ e^{(l)\mathrm{T}} V e^{(l)} - \frac{1}{2}\sum_{l=1}^{n_e}\ w^{(l)\mathrm{T}} w^{(l)} \tag{3.2a}$$

$$e^{(l)} = \Phi^T\ w^{(l)} + b_l 1_N, \tag{3.2b}$$

where $\gamma_l \in R^+$ is the l-th regularization parameter and $V \in R^{N\times N}$ is a diagonal matrix representing the weight of projections.

Thinking of further analysis, we express the above primal formulation in matrix terms, as follows:

$$\max_{E,\, W,\, b} \quad \frac{1}{2N} tr(E^T\ V E \Gamma) - \frac{1}{2} tr(W^T W) \tag{3.3a}$$

$$E = \Phi\ W + 1_N \otimes b^T \tag{3.3b}$$

where $b = \left(b_1, \ldots, b_{n_e}\right)^T$, $b \in R^{n_e}$, $\Gamma = Diag\left((\gamma_1, \ldots, \gamma_{n_e})\right)$, $W = \left(w^{(1)}, \mathrm{L}\ , w^{(n_e)}\right)$, $W \in R^{D_h \times n_e}$ and $E = \left(e^{(1)}, \mathrm{L}\ ,\ e^{(n_e)}\right), E \in R^{N\times n_e}$. Notations $Diag(\cdot), tr(\cdot)$ and $\otimes$ denote the diagonal matrix formed by its argument vector, the trace and the Kronecker product, respectively. By minimizing the previous cost function, the goals of minimizing the weighting variance of E and maximizing the variance of W are reached simultaneously. Let Σ_E the weighting covariance matrix of E and Σ_W the covariance matrix of W. Since matrix V is diagonal, we have that $tr((V^{1/2}E)^T V^{1/2}E) = tr(\Sigma_E)$. In other words, Σ_E is the covariance matrix of weighted projections, i.e., the projections scaled by square root of matrix V. As well, $tr(W^T\ W) = tr(\Sigma_W)$. Then, KSC can be seen as a Kernel WPCA approach.

3.1. Solving the KSC problem

To solve the KSC problem, we form the corresponding Lagrangian of problem from equation (3.2) as follows:

$$\mathrm{L}(E,W,\Gamma,\ A)=\ \frac{1}{2N}tr\ (\Gamma E^T VE)-\frac{1}{2}tr(W^TW)-tr\ (A^T(E-\ \Phi W-\ 1_N\otimes\ b^T)), \tag{3.4}$$

where matrix $A\in R^{N\times n_e}$ holds the Lagrange multiplier vectors $A=\left(\alpha^{(1)},\mathrm{L}\ ,\ \alpha^{(n_e)}\right)$, and $\alpha^{(l)}\in R^N$ is the l-th vector of Lagrange multipliers.

Solving the partial derivatives on $\mathrm{L}(E,W,\Gamma,\ A)$, to determine the Karush-Kuhn-Tucker conditions, we obtain:

$$\frac{\partial \mathrm{L}}{\partial E}=0\ \ \Rightarrow\ E=N\ V^{-1}A\Gamma^{-1},$$
$$\frac{\partial \mathrm{L}}{\partial\ W}=0\ \ \Rightarrow\ W=\Phi^T A,$$
$$\frac{\partial \mathrm{L}}{\partial\ A}=0,\ \ \Rightarrow\ E=\Phi W,$$
$$\frac{\partial \mathrm{L}}{\partial\ b}=0\Rightarrow b^T 1_N=0.$$

Therefore, by eliminating the primal variables from initial problem (3.2) and assuming a kernel trick such that $\Phi\Phi^{\bullet}\ =\ K$, being $K\in R^{N\times N}$, a given kernel matrix (just as any of those explained in section 2) the following eigenvector-based dual solution is obtained:

$$A\Lambda=\ AV(I_N+(1_N\otimes b^T)(K\ )^{-1})K, \tag{3.5}$$

Where $\Lambda=Diag(\lambda),\ \Lambda\in R^{N\times N},\ \lambda\in R^N$ is the vector of eigenvalues with $\lambda_l\in R^+$.

Also, taking into account that the kernel matrix represents the similarity matrix of a graph with K connected components as well as $V=\ D^{-1}$ where $D\in R^{N\times N}$ is the degree matrix defined as $D=Diag(K1_N)$; then the $K-1$ eigenvectors contained in A, associated to the largest eigenvalues, are piecewise constant and become indicators of the corresponding connected parts of the graph. Therefore, value n_e is fixed to be $K-1$ [2].

With the aim of achieving a dual formulation, but satisfying the condition $b^T 1_N=0$ by centering vector b (i.e. with zero mean), the bias term should be chosen in the form:

$$b_l=-1/(1_N^T\ V1_N)1_N^T\ V\ K\ \alpha^{(l)} \tag{3.6}$$

Thus, the solution of problem of equation (3.3) is reduced to the following eigenvector-related problem:

$$A\Lambda=\ VHKA \tag{3.7}$$

where matrix $H\in R^{N\times N}$ is the centering matrix that is defined as

$$H=\ I_N-\frac{1}{1_N^T\ V\ 1_N}1_N\ 1_N^T V$$

where I_N denotes a N-dimensional identity matrix and, $K=[K_{ij}],K\in R^{N\times N}$, being $K_{ij}=\kappa(x_i,x_j),i,j\in\{1,\ldots,N\}$. As a result, the set of projections can be calculated as follows:

$$E=\ KA+\ 1_N\otimes b^T. \tag{3.8}$$

Once projections are calculated, we proceed to carry out the cluster assignment by following an encoding procedure applied on projections. Because each cluster is represented by a single point in the $K-1$-dimensional eigenspace, such that those single points are always in different orthants due

also to the KKT conditions, we can encode the eigenvectors considering that two points are in the same cluster if they are in the same orthant in the corresponding eigenspace [2]. Then, a code book can be obtained from the rows of the matrix containing the $K-1$ binarized leading eigenvectors in the columns, by using $\text{sign}(e^{(l)})$. Then, matrix $E = \text{sgn}(E)$ is the code book being each row a codeword.

3.2. Out-of-sample extension

KSC can be extended to out-of-samples analysis without re-clustering the whole data to determine the assignment cluster membership for new testing data [2]. In particular, defining $z \in R^{n_e}$ as the projection vector of a testing data point x_{test}, and by taking into consideration the training clustering model, the testing projections can be computed as:

$$z = A^T K_{\text{test}} + b \tag{3.9}$$

where $K_{\text{test}} \in R^{n_e}$ is the kernel vector such that

$K_{\text{test}} = [K_{\text{test}_1}, \ldots, K_{\text{test}_N}]^{\cdot}$,

where $K_{\text{test}_i} = \kappa(x_i, x_{\text{test}})$. Once, the test projection vector z is computed, a decoding stage is carried out that consists of comparing the binarized projections with respect to the codewords in the code book $\widetilde{E}$ and assigning cluster membership based on the minimal Hamming distance [2].

3.3. KSC algorithm

Following the pseudo-code (Algorithm 1) to perform KSC is shown

Algorithm 1 Kernel Spectral Clustering: $[q_{\text{train}}, q_{\text{test}}] = \text{KSC}(X, \kappa(\cdot,\cdot), K)$

1: **Input:** $K, X, \kappa(\cdot,\cdot)$
2: Form the kernel matrix K such that $k_{ij} = \kappa(y_i, x_j)$
3: Determine E through (3.8)
4: Form the training codebook by binarizing $\widetilde{E} = sgn(E)$
5: Assign the output training labels q_{train} according to similar codewords
6: Compute the training codewords for testing
7: Assign the output testing labels q_{test} according to the minimal Hamming distance when comparing with training codewords
8: **Output:** $q_{\text{train}}, q_{\text{test}}$

4. EXPERIMENTAL RESULTS

Based upon outcomes reported in previous works [18], this section gathers some experimental results to highlight the benefits of the here-studied KSC. To do so, conventional methods are used for comparison purposes, namely: kernel k-means (KKM) [33], min cuts (Min-cuts) [27] and multi-cluster spectral clustering (MCSC) [31].
They are all performed over the same conditions: kernel matrix and number of clusters. The segmentation performance is quantified by a supervised index noted as Probabilistic Rand Index (PR) explained in [28], such that P R $\in$ [0, 1], being 1 when regions are properly segmented. Images are drawn from the free access Berkeley Segmentation Data Set [12].
To represent each image as a data matrix, we characterize the images by color spaces (RGB, YCbCr, LAbB, LUV) and the xy position of each pixel. To run the experiment, we resize the images at 20%

of the original size due to memory usage restrictions. All the methods are performed with a given number of clusters K manually set as shown in Fig. 3, and using the scaled exponential similarity matrix as described in [33], setting the number of neighbors to be 9.
Also, another real databases collection is considered that is taken from the UCI repository [10]. For quantitative evaluation of compared clustering methods in terms of performance and stability, the estimated mean value of considered measures are shown in Table 2, which are computed after running algorithms 50 times. Methods are performed by setting the number of groups as the original number of classes. Again, a scaled similarity matrix is used with the number of neighbors equals to 15. We use two well-known clustering measures: Fisher's criterion (*J*) and Silhouette (*S*).

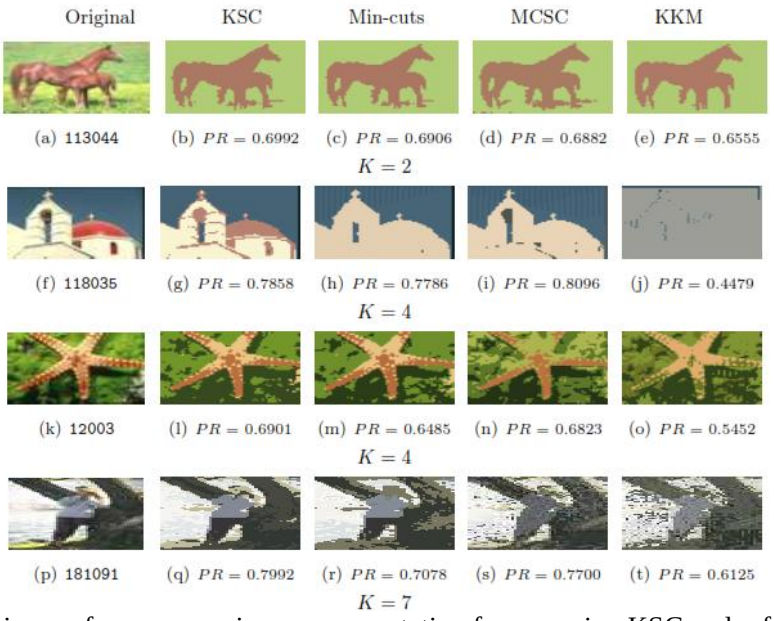

Figure 3: Clustering performance on image segmentation for assessing KSC and reference methods on complex-structure data.

		Method			
Data set	Measure	Min-cuts	MCSC	KKM	KSC
Iris	J	2.8 ± 0.15	2.01 ± 0.1	2.7 ± 0.12	2.9 ± 0.15
	S	0.7 ± 0.009	0.45 ± 0.009	0.55 ± 0.01	0.65 ± 0.011
Biomed	J	1.5 ± 0.09	0.95 ± 0.1	1.1 ± 0.11	1.1 ± 0.3
	S	0.60 ± 0.01	0.25 ± 0.011	0.45 ± 0.014	0.54 ± 0.09
Auto mpg	J	0.69 ± 0.15	0.62 ± 0.13	1.04 ± 0.22	0.80 ± 0.11
	S	0.35 ±0.009	0.30 ± 0.009	0.35 ± 0.009	0.42 ± 0.009
Breast	J	0.85 ± 0.1	0.85 ± 0.12	0.85 ± 0.14	0.85 ± 0.23
	S	0.75 ± 0.0091	0.61 ± 0.01	0.78 ± 0.014	0.78 ± 0.011
Glass	J	0.54 ± 0.11	0.45 ± 0.12	0.50 ± 0.18	0.53 ± 0.22
	S	0.61 ± 0.011	0.41 ± 0.011	0.53 ± 0.012	0.56 ± 0.023
Diabetes	J	0.54 ± 0.091	0.45 ± 0.1	0.50 ± 0.11	0.54 ± 0.51
	S	0.61 ± 0.0091	0.42 ± 0.0092	0.55 ± 0.011	0.59 ± 0.089
Heart	J	0.11 ± 0.26	0.14 ± 0.26	0.14 ± 0.3	0.15 ± 0.54
	S	0.32 ± 0.012	0.37 ± 0.012	0.39 ± 0.017	0.32 ± 0.056

Table 2: Overall performance for clustering methods over real databases by comparing KSC with conventional methods.

As can be appreciated from the results for both real datasets and images, KSC overcomes conventional clustering methods. This is due to the fact that KSC exploits the use of kernels to harness the local structure information. As can be noticed, it works well on image segmentation, which means that complex data can be rightly modeled by KSC. In addition, KSC is also able to deal with real databases where some compactness is guaranteed. Then, KSC is a more flexible and versatile method.

5. FINAL REMARKS

As has been widely reported by literature, the kernel spectral clustering (KSC) method has proved to be a powerful tool for solving pattern recognition problems when labeling is unavailable and clusters are highly non-linearly separable. The suitability of KSC lies in its formulation.

In this work, we presented the formulation (KSC) through elegant and fully matrix statements, which enable its practical use and interpretation, specially thinking of further computational applications. Also, some key aspects for a deep understanding of KSC is provided.

ACKNOWLEDGMENTS: The authors acknowledge the value support given by the SDAS Research Group (www.sdas-group.com).

RECEIVED: APRIL, 2020.
REVISED: SEPTEMBER 2020.

REFERENCES

[1] ALELYANI, S., TANG, J., and LIU, H. (2018**): Feature selection for clustering: A review In "Data Clustering"** 29–60. Chapman and Hall/CRC, city.
[2] ALZATE, C. and SUYKENS, J. A. K. (2010): Multiway spectral clustering with out-of-sample ex-tensions through weighted kernel pca Pattern Analysis and Machine Intelligence, **IEEE Transactions on**, 32, 335–347.
[3] CASTRO-OSPINA, A. E., CASTRO-HOYOS, C., PELUFFO-ORDONEZ, D., and CASTELLANOS-DOMINGUEZ, G. (2013): Novel heuristic search for ventricular arrhythmia detection using normalized cut clustering In 2013 **35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**, pages 7076–7079. IEEE.
[4] DOMENICONI, C., PENG, J., and YAN, B. (2011): **Composite kernels for semi-supervised clustering Knowledge and information systems,** 28(1):99–116.
[5] ESTER, M. (2018): **Density-based clustering In Data Clustering**, pages 111–127. Chapman and Hall/CRC.
[6] FRANK, A. AND ASUNCION, A. (2010): **UCI Machine Learning Repository**. University of California, School of Information and Computer Science. Irvine, CA. http://archive.ics.uci.edu/ml
[7] HERRERA-GRANDA, I. D., LORENTE-LEYVA, L. L., PELFFO-ORDÓÑEZ, D. H., VALENCIA-CHAPI, R. M., MONTERO-SANTOS, Y., CHICAIZA-VACA, J. L., AND CASTRO-OSPINA, A. E. (2018): Optimization of the university transportation by contraction hierarchies method and clustering algorithms In **International Conference on Hybrid Artificial Intelligence Systems**, pages 95–107. Springer.
[8] HERSHEY, J. R., CHEN, Z., LE ROUX, J., and WATANABE, S. (2016): Deep clustering: Discriminative embeddings for segmentation and separation In 2016 **IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, pages 31–35. IEEE.
[9] JAIN, A. K. (2010): Data clustering: 50 years beyond k-means **Pattern recognition letters**, 31(8):651–666.
[10] LI, X. and WONG, K.-C. (2018): Evolutionary multiobjective clustering and its applications to patient stratification **IEEE transactions on cybernetics**, 49(5):1680–1693.
[11] LORENTE-LEYVA, L. L., HERRERA-GRANDA, I. D., ROSERO-MONTALVO, P. D., PONCE-GUEVARA, K. L., CASTRO-OSPINA, A. E ., BECERRA, M. A., PELUFFO-ORDÓÑEZ, D. H., AND RODRÍGUEZ-SOTELO, J. L. (2018): Developments on solutions of the normalized-cut-

clustering problem without eigenvectors In **International Symposium on Neural Networks**, pages 318–328. Springer.
[12] MARTIN, D., FOWLKES, C., TAL, D., AND MALIK, J. (2001): A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics In **Proc. 8th Int'l Conf. Computer Vision**, volume 2, pages 416–423.
[13] MITTAL, M., GOYAL, L. M., HEMANTH, D. J., AND SETHI, J. K. (2019): Clustering approaches for high-dimensional databases: A review Wiley Interdisciplinary Reviews: **Data Mining and Knowledge Discovery**, 9(3):e1300.
[14] MOHEBI, A., AGHABOZORGI, S., YING WAH, T., HERAWAN, T., AND YAHYAPOUR, R. (2016): Iterative big data clustering algorithms: a review Software: **Practice and Experience**, 46(1):107–129.
[15] OÑA-ROCHA, O., RIASCOS-SALAS, J., MARRUFO-RODRÍGUEZ, I., PÁEZ-JAIME, M., MAYORCA-TORRES, D., PONCE-GUEVARA, K., SALAZAR-CASTRO, J., AND PELUff O-ORDÓÑEZ, D. (2019): Kernel-spectral-clustering driven motion segmentation: Rotating-objects first trials In Latin American Workshop on **Computational Neuroscience**, pages 30–40. Springer.
[16] ORMROD J (2004) Human learning, 4th ed. Merrill, Upper Saddle River N.J.
[17] PANAPAKIDIS, I. P. AND CHRISTOFORIDIS, G. C. (2018): Optimal selection of clustering algorithm via multi-criteria decision analysis (mcda) for load profiling applications. **Applied Sciences**, 8(2):237.
[18] PELUFFO, D., ALZATE, C., SUYKENS, J. A., AND CASTELLANOS-DOMÍNGUEZ, G. (2014): Optimal data projection for kernel spectral clustering In **22nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN** 2014-Proceedings, pages 553–558.
[19] PELUFFO-ORDÓÑEZ, D., GARCÍA-VEGA, S., LANGONE, R., SUYKENS, J. A., AND CASTELLANOS-DOMÍNGUEZ, G. (2013): Kernel spectral clustering for dynamic data using multiple kernel learning In **The 2013 International Joint Conference on Neural Networks (IJCNN)**, pages 1–6. IEEE.
[20] RODRÍGUEZ-SOTELO, J., DELGADO-TREJOS, E., PELFFO-ORDÓÑEZ, D., CUESTA-FRAU, D., AND CASTELLANOSDOMÍNGUEZ, G. (2010): Weighted-pca for unsupervised classification of cardiac arrhythmias In **2010 Annual International Conference of the IEEE Engineering in Medicine and Biology**, pages 1906–1909. IEEE.
[21] RODRÍGUEZ-SOTELO, J., PELUFFO-ORDÓÑEZ, D., CUESTA-FRAU, D., AND CASTELLANOS-DOMÍNGUEZ, G. (2009): Nonparametric density-based clustering for cardiac arrhythmia analysis In **2009 36th Annual Computers in Cardiology Conference (CinC)**, pages 569–572. IEEE.
[22] RODRÍGUEZ-SOTELO, J. L., OSORIO-FORERO, A., JIMÉNEZ-RODRÍGUEZ, A., CUESTA-FRAU, D., CIRUGEDA ROLDÁN, E., and PELUFFO, D. (2014): Automatic sleep stages classification using eeg entropy features and unsupervised pattern analysis techniques**. Entropy**, 16(12):6573–6589.
[23] RODRÍGUEZ-SOTELO, J. L., PELUFFO-ORDÓNEZ, D., CUESTA-FRAU, D., AND CASTELLANOS-DOMÍNGUEZ, G. (2012): Unsupervised feature relevance analysis applied to improve ecg heartbeat clustering **Computer methods and programs in biomedicine**, 108(1):250–261.
[24] SATISH, D. and SEKHAR, C. (2004): Kernel based clustering for multiclass data. In **Neural Information Processing**, pages 1266–1272. Springer.
[25] SCHOLKOPF, B. and SMOLA, A. J. (2002**):** Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. **Max Planck Institute for Intelligent Systems**, https://www.is.mpg.de/publications/973
[26] SEELAND, M., KARWATH, A., and KRAMER, S. (2012): A structural cluster kernel for learning on graphs In **Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining**, pages 516–524. ACM.
[27] SHI, J. and MALIK, J. (2000): Normalized cuts and image segmentation Pattern Analysis and
Machine Intelligence, **IEEE Transactions** on, 22(8):888–905.

[28] UNNIKRISHNAN, R., PANTOFARU, C., and HEBERT, M. (2007): Toward objective evaluation of image segmentation algorithms Pattern Analysis and Machine Intelligence, **IEEE Transactions** on, 29(6):929–944.
[29] VALENCIA, X. P. B., BECERRA, M., OSPINA, A. C., ADARME, M. O., MELO, D. V., and ORDÓÑEZ, D. P. (2017): Kernel-based framework for spectral dimensionality reduction and clustering formulation: **A theoretical study ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal**, 6(1):31–40.
[30] VON LUXBURG, U. (2007): A tutorial on spectral clustering. In **Statistics and Computing**, 17, 4, 395–416. https://doi.org/10.1007/s11222-007-9033-z.
[31] X., YU. STELLA. and JIANBO, S. (2003): Multiclass spectral clustering In **ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision**, page 313, Washington, DC, USA. IEEE Computer Society.
[32] YUAN, G., SUN, P., ZHAO, J., LI, D., and WANG, C. (2017): A review of moving object trajectory clustering algorithms. In **Artificial Intelligence Review**, 47(1):123–144.
[33] ZELNIK-MANOR, L. and PERONA, P. (2004): Self-tuning spectral clustering In **Advances in Neural Information Processing Systems** 17, pages 1601–1608. MIT Press.

.