Fuzzy c-means, kernel functions, segmentation, Multiple Sclerosis

Jacek KAWA*, Ewa PIETKA*

KERNELIZED FUZZY C-MEANS METHOD IN SEGMENTATION OF DEMYELINATION PLAQUES IN MULTIPLE SCLEROSIS

In the current study, an alternative approach to a fuzzy clustering in a kernel space has been tested. First, a “kernel trick”' is applied to the fuzzy c-means (FCM) algorithm. Later, the modified method is employed in an automated segmentation of demyelination plaques in Multiple Sclerosis.

1. INTRODUCTION

Many studies have been dedicated to a problem of an automated segmentation of Multiple Sclerosis (MS) demyelination plaques in magnetic resonance images. Algorithms for a segmentation of the global white matter lesion [1], as well as segmentation of plaques in MS [2], [3] (and others) employ both fuzzy and non-fuzzy approaches. In this paper, a new kernel-space variant of a fuzzy c-means clustering method is used.

Fuzzy c-means method (FCM) partitions a set of data vectors into a predefined number of clusters. The standard algorithm performs well on many image processing task, yet it features several weakness. Some of the problems can be addressed by clustering in a high-dimensional kernel space.

The introduction of kernel methods into the c-means algorithm family is attributed [4] [5] to Schölkopf at al. The kernel hard c-means method employs a kernel function in prototypes and partition matrix formulas. The clustering is performed gradually, by introducing a new data vector in each iteration. In [6], Zhan and Cheng present the general FCM-based clustering with partition matrix computation in kernel space. In earlier work, Girolami [7] develops sum-of-square based clustering performed fully in kernel space and presents some stochastic optimization. Another method, also employing full-kernel space computations, has been published by Wu and Xie [8]. In [4], Hathaway at al. present relational clustering, also featuring kernel functions.

* Silesian University of Technology, Department of Biomedical Engineering, ul. Akademicka 16 Gliwice, Poland.

2. KERNELIZED FCM

2.1. FUZZY C-MEANS CLUSTERING

Let $\mathbf{x_k} = (x_i, \ldots, x_n)$ be an observed data vector of $\{\mathbf{x_k}\}_{k=1}^{N}$ data set in feature space $\mathbf{F}$. Standard FCM is derived to minimize the objective function:

$$\mathrm{J}(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m} \left\| \mathbf{x_k} - \mathbf{v_i} \right\|^{2}, \mathbf{x_k}, \mathbf{v_i} \in \mathbf{F} \tag{1}$$

with respect to the partition matrix element u_{ik} and the centre of the i-th cluster - $\mathbf{v_i}$, and for a given fuzzyfication level $m\,(1 \le m < \infty)$.

The partition matrix elements satisfy: $0 \le u_{ik} \le 1, \sum u_{ik} = 1 \forall k$, and $0 < \sum u_{ik} < N \forall i$.

The FCM clustering is performed iteratively, starting with a set of c initially given prototypes and fuzzyfication level m. In each step a new partition matrix $\mathbf{U}$ is created, satisfying:

$$u_{ik}^{new} = \frac{\left\| \mathbf{v_i} - \mathbf{x_k} \right\|^{\frac{-2}{m-1}}}{\sum_{z=1}^{c} \left(\left\| \mathbf{v_z} - \mathbf{x_k} \right\|^{\frac{-2}{m-1}} \right)} \tag{2}$$

The membership matrix $\mathbf{U}$ is later employed to compute a new set of prototypes:

$$v_i^{new} = \frac{\sum_{k=1}^{N} u_{ik}^{m} \mathbf{x_k}}{\sum_{k=1}^{N} u_{ik}^{m}} \tag{3}$$

The procedure is repeated until the desired accuracy of $\mathbf{V}$ is obtained, i.e. $\max\left(\left\| \mathbf{V}_i^{new} - \mathbf{V_i} \right\| \right) < \varepsilon, 0 < i \le c$.

2.2. KERNEL SPACE

The fuzzy c-means algorithm features several weaknesses. Because of the Euclidean norm used to form the objective function, the shape of clusters in feature space is (hyper)spherical, thus data may not be easy separable. The performance of the algorithm in the presence of noise is often poor - outliers influence both membership function, and prototypes calculations, and may change the number of cluster observed in the data set. The improvement is possible in two ways: (1) by modifying the objective function, or (2) by a transformation of the data subjected to the clustering process. The presented kernel FCM method features “kernel trick”' to modify the objective function in order to perform implicit data transformation.

For any given data set a non-linear mapping $\varphi : \mathbf{F} \to \mathbf{K}, \varphi(\mathbf{x_k}) = \boldsymbol{\varphi}_\mathbf{k}$ exist, that transforms the data from $\mathbf{F}$ into a high dimensional space $\mathbf{K}$. A clustering method performed in this space $\mathbf{K}$ yields better results than applied to the data in the original space $\mathbf{F}$. The mapping φ is not always known or computationally feasible. The explicit transformation in specific cases may be avoided by employing the kernel functions.

Let the function $q : \mathbf{F} \times \mathbf{F} \to \mathbf{R}$, be a positive definite kernel over $\mathbf{F}$, satisfying:

$$q(\mathbf{x},\mathbf{y}) = q(\mathbf{y},\mathbf{x}), \forall \mathbf{x},\mathbf{y} \in \mathbf{F} \tag{4}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j q(\mathbf{x_i},\mathbf{x_j}) \geq 0, \forall n \in N^+, \forall c_1 \ldots c_n \in R, \forall(\mathbf{x_1}, \ldots, \mathbf{x_n}) \in \mathbf{F}^n \tag{5}$$

For any given positive definite kernel over $\mathbf{F}$, there exist a Hilbert space $\mathbf{H}$ and a mapping $\varphi : \mathbf{F} \to \mathbf{H}$, such that [9]:

$$q(\mathbf{x},\mathbf{y}) = \varphi(\mathbf{x})^T \varphi(\mathbf{y}), \forall \mathbf{x},\mathbf{y} \in \mathbf{F}$$

This property, often called a “kernel trick”' means, that the value of kernel function q is equal to the inner product in some space $\mathbf{H}$. The space $\mathbf{H}$ is referred to as kernel space.

The “kernel trick”, may be used to increase the robustness of the FCM in two-ways: (1) for the computation of the distance between transformed cluster prototypes and transformed data points in kernel space, and (2) for computation in kernel space with implicit prototypes.

The first approach has been presented by Chen [6]. The introduced kernel function has permitted for a detection of clusters hidden from the standard FCM. The clustering itself has not been performed fully in the kernel space, though, as the cluster prototypes are computed in the original space. The second approach has been available in work of Girolami [7], where a matrix-trace based clustering has been presented.

3. KERNELIZED FCM

In the current study, the “kernel trick”' has been used to modify the original fuzzy c-means algorithm in order to perform the implicit-prototypes-FCM clustering in the kernel space.

Let the transition from a feature space $\mathbf{F}$ to a kernel space $\mathbf{H}$ be obtained through a mapping $\varphi(\mathbf{x})$ and let the corresponding kernel function q be available. This means, that for each data vector $\mathbf{x_k}$, the corresponding $\boldsymbol{\varphi}_\mathbf{k} = \varphi(\mathbf{x}_k)$ is defined, and for each pair of data vectors $(\mathbf{x_k}, \mathbf{x_p})$, the inner product $\boldsymbol{\varphi}_\mathbf{k} \cdot \boldsymbol{\varphi}_\mathbf{p}$ in $\mathbf{H}$ is equal to $q_{kp} = q(\mathbf{x_k}, \mathbf{x_p})$. A kernel space FCM clustering is then possible by minimization of a new objective function:

$$\mathrm{J}(\mathbf{U}, \mathbf{W}(\mathbf{U})) = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \left\| \boldsymbol{\varphi}_\mathbf{k} - \mathbf{w_i} \right\|^2, \boldsymbol{\varphi}_\mathbf{k}, \mathbf{w_i} \in \mathbf{H} \tag{7}$$

where

$$w_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} \boldsymbol{\varphi}_\mathbf{k}}{\sum_{k=1}^{N} u_{ik}^{m}} \tag{8}$$

The optimization of the objective function yields:

$$u_{ik} = \frac{\left\| \mathbf{w_i} - \boldsymbol{\varphi}_\mathbf{k} \right\|^{\frac{-2}{m-1}}}{\sum_{z=1}^{c} \left(\left\| \mathbf{w}_z - \boldsymbol{\varphi}_\mathbf{k} \right\|^{\frac{-2}{m-1}} \right)} \tag{9}$$

Thus

$$u_{ik} = \frac{\left\| \left(\sum_{p=1}^{N} u_{ip}^{m} \boldsymbol{\varphi}_\mathbf{p} \right) \left(\sum_{p=1}^{N} u_{ip}^{m} \right)^{-1} - \boldsymbol{\varphi}_\mathbf{k} \right\|^{\frac{-2}{m-1}}}{\sum_{z=1}^{c} \left(\left\| \left(\sum_{p=1}^{N} u_{zp}^{m} \boldsymbol{\varphi}_\mathbf{p} \right) \left(\sum_{p=1}^{N} u_{zp}^{m} \right)^{-1} - \boldsymbol{\varphi}_\mathbf{k} \right\|^{\frac{-2}{m-1}} \right)} \tag{10}$$

Let $\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}; \mathbf{b} \rangle$, and $C_i = \frac{1}{\sum u_{ip}^{m}}$. The Euclidean norm is given as $\|.\|^2 = \langle .;. \rangle$, thus:

$$u_{ik} = \frac{\left\langle C_i \sum_{p=1}^{N} u_{ip}^{m} \boldsymbol{\varphi}_\mathbf{p} - \boldsymbol{\varphi}_\mathbf{k} ; C_i \sum_{p=1}^{N} u_{ip}^{m} \boldsymbol{\varphi}_\mathbf{p} - \boldsymbol{\varphi}_\mathbf{k} \right\rangle^{\frac{-1}{m-1}}}{\sum_{z=1}^{c} \left(\left\langle C_z \sum_{p=1}^{N} u_{zp}^{m} \boldsymbol{\varphi}_\mathbf{p} - \boldsymbol{\varphi}_\mathbf{k} ; C_z \sum_{p=1}^{N} u_{zp}^{m} \boldsymbol{\varphi}_\mathbf{p} - \boldsymbol{\varphi}_\mathbf{k} \right\rangle^{\frac{-1}{m-1}} \right)} \tag{11}$$

An inner product of selected vectors in kernel space $\mathbf{H}$ may be replaced by corresponding kernel matrix element.

$$\begin{bmatrix} \langle \boldsymbol{\varphi}_1;\boldsymbol{\varphi}_1\rangle & \langle \boldsymbol{\varphi}_1;\boldsymbol{\varphi}_2\rangle & \cdots & \langle \boldsymbol{\varphi}_1;\boldsymbol{\varphi}_N\rangle \\ \langle \boldsymbol{\varphi}_2;\boldsymbol{\varphi}_1\rangle & \langle \boldsymbol{\varphi}_2;\boldsymbol{\varphi}_2\rangle & \cdots & \langle \boldsymbol{\varphi}_2;\boldsymbol{\varphi}_N\rangle \\ \vdots & \vdots & . & \vdots \\ \langle \boldsymbol{\varphi}_N;\boldsymbol{\varphi}_1\rangle & \langle \boldsymbol{\varphi}_N;\boldsymbol{\varphi}_2\rangle & \cdots & \langle \boldsymbol{\varphi}_N;\boldsymbol{\varphi}_N\rangle \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & . & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NN} \end{bmatrix} = \mathbf{Q} \tag{12}$$

Therefore:

$$u_{ik} = \frac{\left(\overbrace{C_i^2 \left(u_{i1}^{2m} q_{11} + \ldots + u_{iN}^{2m} q_{NN}\right)}^{A_i} \overbrace{- 2\mathrm{C}_i \left(u_{i1}^{m} q_{1k} + \ldots + u_{iN}^{m} q_{Nk}\right)}^{B_{ik}} + q_{kk} \right)^{\frac{-1}{m-1}}}{\sum_{z=1}^{c} \left(C_z^2 \left(u_{z1}^{2m} q_{11} + \ldots + u_{zN}^{2m} q_{NN}\right) - 2\mathrm{C}_z \left(u_{z1}^{m} q_{1k} + \ldots + u_{zN}^{m} q_{Nk}\right) + q_{kk} \right)^{\frac{-1}{m-1}}} \tag{13}$$

Denoting:

$$\mathbf{S_i} = \begin{bmatrix} u_{i1}^m \\ u_{i2}^m \\ \vdots \\ u_{iN}^m \end{bmatrix} \cdot [11\ldots1] \tag{14}$$

and introducing element-wise matrix multiplication $\circ$, for a symmetric matrix $\mathbf{Q}$:

$$A_i = C_j^2 \sum \left(\mathbf{S_i} \circ \mathbf{S_i^T} \circ \mathbf{Q} \right) \tag{15}$$

resembles sum of all the elements, and

$$B_{ik} = -2\mathrm{C}_i^2 \sum \left(\mathrm{row}_k \left(\mathbf{S_i} \circ \mathbf{Q} \right) \right) \tag{16}$$

states for a sum of all the elements of row k.

The partition matrix satisfies then:

$$u_{ik} = \frac{\left(A_i + B_{ik} + q_{kk} \right)^{\frac{-1}{m-1}}}{\sum_{z=1}^{c} \left(A_z + B_{zk} + q_{kk} \right)^{\frac{-1}{m-1}}} \tag{17}$$

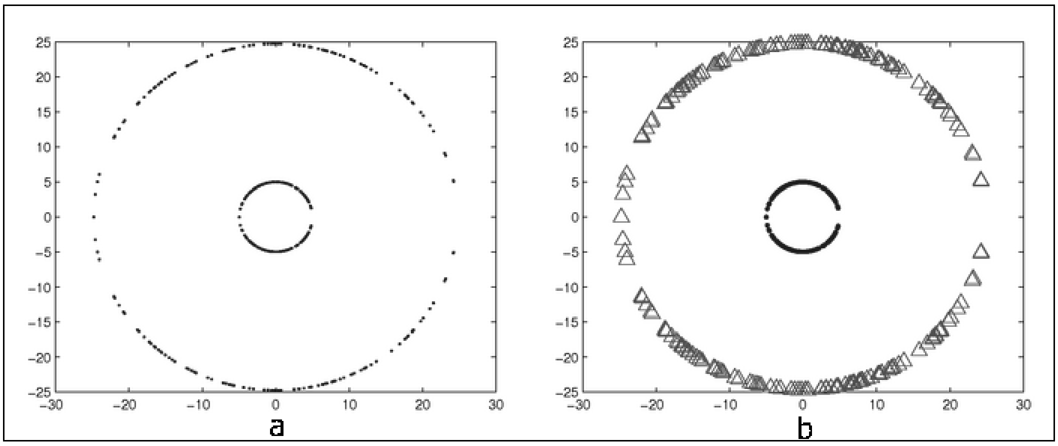

Fig. 1 (a) input data set; (b) clustering results (thresh. 0.5)

Kernelized FCM is performed in following steps:

1. Generate the kernel matrix $\mathbf{Q}$ using a kernel function q. Set the value of c, m and ε
2. Compute a new partition matrix $\mathbf{U}^{\mathbf{new}}$ using (17)
3. $\mathbf{U}^{\mathbf{old}} = \mathbf{U}, \mathbf{U} = \mathbf{U}^{\mathbf{new}}$
4. if $\max\left(\left\|\mathbf{U}^{\mathbf{old}} - \mathbf{U}^{\mathbf{new}}\right\|\right) < \varepsilon$, then break, else: *step 2*

Note: In specific cases, the data for a clustering may be given only by a data-relation-matrix $\mathbf{P}$. If the relation matrix $\mathbf{P}$ assembles relational data for each pair of clustered data points, and satisfies the kernel conditions similar to (4) and (5), it can be directly used in the clustering process instead of the generated kernel matrix $\mathbf{Q}$.

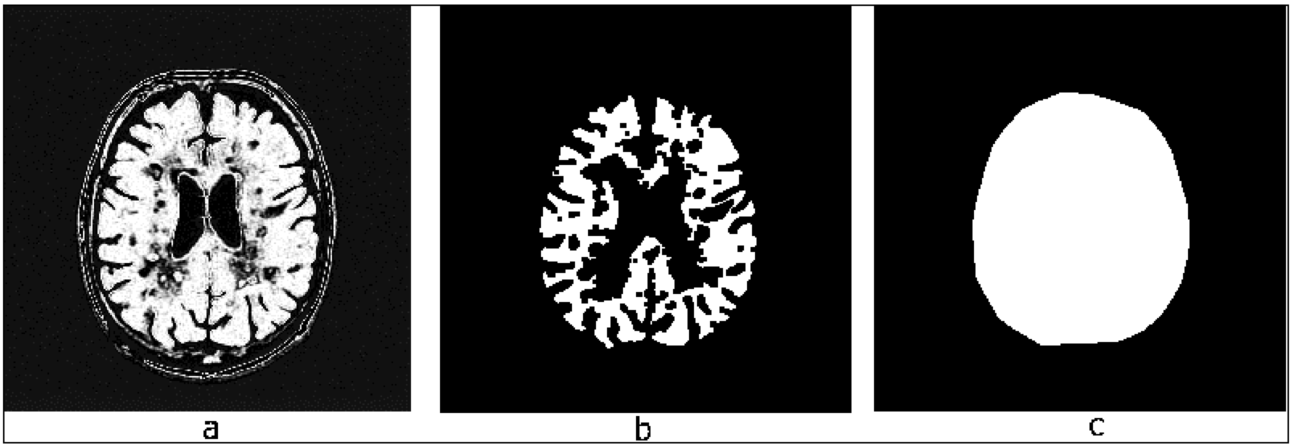

Fig. 2 Brain tissue segmentation (a) second cluster of KFCM, (b) biggest distinct region after thresholding

4. RESULTS AND CONCLUSIONS

The performance of the developed algorithm has been tested on a set of artificial data. The example results obtained for an input set (fig. 1a), and Gaussian (RBF) kernel have been presented in figure 1b. Both circles have been assigned to individual clusters.

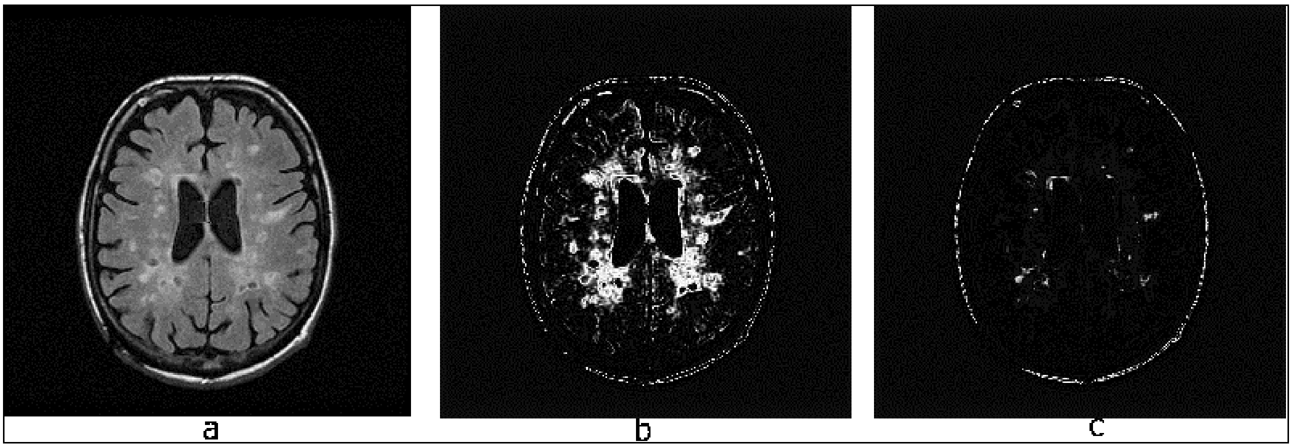

Fig. 3 MS plaques extraction (a) original slice, (b) 4th cluster, (c) 5th cluster

The algorithm has also been implemented to a segmentation of demyelination plaques in Multiple Sclerosis, performed at a CAD workstation [10]. The kernelized FCM has been employed at two phases of the FLAIR-MR (in both cases only the spels with unique signal intensity have been processed) - figure 3a, image analysis. At the preprocessing stage, the FCM performed in the polynomial kernel space has extracted the brain tissue (figure 2abc). Then, the radial basis (RBF) function has been used at the MS plaque segmentation (figure 3bc). The results have been later combined and subjected to further processing (figure 4).

The presented method has been used for segmentation of MR images of 15 patients with advanced Multiple Sclerosis. The results of interobserver comparison [11] with reference set (created with CAD built-in tools), have shown a better performance than a standard FCM algorithm; less false positives rate has been observed.

Fig. 4 Segmentation results

Total time necessary for processing MR-extracted data in the application is similar as for the standard FCM. Generally, the most time and storage consuming operation during the evaluation of the algorithm, is a kernel-matrix computation. This task is performed only once during the clustering process, but the overall evaluation time of the presented method is likely to increase quickly with growing number of input data. One should note, however, that no relation exist between the number of features that describe the data set and the size of kernel-matrix - i.e. the method is expected to perform well for a small data set described by either a small, or a large number of features.

BIBLIOGRAPHY

[1] ALFANO B, BRUNETTI A, LAROBINA M, et al., Automated segmentation and measurement of global white matter lesion volume in patients with multiple sclerosis, Journal of Magnetic Resonance Imaging, No. 12, pp. 799-807, 2000.

[2] ZIJDENBOS A.P, FORGHANI R, EVANS A.C, Automatic “pipeline”' analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis, IEEE Trans. on Medical Imaging , Vol. 21, No. 10, pp. 1280-1291, 2002.

[3] KIKINIS R, GUTTMANN C.R, METACALF D, et al., Quantitative follow-up of patients with multiple sclerosis using MRI: Technical aspects Journal of Magnetic Resonance Imaging, No. 9, pp. 519-530, 1999.

[4] HATHAWAY R.M, HUBAND J.M, BEZDEK J.C, A kernelized non-Euclidean relational fuzzy c-means algorithm, Proc. The 2005 IEEE International Conference on Fuzzy Systems, pp. 414-419, IEEE, 2005.

[5] XU R, WUNSCH D, Survey of clustering algorithms, IEEE Trans. Neural Networks, Vol. 16, No. 3, pp. 645-678, 2005.

[6] CHEN S, ZHANG D, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure, IEEE Trans. System, Man, and Cybernetics – Part B., Vol. 34, No. 4, pp. 1907-1916, 2004.

[7] GIROLAMI M, Mercer kernel-based clustering in feature space, IEEE Trans. Neural Networks, Vol. 13, No. 3, pp. 780-784, 2002.

[8] WU Z, XIE W, YU J, Fuzzy C-Means clustering algorithm based on kernel method, Proc. 5th International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), Xi'an China, pp. 49-54, IEEE, 2003

[9] VERT J, TSUDA K, SCHÖLKOPF B, A Primer on Kernel Methods, In: Kernel Methods in Computational Biology, pp. 35-70, MIT Press, Cambridge, MA, USA, 2004.

[10] PIETKA E, KAWA J, SPINCZYK D, KONOPKA M.N, Computer-aided diagnosis workstation for detection and volumetric measurement of demyelination plaques in multiple sclerosis, In. RSNA: Scientific Assembly and Annual Meeting Program 2005, pp. 855, Radiological Society of North America, Inc., Oak Brook (IL), 2005.

[11] KAWA J, PIETKA E, KIELTYKA A, CAD workstation for inter- and intraobserver studies in detection of demyelination plaques and bone age assessment, Proc. Europacs 2006, http://www.europacs.net/presentations.htm, Europacs, 2006.