Engineering Applications of Artificial Intelligence 68 (2018) 205–213

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Multiple kernel approach to semi-supervised fuzzy clustering algorithm for land-cover classification

Sinh Dinh Mai, Long Thanh Ngo *

Department of Information Systems, Faculty of Information Technology, Le Quy Don Technical University, 236, Hoang Quoc Viet, Bac Tu Liem, Hanoi, Viet Nam

ARTICLE INFO

Keywords:
Semi-supervised clustering
Fuzzy clustering
Kernel fuzzy c-means
Multiple kernel
Land cover classification

ABSTRACT

Clustering is used to detect sound structures or patterns in a dataset in which objects positioned within the same cluster exhibit a substantial level of similarity. In numerous clustering problems, patterns is not easily separable due to the highly complex shaped data. In the previous studies, kernel-based methods have exhibited the effectiveness to partition such data. In this paper, we proposed a semi-supervised clustering method based fuzzy c-means algorithm using multiple kernel technique, called SMKFCM, in which the rudimentary centroids are directly used to the calculating process of centroids. The SMKFCM algorithm is on the basis of combining the labeled and unlabeled data together to improve performance. We used the labeled patterns to calculate the centrality of clusters considered as the rudimentary centroids which are added into the objective functions. The SMKFCM algorithm can be applied to both clustering and classification problems. The experimental results show that SMKFCM algorithm can improve significantly the classification accuracy which comes from comparison with a conventional classification or clustering algorithms such as semi-supervised kernel fuzzy c-means (S2KFCM), semi-supervised fuzzy c-means (SFCM) and Self-trained semi-supervised SVM algorithm (PS3VM).

1. Introduction

Recently, fuzzy clustering methods have been studied and widely used in various applications. One of the most popular fuzzy clustering methods is the Fuzzy C-Means (FCM) algorithm (Hoppner et al., 1999; Bezdek et al., 1984). In the FCM algorithm, a pattern may belong to more than one cluster with individual membership degrees. Although, the FCM clustering algorithm is limited to the discovery of spherical clusters. The limitation of the standard FCM algorithm is the usage of the Euclidian distance in the observation space on which the algorithm is only effective for spherical clusters, but it does not perform well for more general clusters (Shen et al., 2006). Besides, clustering results strongly depend on the characteristics of data-sets, the KFCM algorithm does not produce clusters with the desired accuracy in some datasets (Yu et al., 2011). The real-world clustering problems usually contains some useful features when combining with other ones. For example, image segmentation results could be produced better by combining pixel intensity feature and the local spatial information feature.

To overcome the drawbacks of the conventional FCM technique, kernel fuzzy c-means clustering (KFCM) algorithm was proposed Hathaway et al. (2005) to solve this problem by mapping input data into an appropriate space using a nonlinear function. This approach has received considerable attention because kernels make it possible to map data into a high-dimensional feature space in order to increase the representable capability of a linear clustering. Some proposed improvements KFCM algorithm by adapting a new kernel induced metric in the data space (Graves and Pedrycz, 2007) or transforming the original inputs into a much higher dimensional Hilbert space (Zhang and Chen, 2004, 2003b; Graves and Pedrycz, 2010).

On the basis of using kernel technique in clustering, some studied was done to improve the performance (Shawe-Taylor and Cristianini, 2004; Rakotomamonjy et al., 2007; Varma and Babu, 2009). Girolami (2002) proposed a kernel-based clustering for a wider variety of clusters and Tzortzis and Likas (2009) also introduce an algorithm based on kernel methods to overcome the cluster initialization problem. Later, Zhang and Chen (2003a) proposed the kernel-based fuzzy c-means (KFC) algorithm, which allows for incomplete data as well. Graves and Pedrycz (2010, 2007) launched a comprehensive comparative analysis of kernel-based fuzzy clustering and fuzzy clustering. The Fuzzy C-Means algorithm (FCM) algorithm and Gustafson–Kessel (GK) FCM was compared with two typical kernel-based fuzzy clustering algorithms: one with the prototypes located in the feature space (KFCM-F) and the other where the prototypes are distributed in the kernel space (KFCM-K). Both these kernel clustering algorithms are used the Gaussian kernel

* Corresponding author.
E-mail address: ngotlong@mta.edu.vn (L.T. Ngo).

https://doi.org/10.1016/j.engappai.2017.11.007
Received 1 November 2016; Received in revised form 21 September 2017; Accepted 12 November 2017
0952-1976/

and KFCM-K deals with polynomial kernel. The analyzed datasets in this study show that the kernel-based FCM algorithms are better than the standard FCM and GKFCM.

Because of the advantages of clustering methods based on the kernel techniques, these algorithms have been applied in many different fields, particularly in image processing. Several studies and applications using KFCM in image segmentation include the kernelized fuzzy C-means (KFCM) algorithm (Zhang and Chen, 2004) which used a kernel-induced distance metric and a spatial penalty on the membership functions and a novel modified kernel fuzzy c-means(NMKFCM) (Yu et al., 2011) algorithm based on conventional KFCM which incorporates the neighbor term in its objective function.

Besides, Hathaway et al. (2005) presented a kernel expansion for clustering relational data by producing a kernelized algorithm of the non-Euclidean relational FCM. In addition, Chiang and Hao (2003) also proposed a multiple sphere support vector clustering algorithm based on the adaptive cell growing model which maps the input patterns to a high-dimensional feature space by using the desired kernel function.

Land-cover classification has widely applied to various fields (Torahi and Rai, 2011) and approached according to some methods of soft computing (Stavrakoudis et al., 2010; Stavrakoudis et al., 2011; Gordo et al., 2013). These studies are mainly based on methods of supervised and unsupervised classification (Genitha and Vani, 2013). Further, an approach using an ensemble of semisupervised classifiers were proposed for change detection in remotely sensed images (Roy et al., 2014) by using a multiple classifier system in semi-supervised (learning) framework instead of a single weak classifier. In the manner of semi-supervised change detection, Yuan et al. (2015) proposed a new distance metric learning framework for change detection by abundant spectral information of hyper-spectral image in noisy condition; Liu et al. (2013) proposed a novel semi-supervised SVM (PS3VM) model using selftraining approach to address the problem of remote sensing land cover classification.

Karvelis et al. (2000) presented a semi-unsupervised method for the M-FISH chromosome image classification. Firstly, the separation of foreground and background is performed by using an automated thresholding procedure. Then, these features are normalized. Secondly, the K-Means algorithm was applied to cluster the chromosome pixels into the 24 chromosome classes. Although this algorithm does not require a training, it produces on average, higher accuracy. However, the tests of the algorithm only used a small number of images. The usage of fuzzy C-means method with the labeled information (pre-determined clusters) Mai and Ngo (2015) introduced an approach which exploits local spatial information between the pixel and its neighbors to compute the membership degree. After that, Ngo et al. (2015) has developed algorithms for interval type-2 fuzzy clustering algorithm applying to the problems of satellite image analysis consisting of land cover classification and change detection.

Frigui et al. (2013) provided an overview of several fuzzy kernels clustering algorithms by using partial supervision information to guide the optimization process and avoid local minima. Zhang et al. presented a semi-supervised kernel-based FCM algorithm (Zhang et al., 2004) by introducing semi-supervised learning technique and the kernel method simultaneously into conventional fuzzy clustering algorithm in which the labeled and unlabeled data are used together. A semi-supervised kernel-based FCM algorithm with Pairwise Constraints was proposed by Wang et al. (2008), which incorporates both semi-supervised learning technique and the kernel method into fuzzy clustering algorithm.

Note that kernel methods depend on the usage of a suitable kernel function. If the kernel method only selects a single kernel from a predefined group then sometime it is not insufficient to represent all datasets. In addition, individual features of the selected input data can result in different clusters corresponding to individual kernels. Therefore, combining multiple kernels from a set of basis kernels has been proposed to refine better clusters rather than using single kernel method. The most important key in the kernel method is how to use the formulation of suitable kernel function (Zhao et al., 2009; Ganesh and Palanisamy 2012; Huang et al. 2012; Yugander et al., 2012).

Thus, kernel fuzzy clustering algorithms are necessary to be extended with the aggregation of kernel functions from different sources. The rudimentary information of centroids was also added to the objective function to adjust the centroids through the iterative computing process. The paper introduces new clustering method to show how to apply multiple kernel technique in semi-supervised clustering. There are two problems mentioned in this paper. Firstly, we extracted the characteristics from datasets to estimate the rudimentary centroid of clusters which is one of components of the objective function. Besides, the single kernel and multiple kernel techniques are used in the semi-supervised fuzzy clustering. Secondly, two the proposed algorithms are applied to classification on various datasets and satellite image datasets. Experiments are compared with previous algorithms like semi-supervised kernel fuzzy c-means (S2KFCM), semi-supervised fuzzy c-means (SFCM) and Self-trained semi-supervised SVM algorithm (PS3VM); and validity indices are analyzed in comparison with the survey data.

The remaining parts of this paper are organized as follows: Section 2 provides background on the kernel technique and kernel fuzzy c-mean, Section 3 presents two proposed algorithms i.e semi-supervised kernel or multiple kernel fuzzy C-Means clustering; Section 4 demonstrates experiments on satellite image classification; Section 5 is a conclusion and future works.

2. Prerequisites

2.1. The kernel technique

In machine learning, kernel methods are a class of algorithms for pattern analysis which is to find the general types of relations (for example clusters, rankings, principal components, correlations, classifications) in datasets. In several algorithms solving these tasks, the raw-presented datasets have to be explicitly transformed into the feature vector — presented ones via a user-specified feature map.

The key idea of kernel technique is to invert the series of arguments, i.e. choose a kernel k rather than a mapping before applying a learning algorithm. Note that it is not any symmetric function k is considered as a kernel. Suppose the input space X has a limited number of elements, i.e. $X = \{x_1, x_2, \ldots, x_r\}$. Then, the $r \times r$ kernel matrix K with $K_{ij} = k(x_i, x_j)$ is definitely a symmetric matrix and $k(x_i, x_j)$ is a kernel value between x_i and x_j in X. The necessary and sufficient conditions of $k : \chi \times \chi \to \mathbb{R}$ are a kernel, which are given by Mercers theorem (Girolami, 2002).

Theorem 2.1. *The function $k : \chi \times \chi \to \mathbb{R}$ is a Mercer kernel if, and only if, for each $r \in \mathbb{N}$ and $x = (x_1, x_2 \ldots, x_r) \in \chi^r$ the $r \times r$ matrix $K = \left(k\left(x_i, x_j\right)\right)_{i,j=1}^{r}$ is positive semi definite.*

There exist simple rules for designing kernels on the basis of given kernel functions.

Theorem 2.2. *Kernel functions. Let $k_1 : \chi \times \chi \to \mathbb{R}$ and $k_2 : \chi \times \chi \to \mathbb{R}$ be any two Mercer kernels. Then, the functions $k : \chi \times \chi \to \mathbb{R}$ given by*

1. $k(x, \tilde{x}) = k_1(x, \tilde{x}) + k_2(x, \tilde{x})$
2. $k(x, \tilde{x}) = c.k_1(x, \tilde{x}), \forall c \in \mathbb{R}^+$
3. $k(x, \tilde{x}) = k_1(x, \tilde{x}) + c, \forall c \in \mathbb{R}^+$
4. $k(x, \tilde{x}) = k_1(x, \tilde{x}) \cdot k_2(x, \tilde{x})$
5. $k(x, \tilde{x}) = f(x) \cdot f(\tilde{x}) \forall f : \chi \to \mathbb{R}$.

are also Mercer kernels.

Theorem 2.3. *Let $k_1 : \chi \times \chi \to \mathbb{R}$ be any Mercer kernel. Then, the functions $k : \chi \times \chi \to \mathbb{R}$ given by*

1. $k(x, \tilde{x}) = \left(k_1(x, \tilde{x}) + \theta_1\right)^{\theta_2} \forall \theta_1 \in \mathbb{R}^+, \forall \theta_2 \in \mathbb{N}$.
2. $k(x, \tilde{x}) = \exp\left(\frac{k_1(x, \tilde{x})}{\sigma^2}\right), \forall \sigma \in \mathbb{R}^+$.