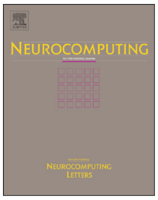
Neurocomputing 198 (2016) 125–134

Contents lists available at ScienceDirect

Neurocomputing

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



Fuzzy clustering with the entropy of attribute weights

Jin Zhou [a,b], Long Chen [b], C.L. Philip Chen [b,*], Yuan Zhang [a], Han-Xiong Li [c]

[a] School of Information Science and Engineering, University of Jinan, China
[b] Faculty of Science and Technology, University of Macau, Macau, China
[c] Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 29 March 2015
Received in revised form
14 September 2015
Accepted 22 September 2015
Available online 8 March 2016

Keywords:
Attribute-weighted clustering
Maximum-entropy regularization
Kernel method
Feature selection

ABSTRACT

For many datasets, it is a difficult work to seek a proper cluster structure which covers the entire feature set. To extract the important features and improve the clustering, the maximum-entropy-regularized weighted fuzzy c-means (EWFCM) algorithm is proposed in this paper. A new objective function is developed in the proposed algorithm to achieve the optimal clustering result by minimizing the dispersion within clusters and maximizing the entropy of attribute weights simultaneously. Then the kernelization of proposed algorithm is realized for clustering the data with 'non-spherical' shaped clusters. Experiments on synthetic and real-world datasets have demonstrated the efficiency and superiority of the presented algorithms.

1. Introduction

Prototype-based partitioning clustering analysis is an efficient tool for data mining, pattern recognition and statistical analysis [1,2]. It partitions the data into clusters according to the similarity of objects, which helps in extraction of new information or discovering new patterns. In the past few decades, various partitioning clustering algorithms have been proposed. K-means algorithm [3] and fuzzy c-means (FCM) algorithm [4] are two well known clustering algorithms. Variants of these two algorithms are further discussed and popularized in [5–8]. However, one drawback of these clustering algorithms is that they treat all features equally in deciding the cluster memberships of objects. This is not desirable in some applications, such as high-dimensions sparse data clustering, where the cluster structure in the dataset is often limited to a subset of features rather than the entire feature set. A better solution is to introduce the proper attribute weight into the clustering process. Weight assignment and feature selection have been a hot research topic in partitioning clustering analysis [9,10].

Desarbo et al. [11] propose the first attribute-weighted method based on K-means algorithm. In [12], Friedmn et al. present the COSA method to calculate a weight for each dimension in each cluster. Jing et al. [13] extend Friedmn's method to optimal variable weighting for high-dimensional sparse data clustering. Recently, Tsai et al. [14] suggest a novel feature weight self-adjustment mechanism embedded into k-means, where feature weight searching is modeled as an optimization problem to simultaneously minimize the separations within clusters and maximize the separations between clusters.

Compared with the hard clustering methods discussed above, the fuzzy clustering technique is peculiarly effective when the boundaries between clusters of data are ambiguous. Moreover, the degree of membership will help us discover complicated relations between a given data object and all clusters [15]. So many attribute-weighted fuzzy clustering methods have been proposed in the past few decades. In [16], Wang et al. use the weighted Euclidean distance to replace the general Euclidean distance in FCM. Borgelt [17] carries out clustering on the selected subspace instead of the full data space by directly assigning zero weights to features that have little information. But one limitation is that they generate the attribute weights from all clusters, rather than each cluster. Keller et al. [18] introduce a basic attribute-weighted FCM algorithm by assigning one influence parameter to each single data dimension for each cluster, while Frigui et al. [19] put forward an approach searching for the optimal prototype parameters and the optimal feature weights simultaneously. Recently, Deng et al. present an enhanced soft subspace clustering (ESSC) algorithm by employing both within-cluster and between-cluster information [20]. In [21], a novel subspace clustering technique has been proposed by introducing the feature interaction using the concepts of fuzzy measures and the Choquet integral. Finally, in [22], Tang et al. give a survey of weighted clustering technologies.

[*] Corresponding author. Tel.: +86 853 83974950.
E-mail addresses: ise_zhouj@ujn.edu.cn (J. Zhou), longchen@umac.mo (L. Chen), Philip.Chen@ieee.org (C.L.P. Chen), yzhang@ujn.edu.cn (Y. Zhang), mehxli@cityu.edu.hk (H.-X. Li).

http://dx.doi.org/10.1016/j.neucom.2015.09.127
0925-2312/© 2016 Elsevier B.V. All rights reserved.

These weighted fuzzy clustering methods are effective in data clustering and feature selection, but there are still some questions. (1) There is still no good criterion, which is not only valid for the attribute weight assignment but also has a clear physical meaning; (2) for some particular applications, such as 'non-spherical' shaped data clustering, traditional weighted clustering algorithms can not show good performance. Different from other weighted fuzzy clustering methods, we propose a maximum-entropy-regularized weighted fuzzy c-means (EWFCM) clustering algorithm, in which the attribute-weight-entropy regularization is defined in the new objective function to achieve the optimal distribution of attribute weights. So that we can simultaneously minimize the dispersion within clusters and maximize the entropy of attribute weights to stimulate important attributes to contribute to the identification of clusters. Then the good clustering result can be yielded and the important attributes can be extracted for cluster identification. Moreover, for the 'non-spherical' shaped data clustering, developments on kernel technique and their applications have emphasized the need to consider kernel method into the data clustering, which is referred to as kernel-based clustering [25–29]. Therefore, the kernelization of the proposed algorithm is realized for clustering the data with 'non-spherical' shaped clusters. Experiments on synthetic and real-world datasets have demonstrated the efficiency of the proposed algorithms compared with traditional clustering approaches.

The rest of the paper is organized as follows. In Section 2, we present the EWFCM clustering algorithm based on attribute-weight-entropy regularization. The kernelization of EWFCM algorithm is given in Section 3. Section 4 illustrates experiment results of different clustering algorithms on synthetic and real world datasets. Finally, conclusions are drawn in Section 5.

2. Maximum-entropy-regularized weighted fuzzy c-means clustering algorithm

2.1. Algorithm description

In our research, the attribute weighted fuzzy clustering problem is considered as a maximum entropy inference problem [23] (the proof is shown in the Appendix), which aims to search for global regularity and obtain the smoothest reconstructions from the available data. We focus on this problem and propose the maximum-entropy-regularized weighted fuzzy c-means (EWFCM) clustering algorithm, in which a new objective function is developed as (1) by combining the weighted dissimilarity measure and an extra term for attribute-weight-entropy regularization.

$$F(\boldsymbol{U},\boldsymbol{C},\boldsymbol{W})=\sum_{k=1}^{K}\sum_{n=1}^{N}u_{kn}^{\alpha}\sum_{m=1}^{M}w_{km}(x_{nm}-c_{km})^2+\gamma^{-1}\sum_{k=1}^{K}\sum_{m=1}^{M}w_{km}\log w_{km} \tag{1}$$

Subject to $\sum_{k=1}^{K}u_{kn}=1, 0\le u_{kn}\le 1$

$$\sum_{m=1}^{M}w_{km}=1, 0\le w_{km}\le 1$$

where N is the number of objects, M is the number of object dimensions/attributes, K is the number of clusters, $\boldsymbol{U}=[u_{kn}]$ is a K-by-N matrix, u_{kn} denotes the degree of membership of the n-th object belonging to the k-th cluster, $\boldsymbol{C}=[c_{km}]$ is a K-by-M matrix, c_{km} denotes cluster center of the k-th cluster defined by u_{kn}, $\boldsymbol{W}=[w_{km}]$ is a K-by-M matrix, w_{km} indicates the attribute weight of the m-th object dimension in the k-th cluster. $\alpha(\alpha>1)$ is the fuzzification coefficient. γ is a positive scalar.

In this new objective function, the first distance-based term controls the shape and size of the clusters and encourages the agglomeration of clusters, while the second term is the negative entropy of attribute weights that regularize the optimal distribution of all attribute weights according with the available data. So that we can simultaneously minimize the dispersion within clusters and maximize the entropy of attribute weights to stimulate important attributes to contribute to the identification of clusters. $\gamma(\gamma>0)$ is a positive regularizing and adjustable parameter. With a proper choice of γ, we can balance the two terms and achieve the optimal and stable solution.

Minimizing $F(\boldsymbol{U}, \boldsymbol{C}, \boldsymbol{W})$ with respect to normalization constraints is a constrained nonlinear optimization problem. Like the traditional FCM algorithm, Picard iteration is also applied to solve this problem. We first fix $\boldsymbol{C}$ and $\boldsymbol{W}$ and find necessary conditions on $\boldsymbol{U}$ to minimize $F(\boldsymbol{U})$. Then we fix $\boldsymbol{W}$ and $\boldsymbol{U}$ and minimize $F(\boldsymbol{C})$ with respect to $\boldsymbol{C}$. Finally, we fix $\boldsymbol{U}$ and $\boldsymbol{C}$ and minimize $F(\boldsymbol{W})$ with respect to $\boldsymbol{W}$. The matrices $\boldsymbol{U}$, $\boldsymbol{C}$ and $\boldsymbol{W}$ are updated according to the Eqs. (2), (3) and (4) respectively.

$$u_{kn}=\frac{1}{\sum_{h=1}^{K}\left(\frac{\sum_{m=1}^{M}w_{km}(x_{nm}-c_{km})^2}{\sum_{m=1}^{M}w_{hm}(x_{nm}-c_{hm})^2}\right)^{\frac{1}{\alpha-1}}} \tag{2}$$

for $1\le k\le K, 1\le n\le N$

$$c_{km}=\frac{\sum_{n=1}^{N}u_{kn}^{\alpha}x_{nm}}{\sum_{n=1}^{N}u_{kn}^{\alpha}} \tag{3}$$

for $1\le k\le K, 1\le m\le M$

$$w_{km}=\frac{\exp\left(-\gamma\sum_{n=1}^{N}u_{kn}^{\alpha}(x_{nm}-c_{km})^2\right)}{\sum_{s=1}^{M}\exp\left(-\gamma\sum_{n=1}^{N}u_{kn}^{\alpha}(x_{ns}-c_{ks})^2\right)} \tag{4}$$

for $1\le k\le K, 1\le m\le M$

The process is repeated until the stop criteria are satisfied like the improvement in the objective function values is small enough. The pseudo-code of EWFCM algorithm is summarized as follows.

Algorithm. EWFCM Clustering Algorithm

```
Input: The number of objects $N$, object attributes $M$, clusters $K$;
parameter $\alpha$ and $\gamma$, and a small enough error $\varepsilon$.
Randomly choose $K$ objects as initial cluster centers and set all
initial weights of each attribute to $1/M$, $t$=0;
Repeat
    Update the partition matrix $\boldsymbol{U}$ by (2);
    Update the cluster center matrix $\boldsymbol{C}$ by (3);
    Update the attribute weight matrix $\boldsymbol{W}$ by (4);
    Calculate the objective function $F^{(t)}$ by (1);
    $t++$;
Until $\left|F^{(t)}-F^{(t-1)}\right|<\varepsilon$
```

According to the definitions above, the computational complexity of EWFCM is $O(tKNM)$, where t is the total number of iterations required, and N, M, K indicate the number of data objects, data dimensions and clusters respectively. For the storage, we need $O(NM + 2KM+KN)$ space to keep the data objects (NM), the cluster centers (KM), the membership matrix (KN) and the