Neurocomputing 136 (2014) 49–55

Contents lists available at ScienceDirect

Neurocomputing

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

Locally incremental visual cluster analysis using Markov random field

Zhen Zhou, Li Zhong, Liang Wang *

Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Article history:
Received 22 October 2013
Received in revised form
7 January 2014
Accepted 15 January 2014
Communicated by L. Shao
Available online 17 February 2014

Keywords:
Visual cluster analysis
Markov random field
Visual assessment tendency

ABSTRACT

Clustering methods are widely deployed in the fields of data mining and pattern recognition. Many of them require the number of clusters as the input, which may not be practical when it is totally unknown. Several existing visual methods for cluster tendency assessment can be used to estimate the number of clusters by displaying the pairwise dissimilarity matrix into an intensity image where objects are reordered to reveal the hidden data structure as dark blocks along the diagonal. A major limitation of the existing methods is that they are not capable to highlight cluster structure with complex clusters. To address this problem, this paper proposes an effective approach by using Markov Random Fields, which updates each object with its local information dynamically and maximizes the global probability measure. The proposed method can be used to determine the cluster tendency and partition data simultaneously. Experimental results on synthetic and real-world datasets demonstrate the effectiveness of the proposed method.

1. Introduction

Clustering is a fundamental technique in pattern recognition and data mining [1,2]. The principle is to maximize inter-cluster difference and minimize intra-cluster difference. Given a set of data points, there are two major clustering approaches. One is hard clustering, for example, *K*-means [3]. The other is fuzzy clustering, for example, fuzzy *C*-means [4]. For those methods requiring the number of clusters as the input, the first step is to estimate the cluster tendency.

Visual clustering methods are widely used to assess the cluster tendency, which generally require the pairwise dissimilarity matrix as the input. A popular method is *V*isual *A*ssessment of cluster *T*endency (VAT) [5]. The method produces an intensity image, called *O*rdered *D*issimilarity *I*mage (RDI) or *Re*ordered *D*issimilarity *I*mage (RDI) [6], which is able to reveal cluster tendency after suitably reordering the dissimilarity matrix [7].

Several algorithms extend the VAT method. bigVAT [8] and sVAT [9] offer different ways to approximate RDIs for large scale datasets. CCE [10] and DBE [11] use different schemes to automatically estimate the number of clusters in RDIs. It has been found that VAT is effective when data contains compact structures. However, many real datasets involve highly irregular structures.

In this paper, we propose a VAT based approach by using *M*arkov *R*andom *F*ield (MRF) [12] to handle with complex data structures. Meanwhile, the membership matrix is computed simultaneously, which can be directly used for fuzzy clustering. Experimental results on synthesis and real datasets have demonstrated the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 2 briefly reviews the VAT algorithm and its extensions. The proposed method is introduced in Section 3. The experimental results are displayed and analyzed in Section 4, prior to the conclusion in Section 5.

2. VAT and its extensions

Suppose $X=\{x_i\}_n$ denotes the set of n data points and $D=[d_{ij}]_{n\times n}$ is the dissimilarity matrix, of which each element denotes the difference between two data points, satisfying $d_{ij}=d_{ji}$ and $d_{ii}=0$. The goal of the VAT method is to find a permutation rule to reorder the rows of the dissimilarity matrix. The reordered dissimilarity matrix, $\tilde{D}$, displays as an intensity image. Such $\tilde{D}$ is expected to have a block diagonal form. If x_i is a member of a cluster, the corresponding d_{ij} should be in a sub-matrix with low dissimilarity values, which appears as one of the dark blocks along the diagonal of the VAT image. Each dark block can be regard as a potential group. The VAT algorithm can be found in Table 1 taken from [13].

* Corresponding author.
E-mail addresses: zzhou@nlpr.ia.ac.cn (Z. Zhou), rzonelee@mail.ustc.edu.cn (L. Zhong), wangliang@nlpr.ia.ac.cn (L. Wang).

http://dx.doi.org/10.1016/j.neucom.2014.01.032
0925-2312

Table 1
VAT algorithm.

Input: An $n \times n$ scaled matrix of pairwise dissimilarities $D=[d_{ij}]$, with $1 \geq d_{ij} \geq 0; d_{ij}=d_{ji}; d_{ii}=0$, for $1 \leq i,j \leq n$
(1): Set $I=\varnothing$, $J=\{1,2,\ldots,n\}$ and $\pi=(0,0,\ldots,0)$.
Select $(i,j) \in \arg_{p \in I, q \in J} \max \{d_{pq}\}$.
Set $\pi(1)=i$, $I \leftarrow \{i\}$ and $J \leftarrow J-\{i\}$.
(2): Repeat for $t=2,3,\ldots,n$
Select $(i,j) \in \arg_{p \in I, q \in J} \min \{d_{pq}\}$.
Set $\pi(t)=j$, update $I \leftarrow I \cup \{j\}$ and $J \leftarrow J-\{j\}$.
(3): Form the reordered matrix $\tilde{D}=[\tilde{d_{ij}}]=[d_{\pi(i)\pi(j)}]$, for $1 \leq i,j \leq n$.
Output: A scaled gray-scale image $I(\tilde{D})$, in which $\max\{\tilde{d_{ij}}\}$ corresponds to *white* and $\min\{\tilde{d_{ij}}\}$ to *black*.

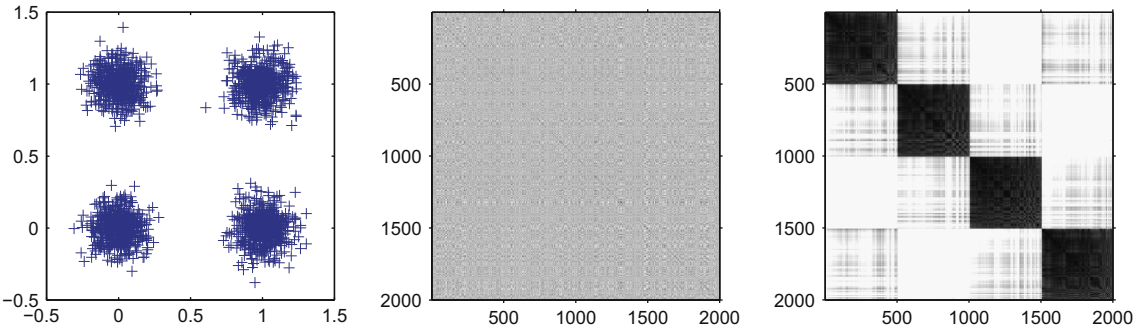

Fig. 1. The original scatter (*left*) includes four groups. The VAT algorithm uses the pairwise dissimilarity matrix (*middle*) of the scatter as the input and generates a reordered pairwise matrix (*right*). There exists 4 distinct black squares along the diagonal, which indicates 4 clusters in the original scatter.

An example of VAT is shown in Fig. 1, in which (*left*) the scatter plots of data consist of $n=4000$ points in $\mathcal{R}^2$. These data points are generated from a mixture of bi-variate Gaussian distributions. The dissimilarity metric is computed by the Euclidean distance between every two points. The 4 visually apparent clusters in (*left*) are displayed by 4 distinct dark blocks along the diagonal in (*right*), which is the VAT image of the data. Therefore, reordering is necessary to reveal the underlying cluster structure of data, in contrast to the image (*middle*) of the input pairwise dissimilarity matrix in an original order.

The VAT algorithm performs well on datasets with compact structures. However, such assumption is often hard to satisfy in the real-world task. Fig. 2 is a toy example of this case. To address this problem, several extensions have been proposed, e.g., improved VAT (iVAT) [6] and Spectral VAT (SpecVAT) [13]. The iVAT algorithm is based on the path-transformation (Eq. (1)). Here, d_{ij} represents the weight of the edge between x_i and x_j, P_{ij} is the set of all possible paths from x_i to x_j. $|p|$ is the number of vertices along path p and $p[h]$ is the index of the hth vertex along path p. For each path $p \in P_{ij}$, the effective dissimilarity between x_i and x_j along p is the maximum of all weights of this path. Reordering the new matrix $D'=[d'_{ij}]_{n \times n}$ obtains the iVAT image (Fig. 2 *right*).

$$d'_{ij}=\min_{p \in P_{ij}}\left\{\max_{1 \leq h < |p|} d_{p[h]p[h+1]}\right\}. \quad (1)$$

The SpecVAT algorithm is based on spectral transformation [13]. The key of SpecVAT is graph embedding, which first calculates a weighted affinity matrix and uses the Laplacian Eigenmap to transform the pairwise dissimilarity matrix, and then applies the VAT algorithm on this transformed matrix.

3. Markov random field VAT (MrfVAT)

The VAT algorithm often fails on datasets with highly complicated structure. And iVAT has limitations when there is too much noise along the paths when conducting path-transformation. SpecVAT, which uses graph-embedding techniques, has also two major weaknesses: one is that the qualities of reordered images by SpecVAT largely depend on the selection of the number of eigenvectors; the other is that SpectVAT fails on sparse and uneven datasets, due to its dimensionality reduction method. Besides, these algorithms can be generally used for cluster tendency assessment, not capable for data partition and labeling. Based on those intuitions, we propose a novel method for visual cluster analysis with the Markov Random Field model [12].

The first step of our model is to construct a graph based on the input dataset. In our case, it is a full connected graph, where each data point is regarded as a node and the edge between two nodes is their dissimilarity computed by the Euclidean distance, thus leading to an undirected graphical model. Meanwhile, it is natural to think that the neighbor of a node will provide extra information. If we change the range of the neighbor from several nearest nodes to the entire graph, it obviously provides the spatial information. This is the way to compute the global probability. With more exploration, we find that the graph holds the local Markov property [14]. Second, we construct a scatter graph representation for the dissimilarity matrix. The next step is to build the local update system by the Markov property. Meanwhile, a criterion is set to stop the iteration. When we find the optimal k nearest neighborhood system (later will explain), we will complete data partition.

Here is the details of our method. At first, there are n groups, written as $\Omega=\{\omega_i=i\}_n$ where ω_i is the label of x_i. The membership matrix $B=[b_{ij}]_{n \times n}$ is set to identity matrix, where $b_{ij}=p(\omega_i=j)$. Definition 1 is the neighborhood system, which will be used to compute the global probability in Definition 2.

Definition 1. $X=\{x_i\}_n$ is the set of nodes and $S=\{S_{x_i} \subseteq X | x_i \in X\}$ is a family of subsets of X. Note that S_{x_i} is the neighborhood of x_i and $x_i \notin S_{x_i}$. S is called a neighborhood system, if $\forall r,t \in X$, $t \in S_r \iff r \in S_t$. That means the neighborhood relation is anti-reflexive and symmetric, but not transitive [15].