Pattern Recognition Letters 52 (2015) 9–16

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Multiple circle detection based on center-based clustering ☆

Rudolf Scitovski, Tomislav Marošević*

Department of Mathematics, University of Osijek, Trg Lj. Gaja 6, HR-31 000 Osijek, Croatia

ARTICLE INFO

Article history:
Received 1 October 2013
Available online 30 September 2014

Keywords:
Multiple circle detection
Center-based clustering
Globally optimal partition
Approximate optimization
DIRECT

ABSTRACT

The multiple circle detection problem has been considered in the paper on the basis of given data point set $\mathcal{A} \subset \mathbb{R}^2$. It is supposed that all data points from the set $\mathcal{A}$ come from k circles that should be reconstructed or detected. The problem has been solved by the application of center-based clustering of the set $\mathcal{A}$, i.e. an optimal k-partition is searched for, whose clusters are determined by corresponding circle-centers. Thereby, the algebraic distance from a point to the circle is used. First, an adaptation of the well-known k-means algorithm is given in the paper. Also, the incremental algorithm for searching for an approximate globally optimal k-partition is proposed. The algorithm locates either a globally optimal k-partition or a locally optimal k-partition close to the global one. Since optimal partitions with $2, 3, \ldots$ clusters are determined successively in the algorithm, several well-known indexes for determining an appropriate number of clusters in a partition are adopted for this case. Thereby, the Hausdorff distance between two circles is used and adopted. The proposed method and algorithm are illustrated and tested on several numerical examples.

1. Introduction

Clustering or grouping a data set into conceptually meaningful clusters is a well-studied problem in recent literature, and it has practical importance in a wide variety of applications such as medicine, biology, pattern recognition, facility location problem, text classification, information retrieval, earthquake investigation, understanding the Earth's climate, psychology, ranking of municipalities for financial support, business, etc. [21,23,26,29,28,35].

A multiple circle detection problem is considered in the paper based on given data set $\mathcal{A} \subset \mathbb{R}^2$. Let us assume that all data from the set $\mathcal{A} = \{a_i = (x_i, y_i) \in \mathbb{R}^2 : i = 1, \ldots, m\} \subset \mathbb{R}^2$ come from k circles that should be reconstructed or detected. There are many different approaches for solving this problem in literature, which are most often based on Hough transformation [2], different heuristic methods [8,9,18,27], RANSAC [11] or fuzzy clustering techniques, that search for the so-called soft or fuzzy partitions [5,33,37].

A center-based clustering method is applied to solving this problem [19,28,36]. The set $\mathcal{A}$ will be grouped into k nonempty disjoint subsets $\pi_1, \ldots, \pi_k$, $1 \le k \le m$, such that

$$\bigcup_{i=1}^{k} \pi_i = \mathcal{A}, \quad \pi_r \cap \pi_s = \emptyset,\ r \neq s, \quad |\pi_j| \ge 1,\ j = 1, \ldots, k. \tag{1}$$

This partition will be denoted by $\Pi(\mathcal{A}) = \{\pi_1, \ldots, \pi_k\}$, and the elements $\pi_1, \ldots, \pi_k$ of such a partition are called *clusters* in $\mathbb{R}^2$.

To each cluster $\pi_j \in \Pi$ a corresponding circle-center $C_j^\star(S_j^\star, r_j^\star)$ with center $S_j^\star = (p_j^\star, q_j^\star)$ and radius $r_j^\star$ is associated by solving the following global optimization problem (GOP)

$$(p_j^\star, q_j^\star, r_j^\star) = \underset{p,q,r \in \mathbb{R}}{\operatorname{argmin}} \sum_{a_i \in \pi_j} D(C(p, q, r), a_i), \tag{2}$$

where $D(C(p, q, r), a_i)$ represents the distance from the point a_i to the circle C (see Section 2.1).

If the objective function $\mathcal{F} : \mathcal{P}(\mathcal{A}; m, k) \to \mathbb{R}_+$ is defined on the set of all partitions $\mathcal{P}(\mathcal{A}; m, k)$ of the set $\mathcal{A}$ containing k clusters, then the quality of a partition may be defined and one can search for the *globally optimal k-partition* by solving the following GOP

$$\underset{\Pi \in \mathcal{P}(\mathcal{A}; m,k)}{\operatorname{argmin}} \mathcal{F}(\Pi), \quad \mathcal{F}(\Pi) = \sum_{j=1}^{k} \sum_{a_i \in \pi_j} D(C_j(p_j, q_j, r_j), a_i). \tag{3}$$

Conversely, for a given set of different circles $C_1, \ldots, C_k \subset \mathbb{R}^2$, applying the minimal distance principle, the partition $\Pi = \{\pi_1, \ldots, \pi_k\}$ of the set $\mathcal{A}$ can be defined in the following way:

$$\pi_j = \{a \in \mathcal{A} : D(C_j, a) \le D(C_s, a),\ \forall s = 1, \ldots, k,\ s \neq j\},\ j = 1, \ldots, k, \tag{4}$$

where one has to pay attention to the fact that every element of the set $\mathcal{A}$ occurs in one and only one cluster. Therefore, the problem of finding an optimal partition of the set $\mathcal{A}$ can be reduced to the

☆ This paper has been recommended for acceptance by S. Wang.

* Corresponding author. Tel.: +385 31 224 800; fax: +385 31 224 801.
E-mail address: tmarosev@mathos.hr (T. Marošević).

http://dx.doi.org/10.1016/j.patrec.2014.09.010
0167-8655/

following optimization problem

$$\underset{C_1,\ldots,C_k\subset\mathbb{R}^2}{\operatorname{argmin}} F(C_1,\ldots,C_k),$$

$$F(C_1,\ldots,C_k)=\sum_{i=1}^{m}\min_{j=1,\ldots,k} D(C_j,a_i). \tag{5}$$

The solution of (3) and (5) coincides [34]. In general, the function F can have a great number of independent variables, it does not have to be either convex or differentiable and usually it has several local and global minima. Hence, this becomes a complex GOP.

The adaptation of the well-known k-means algorithm for searching for a locally optimal partition is given in Section 2, where clusters are determined by corresponding circle-centers. In Section 3, a new algorithm for searching for an optimal partition is proposed and an illustrative example is shown. In Section 4, the problem of determining an appropriate number of clusters in a partition is considered. The well-known Davies-Bouldin index, Calinski-Harabasz index, and Simplified Silhouette Width criterion are adopted and used for the case observed in the paper. Finally, in Section 5, several numerical examples are shown.

2. Adaptation of the k-means algorithm

The well-known k-means algorithm [19,20,36] will be adapted for searching for a locally optimal partition with circles as clusters-centers. The algorithm can be described in two steps which are repeated iteratively:

Algorithm 1. (k closest circles algorithm (KCC))

Step A: For each set of mutually different circles $C_1,\ldots,C_k$ the set $\mathcal{A}$ should be divided into k disjoint unempty clusters $\pi_1,\ldots,\pi_k$ by using the minimal distance principle (4);

Step B: Given a partition $\Pi=\{\pi_1,\ldots,\pi_k\}$ of the set $\mathcal{A}$, one can define the corresponding circle-centers $C_1^\star(p_1^\star,q_1^\star,r_1^\star),\ldots,C_k^\star(p_k^\star,q_k^\star,r_k^\star)$ by solving the following GOPs ($j=1,\ldots,k$)

$$(p_j^\star,q_j^\star,r_j^\star)=\underset{p,q,r\in\mathbb{R}}{\operatorname{argmin}}\sum_{a_i\in\pi_j} D(C(p,q,r),a_i). \tag{6}$$

Knowing a good initial approximation, this algorithm can provide an acceptable solution, but in case we do not have a good initial approximation, the KCC-algorithm can be restarted several times with various random initializations [20].

2.1. The distance from the point to the circle

In order to solve the multiple circle detection problem, it is very important for the distance between a point and a circle to be well defined. This distance will be used by applying the minimal distance principle (4), by determining circle-centers of clusters (6), and by defining the most appropriate number of clusters in a partition in Section 4. Several approaches to determining the distance from the point $a_i=(x_i,y_i)\in\mathbb{R}^2$ to the circle $C(S,r)$ with center $S=(p,q)$ and radius r have been proposed in literature [1,7,10,25,33]:

$$D_1(C(S,r),a_i)=|\|S-a_i\|-r|, \tag{7}$$

$$D_2(C(S,r),a_i)=(\|S-a_i\|-r)^2, \tag{8}$$

$$D(C(S,r),a_i)=(\|S-a_i\|^2-r^2)^2. \tag{9}$$

The last possibility (9) is called the *algebraic distance*. It is very often used in practical applications (see e.g. [25,33]) and for that reason that possibility is also used in our paper.

2.2. Searching for the circle-center of a cluster

The GOPs (6) can have several local and global minima and the corresponding minimizing function is continuous. Note that, if (7) or (8) were used as the distance from the point a_i to the circle C, then the corresponding minimizing function would be a Lipschitz continuous function (see e.g. [32]). Therefore, for solving GOPs (6) in these cases, some global optimization methods (see e.g. [15,24,26,31] and corresponding software http://www.pinterconsulting.com) can be used. One of the most popular algorithms for solving a GOP for the Lipschitz continuous function is the DIRECT (DIvidingRECTangles) algorithm [12,15,14].

In the case of using the algebraic distance (9), exact solutions of GOPs (6) can be obtained [7]. In the case of using the distance (7) or (8), special methods can also be applied for solving GOPs (6) [25]. The random circle consensus technique—RANSAC [11] can also be used for circle fitting.

3. A method for searching for a globally optimal partition

Let $\mathcal{A}\subset[a,b]\times[c,d]\subset\mathbb{R}^2$ be a data points set. A globally optimal k-partition $\Pi^\star=\{\pi_1^\star,\ldots,\pi_k^\star\}$ with circle-centers $C_1^\star,\ldots,C_k^\star$ should be determined as a solution of GOP (3), or equivalently (5).

Since objective function (5) is a Lipschitz continuous function, the aforementioned DIRECT algorithm for global optimization can be used. Because of the symmetry property of the function F defined by (5), there are at least $k!$ solutions of this problem. A very efficient special version of the DIRECT algorithm for symmetric functions is developed in [14]. However, complexity of this problem is particularly emphasized if the number of clusters or data points m is large. Instead of solving the GOP (5), motivated by [3,4,22,30], an efficient method for searching for a very acceptable approximate of an optimal k-partition is proposed.

First, a sequence of objective functions $F_k:\underbrace{\mathbb{R}^3\times\cdots\times\mathbb{R}^3}_{k}\to\mathbb{R}_+$,

$$F_k(C_1,\ldots,C_k)=\sum_{i=1}^{m}\min\{D(C_1,a_i),\ldots,D(C_k,a_i)\}, \tag{10}$$

is defined, where $C_j:=(S_j,r_j)$, $S_j=(p_j,q_j)$.

For $k=1$, the problem is reduced to a simple problem of locating a circle on the basis of a given set of points in the plane. The corresponding function $F_1:\mathbb{R}^3\to\mathbb{R}_+$ is of the form

$$\begin{aligned}F_1(p_1,q_1,r_1)&=\sum_{i=1}^{m}D(C_1(p_1,q_1,r_1),a_i)\\&=\sum_{i=1}^{m}(\|S_1-a_i\|^2-r_1^2)^2,\end{aligned} \tag{11}$$

where $S_1=(p_1,q_1)$. There exists a number of algorithms and methods for solving this problem (see e.g. [7,10,25]).

For $k>1$, an optimal k-partition with circle-centers $C_1^{\star(k)},\ldots,C_k^{\star(k)}$ is determined by the following incremental algorithm.

Algorithm 2. (Searching for an optimal k-partition)

Step 1: Let $\hat{C}_1,\ldots,\hat{C}_{k-1}$ be the solution to the $k-1$-circle detecting problem and let

$$\begin{aligned}F_{k-1}(\hat{C}_1,\ldots,\hat{C}_{k-1})&=\sum_{i=1}^{m}\delta_{k-1}^i,\\ \delta_{k-1}^i&=\min\{D(\hat{C}_1,a_i),\ldots,D(\hat{C}_{k-1},a_i)\},\end{aligned} \tag{12}$$

$$\begin{aligned}\Phi_k(p,q,r)&:=F_k(\hat{C}_1,\ldots,\hat{C}_{k-1},C(p,q,r))\\&=\sum_{i=1}^{m}\min\{\delta_{k-1}^i,D(C(p,q,r),a_i)\}.\end{aligned} \tag{13}$$