Pattern Recognition Letters 71 (2016) 78–85

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

An angle-based neighborhood graph classifier with evidential reasoning☆

Yi Yang [a], De-Qiang Han [b,*], Jean Dezert [c]

[a] SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[b] Center for Information Engineering Science Research, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[c] ONERA, The French Aerospace Lab, Chemin de la Hunière, F-91761 Palaiseau, France

ARTICLE INFO

Article history:
Received 24 April 2015
Available online 29 December 2015

Keywords:
Neighborhood classifier
Neighborhood graph
Belief functions
Pattern classification
Geometrical relation

ABSTRACT

A classification approach called angle-based neighborhood graph (ANG) is proposed in this paper, which can flexibly define the neighborhood of a given query sample based on the geometrical relation established using an angle parameter. The proposed ANG is geometrically intuitive and can be readily implemented. Compared with the traditional neighborhood graph classifiers, ANG can adjust the size of the neighborhood by tuning the angle parameter to obtain better classification accuracy. To deal with the parameter selection in ANG, an evidential reasoning based approach is proposed. Experimental results are provided for comparing ANG and the traditional neighborhood graph classifiers, including Gabriel Graph (GG), Relative Neighborhood Graph (RNG), β skeletons, and adaptive weighted k nearest neighbors classifiers. It can be concluded that ANG is a simple yet flexible and effective classifier, and the evidential reasoning based parameter selection approach for ANG is also effective.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Important and efficient techniques developed in the supervised pattern classification exploit neighbors in the data (or attribute) space of the data to classify [4]. They are referred as neighborhood-based classifiers. The simple yet effective nearest neighbor (NN) and the k nearest neighbor classifiers (k-NN) are typical representatives [4,16], where the neighbors are defined using the distance between the query sample and the training samples. Obviously, they only take into account that the neighbors should be as close to a query sample as possible.

Traditional k-NN adopts a fixed k for all query samples regardless of their geometric location and related specialties. Furthermore, those k nearest neighbors may not distribute symmetrically around the query sample if the neighborhood in the training set is not spatially homogeneous. The geometrical placement might be more important than the actual distance to depict a query sample's neighborhood. Therefore, to improve neighborhood based classifiers, it is appropriate to adopt the graph techniques to fully use the geometric information. Some graph based neighborhood classifiers have been proposed accordingly. For example, the adaptive graph based k-NN [9], which adaptively uses different k values for different query samples based on geometric graph. The other type of graph-based classifiers use the geometrical relationship between the query sample and the training samples [2,11]. Such geometric relationships are defined based on the geometric graph such as Gabriel Graph (GG) [6], the Relative Neighborhood Graph (RNG) [6], and β-skeleton [7,17]. These graph-based neighborhood definitions consider both the distance and the geometrical placement, which are more comprehensive. Note that GG and RNG based classifiers are non-parametric [2]. Once the training set is given, the surrounding neighbors of a given query sample will be automatically determined. This is an advantage and also a disadvantage of GG and RNG. The "advantage" means that there is no problem of parameter selection such as the selection of k in k-NN. The "disadvantage" means that the number of neighbors are fixed for a given query sample, therefore, it is impossible to make further the optimization of classification performance. β-skeleton has a parameter of k, based on which, the size and shape of the graph can be changed. This makes the further optimization possible; however, there exists the problem of parameter selection.

In this paper, we implement a graph based classifier with a comprehensive neighborhood definition, and meanwhile with a parameter to make the optimization possible. An angle parameter is used to define the neighborhood based on the geometrical relationship between the interior angle and the circle angle. The size of the neighborhood can be enlarged or reduced by adjusting the angle. As aforementioned, having a parameter is a "double-edged

☆ This paper has been recommended for acceptance by F. Tortorella.

* Corresponding author. Tel.: +86 29 8266 8775; fax: +86 29 8266 8775.
E-mail address: deqhan@gmail.com, deqhan@mail.xjtu.edu.cn (D.-Q. Han).

http://dx.doi.org/10.1016/j.patrec.2015.11.027
0167-8655/© 2015 Elsevier B.V. All rights reserved.

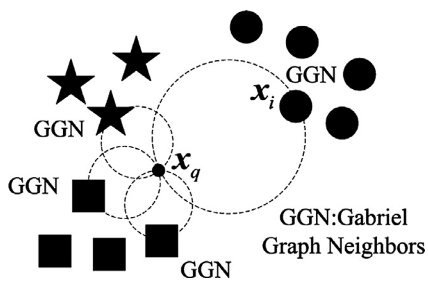

Fig. 1. GGN.

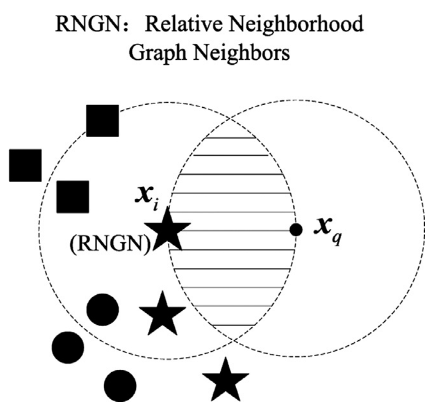

Fig. 2. RNGN.

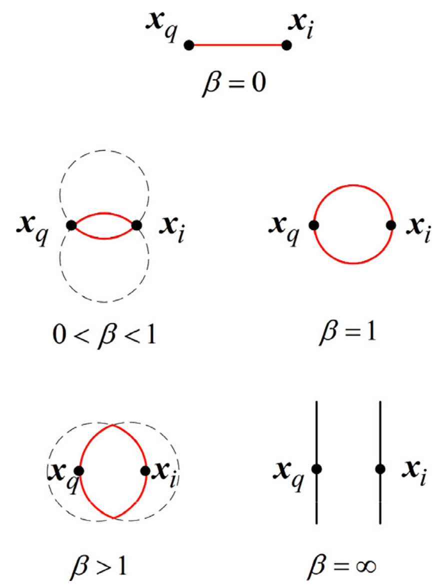

Fig. 3. Beta skeletons.

sword". Therefore, we accordingly propose an evidential reasoning [12] based approach to simplify the parameter selection. Experimental results based on some artificial and public data sets show the efficiency of the proposed angle based neighborhood classifier and the related evidential reasoning based parameter selection approach.

2. Typical available graph neighborhood classifiers

In classification, geometrical placement can be much more important than actual distances [11]. Some geometric graphs [6,10] such as Gabriel graph neighbors (GGN) , relative neighborhood graph neighbors (RNGN), and β-skeletons [7,17] can be used to establish classifiers which consider both distance and spatial distribution according to the query sample using sample pair.

2.1. Gabriel graph based classifier

$X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ is a training sample set. $\boldsymbol{x}_q$ is a testing sample and $d(\cdot, \cdot)$ is Euclidean distance. If the condition

$$d^2(\boldsymbol{x}_q, \boldsymbol{x}_i) \le d^2(\boldsymbol{x}_q, \boldsymbol{x}_j) + d^2(\boldsymbol{x}_i, \boldsymbol{x}_j), \forall \boldsymbol{x}_j \in X, i \ne j \tag{1}$$

is satisfied, $\boldsymbol{x}_i$ is called the Gabriel graph neighbors (GGN) of $\boldsymbol{x}_q$. It means that there is no other point from X lying in the hypersphere centered at their middle point and whose diameter is the distance between them, as illustrated in Fig. 1.

Find all the GGNs of the testing sample $\boldsymbol{x}_q$, then assign the dominant class of all the GGNs to $\boldsymbol{x}_q$.

2.2. Relative neighborhood graph based classifier

If the condition in (2) is satisfied, $\boldsymbol{x}_i$ is called the relative neighborhood graph neighbors (RNGN) of $\boldsymbol{x}_q$.

$$d(\boldsymbol{x}_q, \boldsymbol{x}_i) \le \max(d(\boldsymbol{x}_q, \boldsymbol{x}_j), d(\boldsymbol{x}_i, \boldsymbol{x}_j)), \ \forall \boldsymbol{x}_j \in X, i \ne j \tag{2}$$

A geometric interpretation of RNGN is illustrated in Fig. 2. A lune is defined as the intersection between two hyper-spheres centered at $\boldsymbol{x}_i$ and $\boldsymbol{x}_q$. Both the hyper-spheres' radius are distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_q$. If there is no other training point lying in the lune defined, $\boldsymbol{x}_i$ is called the relative neighbor (RN) of $\boldsymbol{x}_q$. The distance adopted is always Euclidean distance.

Find all RNGNs of the query sample $\boldsymbol{x}_q$, then assign the dominant class of all RNGNs to $\boldsymbol{x}_q$. As we can see, GGN and RNGN use sample pair (a training sample and a testing sample) to define geometric relationship and implement the classification.

Note that both GGN and RNGN have comprehensive definition of neighborhood, i.e., the neighbors are close to the query sample and their placement are surrounding the query sample. They are both non-parametric classifiers [11], which has no selection of parameters. However, since they are non-parametric, once the training set is given, the neighborhood will be determined, therefore, it is impossible to optimize the classification performance by adjusting some parameters.

2.3. β-skeletons based neighborhood classifier

β-skeletons are as shown in Fig. 3. For a query sample x_q and a training sample x_i, the region $I_\beta(x_q, x_i)$ containing no any other sample x_j, is defined as

- For $\beta = 0$, $I_\beta(x_q, x_i)$ is the line segment $x_q x_i$;
- For $0 < \beta < 1$, $I_\beta(x_q, x_i)$ is the intersection of the two disks of radius $d(x_q, x_i)/(2\beta)$ passing through both x_q and x_i. Here d is the Euclidean distance;
- For $1 < \beta < inf$, $I_\beta(x_q, x_i)$ is the intersection of the two disks of radius $\beta d(x_q, x_i)/2$ and centered at the points $(1 - \beta/2)x_q + (\beta/2)x_i$ and $(\beta/2)x_q + (1 - \beta/2)x_i$, respectively;
- For $\beta = Inf$, $I_\beta(x_q, x_i)$ is the infinite strip perpendicular to the line segment from x_q to x_i

Here x_i is called a β neighbor of x_q. Find all the β neighbors of the testing sample $\boldsymbol{x}_q$, then assign the dominant class of all the β neighbors to $\boldsymbol{x}_q$.

Note that when $\beta = 1$, β-skeletons becomes the GGN; and when $\beta = 2$, β-skeletons becomes the RNGN.

GGN, RNGN, and β-skeletons can all be categorized as empty region graphs [7], i.e., there is no other sample inside the region

Download English Version:

https://daneshyari.com/en/article/533950

Download Persian Version:

https://daneshyari.com/article/533950

Daneshyari.com