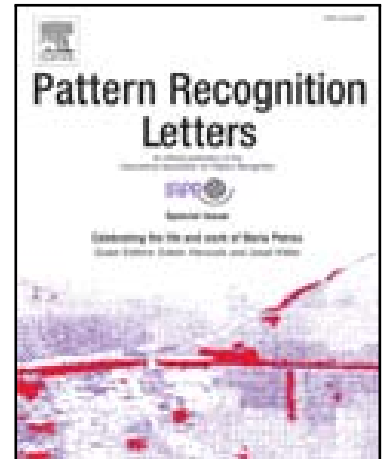


## Accepted Manuscript

Adaptive clustering algorithm based on kNN and density

Bing Shi , Lixin Han , Hong Yan

PII: S0167-8655(18)30026-6  
DOI: [10.1016/j.patrec.2018.01.020](https://doi.org/10.1016/j.patrec.2018.01.020)  
Reference: PATREC 7064



To appear in: *Pattern Recognition Letters*

Received date: 17 May 2017  
Revised date: 12 November 2017  
Accepted date: 22 January 2018

Please cite this article as: Bing Shi , Lixin Han , Hong Yan , Adaptive clustering algorithm based on kNN and density, *Pattern Recognition Letters* (2018), doi: [10.1016/j.patrec.2018.01.020](https://doi.org/10.1016/j.patrec.2018.01.020)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



## Adaptive clustering algorithm based on $k$ NN and density

Bing Shi<sup>a,b</sup>, Lixin Han<sup>a,\*</sup> and Hong Yan<sup>c</sup>

<sup>a</sup>School of Computer and Information, Hohai University, Nanjing 211100, China

<sup>b</sup>School of Computer and Information, Anqing Normal University, Anqing 246011, China

<sup>c</sup>Department of Electronic Engineering, City University of Hong Kong, Kowloon 999077, Hong Kong

### ABSTRACT

Although many clustering algorithms have been proposed, they all have various limitations. Existing clustering algorithms usually require the user to set the appropriate threshold parameters and those parameters are usually not adaptive. In this paper, a new clustering method is proposed. Firstly, the  $k$  nearest neighbors of all samples is calculated, and then a density method based on  $k$ NN is used to complete the clustering process. In order to achieve this goal, a statistics  $\alpha$  is proposed for measuring the degree of denseness or sparseness of distribution of the entire samples, and a statistical model is proposed to calculate the local parameters,  $\beta$  and  $\gamma$ . That is, the category of each sample is determined by the combination of  $k$ NN with density. That is to say, the total number of samples determines an appropriate  $k$  value to obtain the ordered  $k$ NN of every sample, and then the radius of surrounding region (SR) is calculated by 5NN of all samples. Furthermore, the density of every sample is calculated. In the whole process of clustering, the global threshold is determined by the density distribution of all samples, and then the local threshold is self-adaptive. All sample density is sorted to search automatically for clusters from the highest point of density of the distribution of all samples. The algorithm can not only discover clusters of arbitrary shapes and automatically remove noise and outliers, but also it can find clusters with different densities and those with internal density variation. The results of experiments on the synthetic data and human face image data demonstrate that the algorithm is effective. The code is available under the following URL: <https://github.com/bing-shi/Acnd>.

KEYWORDS:  $k$ NN, SR, density, adaptive, clustering

2017 Elsevier Ltd. All rights reserved.

\* Corresponding author. Tel.: +0-86-18915973008; fax: +0-000-000-0000; e-mail: [lhhan@hhu.edu.cn](mailto:lhhan@hhu.edu.cn)

Download English Version:

<https://daneshyari.com/en/article/6940620>

Download Persian Version:

<https://daneshyari.com/article/6940620>

[Daneshyari.com](https://daneshyari.com)