

Author's Accepted Manuscript

Clustering by fast search and find of density peaks
via heat diffusion

Rashid Mehmood, Guangzhi Zhang, Rongfang Bie,
Hassan Dawood, Haseeb Ahmad



PII: S0925-2312(16)30482-9
DOI: <http://dx.doi.org/10.1016/j.neucom.2016.01.102>
Reference: NEUCOM17107

To appear in: *Neurocomputing*

Received date: 27 October 2015
Revised date: 29 December 2015
Accepted date: 24 January 2016

Cite this article as: Rashid Mehmood, Guangzhi Zhang, Rongfang Bie, Hassan Dawood and Haseeb Ahmad, Clustering by fast search and find of density peak via heat diffusion, *Neurocomputing*
<http://dx.doi.org/10.1016/j.neucom.2016.01.102>

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 galley proof before it is published in its final citable 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

Clustering by fast search and find of density peaks via heat diffusion

Rashid Mehmood^{a,b}, Guangzhi Zhang^a, Rongfang Bie ^{*a}, Hassan Dawood^a, Haseeb Ahmad^c

^aCollege of Information Science and Technology, Beijing Normal University, Beijing 100875, China

^bDepartment of Computer Science and Information Technology, University of Management Sciences and Information Technology, Kotli (AJK, Pakistan)

^cInformation Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, Peoples Republic of China.

Abstract

Clustering by fast search and find of density peaks (CFSFDP) is a novel algorithm that efficiently discovers the centers of clusters by finding the density peaks. The accuracy of CFSFDP depends on the accurate estimation of densities for a given dataset and also on the selection of the cutoff distance (d_c). Mainly, d_c is used to calculate the density of each data point and to identify the border points in the clusters. CFSFDP necessitates using different methods for estimating the densities of different datasets and the estimation of d_c largely depends on subjective experience. To overcome the limitations of CFSFDP, this paper presents a method for CFSFDP via heat diffusion (CFSFDP-HD). CFSFDP-HD proposes a nonparametric method for estimating the probability distribution of a given dataset. Based on heat diffusion in an infinite domain, this method accounts for both selection of the cutoff distance and boundary correction of the kernel density estimation. Experimental results on standard clustering benchmark datasets validate the robustness and effectiveness of the proposed approach over CFSFDP, AP, mean shift, and K-means methods.

Keywords: Clustering, Probability density estimation, Kernel density estimation, Heat equation

1. Introduction

Clustering plays an important role in the fields of knowledge discovery and data mining. Clustering algorithms attempt to organize data into different disjoint categories, with more similar data points organized into the same cluster, while dissimilar data points are grouped into different clusters.

Clustering has been successfully applied in different fields such as bioinformatics [1–3], cyber security [4, 5], image processing [6–11], astronomy [12], social networks [13, 14] etc.

Email addresses: gulkhan007@gmail.com (Rashid Mehmood), zgz@mail.bnu.edu.cn (Guangzhi Zhang), rfbie@bnu.edu.cn (Rongfang Bie *), hasandawod@yahoo.com (Hassan Dawood), haseeb_ad@hotmail.com (Haseeb Ahmad)

Preprint submitted to Neurocomputing

May 30, 2016

Download English Version:

<https://daneshyari.com/en/article/4948618>

Download Persian Version:

<https://daneshyari.com/article/4948618>

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