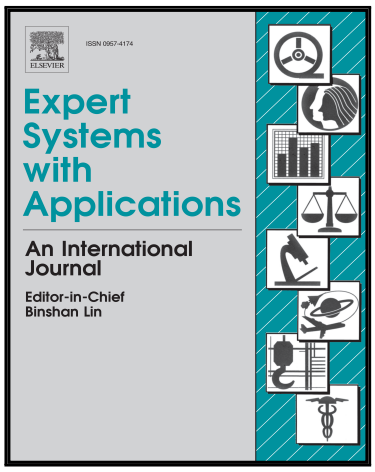

A Novel and Efficient Data Point Neighborhood Construction Algorithm based on Apollonius Circle

Shahin Pourbahrami , Leyli Mohammad Khanli , Sohrab Azimpour

PII: S0957-4174(18)30486-X
DOI: https://doi.org/10.1016/j.eswa.2018.07.066
Reference: ESWA 12120

To appear in: *Expert Systems With Applications*

Received date: 4 April 2018
Revised date: 23 July 2018
Accepted date: 29 July 2018

Please cite this article as: Shahin Pourbahrami , Leyli Mohammad Khanli , Sohrab Azimpour , A Novel and Efficient Data Point Neighborhood Construction Algorithm based on Apollonius Circle, *Expert Systems With Applications* (2018), doi: https://doi.org/10.1016/j.eswa.2018.07.066

A Novel and Efficient Data PointNeighborhoodConstructionAlgorithm based on Apollonius Circle

Shahin Pourbahrami[1], Leyli Mohammad Khanli[*1], Sohrab Azimpour[2]

Computer Engineering Department, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran[1]
Mathematic Department, Faculty of Mathematic, University of Farhangian, Tehran, Iran[2]

Abstract

Neighborhood construction models are important in finding connection among the data points, which helps demonstrateinterrelations among the information. Hence, employing a new approach to find neighborhood among the data points is a challenging issue. The methods, suggested so far, are not useful for simultaneous analysis of distances and precise examination of the geometric position of the data as well as their geometric relationships. Moreover, most of the suggested algorithms depend on regulating parameters including number of neighborhoods and limitations in fixed regions.The purpose of the proposed algorithm is to detect and offer an applied geometric pattern among the data through data mining. Precise geometric patterns are examined according to the relationships among the data in neighborhood space. These patterns can reveal the behavioural discipline and similarity across the data. It is assumed that there is no prior information about the data sets at hand. The aimof the present research study is to locate the precise neighborhood using Apollonius circle, which can help us identify the neighborhood state of data points. High efficiency of Apollonius structure in assessing local similarities among the observations has opened a new field of the science of geometry in data mining. In order to assess the proposed algorithm, its precision is compared with the state-of-the-art and well-known (k-Nearest Neighborand epsilon-neighborhood) algorithms.

Keywords: Apollonius circle, geometric patterns, neighborhoodconstruction

1. Introduction

Neighborhood includes a group of data points, which are locally similar to each other, being defined according to their local inter-relationships in a database (İnkaya et al., 2015). One of the main features of data mining is analysis of the data and their categorization in similar groups by a precise examination of the neighborhood features of the data. Neighborhood is a criterion to make precise neighborhood models by changing the data tendencies towards similar categories. Similarity, here it refers to the intrinsic dependency inside the data and automatic connection of similar sets of data.In other words, neighborhood for each set of data is defined in terms of similarity and connection among the data in such a way that the obtained precision can satisfy the current challenges for data mining in different areas. In fact, the purpose is to discover the hidden information and the interconnections in the current data and to predict the unclear or unobserved cases.

The *k*-Nearest Neighbor(*k*-NN) is one of the most commonly used algorithms in defining neighborhood's data (graph-based) and clustering them, according to the Euclidean distance. First, it decides to select the similar data sets. Then, the clusters are determined(Qin et al., 2018). Although this algorithm is simple and effective, determining neighborhood in this algorithm depends only on the distance and geometric location of the points, and statistical rules are not considered. Meanwhile, parameter *k* has an important role in defining neighborhood(Stork et al., 2001, Maillo et al., 2017,Pan et al., 2015, Güney & Atasoy, 2012,Mohammadi et al., 2015, García-Pedrajas et al., 2017).

The simple epsilon (ε) method determines neighborhood according to a small radius and the parameter of epsilon distance (Pedrycz, 2010). In this area of neighborhood, the data are clustered based on parameter of epsilon. If an inappropriate parameter is selected, the efficiency epsilon of the approach will be decreased. Therefore, this method lacks a strong and precise neighborhood structure. In this method, if the value of the selected epsilon is small, the concerned point in the assigned radius

[*] Corresponding Author.
*Email Addresses:*sh.pourbahrami@tabrizu.ac.ir (Shahin Pourbahrami), l-khanli@tabrizu.ac.ir (Leyli Mohammad Khanli), azimpour@cfu.ac.ir (Sohrab Azimpour).