Information Sciences 279 (2014) 498–511

Contents lists available at ScienceDirect

Information Sciences

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

Fuzzy granular gravitational clustering algorithm for multivariate data

Mauricio A. Sanchez [a], Oscar Castillo [b,*], Juan R. Castro [a], Patricia Melin [b]

[a] Autonomous University of Baja California, Tijuana, Mexico
[b] Tijuana Institute of Technology, Tijuana, Mexico

ARTICLE INFO

Article history:
Received 18 January 2013
Received in revised form 21 March 2014
Accepted 8 April 2014
Available online 23 April 2014

Keywords:
Gravitational algorithm
Clustering
Granular computing
Fuzzy system
Takagi–Sugeno–Kang

ABSTRACT

A new method for finding fuzzy information granules from multivariate data through a gravitational inspired clustering algorithm is proposed in this paper. The proposed algorithm incorporates the theory of granular computing, which adapts the cluster size with respect to the context of the given data. Via an inspiration in Newton's law of universal gravitation, both conditions of clustering similar data and adapting to the size of each granule are achieved. This paper compares the Fuzzy Granular Gravitational Clustering Algorithm (FGGCA) against other clustering techniques on two grounds: classification accuracy, and clustering validity indices, e.g. Rand, FM, Davies–Bouldin, Dunn, Homogeneity, and Separation. The FGGCA is tested with multiple benchmark classification datasets, such as Iris, Wine, Seeds, and Glass identification.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

In this information age, the acquisition of information far surpasses our ability to understand it. Classic methods of forming mathematical models from the said information are no longer viable. This breeds a need for new methods of modeling the information and/or finding hidden relationships between the information; this is now relevant and also a necessity. With mathematical techniques [3,24,42], the designers of such models must be able to find relationships between the data itself in order to create a suitable representation, yet as the number of variables increase so does the possibility that these methods will not be able to accurately represent said information, if at all.

One way to find relationships in data is to use clustering algorithms, these algorithms look through each piece of given data and find similarities between them, grouping them into clusters, serving to identify relationships which could be impossible for a human to find. The areas of application for these types of algorithms are vast; such as, business failure prediction [10], grouping similar opinions in the analysis of judicial prose [16], medical ECG beat type identification [18], find relations in the data between multidatabase systems [20], control de desulphurization process of steel [23], image segmentation [36], analysis of human motion for recreation in autonomous systems [43], pattern recognition [49], diagnosis decision support for Human Papilloma virus identification [51], sales forecasting of new apparel items [62], and molecular sequence analysis and taxonomy analysis [65].

* Corresponding author. Tel.: +52 6646236318.
E-mail addresses: mauricio.sanchez@uabc.edu.mx (M.A. Sanchez), ocastillo@tectijuana.mx (O. Castillo), jrcr@uabc.edu.mx (J.R. Castro), pmelin@tectijuana.mx (P. Melin).

http://dx.doi.org/10.1016/j.ins.2014.04.005
0020-0255/© 2014 Elsevier Inc. All rights reserved.

There are distinct differences between existing clustering algorithms, mainly focused in how clusters are found, such as cluster accumulation [4], multi-objective nature-inspired [6], cluster estimation [11], density-based [21], fuzzy neural networks [33], general Type-2 fuzzy c-means [35], ant colonies [37], weight clustering [45], condition fuzzy c-means [46], simultaneous clustering [50], and constrained clustering with active query selection [63].

Granular computing, related to how information is grouped together and how these groups can be utilized to make decisions [2,47] is inspired by how human cognition manages information. It groups similar information based on the context of the situation and dynamically modifies such groups in order to simplify its representation. Granular computing is used to improve the final representation of each cluster by forming information granules which better adapt to the numerical evidence.

Although granular computing can express groups, more commonly known as information granules, it can use a variety of representations to express such granules, which could be fuzzy sets [74], rough sets [55], shadowed sets [48], and quotient space [71]. Fuzzy sets [70] being one of the most common form of information granule representation.

Based on previous research [56] this paper is a general improvement of the original algorithm. Inspired by gravitational forces and how its interactions form clusters of bodies in space, the proposed approach seeks to find a solution clustering data by reproducing this behavior using multivariate data, where each data point represents a body in space that has mass. A replication of gravitational interactions takes place to find the clusters which best represent the data, and ultimately obtain fuzzy information granules.

This paper is organized into four sections, the first section reviews existing gravitational clustering techniques, the second section includes the description and explanation of the FGGCA, the third section deals with application examples of synthetics datasets and various benchmark datasets, and the last section concludes the paper.

2. A review of existing gravitational clustering approaches

Gravitational heuristic clustering algorithms are not new; as there have been many different approaches on how gravity is used to find the optimal placement of clusters.

The first instance of gravitational forces in clustering was in 1977 [66], known as Gravitational Clustering Algorithm (GCA), where a non-Newtonian gravitational physical model was used, which was then reformed to build a Markovian model for the gravitational clustering to use. It also has the characteristic of updating the position for each data point while each iteration of the algorithm is being performed. This approach follows a gravitational function as shown in Eq. (1), where, g is a gravitational function which calculates the next position of particle i in time t, x is the position of the data point, and dt is a small discrete time interval.

$$g(i,t,dt) = dt^2 \sum_{j \in N(t), j \neq i} \frac{1}{m_i(t)} \frac{x_j(t) - x_i(t)}{|x_j(t) - x_i(t)|^3} \tag{1}$$

Another instance is from Gomez et al. [28]. This approach uses a modified version of Newton's original equation of universal gravitation. This modification simplifies how the gravitational force is calculated; leaving only Eq. (2), which at the same time calculates the next position of each data point. The value of the universal gravitational constant G is reduced after each iteration, which serves to eliminate the big crunch effect of all data points, which serves as a mechanism to not end up with only one cluster. Here, x is the position of the data point, and $\vec{d}$ the vector direction of such data point.

$$x(t+1) = x(t) + \vec{d}\frac{G}{\|\vec{d}\|^3} \tag{2}$$

The previous approach was later improved by the same authors [29]. They modified a part of the equation for updating the positions of each point, more specifically, how the gravitational force is calculated, thus obtaining Eq. (3). Where, f is a decreasing function, and $\hat{d}$ is the rough estimate of maximum distance between closest points.

$$x(t+1) = x(t) + G * \vec{d} * f\left(\frac{\|\vec{d}\|}{\hat{d}}\right) \tag{3}$$

A different approach by Kundu [34] uses a simplified version of Newton's gravitational equation, as shown in Eq. (4). They propose their own version of how the position of each data point is updated and this is seen in Eq. (5). It integrates an idea of heights in each data point, which aids the algorithm in determining the choices of good clusters.

$$F_{ij} = \frac{1}{d_{ij}^u}; \quad F_i = \sum_{j \neq i} [m_j F_{ij}] \tag{4}$$

$$x_i^{new}(\eta) = x_i^{old} + \eta F_i \tag{5}$$

Long and Jin [38] proposes a minimalistic version of Newtońs gravitational equation, which finds a pair of points which are most likely to meet and merge first, as shown in Eq. (6), and then updates their position by calculating their centroid, as shown in Eq. (7). This algorithm requires a user given parameter for the amount of clusters to find.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات