(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107798338 B
(45)授权公告日 2021.03.26

(21)申请号 201710898511.3

(22)申请日 2017.09.28

(65)同一申请的已公布的文献号
申请公布号 CN 107798338 A

(43)申请公布日 2018.03.13

(73)专利权人 佛山科学技术学院
地址 528000 广东省佛山市禅城区江湾一路18号

(72)发明人 许红龙

(74)专利代理机构 广州嘉权专利商标事务所有限公司 44205
代理人 王国标

(51)Int.Cl.
*G06K 9/62*(2006.01)

(56)对比文件
CN 106528790 A,2017.03.22
CN 106503245 A,2017.03.15
CN 104281652 A,2015.01.14
CN 105117485 A,2015.12.02
CN 105975519 A,2016.09.28
CN 104462379 A,2015.03.25
许红龙等.基于多种支撑点的度量空间离群检测算法.《计算机学报》.2017,第40卷(第12期),
Honglong Xu等.Hilbert Index-based Outlier Detection Algorithm in Metric Space.《International Journal of Grid and High Performance Computing》.2016,第8卷(第4期),
李兴亮.度量空间索引支撑点选择问题研究.《万方数据知识服务平台》.2017,
许红龙等.改进密度峰值支撑点选取及其在度量空间离群检测的应用.《小型微型计算机系统》.2017,第38卷(第5期),
Kewei Ma等.Pivot Selection Methods Based on Covariance and Correlation for Metric-space Indexing.《National Conference on Information Technology and Computer Science (CITCS 2012)》.2012,
Kewei Ma等.LLE Based Pivot Selection for Similarity Search of Biological Data.《National Conference on Information Technology and Computer Science (CITCS 2012)》.2012,

审查员 刘素兵

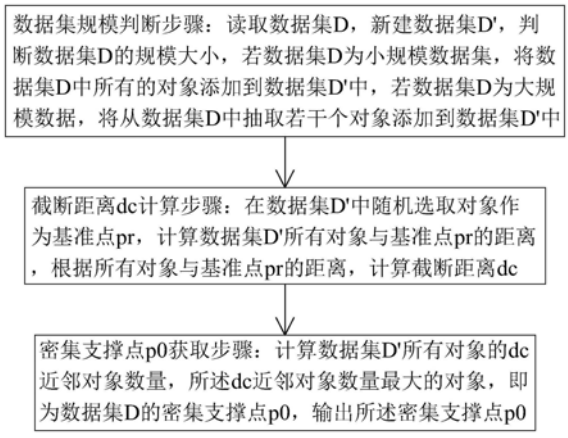
权利要求书2页 说明书5页 附图4页

(54)发明名称

一种大数据密集支撑点快速选取方法

(57)摘要

本发明公开了一种大数据密集支撑点快速选取方法,包括以下步骤:数据集规模判断步骤;截断距离dc计算步骤;密集支撑点p0获取步骤。本发明首先通过对数据集D的规模进行判断,对大规模的数据集D进行缩减有效减少后续的运算次数;其中本发明在整个运算过程中,完全是基于对象间的距离,实际设计时实现难度低,通用性强,能从数据集中密集的区域获取密集支撑点。本发明用于从数据集中获取密集支撑点。




CN 107798338 B

1.一种大数据密集支撑点快速选取方法,其特征在于,包括以下步骤:

数据集规模判断步骤:读取数据集D,新建数据集D',判断数据集D的规模大小,若数据集D为小规模数据集,将数据集D中所有的对象添加到数据集D'中,若数据集D为大规模数据,将从数据集D中抽取若干个对象添加到数据集D'中;

截断距离dc计算步骤:在数据集D'中随机选取对象作为基准点pr,计算数据集D'所有对象与基准点pr的距离,根据所有对象与基准点pr的距离,计算截断距离dc;

密集支撑点p0获取步骤:计算数据集D'所有对象的dc近邻对象数量,所述dc近邻对象数量最大的对象,即为数据集D的密集支撑点p0,输出所述密集支撑点p0。

2.根据权利要求1所述的一种大数据密集支撑点快速选取方法,其特征在于,所述数据集规模判断步骤包括以下步骤:

步骤A1:读取数据集D,新建数据集D';

步骤A2:设定规模界限,若数据集D中对象数量大于规模界限,则将数据集D定义为大规模数据集,否则将数据集D定义为小规模数据集;

步骤A3:若数据集D为小规模数据集,则将数据集D中所有的对象添加到数据集D'中,若数据集D为大规模数据集,则通过均匀抽样或者随机抽样的方式,从数据集D中抽取对象添加到数据集D'中,抽取对象的数量与规模界限数值一致。

3.根据权利要求2所述的一种大数据密集支撑点快速选取方法,其特征在于,所述截断距离dc计算步骤包括以下步骤:

步骤B1:设定截断距离参数u,所述截断距离参数u数值范围为0.1至0.2之间;

步骤B2:在数据集D'中随机选取一对象作为基准点pr,计算数据集D'所有对象与基准点pr的距离,记为第一距离;

步骤B3:设定参数r,取数据集D'所有对象的第一距离中的最大值赋给参数r,所述截断距离参数dc=ur。

4.根据权利要求3所述的一种大数据密集支撑点快速选取方法,其特征在于,所述密集支撑点p0获取步骤包括以下步骤:

步骤C1:逐一获取数据集D'的各个对象O,定义对象O的dc近邻对象数量为M,并初始化为0;

步骤C2:逐一读取数据集D'的其他对象X,计算对象O第一距离与对象X第一距离的差的绝对值;

步骤C3:若所述对象O第一距离与对象X第一距离的差的绝对值小于截断距离dc,计算对象O与对象X间的距离,记为第二距离,若所述对象O第一距离与对象X第一距离的差的绝对值大于截断距离dc,则认为对象X不可能是对象O的dc近邻对象,无需计算对象O与对象X间的距离,跳转到步骤C5;

步骤C4:若第二距离小于截断距离dc,对象O的dc近邻对象数量自加1;

步骤C5:获取下一个对象X,返回步骤C2,直到数据集D'全部对象读取完毕;

步骤C6:获取下一个对象O,返回步骤C1;

步骤C7:输出dc近邻对象数量最多的对象O,即为数据集D的密集支撑点p0。

5.根据权利要求2所述的一种大数据密集支撑点快速选取方法,其特征在于,所述截断距离dc计算步骤包括以下步骤:

步骤b1:设定截断距离参数u,所述截断距离参数u数值范围为0.1至0.2之间;

步骤b2:在数据集D'中随机选取多个对象作为基准点pr1、基准点pr2……基准点prn,建立第一数组,计算各个基准点与数据集D'所有对象的距离,记为第三距离,将各个基准点的第三距离的最大值存入第一数组中;

步骤b3:设定参数r,取第一数组的最小值赋给参数r,所述截断距离参数dc=ur。

6.根据权利要求5所述的一种大数据密集支撑点快速选取方法,其特征在于,所述密集支撑点p0获取步骤包括以下步骤:

步骤c1:逐一获取数据集D'的各个对象O,定义对象O的dc近邻对象数量为M,并初始化为0;

步骤c2:逐一读取数据集D'的其他对象X,对于同一个基准点,计算对象O第三距离与对象X第三距离的差的绝对值;

步骤c3:对于所有的基准点,若所述对象O第三距离与对象X第三距离的差的绝对值均小于截断距离dc,计算对象O与对象X间的距离,记为第四距离,若对于某个基准点,所述对象O第三距离与对象X第三距离的差的绝对值大于截断距离dc,则认为对象X不可能是对象O的dc近邻对象,无需计算对象O与对象X间的距离,跳转到步骤c5;

步骤c4:若第四距离小于截断距离dc,对象O的dc近邻对象数量自加1;

步骤c5:获取下一个对象X,返回步骤c2,直到数据集D'全部对象读取完毕;

步骤c6:获取下一个对象O,返回步骤c1;

步骤c7:输出dc近邻对象数量最多的对象O,即为数据集D的密集支撑点p0。

# 一种大数据密集支撑点快速选取方法

## 技术领域

[0001] 本发明涉及数据挖掘领域,更具体地说涉及一种大数据的密集支撑点选取方法。

## 背景技术

[0002] 现有的很多数据处理技术,都是面向多维空间的,仅适用于多维数据,难以应用于图像、音频视频、蛋白质等复杂的数据类型,这正是大数据时代常见的多样性挑战。

[0003] 度量空间算法是一种面向于上述复杂数据类型的数据处理算法,其中所述度量空间算法又包括密集支撑点选取步骤,良好的密集支撑点有利于建立更高效的索引,加快搜索过程,更有效地排除非目标对象或者非离群点等。

[0004] 常用的密集支撑点选取方法有两种,第一种是近似密集区域支撑点选取算法,该算法随机选取临时参考点,搜索数据集中与其距离最远的对象,以该对象为基点,计算数据集中各个对象与参考点的距离,按照从小到大的顺序排序,采用“等距划分+数量中点”的方法,取各段中位点加入支撑点候选集。计算每个段的对象数量,再对对象数量按从大到小的顺序排序。对于对象数量相等的分段,比较获得这些分段之中与参考点距离最近的分段,取其数量中点作为第一个支撑点,但是这种算法具体选取过程决定了其选取结果不够准确,可能把密集程度不高的支撑点也作为密集支撑点选取;第二种是暴力精确计算方法,即在确定密集的标准,或称密度值(例如以某给定距离值的范围内近邻数量多者为密集)之后,计算数据集里每个对象的密度值,最终得出最密集的对象(即给定距离值的范围内近邻数量最多者)。这种方法显然最为精确,但是时间开销也是最大的。

## 发明内容

[0005] 本发明要解决的技术问题是:提供一种时间开销小的大数据密集支撑点精确选取方法。

[0006] 本发明解决其技术问题的解决方案是:

[0007] 一种大数据密集支撑点快速选取方法,包括以下步骤:

[0008] 数据集规模判断步骤:读取数据集D,新建数据集D',判断数据集D的规模大小,若数据集D为小规模数据集,将数据集D中所有的对象添加到数据集D'中,若数据集D为大规模数据,将从数据集D中抽取若干个对象添加到数据集D'中;

[0009] 截断距离dc计算步骤:在数据集D'中随机选取对象作为基准点pr,计算数据集D'所有对象与基准点pr的距离,根据所有对象与基准点pr的距离,计算截断距离dc;

[0010] 密集支撑点p0获取步骤:计算数据集D'所有对象的dc近邻对象数量,所述dc近邻对象数量最大的对象,即为数据集D的密集支撑点p0,输出所述密集支撑点p0。

[0011] 作为上述技术方案的进一步改进,所述数据集规模判断步骤包括以下步骤:

[0012] 步骤A1:读取数据集D,新建数据集D';

[0013] 步骤A2:设定规模界限,若数据集D中对象数量大于规模界限,则将数据集D定义为大规模数据集,否则将数据集D定义为小规模数据集;

[0014] 步骤A3:若数据集D为小规模数据集,则将数据集D中所有的对象添加到数据集D'中,若数据集D为大规模数据集,则通过均匀抽样或者随机抽样的方式,从数据集D中抽取对象添加到数据集D'中,抽取对象的数量与规模界限数值一致。

[0015] 作为上述技术方案的进一步改进,所述截断距离dc计算步骤的第一实施方式,包括以下步骤:

[0016] 步骤B1:设定截断距离参数u,所述截断距离参数u数值范围为0.1至0.2之间;

[0017] 步骤B2:在数据集D'中随机选取一对象作为基准点pr,计算数据集D'所有对象与基准点pr的距离,记为第一距离;

[0018] 步骤B3:设定参数r,取数据集D'所有对象的第一距离中的最大值赋给参数r,所述截断距离参数dc=ur。

[0019] 基于上述实施方式,所述密集支撑点p0获取步骤包括以下步骤:

[0020] 步骤C1:逐一获取数据集D'的各个对象O,定义对象O的dc近邻对象数量为M,并初始化为0;

[0021] 步骤C2:逐一读取数据集D'的其他对象X,计算对象O第一距离与对象X第一距离的差的绝对值;

[0022] 步骤C3:若所述对象O第一距离与对象X第一距离的差的绝对值小于截断距离dc,计算对象O与对象X间的距离,记为第二距离,若大于截断距离dc,则认为对象X不可能是对象O的dc近邻对象,无需计算对象O与对象X间的距离,跳转到步骤C5;

[0023] 步骤C4:若第二距离小于截断距离dc,对象O的dc近邻对象数量自加1;

[0024] 步骤C5:获取下一个对象X,返回步骤C2,直到数据集D'全部对象读取完毕;

[0025] 步骤C6:获取下一个对象O,返回步骤C1;

[0026] 步骤C7:输出dc近邻对象数量最多的对象O,即为数据集D的密集支撑点p0。

[0027] 作为上述技术方案的进一步改进,所述截断距离dc计算步骤的第二实施方式,包括以下步骤:

[0028] 步骤b1:设定截断距离参数u,所述截断距离参数u数值范围为0.1至0.2之间;

[0029] 步骤b2:在数据集D'中随机选取多个对象作为基准点pr1、基准点pr2……基准点prn,建立第一数组,计算各个基准点与数据集D'所有对象的距离,记为第三距离,将各个基准点的第三距离的最大值存入第一数组中;

[0030] 步骤b3:设定参数r,取第一数组的最小值赋给参数r,所述截断距离参数dc=ur。

[0031] 基于上述实施方式,所述密集支撑点p0获取步骤包括以下步骤:

[0032] 步骤c1:逐一获取数据集D'的各个对象O,定义对象O的dc近邻对象数量为M,并初始化为0;

[0033] 步骤c2:逐一读取数据集D'的其他对象X,对于同一个基准点,计算对象O第三距离与对象X第三距离的差的绝对值;

[0034] 步骤c3:对于所有的基准点,若所述对象O第三距离与对象X第三距离的差的绝对值均小于截断距离dc,计算对象O与对象X间的距离,记为第四距离,若对于某个基准点,所述对象O第三距离与对象X第三距离的差的绝对值大于截断距离dc,则认为对象X不可能是对象O的dc近邻对象,无需计算对象O与对象X间的距离,跳转到步骤c5;

[0035] 步骤c4:若第四距离小于截断距离dc,对象O的dc近邻对象数量自加1;

[0036] 步骤c5:获取下一个对象X,返回步骤c2,直到数据集D'全部对象读取完毕;

[0037] 步骤c6:获取下一个对象O,返回步骤c1;

[0038] 步骤c7:输出dc近邻对象数量最多的对象O,即为数据集D的密集支撑点p0。

[0039] 在步骤c2和步骤c3中,计算对象O第三距离与对象X第三距离的差时,对象O的第三距离与对象X的第三距离,是基于同一个基准点(如基准点pr1)计算的。

[0040] 本发明的有益效果是:本发明首先通过对数据集D的规模进行判断,对大规模的数据集D进行缩减有效减少后续的运算次数;其中本发明在整个运算过程中,完全是基于对象间的距离,实际设计时实现难度低,通用性强,能从数据集中密集的区域获取密集支撑点。本发明用于从数据集中获取密集支撑点。

## 附图说明

[0041] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单说明。显然,所描述的附图只是本发明的一部分实施例,而不是全部实施例,本领域的技术人员在不付出创造性劳动的前提下,还可以根据这些附图获得其他设计方案和附图。

[0042] 图1是本发明的步骤流程图;

[0043] 图2是本发明的数据集规模判断步骤实施例流程图;

[0044] 图3是本发明的截断距离dc计算步骤以及密集支撑点p0获取步骤的第一实施方式流程图;

[0045] 图4是本发明的截断距离dc计算步骤以及密集支撑点p0获取步骤的第二实施方式流程图。

## 具体实施方式

[0046] 以下将结合实施例和附图对本发明的构思、具体结构及产生的技术效果进行清楚、完整的描述,以充分地理解本发明的目的、特征和效果。显然,所描述的实施例只是本发明的一部分实施例,而不是全部实施例,基于本发明的实施例,本领域的技术人员在不付出创造性劳动的前提下所获得的其他实施例,均属于本发明保护的范围。

[0047] 参照图1～图4,本发明创造公开了一种大数据密集支撑点快速选取方法,包括以下步骤:

[0048] 数据集规模判断步骤:读取数据集D,新建数据集D',判断数据集D的规模大小,若数据集D为小规模数据集,将数据集D中所有的对象添加到数据集D'中,若数据集D为大规模数据,将从数据集D中抽取若干个对象添加到数据集D'中;

[0049] 截断距离dc计算步骤:在数据集D'中随机选取对象作为基准点pr,计算数据集D'所有对象与基准点pr的距离,根据所有对象与基准点pr的距离,计算截断距离dc;

[0050] 密集支撑点p0获取步骤:计算数据集D'所有对象的dc近邻对象数量,所述dc近邻对象数量最大的对象,即为数据集D的密集支撑点p0,输出所述密集支撑点p0。

[0051] 具体地,本支撑点选取方法首先限制了数据集D的规模,若数据集D规模较小,则在后续的步骤中可逐个计算每个对象的dc近邻对象数量,而实际应用中,数据集D是小规模数据集的概率极低,此时需要从数据集D中以均匀抽样或者随机抽样的方式抽取若干个对象,

组成新的数据集D'，从而在最大程度上减少后续步骤的运算量，极大地降低时间开销；之后通过计算数据集D'中各个对象基于基准点pr的距离，得到参数截断距离dc，最后计算数据集D'中所有对象的dc近邻数据数量，dc近邻数据数量最大对象即为数据集D'的密集支撑点p0，其中所述dc近邻数据数量表示以数据集D'一对象为中心，以截断距离dc为半径的区域内其他对象的数量，本发明通过所述截断距离dc作为定位数据集D'密集区域的依据，避免所选取的密集支撑点p0出现在非密集区域。

[0052] 本发明首先通过对数据集D的规模进行判断，对大规模的数据集D进行缩减有效减少后续的运算次数；其中本发明在整个运算过程中，完全是基于对象间的距离，实际设计时实现难度低，通用性强，能从数据集中密集的区域获取密集支撑点。

[0053] 参照图2，进一步作为优选的实施方式，本发明创造具体实施方式中，通过如下步骤对数据集D'的规模大小进行判断，所述数据集规模判断步骤包括以下步骤：

[0054] 步骤A1：读取数据集D，新建数据集D'；

[0055] 步骤A2：设定规模界限，若数据集D中对象数量大于规模界限，则将数据集D定义为大规模数据集，否则将数据集D定义为小规模数据集；

[0056] 步骤A3：若数据集D为小规模数据集，则将数据集D中所有的对象添加到数据集D'中，若数据集D为大规模数据集，则通过均匀抽样或者随机抽样的方式，从数据集D中抽取对象添加到数据集D'中，抽取对象的数量与规模界限数值一致。

[0057] 具体地，所述本发明创造具体实施例中，所述规模界限为1000，当数据集D中对象数量小于1000，则认为是小规模数据集，即使逐个计算数据集D各个对象的dc近邻对象数量，计算次数也不会太多；但数据集D中对象数量大于1000，则认为是大规模数据集，若逐个计算数据集D各个对象的dc近邻对象数量，计算次数过多，需要从数据集D中抽取通过均匀抽样或者随机抽样的方式抽取固定数量的对象，以减少数据集D的规模大小，降低计算时间开销。

[0058] 参照图3，本发明创造中所述截断距离dc计算步骤的第一实施方式，包括以下步骤：

[0059] 步骤B1：设定截断距离参数u，所述截断距离参数u数值范围为0.1至0.2之间；

[0060] 步骤B2：在数据集D'中随机选取一对象作为基准点pr，计算数据集D'所有对象与基准点pr的距离，记为第一距离；

[0061] 步骤B3：设定参数r，取数据集D'所有对象的第一距离中的最大值赋给参数r，所述截断距离参数dc＝ur。

[0062] 基于上述截断距离dc计算步骤的第一实施方式，所述密集支撑点p0获取步骤包括以下步骤：

[0063] 步骤C1：逐一获取数据集D'的各个对象O，定义对象O的dc近邻对象数量为M，并初始化为0；

[0064] 步骤C2：逐一读取数据集D'的其他对象X，计算对象O第一距离与对象X第一距离的差的绝对值；

[0065] 步骤C3：若所述对象O第一距离与对象X第一距离的差的绝对值小于截断距离dc，计算对象O与对象X间的距离，记为第二距离，若大于截断距离dc，则认为对象X不可能是对象O的dc近邻对象，无需计算对象O与对象X间的距离，跳转到步骤C5；

[0066] 步骤C4:若第二距离小于截断距离dc,对象O的dc近邻对象数量自加1;

[0067] 步骤C5:获取下一个对象X,返回步骤C2,直到数据集D'全部对象读取完毕;

[0068] 步骤C6:获取下一个对象O,返回步骤C1;

[0069] 步骤C7:输出dc近邻对象数量最多的对象O,即为数据集D的密集支撑点p0。

[0070] 具体地,本发明创造第一实施方式,在计算数据集D'中的一个对象的dc近邻数量时,定义该对象为对象O,对象O以外的其他对象定义为对象X。本步骤中,首先逐次选定一个对象O,再计算所有的对象X与对象O的距离,计算对象O的dc近邻数量。但是本方法在计算对象X与对象O的距离之前,首先通过步骤C2和步骤C3判断对象X是否是对象O的dc近邻对象,如果不是,就无需计算对象X与对象O的距离,减少整个过程的计算次数,降低计算时间开销。

[0071] 参照图4,本发明创造中所述截断距离dc计算步骤的第二实施方式,包括以下步骤:

[0072] 步骤b1:设定截断距离参数u,所述截断距离参数u数值范围为0.1至0.2之间;

[0073] 步骤b2:在数据集D'中随机选取多个对象作为基准点pr1、基准点pr2……基准点prn,建立第一数组,计算各个基准点与数据集D'所有对象的距离,记为第三距离,将各个基准点的第三距离的最大值存入第一数组中;

[0074] 步骤b3:设定参数r,取第一数组的最小值赋给参数r,所述截断距离参数dc=ur。

[0075] 基于上述截断距离dc计算步骤的第二实施方式,所述密集支撑点p0获取步骤包括以下步骤:

[0076] 步骤c1:逐一获取数据集D'的各个对象O,定义对象O的dc近邻对象数量为M,并初始化为0;

[0077] 步骤c2:逐一读取数据集D'的其他对象X,对于同一个基准点,计算对象O第三距离与对象X第三距离的差的绝对值;

[0078] 步骤c3:对于所有的基准点,若所述对象O第三距离与对象X第三距离的差的绝对值均小于截断距离dc,计算对象O与对象X间的距离,记为第四距离,若对于某个基准点,所述对象O第三距离与对象X第三距离的差的绝对值大于截断距离dc,则认为对象X不可能是对象O的dc近邻对象,无需计算对象O与对象X间的距离,跳转到步骤c5;

[0079] 步骤c4:若第四距离小于截断距离dc,对象O的dc近邻对象数量自加1;

[0080] 步骤c5:获取下一个对象X,返回步骤c2,直到数据集D'全部对象读取完毕;

[0081] 步骤c6:获取下一个对象O,返回步骤c1;

[0082] 步骤c7:输出dc近邻对象数量最多的对象O,即为数据集D的密集支撑点p0。

[0083] 具体地,本发明创造中,所述选取方法第一实施方式和第二实施方式之间的区别在与,第一实施方式中选取的是一个基准点,而第二实施方式中选取多个基准点,相比较而言,本方法第二实施方式选取多个基准点,虽然在一定程度上增加了第三距离的计算次数,但是却能更大幅度地减少第四距离的计算次数。选取多个基准点,通常能减少总的距离计算次数。

[0084] 以上对本发明的较佳实施方式进行了具体说明,但本发明创造并不限于所述实施例,熟悉本领域的技术人员在不违背本发明精神的前提下还可作出种种的等同变型或替换,这些等同的变型或替换均包含在本申请权利要求所限定的范围内。

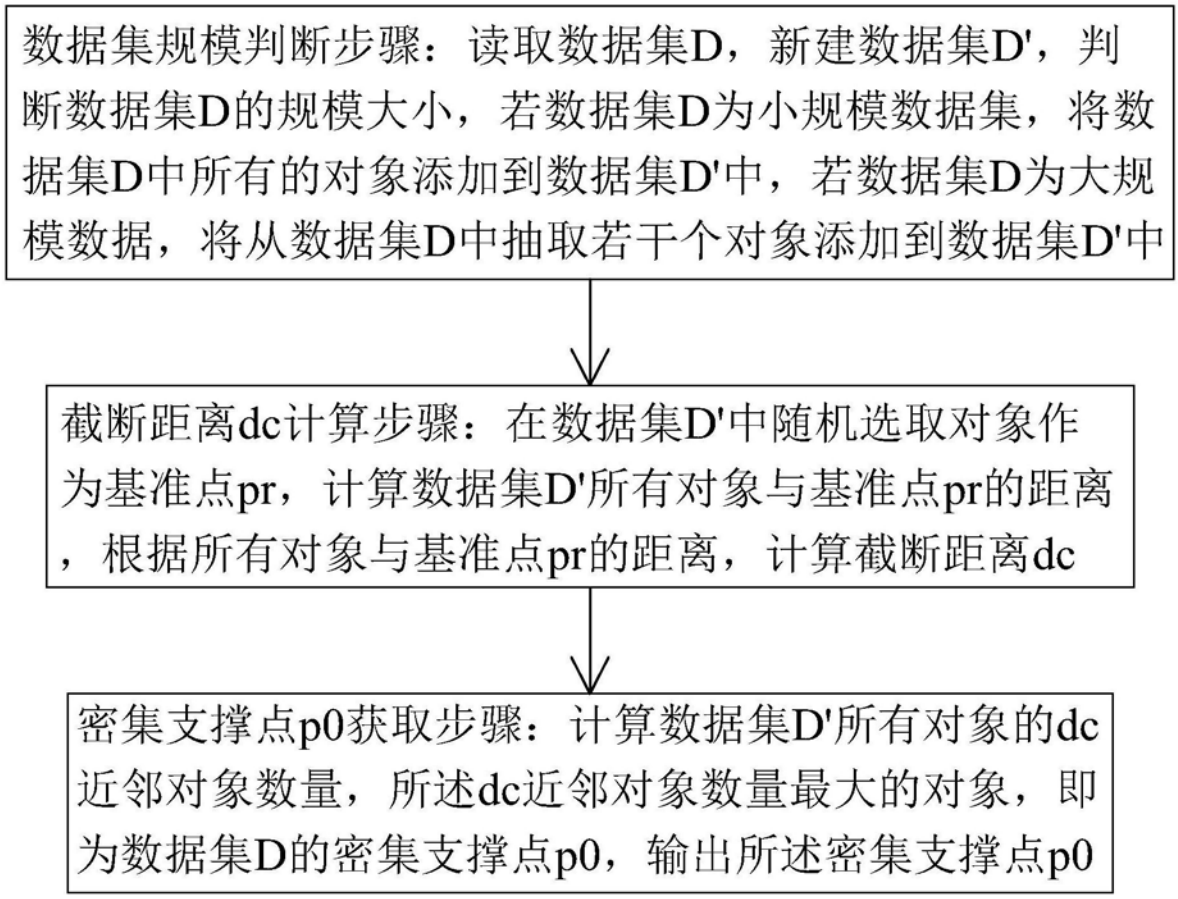

数据集规模判断步骤：读取数据集D，新建数据集D'，判断数据集D的规模大小，若数据集D为小规模数据集，将数据集D中所有的对象添加到数据集D'中，若数据集D为大规模数据，将从数据集D中抽取若干个对象添加到数据集D'中
截断距离dc计算步骤：在数据集D'中随机选取对象作为基准点pr，计算数据集D'所有对象与基准点pr的距离，根据所有对象与基准点pr的距离，计算截断距离dc
密集支撑点p0获取步骤：计算数据集D'所有对象的dc近邻对象数量，所述dc近邻对象数量最大的对象，即为数据集D的密集支撑点p0，输出所述密集支撑点p0

图1

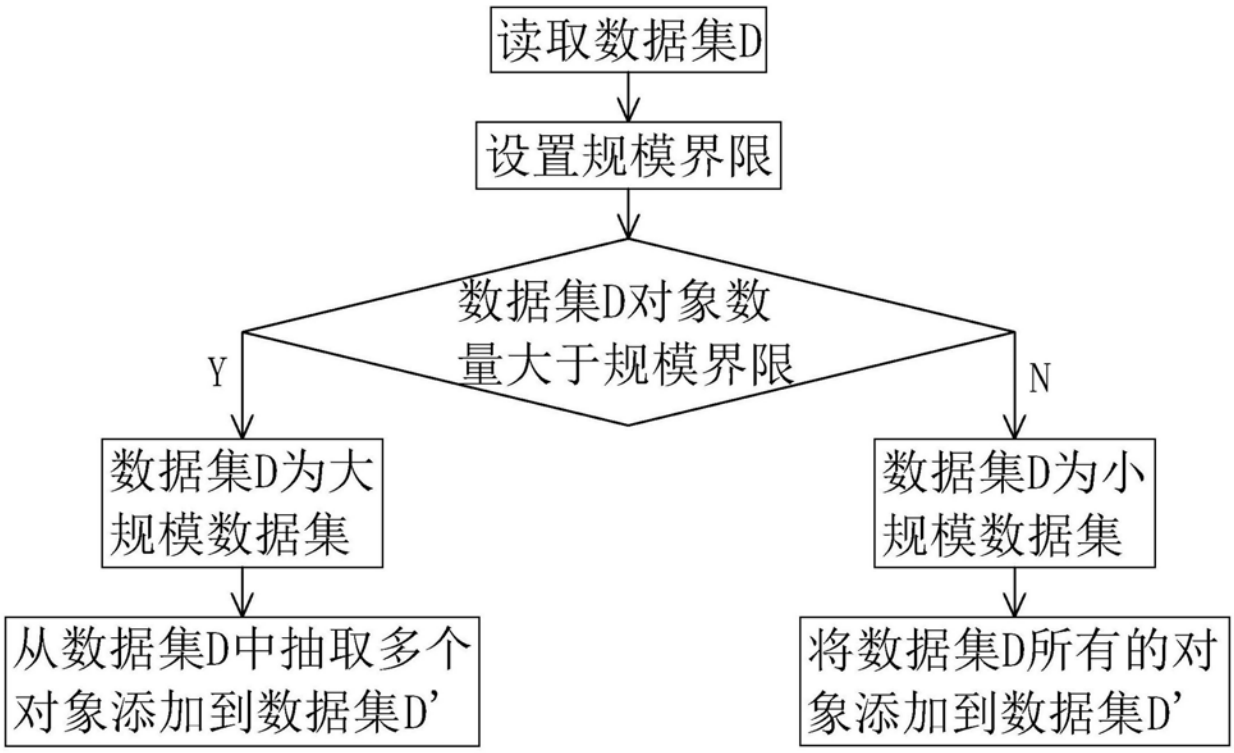

读取数据集D
设置规模界限
数据集D对象数量大于规模界限
Y
N
数据集D为大规模数据集
数据集D为小规模数据集
从数据集D中抽取多个对象添加到数据集D'
将数据集D所有的对象添加到数据集D'

图2

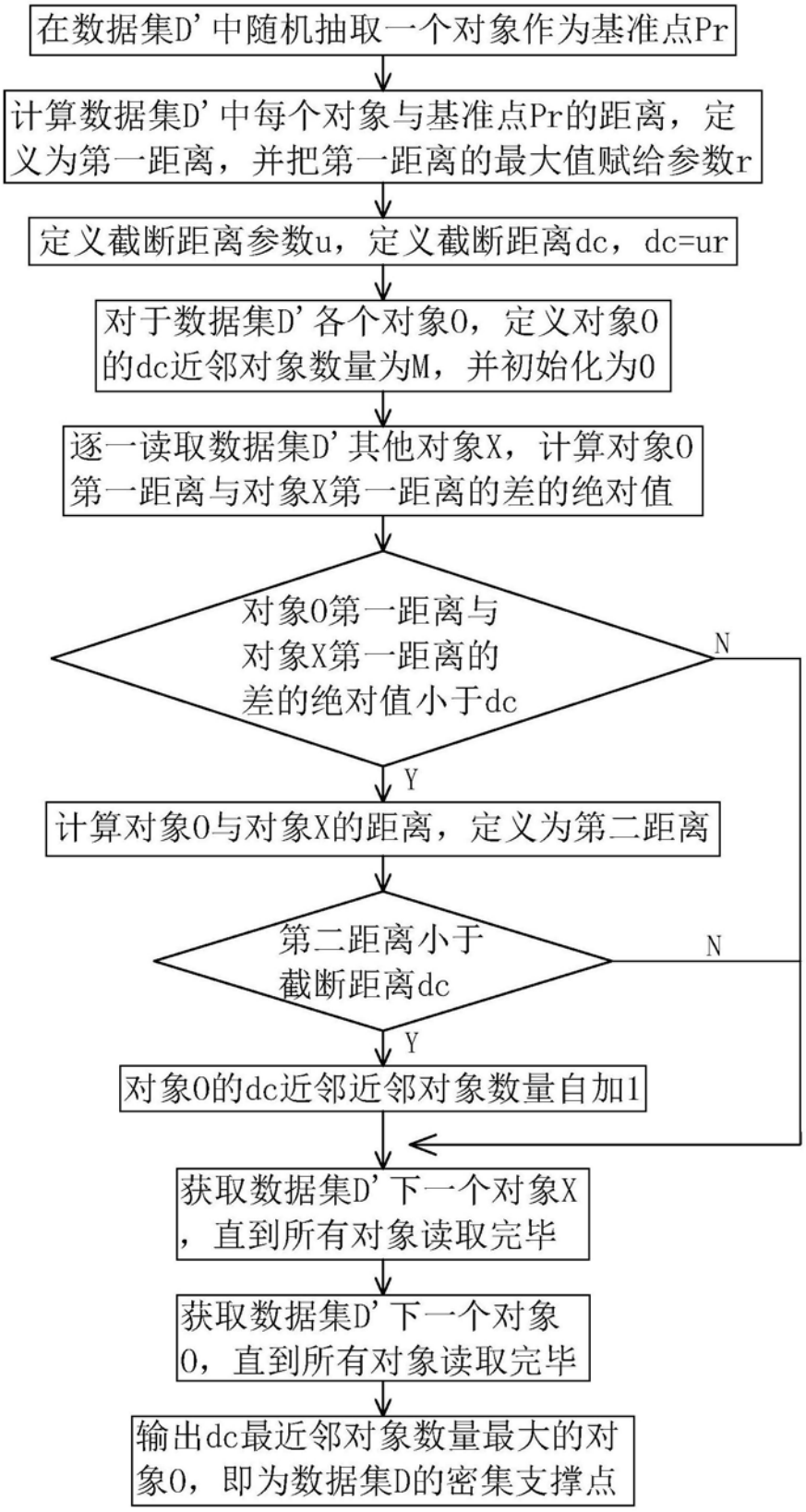

在数据集D'中随机抽取一个对象作为基准点Pr
计算数据集D'中每个对象与基准点Pr的距离，定义为第一距离，并把第一距离的最大值赋给参数r
定义截断距离参数u，定义截断距离dc，dc=ur
对于数据集D'各个对象O，定义对象O的dc近邻对象数量为M，并初始化为0
逐一读取数据集D'其他对象X，计算对象O第一距离与对象X第一距离的差的绝对值
对象O第一距离与对象X第一距离的差的绝对值小于dc
N
Y
计算对象O与对象X的距离，定义为第二距离
第二距离小于截断距离dc
N
Y
对象O的dc近邻近邻对象数量自加1
获取数据集D'下一个对象X，直到所有对象读取完毕
获取数据集D'下一个对象O，直到所有对象读取完毕
输出dc最近邻对象数量最大的对象O，即为数据集D的密集支撑点

图3

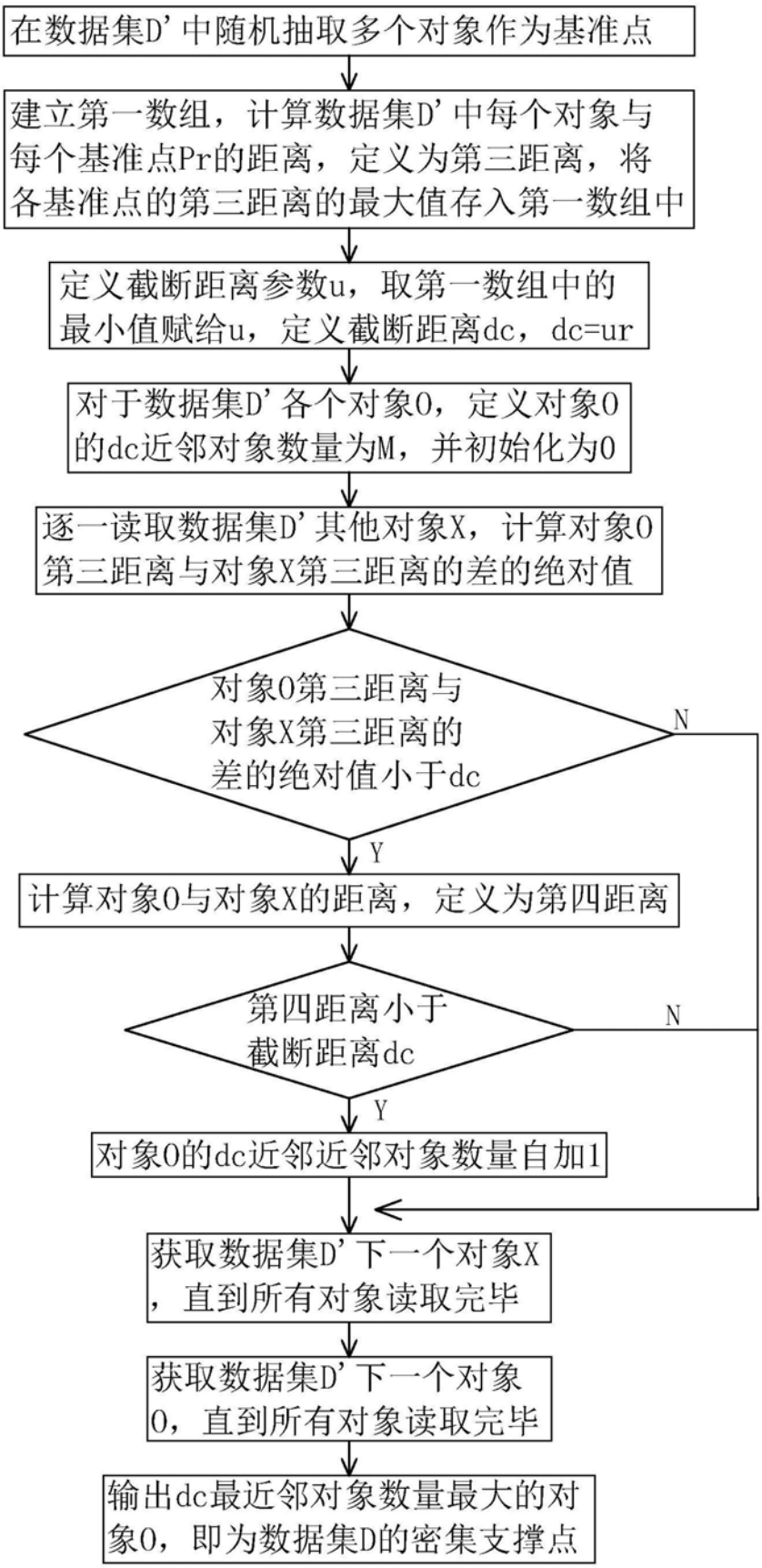

在数据集D'中随机抽取多个对象作为基准点
建立第一数组，计算数据集D'中每个对象与每个基准点Pr的距离，定义为第三距离，将各基准点的第三距离的最大值存入第一数组中
定义截断距离参数u，取第一数组中的最小值赋给u，定义截断距离dc，dc=ur
对于数据集D'各个对象O，定义对象O的dc近邻对象数量为M，并初始化为0
逐一读取数据集D'其他对象X，计算对象O第三距离与对象X第三距离的差的绝对值
对象O第三距离与对象X第三距离的差的绝对值小于dc
N
Y
计算对象O与对象X的距离，定义为第四距离
第四距离小于截断距离dc
N
Y
对象O的dc近邻近邻对象数量自加1
获取数据集D'下一个对象X，直到所有对象读取完毕
获取数据集D'下一个对象O，直到所有对象读取完毕
输出dc最近邻对象数量最大的对象O，即为数据集D的密集支撑点

图4