

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 12 月 30 日 (30.12.2020)



(10) 国际公布号
WO 2020/258772 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2019/126716

(22) 国际申请日: 2019 年 12 月 19 日 (19.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910551643.8 2019 年 6 月 24 日 (24.06.2019) CN

(71) 申请人: 广州视源电子科技股份有限公司 (GUANGZHOU SHIYUAN ELECTRONICS CO., LTD.) [CN/CN]; 中国广东省广州市黄埔区云埔四路 6 号, Guangdong 510530 (CN)。

(72) 发明人: 熊凯(XIONG, Kai); 中国广东省广州市黄埔区云埔四路 6 号, Guangdong 510530 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: SAMPLE CLUSTERING METHOD, APPARATUS AND DEVICE AND STORAGE MEDIUM

(54) 发明名称: 样本聚类方法、装置、设备及存储介质

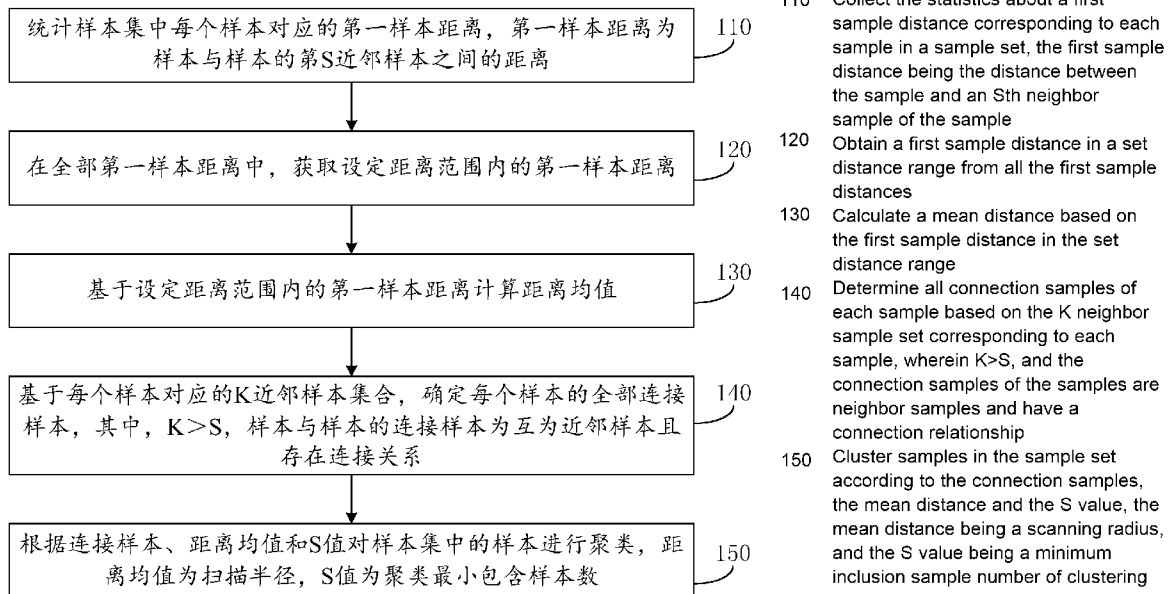


图 2

(57) Abstract: The present invention provides a sample clustering method, apparatus and device, and a storage medium, and relates to the field of data processing. The method comprises: collecting the statistics about a first sample distance corresponding to each sample in a sample set, the first sample distance being the distance between the sample and an Sth neighbor sample of the sample; obtaining a first sample distance in a set distance range from all the first sample distances; calculating a mean distance based on the first sample distance in the set distance range; determining all connection samples of each sample based on the K neighbor sample set corresponding to each sample, wherein $K > S$, and the connection samples of the samples are neighbor samples and have a connection relationship; and clustering samples in the sample set according to the connection samples, the mean distance and the S value, the mean distance being

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

a scanning radius, and the S value being a minimum inclusion sample number of clustering. By adopting the method, the technical problem that a DBSCAN algorithm in the prior art cannot reasonably cluster a sample set with uneven density can be solved.

(57) 摘要: 本公开提供了一种样本聚类方法、装置、设备及存储介质, 涉及数据处理领域, 其包括: 统计样本集中每个样本对应的第一样本距离, 第一样本距离为样本与样本的第S近邻样本之间的距离; 在全部第一样本距离中, 获取设定距离范围内的第一样本距离; 基于设定距离范围内的第一样本距离计算距离均值; 基于每个样本对应的K近邻样本集合, 确定每个样本的全部连接样本, 其中, $K > S$, 样本与样本的连接样本为互为近邻样本且存在连接关系; 根据连接样本、距离均值和S值对样本集中的样本进行聚类, 距离均值为扫描半径, S值为聚类最小包含样本数。采用上述方法可以解决相关技术中DBSCAN算法对于密度不均的样本集无法合理聚类的技术问题。

样本聚类方法、装置、设备及存储介质

本申请要求在 2019 年 6 月 24 日提交中国专利局、申请号为 201910551643.8 的中国专利申请的优先权，以上申请的全部内容通过引用结合在本申请中。

技术领域

本公开涉及数据处理技术领域，例如涉及一种样本聚类方法、装置、设备及存储介质。

背景技术

聚类分析指将物理或抽象对象的集合分组为由类似的对象组成的多个类的分析过程。如今，聚类分析被广泛用于各类领域中，并且随着聚类分析的广泛应用，各类聚类算法应运而生。例如，K-MEANS 算法、K-MEDOIDS 算法、BIRCH 算法、CURE 算法、DBSCAN 算法、OPTICS 算法等。其中，DBSCAN 算法是一个比较有代表性的基于密度的聚类算法，其需要人工输入两个参数：一个为扫描半径，记为 ϵ ；另一个为最小包含点数，记为 $\minPts$ ，并通过两个参数在样本集中找到密度相连对象的最大集合。发明人在实现本公开的过程中，发现相关技术存在如下缺陷：基于 DBSCAN 算法对样本集进行聚类时，对于密度不均的样本集而言，若扫描半径较小，则对于密度稀疏的样本而言，容易被认为是噪声点而剔除，若扫描半径较大，则易将距离较远的样本聚为一类，此时，无法保证样本聚类的准确性。

综上，如何在 DBSCAN 算法下，对于密度不均的样本集进行合理聚类，成为了亟待解决的问题。

发明内容

本公开提供了一种样本聚类方法、装置、设备及存储介质，以解决相关技术中 DBSCAN 算法对于密度不均的样本集无法合理聚类的技术问题。

第一方面，本公开实施例提供了一种样本聚类方法，包括：

统计样本集中每个样本对应的第一样本距离，所述第一样本距离为所述样本与所述样本的第 S 近邻样本之间的距离；

在全部所述第一样本距离中，获取设定距离范围内的第一样本距离；

基于所述设定距离范围内的第一样本距离计算距离均值；

基于每个所述样本对应的 K 近邻样本集合，确定每个样本的全部连接样本，其中， $K > S$ ，所述样本与所述样本的连接样本为互为近邻样本且存在连接关系；

根据所述连接样本、所述距离均值和 S 值对所述样本集中的样本进行聚类，所述距离均值为扫描半径，所述 S 值为聚类最小包含样本数。

在一实施方式中，所述根据所述连接样本、所述距离均值和 S 值对所述样本集中的样本进行聚类包括：

基于所述距离均值对全部所述连接样本进行过滤，以滤除第二样本距离大于所述距离均值的连接样本，所述第二样本距离为样本与所述样本的连接样本之间的距离；

基于 S 值和过滤后得到的连接样本对所述样本集中的样本进行聚类。

在一实施方式中，所述基于 S 值和过滤后得到的连接样本对所述样本集中的样本进行聚类包括：

依次统计每个样本的连接样本总数量；

将所述连接样本总数量大于 S 值的样本作为核心样本；

在得到的全部核心样本中，选择任一核心样本作为当前样本；

访问所述当前样本的全部连接样本；

将访问得到的每个连接样本分别作为顶点，并访问所述顶点对应的全部连接样本；

重复将访问得到的每个连接样本分别作为顶点，并访问所述顶点对应的全部连接样本的操作，直到访问不到新的连接样本为止；

将未被访问过的任一核心样本更新为当前样本，并返回执行访问所述当前样本的全部连接样本的操作，直到全部核心样本均被访问为止；

将所述当前样本及基于当前样本访问得到的连接样本聚类为簇。

在一实施方式中，所述基于每个所述样本对应的 K 近邻样本集合，确定每个样本的全部连接样本包括：

获取每个样本对应的 K 近邻样本集合；

根据全部所述 K 近邻样本集合，构建邻接矩阵，所述邻接矩阵中每个元素代表对应两个样本间的近邻关系；

统计所述邻接矩阵中非零元素，以确定每个样本的全部连接样本。

在一实施方式中，所述统计所述邻接矩阵中非零元素，以确定每个样本的全部连接样本包括：

在所述邻接矩阵中，获取处于对称位置的元素组，所述元素组包括第 i 行第 j 列的第一元素和第 j 行第 i 列的第二元素；

若所述第一元素和所述第二元素中包含至少一个零元素，则将所述第一元素和第二元素均设置为零元素；

遍历所述邻接矩阵的全部元素组后，更新所述邻接矩阵；

统计更新后的邻接矩阵中非零元素，并将所述非零元素对应的两个样本确定为互为近邻样本且具有连接关系；

基于所述互为近邻样本，得到每个样本的全部连接样本。

在一实施方式中，所述在全部所述第一样本距离中，获取设定距离范围内的第一样本距离包括：

基于全部所述第一样本距离，构建频数分布直方图；

统计所述频数分布直方图中各 bin 的频数，以确定设定距离范围；

获取设定距离范围内的第一样本距离。

在一实施方式中，所述统计所述频数分布直方图中各 bin 的频数，以确定设定距离范围包括：

获取所述频数分布直方图中频数最大 bin；

计算相邻后位 bin 之间的频数落差，所述后位 bin 为所述频数分布直方图中位于频数最大 bin 后方的 bin；

确认频数落差最大的相邻后位 bin，并在所述最大的相邻后位 bin 中选择位于后方的 bin；

将所述频数最大 bin 对应的第一样本距离和所述位于后方的 bin 对应的第一样本距离作为设定距离范围的距离阈值。

在一实施方式中，所述基于所述设定距离范围内的第一样本距离计算距离均值包括：

获取所述第一样本距离处于所述设定距离范围内的样本数量；

对所述设定距离范围内每个第一样本距离进行相加，以得到样本总距离；

将所述样本总距离与所述样本数量的商值作为距离均值。

在一实施方式中，所述统计样本集中每个样本对应的第一样本距离之前，还包括：

构建样本集中每个样本的K近邻图，所述K近邻图中每条边的权值为对应样本间的距离。

第二方面，本公开实施例还提供了一种样本聚类装置，包括：

距离统计模块，用于统计样本集中每个样本对应的第一样本距离，所述第一样本距离为所述样本与所述样本的第S近邻样本之间的距离；

距离获取模块，用于在全部所述第一样本距离中，获取设定距离范围内的第一样本距离；

均值计算模块，用于基于所述设定距离范围内的第一样本距离计算距离均值；

连接确定模块，用于基于每个所述样本对应的K近邻样本集合，确定每个样本的全部连接样本，其中， $K > S$ ，所述样本与所述样本的连接样本为互为近邻样本且存在连接关系；

样本聚类模块，用于根据所述连接样本、所述距离均值和S值对所述样本集中的样本进行聚类，所述距离均值为扫描半径，所述S值为聚类最小包含样本数。

第三方面，本公开实施例还提供了一种样本聚类设备，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如第一方面所述的样本聚类方法。

第四方面，本公开实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如第一方面所述的样本聚类方法。

上述样本聚类方法、装置、设备及存储介质，通过统计样本集中每个样本与其第S近邻样本之间的第一样本距离，并基于第一样本距离得到距离均值，同时，基于每个样本的K ($K > S$) 近邻样本集合，确定每个样本对应的连接样本，该连接样本与样本之间为互为近邻样本且具有连接关系，之后，基于聚类最小包含样本数 (S 值) 以及扫描半径 (距离均值) 对具有连接关系的样本进行聚类的技术手段，解决了相关技术中 DBSCAN 算法对于密度不均的样本集无法合理聚类的技术问题，通过第一样本距离确定合理的扫描半径，之后，基于扫描半径对互为近邻样本进行聚类，保证了聚类合理性，当样本集中的样本分布密度不均时，通过互为近邻样本可以避免将稀疏分布的样本与密集分布的样本聚类成簇。同时，通过频数分布直方图确定距离均值，无需用户输入，减小了手动调参的工作量。

附图说明

图1为本公开实施例一提供的一种样本集分布示意图；

图2为本公开实施例一提供的一种样本聚类方法的流程图；

图3为本公开实施例一提供的另一种样本集分布示意图；

图4为本公开实施例二提供的一种样本聚类方法的流程图；

图5为本公开实施例二提供的一种K近邻图；

图6为本公开实施例二提供的一种频数分布直方图；

图7为本公开实施例二提供的一种邻接矩阵示意图；

图8为本公开实施例二提供的另一种邻接矩阵示意图；

图9为本公开实施例三提供的一种样本聚类装置的结构示意图；

图10为本公开实施例四提供的一种样本聚类设备的结构示意图。

具体实施方式

下面结合附图和实施例对本公开作详细说明。可以理解的是，此处所描述的实施例用于解释本公开，而非对本公开的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本公开相关的部分而非全部结构。

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) 是一个比较有代表性的基于密度的聚类算法。通常，DBSCAN 需要人工输入两个参数：eps 和 minPts。针对样本集中的某个对象（即某个样本）而言，若 eps 的值为 E，则将该对象的扫描半径 E 内的区域记为该对象的 E 邻域。若该对象的 E 邻域内的样本点数大于或等于 minPts，则将该对象记为核心对象。对于样本 P 和样本 Q，若样本 Q 在样本 P 的 E 邻域内，且样本 P 为核心对象，那么，样本 Q 从样本 P 直接密度可达。对于样本 P_1 、样本 P_2 、……、样本 P_n ，若样本 P_i 从样本 P_{i-1} 直接密度可达，那么样本 P_n 从样本 P_1 密度可达。设定样本 O 到样本 P 密度可达，样本 O 到样本 Q 密度可达，那么样本 P 和样本 Q 密度相连。对于 DBSCAN 而言，其目的是找到密度相连对象的最大集合。

对于密度不均的样本集，如图 1 所示的样本集，其中，图 1 为本公开实施例一提供的一种样本集分布示意图，参考图 1，左半部分的样本较为密集，右半部分的样本较为稀疏。此时，若将 eps 设置较小的值，例如，将 eps 设置为图 1 中圆圈 11 对应的半径，此时，对于右半部分的样本而言，由于扫描不到其他样本，因此，会被认为是噪声点并滤除。这样会导致右半部分的样本大量被滤除，影响聚类准确性。若将 eps 设置较大的值，例如，将 eps 设置为图 1 中圆圈 12 对应的半径，那么，在扫描时会将左半部分和右半部分的样本聚为一类，或者，使得左半部分的大量样本聚为一类，进而影响聚类准确性。

综上，本公开实施例提供一种样本集聚类方法，以解决对于密度不均的样本集无法合理聚类的问题。

实施例一

图 2 为本公开实施例一提供的一种样本聚类方法的流程图。实施例中提供的样本聚类方法可以由样本聚类设备执行，该样本聚类设备可以通过软件和/或硬件的方式实现，该样本聚类设备可以是两个或多个物理实体构成，也可以是一个物理实体构成。例如，样本聚类设备可以是电脑、手机、平板或交互智能平板等具有数据运算、分析能力的智能设备。

在一实施方式中，参考图 2，该样本聚类方法包括：

步骤 110、统计样本集中每个样本对应的第一样本距离，第一样本距离为样本与样本的第 S 近邻样本之间的距离。

示例性的，样本集中包括多个样本，每个样本的数据类型相同。其中，样本的数据类型可以根据实际情况设定，实施例对此不作限定。例如，实施例中以对样本集中的样本进行聚类为例描述样本聚类方法。在一实施方式中，样本集的获取方式实施例不作限定，其可以是样本聚类设备自行采集的数据，也可以是用户输入的数据，还可以是对特定数据进行处理后得到的数据。通常，样本集中每个样本代表一个数据特征。例如，样本集中每个样本表示同一用户在设定周期内每天不同时间段内的位置数据。

典型的，以图 1 为例，样本集中的样本散落在特征空间不同的位置。通常，样本位置与样本自身代表的特征有关，特征越相似，样本间的距离越近。例如，预先设定特征放置规则，进而根据该规则确定样本位置。其中，特征放置规则的具体内容可以根据实际情况设定。例如，针对位置数据而言，在特征空间内划分经纬度，之后，基于各样本的位置数据包含的经

纬度确定各样本在特征空间的位置。

在一实施方式中，获取样本集后，样本聚类设备可以计算样本集中任意两个样本之间的距离。其中，距离的计算方式实施例不作限定，例如，采用欧式距离、闵可夫斯基距离、曼哈顿距离等方式确定样本间的距离。通常，两个样本间距离越近，表明两个样本越相似。

示例性的，样本的第 S 近邻样本是指距离该样本第 S 近的样本。针对任一样本，样本聚类设备可以计算该样本与其他样本的距离，并根据各距离，确定距离该样本第 S 近的样本，并记为第 S 近邻样本。其中， S 为正整数，其具体数值可以结合实际情况设定。在一实施方式中， S 值表示聚类最小包含样本数。即对样本集聚类后得到多个簇，每个簇最少包含 S 个样本。一般而言，第 S 近邻样本的数量为 1，有些情况下，第 S 近邻样本的数量可能是多个，此时，可以任选一个第 S 近邻样本。

在一实施方式中，在确定第 S 近邻样本前，对每个样本绘制对应的 K 近邻图，其中， K 为正整数且大于 S ，通常， K 的具体数值可以根据实际情况设定。在一实施方式中，对于某个样本的 K 近邻图而言，顶点为该样本，且图中包含距离该样本最近的 K 个近邻样本，同时，将样本与 K 个近邻样本分别相连，且任一连线的权值为该连线的两个样本之间的距离。举例而言， K 等于 8 时，获取距离样本最近的 8 个样本，并将其分别连线，此时，连线两端的样本可以认为有连接关系，且其连线的权值为两个样本之间的距离。确定 K 近邻图后，便可以根据 K 近邻图得到样本的第 S 近邻样本，以及相应的距离。

在一实施方式中，将样本与第 S 近邻样本间的距离记为第一样本距离。

步骤 120、在全部第一样本距离中，获取设定距离范围内的第一样本距离。

在一实施方式中，设定距离范围是统计第一样本距离后得到的，用于计算扫描半径时的参考数据。通常，对于样本密度不均的样本集而言，样本稀疏的区域内，第一样本距离通常较大，样本密集的区域，第一样本距离通常较小。此时，为了保证扫描半径的准确性，需要参考设定距离范围，仅获取设定距离范围内的第一样本距离得到扫描半径。一般而言，设定距离范围内的第一样本距离是全部第一样本距离中具有代表性的数据。

示例性的，统计每个第一样本距离对应的样本数量。其中，样本数量为 50，表示存在 50 个样本的第一样本距离相同。在一实施方式中，根据样本数量确定设定距离范围。例如，基于样本数量构建频数分布直方图，其中，频数分布直方图的 bin 的具体数量可以结合样本集的样本总数量确定。在一实施方式中，频数分布直方图中横轴代表第一样本距离，纵轴代表第一样本距离的样本数量。获取频数分布直方图中频数最大 bin，其中，频数最大 bin 对应的样本数量最多。之后，在频数最大 bin 的后位 bin 中，计算任意相邻两个 bin 之间的样本数量差值，选择差值最大的相邻两个 bin。其中，后位 bin 是指从横轴而言位于频数最大 bin 后方的 bin。在一实施方式中，选择相邻两个 bin 中位于后方的 bin，并将该后方的 bin 和频数最大 bin 对应的两个第一样本距离确定为设定距离范围的两个距离阈值。或者是，获取频数最大 bin 后设定位数的后位 bin，并将设定位数的后位 bin 和频数最大 bin 对应的两个第一样本距离确定为设定距离范围的两个距离阈值。再如，统计每个第一样本距离对应的样本数量后，由人工基于样本数量确定设定距离范围。实施例中，以将差值最大的后方 bin 和频数最大 bin 对应的两个第一样本距离作为设定距离范围的两个距离阈值为例进行描述。其中，频数最大 bin 对应的第一样本距离较小，因此，将其作为设定距离范围的较小距离阈值，以在聚类时，保证扫描到足够数量特征相似的样本。差值最大的后方 bin 对应的第一样本距离较大，因此，将其作为设定距离范围的较大距离阈值。通常，差值最大的后方 bin 表明该 bin 对应的样本数量减少幅度最大，即说明从该 bin 对应的第一样本距离开始，样本数量越来越少，相应的，

样本和其第 S 近邻样本间特征差异越来越大, 因此, 将该 bin 对应的第一样本距离作为设定距离范围的较大阈值, 可以在计算扫描半径时, 忽略差异较大的样本, 进而保证聚类准确度。

步骤 130、基于设定距离范围内的第一样本距离计算距离均值。

在一实施方式中, 统计在设定距离范围内的全部样本及相应的第一样本距离。之后, 将设定距离范围内的全部第一样本距离相加, 并将相加结果记为样本总距离。同时, 统计设定距离范围内的样本总数量, 并将样本总距离除以样本总数量, 进而将得到的商记为距离均值。此时, 该距离均值表示扫描半径。

步骤 140、基于每个样本对应的 K 近邻样本集合, 确定每个样本的全部连接样本, 其中, $K > S$, 样本与样本的连接样本为互为近邻样本且存在连接关系。

在一实施方式中, 计算每个样本与其他各样本之间的距离后, 获取与该样本距离最近的 K 个样本, 组成该样本对应的 K 近邻样本集合, 其中, K 近邻样本集合中的每个样本可以记为近邻样本, 即样本与近邻样本间存在近邻关系。如果预先构建样本的 K 近邻图, 则本步骤可以直接获取 K 近邻图中的样本作为 K 近邻样本集合。

在一实施方式中, 连接样本是指与当前样本具有连接关系的样本, 通常, 具有连接关系的两个样本可以记为互为近邻样本, 互为近邻样本可以理解为两个样本对应的 K 近邻样本集合中均包含另一个样本。在一实施方式中, 获取当前样本的 K 近邻样本集合, 之后, 再获取 K 近邻样本集合中每个近邻样本的 K 近邻样本集合。确定当前样本是否包含在每个近邻样本的 K 近邻样本集合中, 若当前样本包含在某个近邻样本的 K 近邻样本集合中, 则将该近邻样本和当前样本确定为互为近邻样本, 并保存近邻样本和当前样本间连接关系, 此时, 将该近邻样本记为当前样本的连接样本。其中, 保存连接关系的方式可以是在样本集中, 绘制两个样本之间的连线。在一实施方式中, 按照上述步骤便可以确定每个样本对应的全部连接样本。相应的, 对于 K 近邻样本集合中的非连接样本, 可以不保存其连接关系。

此外, 还可以通过邻接矩阵的方式确定连接样本。在一实施方式中, 基于 K 近邻样本集合构建邻接矩阵, 其中, 邻接矩阵中每个元素用于表示对应的两个样本是否为近邻关系。若是近邻关系, 则对应的元素为非零元素, 若不是近邻关系, 则对应的元素为零元素。在一实施方式中, 确定邻接矩阵中任一组位置对称的元素是否均为非零元素, 若是, 则说明两个样本对应的 K 近邻样本集合中均包含另一个样本, 即两个样本为互为近邻样本且具有连接关系。遍历邻接矩阵的全部对称元素后, 便可以确定每个样本的连接样本。

举例而言, 图 3 为本公开实施例一提供的另一种样本集示意图。参考图 3, K 为 5, 对于样本 A, 其 K 近邻样本集合包含与样本 A 为实线连接的 5 个样本, 其中, 有两个样本与样本 A 的连线部分重合。对于样本 B, 其 K 近邻样本集合包含与样本 B 为实线连接的 4 个样本以及样本 A。虽然样本 B 的 K 近邻样本集合中包含样本 A, 但是样本 A 的 K 近邻样本集合中不包含样本 B, 因此, 样本 B 和样本 A 不是互为近邻样本, 此时不保存样本 A 和样本 B 的连接关系。按照上述方式遍历全部样本后, 便可以得到与每个样本具有连接关系的连接样本。

需要说明的是, 实施例不限定步骤 140 与步骤 110-步骤 130 的执行顺序, 实际应用中, 也可以先执行步骤 140, 再执行步骤 110-步骤 130, 或者是步骤 140 与步骤 110-步骤 130 同步执行。

步骤 150、根据连接样本、距离均值和 S 值对样本集中的样本进行聚类, 距离均值为扫描半径, S 值为聚类最小包含样本数。

其中, 聚类最小包含样本数为 minPts 。示例性的, 将距离均值作为扫描半径, 将 S 值作为 minPts 对样本集中的样本进行 DBSCAN 聚类。在一实施方式中, 选择某个样本作为当前

样本，之后，以距离均值作为扫描半径对当前样本的周围区域进行扫描。此时，在扫描过程中，仅获取当前样本的连接样本，之后，若连接样本的数量大于 S ，则将当前样本和扫描得到的连接样本聚类为簇。

或者是，确定每个样本的连接样本总数量，若连接样本总数量大于 S ，则将该样本作为核心样本。遍历样本集中的全部样本后，找到全部的核心样本。之后，任选一核心样本作为当前样本，并获取当前样本的全部连接样本。在一实施方式中，将获取到的连接样本作为顶点，继续获取每个连接样本的全部连接样本，之后，再将获取到的连接样本作为顶点，并继续获取其对应的全部连接样本，重复此操作，直到遍历不到新的连接样本为止，此时，将得到的全部连接样本和当前样本聚类为簇。之后，获取未被处理的核心样本，并继续按照上述步骤得到连接样本，进而形成新的簇。当确定每个核心样本均被处理后，确定聚类结束。可以理解的是，在聚类过程中，若获取的某个连接样本为核心样本，则在后续过程中，不会再对该核心样本进行任何处理。

此时，针对于图 3 的样本集而言，在聚类的过程，样本 A 和样本 B 不会聚类成簇，样本 A 会与其特征相似且密度较密集的其他样本进行聚类，样本 B 会与其特征相似且密度较稀疏的其他样本进行聚类，保证了聚类合理性。

上述，通过统计样本集中每个样本与其第 S 近邻样本之间的第一样本距离，并基于第一样本距离得到距离均值，同时，基于每个样本的 K ($K > S$) 近邻样本集合，确定每个样本对应的连接样本，该连接样本与样本之间为互为近邻样本且具有连接关系，之后，基于聚类最小包含样本数 (S 值) 以及扫描半径 (距离均值) 对具有连接关系的样本进行聚类的技术手段，解决了相关技术中 DBSCAN 算法对于密度不均的样本集无法合理聚类的技术问题，通过第一样本距离确定合理的扫描半径，之后，基于扫描半径对互为近邻样本进行聚类，保证了聚类合理性，当样本集中的样本分布密度不均时，通过互为近邻样本可以避免将稀疏分布的样本与密集分布的样本聚类成簇，进而保证聚类准确性。

实施例二

图 4 为本公开实施例二提供的一种样本聚类方法的流程图。本实施例是在上述实施例的基础上进行具体化。参考图 4，本实施例提供的样本聚类方法包括：

步骤 201、构建样本集中每个样本的 K 近邻图， K 近邻图中每条边的权值为对应样本间的距离。

在一实施方式中，计算每个样本与其他各样本之间的距离后，绘制某个样本的 K 近邻图时，将该样本作为顶点，并根据样本之间的距离获取距该样本最近的 K 个样本以及对应的距离。之后，分别绘制顶点与 K 个样本的连线，并将顶点与对应样本的距离作为该连线的权值。举例而言，图 5 为本公开实施例二提供的一种 K 近邻图。参考图 5，其为图 3 中样本 C 的 K 近邻图，其中， K 为 6，根据样本 C 与其他样本的距离获取距离样本 C 最近的 K 个样本以及距离，之后，绘制样本 C 与 K 个样本的连线，并将距离作为连线的权值。需要说明的是，图 5 中示出的连线的权值仅用于描述 K 近邻图，并非对距离或距离计算的限定。实际应用中，权值可以不示于 K 近邻图中。一般而言，基于 K 近邻图，可以得到样本集中各样本之间的连线关系，即近邻关系。

需要说明的是，构建 K 近邻图的好处是便于后续确定第一样本距离、连接样本等，即便于后续计算。

步骤 202、统计样本集中每个样本对应的第一样本距离，第一样本距离为样本与样本的第 S 近邻样本之间的距离。

由于 $S < K$ ，因此，可以在每个样本的 K 近邻图中，获取距离其最近的第 S 近邻样本，并根据样本和第 S 近邻样本的连线的权值确定第一样本距离。

步骤 203、基于全部第一样本距离，构建频数分布直方图。

在一实施方式中，统计每个第一样本距离的出现次数，并将该次数作为对应第一样本距离的频数，之后，基于各频数构建频数分布直方图。实际应用中，考虑到存在数值很接近的第一样本距离，例如，两个第一样本距离分别为 0.585 和 0.593，其具体的数值比较接近，此时，若每个第一样本距离对应一个频数，会增加统计复杂度及频数分布直方图复杂度，因此。实施例中，预先对第一样本距离进行分组，之后，统计在每个分组内第一样本距离的出现次数，并将出现次数作为该分组对应的频数。例如，距离 0.55-0.65 的分组内有 1100 个第一样本距离，因此，该分组对应的频数为 1100。需要说明的是，分组规则实施例不作限定。在一实施方式中，在建立频数分布直方图时，用横坐标表示距离（本示例中为第一样本距离），纵坐标表示样本数（即频数）。此时，每个距离分组可以记为一个 bin。通常，bin 的数量跟样本集的样本总数量相关。例如，样本集的样本总数量在 5000 以下，此时，bin 的数量可以设为 10，样本总数量超过 5000 后，设定每增加 500 个样本，bin 的数量增加 1。在一实施方式中，考虑到有些分组内频数很少，在后续计算时，其参考意义较低，因此，可以结合实际情况，忽略该距离分组的频数。

例如，图 6 为本公开实施例二提供的一种频数分布直方图。参考图 6，该频数分布直方图的横坐标为距离，纵坐标为样本数，即频数，且 bin 为 10。由图 6 可知每个 bin 对应的距离范围以及频数，并且，bin 从 1 至 10 顺序排列。

设置频数分布直方图的好处是，清楚显示各距离范围内频数分布情况，以及易于显示各距离范围之间频数的差别。

步骤 204、统计频数分布直方图中各 bin 的频数，以确定设定距离范围。

实施例中，设定基于各 bin 的频数确定设定距离范围。此时，该步骤包括步骤 2041-步骤 2044：

步骤 2041、获取频数分布直方图中频数最大 bin。

在一实施方式中，统计各距离范围对应的频数，之后，获取最大的频数对应的 bin，并记为频数最大 bin。例如，参考图 6，基于纵坐标确定频数最大 bin 为 5。

步骤 2042、计算相邻后位 bin 之间的频数落差，后位 bin 为频数分布直方图中位于频数最大 bin 后方的 bin。

示例性的，根据 bin 的排列顺序，将位于某个 bin 后方的任一 bin 记为该 bin 的后位 bin。实施例中，获取频数最大 bin 的全部后位 bin。以图 6 为例，频数最大 bin 的后位 bin 为第 6 个 bin 至第 10 个 bin。

在一实施方式中，相邻 bin 是指顺序相邻的两个 bin，例如，图 6 中，第 1 个 bin 和第 2 个 bin 为相邻 bin，第 2 个 bin 和第 3 个 bin 为相邻 bin，依次类推。实施例中，获取后位 bin 中的相邻 bin，并计算相邻 bin 中两个 bin 的频数差值，其中，频数差值为正数且记为频数落差。在计算频数差值时，可以是对相邻的两个 bin 对应的频数做减法，若结果为正数，则直接将该结果作为频数落差，若结果为负数，则将该结果的绝对值作为频数落差。

实施例中，统计频数落差时，还可以计算频数最大 bin 与其相邻后位 bin 之间的频数落差，并将该频数落差与各相邻后位 bin 之间的频数落差一同作为本步骤的计算结果。

步骤 2043、确认频数落差最大的相邻后位 bin，并在最大的相邻后位 bin 中选择位于后方的 bin。

通常, 频数落差越大, 表明相邻 bin 中位于后方的 bin 与位于前方的 bin 之间的频数相差越多, 进而确定相邻 bin 中位于后方的 bin 对应的样本数量明显减少。实施例中, 统计各频数落差后, 选择频数落差最大的相邻后位 bin, 此时, 在该相邻后位 bin 中, 位于后方的 bin 对应的样本数量明显减少, 且位于该 bin 后方的 bin 对应的样本数量很少, 其代表性较低, 在计算距离均值时, 对准确性贡献较小, 因此, 可以在计算时可以忽略不计。据此, 实施例中设定选择频数落差最大的相邻后位 bin 中位于后方的 bin 作为本步骤的计算结果。举例而言, 参考图 6, 分别计算第 6 个 bin 与第 7 个 bin 之间的频数落差、第 7 个 bin 与第 8 个 bin 之间的频数落差、第 8 个 bin 与第 9 个 bin 之间的频数落差以及第 9 个 bin 与第 10 个 bin 之间的频数落差。计算各频数落差后, 确定, 第 6 个 bin 与第 7 个 bin 之间频数落差的频数落差最大, 此时, 选择位于后方的第 7 个 bin。

步骤 2044、将频数最大 bin 对应的第一样本距离和位于后方的 bin 对应的第一样本距离作为设定距离范围的距离阈值。

在一实施方式中, 由于每个 bin 对应一个距离范围, 因此, 确定设定距离范围时, 可以是选择每个距离范围的最小值作为距离阈值。例如, 频数最大 bin 对应的距离范围的最小值作为设定距离范围中小的距离阈值, 将后方的 bin 对应的距离范围的最小值作为设定距离范围中大的距离阈值。也可以是选择每个距离范围的最大值作为距离阈值。例如, 频数最大 bin 对应的距离范围的最大值作为设定距离范围中小的距离阈值, 将后方的 bin 对应的距离范围的最大值作为设定距离范围中大的距离阈值。还可以是结合实际情况, 选择频数最大 bin 对应的距离范围的最小值作为设定距离范围中小的距离阈值, 将后方的 bin 对应的距离范围的最大值作为设定距离范围中大的距离阈值。例如, 参考图 6, 将第 5 个 bin 对应的距离范围中的最小值作为设定距离范围中的小的距离阈值, 将第 7 个 bin 对应的距离范围中的最大值作为设定距离范围中的大的距离阈值。也可以是确定每个距离范围的居中距离值, 并将居中距离值作为距离阈值。

通常, 将频数最大 bin 对应的第一样本距离和位于后方的 bin 对应的第一样本距离作为设定距离范围的距离阈值的好处是, 后续计算距离均值后, 可以保证通过该距离均值聚类到较多样本, 且滤除距离较远的样本, 即保证了距离均值的合理性。

可以理解的是, 步骤 2041-步骤 2044 仅为确定设定距离范围的一种方式。实际应用中, 还可以结合频数分布直方图采用其他方式确定设定距离范围。

步骤 205、获取设定距离范围内的第一样本距离。

步骤 206、获取第一样本距离处于设定距离范围内的样本数量。

在一实施方式中, 由于一个样本对应一个第一样本距离。因此, 可以统计在设定距离范围内的第一样本距离的总个数, 并将总个数作为样本数量。也可以结合频数分布直方图确定样本数量, 此时, 将频数最大 bin 以及位于后方的 bin 之间的各 bin 对应的频数相加, 以得到样本数量。例如, 参考图 6, 将第 5 个 bin、第 6 个 bin 以及第 7 个 bin 对应的频数相加, 便可以得到样本数量。

步骤 207、对设定距离范围内每个第一样本距离进行相加, 以得到样本总距离。

在一实施方式中, 将设定距离范围内的每个第一样本距离相加, 并将结果记为样本总距离。或者是, 基于频数分布直方图得到样本总距离, 此时, 可以将对应 bin 的频数乘以对应距离范围的居中距离值, 之后, 在再将得到的结果相加, 进而得到样本总距离。例如, 参考图 6, 获取第 5 个 bin 对应距离范围, 之后, 选择距离范围的居中距离值。将第 5 个 bin 对应的居中距离值和频数相乘得到第一结果, 之后, 按照同样的计算方式得到第 6 个 bin 对应的

第二结果以及第 7 个 bin 对应的第三结果，之后，将三个结果相加，以得到样本总距离。

步骤 208、将样本总距离与样本数量的商值作为距离均值。

在一实施方式中，用样本总距离除以样本数量，便可以得到设定距离范围内第一样本距离的平均距离。实施例中，将平均距离记为距离均值，并将距离均值设置为扫描半径。相比于 DBSCAN 算法中人为指定扫描半径，本实施例可以集合样本集的实际情况自适应得到扫描半径，且利用了统计学原理确定扫描半径，保证了扫描半径的合理性。

步骤 209、获取每个样本对应的 K 近邻样本集合。

在一实施方式中，基于每个样本的 K 近邻图，获取每个样本的 K 个近邻样本，并组成 K 近邻样本集合。即将 K 近邻图中除顶点外的全部样本组成顶点的 K 近邻样本集合。

步骤 210、根据全部 K 近邻样本集合，构建邻接矩阵，邻接矩阵中每个元素代表对应两个样本间的近邻关系。

其中，邻接矩阵是用一个一维数组存放样本集所有样本；用一个二维数组存放样本集间关系的数据。邻接矩阵可以分为有向图邻接矩阵和无向图邻接矩阵。实施例中，以无向图邻接矩阵为例。在一实施方式中，将每个样本按顺序排列，排列后，每个样本对应一个编号。其中，排列规则实施例不做限定。在一实施方式中，将排列后的样本分别作为矩阵的横轴和纵轴，之后，用横轴和纵轴的交点元素表明对应两个样本之间是否为近邻关系。在一实施方式中，若某个样本包含在另一个样本的 K 近邻样本集合中，则该样本与另一个样本为近邻关系，此时，将以该样本为纵坐标，另一个样本为横坐标的交点元素记为非零元素。其中，非零元素的具体值可以根据实际情况设定。实施例中，以非零元素为 1 为例。实际应用中，非零元素还可以为对应的距离值，或其他数值。相应的，若某个样本不包含在另一个样本的 K 近邻样本集合中，则该样本与另一个样本为非近邻关系，此时，将以该样本为纵坐标，另一个样本为横坐标的交点元素记为零元素，即记为 0。举例而言，图 7 为本公开实施例二提供的一种邻接矩阵示意图。参考图 7，当前共有 8 个样本，此时，将 8 个样本按顺序赋予 1-8 编号。之后，在横轴和纵轴上排列的 8 个样本，即将 8 个样本编号作为横坐标和纵坐标，之后，根据样本间的近邻关系构建二维矩阵。此时，二维矩阵中第 i 行第 j 列的元素表明第 j 个样本是否为第 i 个样本的近邻样本。例如，第 2 行第 3 列的元素为 1，则表明编号为 2 的样本对应的 K 近邻样本集合中包含编号为 3 的样本。再如，第 7 行第 1 列的元素为 0，则表明编号为 7 的样本对应的 K 近邻样本集合中不包含编号为 1 的样本。通常，确定 K 近邻样本集合后，便可以根据 K 近邻样本集合为邻接矩阵内的元素进行赋值。通常，第 i 行第 i 列的元素表明样本与自身的连接关系，实施例中，将该元素记为 1。可以理解的是，图 7 中横向的 1-8 编号和纵向的 1-8 编号为样本编号，不计在行数和列数中。

在一实施方式中，实际应用中，样本集包含很多的样本，此时，可以基于每个样本构建一个邻接矩阵，并将相应样本及对应的 K 近邻样本集合中的样本作为邻接矩阵的横轴和纵轴，通过该邻接矩阵可以确定样本与 K 个近邻样本间的近邻关系。或者是，基于样本集构建一个邻接矩阵。此时，通过该邻接矩阵可以确定样本集中各样本间的近邻关系。

步骤 211、统计邻接矩阵中非零元素，以确定每个样本的全部连接样本。

在一实施方式中，若某个样本的 K 近邻样本集合包含另一样本，而另一样本的 K 近邻样本集合不包含该样本，则确定两个样本为非互为近邻样本，在聚类时可以忽略。因此，需要找到全部的互为近邻样本，并基于互为近邻样本确定各样本的连接样本。实施例中，通过邻接矩阵中的非零元素确定互为近邻样本。例如，两个样本的编号分别为 5 和 6，并记为样本 5 和样本 6。此时，邻接矩阵中，第 5 行第 6 列的元素表明样本 5 的 K 近邻样本集合中是否包

含样本 6，第 6 行第 5 列的元素表明样本 6 的 K 近邻样本集合中是否包含样本 5。若两个元素均为非零元素，则说明样本 5 和样本 6 相互包含，即样本 5 和样本 6 为互为近邻样本，且保存样本 5 和样本 6 的连接关系，此时，将样本 5 记为样本 6 的连接样本，同样，将样本 6 记为样本 5 的连接样本。按照上述方式，统计每个非零元素，便可以得到每个样本的连接样本。

实施例中，在统计非零元素时，设定该步骤包括步骤 2111-步骤 2115：

步骤 2111、在邻接矩阵中，获取处于对称位置的元素组，元素组包括第 i 行第 j 列的第一元素和第 j 行第 i 列的第二元素。

在一实施方式中，将邻接矩阵中处于对称位置的两个元素记为元素组。其中，对称位置是指横纵坐标相反的两个位置。例如，第 i 行第 j 列和第 j 行第 i 列为对称位置，此时，将两个对称位置对应的元素记为一个元素组。在一实施方式中，将第 i 行第 j 列的元素记为第一元素，其表明第 i 个样本对应的 K 近邻样本集合中是否包含第 j 个样本。将第 j 行第 i 列的元素记为第二元素，其表明第 j 个样本对应的 K 近邻样本集合中是否包含第 i 个样本。通过获取对称位置的元素组，便可以得到对应两个样本间的近邻关系。

步骤 2112、若第一元素和第二元素中包含至少一个零元素，则将第一元素和第二元素均设置为零元素。

在一实施方式中，确定第一元素和第二元素中是否包含至少一个零元素，若第一元素和第二元素中包含至少一个零元素，则将第一元素和第二元素均设置为零元素，否则，保持第一元素和第二元素不变，并执行步骤 2123。其中，第一元素和第二元素中包含至少一个零元素表明对应的两个样本中至少有一个样本的 K 近邻样本集合中不包含另一个样本，即两个样本为非互为近邻样本。此时，将第一元素和第二元素均修改为零元素，即取消两个元素间的近邻关系。举例而言，参考图 7，第 1 行第 8 列的第一元素为 1，第 8 行第 1 列的第二元素为 0，且第一元素和第二元素属于对称位置的元素组，因此，将第 1 行第 8 列的第一元素修改为 0，即取消样本 1 和样本 8 的近邻关系。

步骤 2113、遍历邻接矩阵的全部元素组后，更新邻接矩阵。

在一实施方式中，遍历处于对称位置的全部元素组后，更新邻接矩阵。以图 7 中所示的邻接矩阵为例，此时，遍历图 7 中的全部元素组后，更新该邻接矩阵，并得到图 8 所示的邻接矩阵。其中，图 8 为本公开实施例二提供的另一种邻接矩阵示意图。

步骤 2114、统计更新后的邻接矩阵中非零元素，并将非零元素对应的两个样本确定为互为近邻样本且具有连接关系。

在一实施方式中，更新后的邻接矩阵中任一非零元素对应的两个样本为互为近邻样本。因此，基于更新后的邻接矩阵中的非零元素，边可以确定全部互为近邻样本。在一实施方式中，考虑到更新后的邻接矩阵为对称化的邻接矩阵，因此，在获取非零元素时，可以仅获取对称位置中的一个非零元素，并根据该非零元素确定两个样本为互为近邻样本。确定互为近邻样本后，可以在样本集中保留互为近邻样本间的连线。

步骤 2115、基于互为近邻样本，得到每个样本的全部连接样本。

在一实施方式中，获取包含某个样本的全部互为近邻样本，并将获取的全部互为近邻样本中的另一样本作为该样本的全部连接样本。

步骤 212、基于距离均值对全部连接样本进行过滤，以滤除第二样本距离大于距离均值的连接样本，第二样本距离为样本与样本的连接样本之间的距离。

在一实施方式中，实施例中将样本与其对应的连接样本之间的距离记为第二样本距离，

即将互为近邻样本间的距离记为第二样本距离。在一实施方式中，若第二样本距离大于距离均值，则说明对应的两个样本虽然为互为近邻样本，但是其具体的特征差异较大，若聚类在一起会影响聚类结果的准确性。因此，实施例设定剔除两个样本的连接关系，即将两个样本确定为非互为近邻样本。此时，可以在样本集中删除两个样本间的连线。同时，将邻接矩阵中对应的元素调整为零元素。可以理解的是，按照上述方式便可以过滤每个样本对应的连接样本，且仅保留小于距离均值的连接样本。此时，该步骤也可以理解为基于扫描半径对样本集中各样本的连接关系进行扫描，以得到准确的连接关系。

步骤 213、基于 S 值和过滤后得到的连接样本对样本集中的样本进行聚类。

在一实施方式中，该步骤包括步骤 2131-步骤 2139：

步骤 2131、依次统计每个样本的连接样本总数量。

在一实施方式中，根据样本集中各样本间的连线，确定每个样本的连线总数量，进而得到连接样本总数量。一般而言，每个样本对应的连接样本总数量为基于距离均值过滤后保留的连接样本的总数量。可以理解的是，实际应用中，也可以仅记录互为相邻样本间的连接关系，而不在样本集中体现。此时，可以根据记录的连接关系确定每个样本的连接样本，进而得到连接样本总数量。

步骤 2132、将连接样本总数量大于 S 值的样本作为核心样本。

在一实施方式中，将每个样本的连接样本总数量与 S 值进行比较，若连接样本总数量大于 S 值，则将对应该样本记为核心样本。按照上述方式，遍历每个样本后，便可以得到样本集中的全部核心样本。其中，核心样本可以理解为聚类过程中，可以作为起始点的样本。通常，每个样本及其连接样本因为特征相似度较高，通常会被聚类为一簇，若某个样本的总数量小于 S 值，则说明后续聚类得到的簇中存在样本数量低于聚类最小包含样本数的可能，因此，在聚类时，不会选择该样本作为起始点，即不会选择该样本作为核心样本。

步骤 2133、在得到的全部核心样本中，选择任一核心样本作为当前样本。

在一实施方式中，任选一个核心样本作为本次聚类的起始点，并记为当前样本。需要说明的是，实施例以随机方式确定当前样本。实际应用中，还可以设定当前样本选择规则，并通过该规则选择当前样本。通常，确定当前样本后，将该核心样本标记为被访问过。

步骤 2134、访问当前样本的全部连接样本。

在一实施方式中，根据当前保留的连接关系，获取与当前样本具有连接关系的全部连接样本，并将获取的全部连接样本标记为被访问过。

步骤 2135、将访问得到的每个连接样本分别作为顶点，并访问顶点对应的全部连接样本。

在一实施方式中，将当前得到的每个连接样本分别作为一个顶点，之后，根据当前保留的连接关系，继续获取与每个顶点具有连接关系的全部连接样本。此时，每个顶点也可以认为是一次聚类中的子起始点。

步骤 2136、确认是否访问得到新的连接样本。若访问得到新的连接样本，则返回执行步骤 2135，若访问不到新的连接样本，则执行步骤 2137。

在一实施方式中，根据当前保留的连接关系，继续获取与每个顶点具有连接关系的全部连接样本时，确定是否得到新的连接样本，即是否得到没有标记为被访问过的连接样本。若得到新的连接样本，则说明当前还有新的特征相似的样本，此时，可以返回执行步骤 2135，即将新得到的连接样本作为顶点，继续访问该顶点的全部连接样本，直到得不到新的连接样本为止。若得不到新的连接样本，则说明当前已经找到基于核心样本的全部特征相似的样本。此时，可以认为本次聚类结束，并执行 2137。

需要说明的是,在本次聚类过程中,若某个核心样本被认为是连接样本,则将该核心样本标记为被访问过的样本。

步骤 2137、确认是否还存在未被访问的核心样本。若存在,则执行步骤 2138,否则,执行 2139。

在一实施方式中,确定当前是否还有未被访问的核心样本,即确定当前是否还有未被标记为被访问过的核心样本。若存在未被访问的核心样本,则将该核心样本更新为当前样本,并开始新一次的聚类过程,即执行步骤 2138。若确认核心样本均被访问过,则说明当前已经访问了全部可被聚类的起始点,无法再找到聚类起始点,因此,执行步骤 2139。

步骤 2138、将未被访问过的任一核心样本更新为当前样本。返回执行步骤 2134。

在一实施方式中,若未被访问过的核心样本数量大于 1,则采用随机的方式选定当前样本。若未被访问过的核心样本数量为 1,则将该核心样本作为当前样本。之后,返回执行步骤 2134,即开始一次新的聚类。

步骤 2139、将当前样本及基于当前样本访问得到的连接样本聚类为簇。

在一实施方式中,一次聚类过程可以认为是一个访问过程,此时,设定将每次聚类得到的全部连接样本和当前样本聚类为一簇。通常,对于样本集而言,存在几次聚类过程便可以得到相应数量的簇。例如,将始终未被访问过的样本记为噪声点。

举例而言,针对于图 3 的样本集而言,样本 B 的 K ($K=5$) 近邻样本集合中包含样本 A,样本 A 的 K 近邻样本集合中包含样本 B,此时,样本 A 和样本 B 间的连接关系会被踢出,即删除样本 A 和样本 B 间虚线。在后续聚类过程中,无论访问样本 A 的连接样本还是访问样本 B 的连接样本,均不会将样本 A 和样本 B 聚类成簇,进而保证了聚类合理性。

上述,通过构建每个样本的 K 近邻图,基于 K 近邻图获取第一样本距离,其中,第一样本距离为样本与第 S ($S < K$) 近邻样本之间的距离,之后,基于第一样本距离构建频数分布直方图,并根据频数分布直方图确定距离均值,同时,基于 K 近邻图构建邻接矩阵,并对称邻接矩阵,以确定具有连接关系的样本,之后,通过距离均值对具有连接关系的样本进行过滤,并基于过滤后的连接关系以及 S 值对样本集进行聚类的技术手段,解决了相关技术中 DBSCAN 算法对于密度不均的样本集无法合理聚类的技术问题,通过对称邻接矩阵及基于距离均值过滤具有连接关系的样本的方式,可以在样本的分布密度不均时,避免不同密度的样本聚成一个簇,影响聚类准确性。同时,通过频数分布直方图确定距离均值,无需用户输入,减小了手动调参的工作量,并且通过统计的方式确定距离均值,保证了距离均值的合理性,进而保证聚类准确性。

实施例三

图 9 为本公开实施例三提供的一种样本聚类装置的结构示意图。参考图 9,该样本聚类装置包括:距离统计模块 301、距离获取模块 302、均值计算模块 303、连接确定模块 304 以及样本聚类模块 305。

其中,距离统计模块 301,用于统计样本集中每个样本对应的第一样本距离,所述第一样本距离为所述样本与所述样本的第 S 近邻样本之间的距离;距离获取模块 302,用于在全部所述第一样本距离中,获取设定距离范围内的第一样本距离;均值计算模块 303,用于基于所述设定距离范围内的第一样本距离计算距离均值;连接确定模块 304,用于基于每个所述样本对应的 K 近邻样本集合,确定每个样本的全部连接样本,其中, $K > S$,所述样本与所述样本的连接样本为互为近邻样本且存在连接关系;样本聚类模块 305,用于根据所述连接样本、所述距离均值和 S 值对所述样本集中的样本进行聚类,所述距离均值为扫描半径,所

述 S 值为聚类最小包含样本数。

上述，通过统计样本集中每个样本与其第 S 近邻样本之间的第一样本距离，并基于第一样本距离得到距离均值，同时，基于每个样本的 K ($K > S$) 近邻样本集合，确定每个样本对应的连接样本，该连接样本与样本之间为互为近邻样本且具有连接关系，之后，基于聚类最小包含样本数 (S 值) 以及扫描半径 (距离均值) 对具有连接关系的样本进行聚类的技术手段，解决了相关技术中 DBSCAN 算法对于密度不均的样本集无法合理聚类的技术问题，通过第一样本距离确定合理的扫描半径，之后，基于扫描半径对互为近邻样本进行聚类，保证了聚类合理性，当样本集中的样本分布密度不均时，通过互为近邻样本可以避免将稀疏分布的样本与密集分布的样本聚类成簇，进而保证聚类准确性。

在上述实施例的基础上，样本聚类模块 305 包括：样本过滤子模块，用于基于所述距离均值对全部所述连接样本进行过滤，以滤除第二样本距离大于所述距离均值的连接样本，所述第二样本距离为样本与所述样本的连接样本之间的距离；聚类子模块，用于基于 S 值和过滤后得到的连接样本对所述样本集中的样本进行聚类。

在上述实施例的基础上，聚类子模块包括：总数量统计单元，用于依次统计每个样本的连接样本总数量；核心样本确定单元，用于将所述连接样本总数量大于 S 值的样本作为核心样本；当前样本选择单元，用于在得到的全部核心样本中，选择任一核心样本作为当前样本；第一访问单元，用于访问所述当前样本的全部连接样本；第二访问单元，用于将访问得到的每个连接样本分别作为顶点，并访问所述顶点对应的全部连接样本；第三访问单元，用于重复将访问得到的每个连接样本分别作为顶点，并访问所述顶点对应的全部连接样本的操作，直到访问不到新的连接样本为止；样本更新单元，用于将未被访问过的任一核心样本更新为当前样本，并返回执行访问所述当前样本的全部连接样本的操作，直到全部核心样本均被访问为止；簇聚类单元，用于将所述当前样本及基于当前样本访问得到的连接样本聚类为簇。

在上述实施例的基础上，连接确定模块 304 包括：集合获取子模块，用于获取每个样本对应的 K 近邻样本集合；邻接矩阵构建子模块，用于根据全部所述 K 近邻样本集合，构建邻接矩阵，所述邻接矩阵中每个元素代表对应两个样本间的近邻关系；非零元素统计子模块，用于统计所述邻接矩阵中非零元素，以确定每个样本的全部连接样本。

在上述实施例的基础上，非零元素统计子模块包括：元素组获取单元，用于在所述邻接矩阵中，获取处于对称位置的元素组，所述元素组包括第 i 行第 j 列的第一元素和第 j 行第 i 列的第二元素；置零单元，用于若所述第一元素和所述第二元素中包含至少一个零元素，则将所述第一元素和第二元素均设置为零元素；矩阵更新单元，用于遍历所述邻接矩阵的全部元素组后，更新所述邻接矩阵；近邻关系确定单元，用于统计更新后的邻接矩阵中非零元素，并将所述非零元素对应的两个样本确定为互为近邻样本且具有连接关系；连接样本确定单元，用于基于所述互为近邻样本，得到每个样本的全部连接样本。

在上述实施例的基础上，距离获取模块 302 包括：直方图构建子模块，用于基于全部所述第一样本距离，构建频数分布直方图；频数统计子模块，用于统计所述频数分布直方图中各 bin 的频数，以确定设定距离范围；第一距离获取子模块，用于获取设定距离范围内的第一样本距离。

在上述实施例的基础上，频数统计子模块包括：最大 bin 获取单元，用于获取所述频数分布直方图中频数最大 bin；落差计算单元，用于计算相邻后位 bin 之间的频数落差，所述后位 bin 为所述频数分布直方图中位于频数最大 bin 后方的 bin；bin 确认单元，用于确认频数落差最大的相邻后位 bin，并在所述最大的相邻后位 bin 中选择位于后方的 bin；阈值确定单

元,用于将所述频数最大 bin 对应的第一样本距离和所述位于后方的 bin 对应的第一样本距离作为设定距离范围的距离阈值。

在上述实施例的基础上,均值计算模块 303 包括:样本数量获取子模块,用于获取所述第一样本距离处于所述设定距离范围内的样本数量;总距离子模块,用于对所述设定距离范围内每个第一样本距离进行相加,以得到样本总距离;商值计算子模块,用于将所述样本总距离与所述样本数量的商值作为距离均值。

在上述实施例的基础上,还包括:K 近邻图构建模块,用于统计样本集中每个样本对应的第一样本距离之前,构建样本集中每个样本的 K 近邻图,所述 K 近邻图中每条边的权值为对应样本间的距离。

本公开实施例提供的样本聚类装置包含在样本聚类设备中,且可用于执行上述任意实施例提供的样本聚类方法,具备相应的功能和有益效果。

实施例四

图 10 为本公开实施例四提供的一种样本聚类设备的结构示意图。如图 10 所示,该样本聚类设备包括处理器 40、存储器 41、输入装置 42 以及输出装置 43;样本聚类设备中处理器 40 的数量可以是一个或多个,图 10 中以一个处理器 40 为例;样本聚类设备中的处理器 40、存储器 41、输入装置 42 以及输出装置 43 可以通过总线或其他方式连接,图 10 中通过总线连接为例。

存储器 41 作为一种计算机可读存储介质,可用于存储软件程序、计算机可执行程序以及模块,如本公开实施例中的样本聚类方法对应的程序指令/模块(例如,样本聚类装置中的距离统计模块 301、距离获取模块 302、均值计算模块 303、连接确定模块 304 和样本聚类模块 305)。处理器 40 通过运行存储在存储器 41 中的软件程序、指令以及模块,从而执行样本聚类设备的各种功能应用以及数据处理,即实现上述的样本聚类方法。

存储器 41 可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序;存储数据区可存储根据样本聚类设备的使用所创建的数据等。此外,存储器 41 可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中,存储器 41 可包括相对于处理器 40 远程设置的存储器,这些远程存储器可以通过网络连接至样本聚类设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

输入装置 42 可用于接收输入的数字或字符信息,以及产生与样本聚类设备的用户设置以及功能控制有关的键信号输入。输出装置 43 可包括显示屏等显示设备。

上述样本聚类设备包含样本聚类装置,可以用于执行任意样本聚类方法,具备相应的功能和有益效果。

实施例五

本公开实施例还提供一种包含计算机可执行指令的存储介质,所述计算机可执行指令在由计算机处理器执行时用于执行一种样本聚类方法,该方法包括:

统计样本集中每个样本对应的第一样本距离,所述第一样本距离为所述样本与所述样本的第 S 近邻样本之间的距离;

在全部所述第一样本距离中,获取设定距离范围内的第一样本距离;

基于所述设定距离范围内的第一样本距离计算距离均值;

基于每个所述样本对应的 K 近邻样本集合,确定每个样本的全部连接样本,其中, $K > S$,所述样本与所述样本的连接样本为互为近邻样本且存在连接关系;

根据所述连接样本、所述距离均值和 S 值对所述样本集中的样本进行聚类，所述距离均值为扫描半径，所述 S 值为聚类最小包含样本数。

当然，本公开实施例所提供的一种包含计算机可执行指令的存储介质，其计算机可执行指令不限于如上所述的方法操作，还可以执行本公开任意实施例所提供的样本聚类方法中的相关操作。

通过以上关于实施方式的描述，所属领域的技术人员可以清楚地了解到，本公开可借助软件及必需的通用硬件来实现，当然也可以通过硬件实现，但很多情况下前者是更合适的实施方式。基于这样的理解，本公开的技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如计算机的软盘、只读存储器 (Read-Only Memory, ROM)、随机存取存储器 (Random Access Memory, RAM)、闪存 (FLASH)、硬盘或光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）执行本公开各个实施例所述的方法。

值得注意的是，上述样本聚类装置的实施例中，所包括的各个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，各功能单元的具体名称也只是为了便于相互区分，并不用于限制本公开的保护范围。

注意，上述仅为本公开的部分实施例及所运用技术原理。

1、一种样本聚类方法，包括：

统计样本集中每个样本对应的第一样本距离，所述第一样本距离为所述样本与所述样本的第 S 近邻样本之间的距离；

在全部所述第一样本距离中，获取设定距离范围内的第一样本距离；

基于所述设定距离范围内的第一样本距离计算距离均值；

基于每个所述样本对应的 K 近邻样本集合，确定每个样本的全部连接样本，其中， $K > S$ ，所述样本与所述样本的连接样本为互为近邻样本且存在连接关系；

根据所述连接样本、所述距离均值和 S 值对所述样本集中的样本进行聚类，所述距离均值为扫描半径，所述 S 值为聚类最小包含样本数。

2、根据权利要求 1 所述的样本聚类方法，其中，所述根据所述连接样本、所述距离均值和 S 值对所述样本集中的样本进行聚类包括：

基于所述距离均值对全部所述连接样本进行过滤，以滤除第二样本距离大于所述距离均值的连接样本，所述第二样本距离为样本与所述样本的连接样本之间的距离；

基于 S 值和过滤后得到的连接样本对所述样本集中的样本进行聚类。

3、根据权利要求 2 所述的样本聚类方法，其中，所述基于 S 值和过滤后得到的连接样本对所述样本集中的样本进行聚类包括：

依次统计每个样本的连接样本总数量；

将所述连接样本总数量大于 S 值的样本作为核心样本；

在得到的全部核心样本中，选择任一核心样本作为当前样本；

访问所述当前样本的全部连接样本；

将访问得到的每个连接样本分别作为顶点，并访问所述顶点对应的全部连接样本；

重复将访问得到的每个连接样本分别作为顶点，并访问所述顶点对应的全部连接样本的操作，直到访问不到新的连接样本为止；

将未被访问过的任一核心样本更新为当前样本，并返回执行访问所述当前样本的全部连接样本的操作，直到全部核心样本均被访问为止；

将所述当前样本及基于当前样本访问得到的连接样本聚类为簇。

4、根据权利要求 1 所述的样本聚类方法，其中，所述基于每个所述样本对应的 K 近邻样本集合，确定每个样本的全部连接样本包括：

获取每个样本对应的 K 近邻样本集合；

根据全部所述 K 近邻样本集合，构建邻接矩阵，所述邻接矩阵中每个元素代表对应两个样本间的近邻关系；

统计所述邻接矩阵中非零元素，以确定每个样本的全部连接样本。

5、根据权利要求 4 所述的样本聚类方法，其中，所述统计所述邻接矩阵中非零元素，以确定每个样本的全部连接样本包括：

在所述邻接矩阵中，获取处于对称位置的元素组，所述元素组包括第 i 行第 j 列的第一元素和第 j 行第 i 列的第二元素；

若所述第一元素和所述第二元素中包含至少一个零元素，则将所述第一元素和第二元素均设置为零元素；

遍历所述邻接矩阵的全部元素组后，更新所述邻接矩阵；

统计更新后的邻接矩阵中非零元素，并将所述非零元素对应的两个样本确定为互为近邻样本且具有连接关系；

基于所述互为近邻样本，得到每个样本的全部连接样本。

6、根据权利要求1所述的样本聚类方法，其中，所述在全部所述第一样本距离中，获取设定距离范围内的第一样本距离包括：

基于全部所述第一样本距离，构建频数分布直方图；

统计所述频数分布直方图中各bin的频数，以确定设定距离范围；

获取设定距离范围内的第一样本距离。

7、根据权利要求6所述的样本聚类方法，其中，所述统计所述频数分布直方图中各bin的频数，以确定设定距离范围包括：

获取所述频数分布直方图中频数最大bin；

计算相邻后位bin之间的频数落差，所述后位bin为所述频数分布直方图中位于频数最大bin后方的bin；

确认频数落差最大的相邻后位bin，并在所述最大的相邻后位bin中选择位于后方的bin；

将所述频数最大bin对应的第一样本距离和所述位于后方的bin对应的第一样本距离作为设定距离范围的距离阈值。

8、根据权利要求1所述的样本聚类方法，其中，所述基于所述设定距离范围内的第一样本距离计算距离均值包括：

获取所述第一样本距离处于所述设定距离范围内的样本数量；

对所述设定距离范围内每个第一样本距离进行相加，以得到样本总距离；

将所述样本总距离与所述样本数量的商值作为距离均值。

9、根据权利要求1所述的样本聚类方法，所述统计样本集中每个样本对应的第一样本距离之前，还包括：

构建样本集中每个样本的K近邻图，所述K近邻图中每条边的权值为对应样本间的距离。

10、一种样本聚类装置，包括：

距离统计模块，用于统计样本集中每个样本对应的第一样本距离，所述第一样本距离为所述样本与所述样本的第S近邻样本之间的距离；

距离获取模块，用于在全部所述第一样本距离中，获取设定距离范围内的第一样本距离；

均值计算模块，用于基于所述设定距离范围内的第一样本距离计算距离均值；

连接确定模块，用于基于每个所述样本对应的K近邻样本集合，确定每个样本的全部连接样本，其中， $K > S$ ，所述样本与所述样本的连接样本为互为近邻样本且存在连接关系；

样本聚类模块，用于根据所述连接样本、所述距离均值和S值对所述样本集中的样本进行聚类，所述距离均值为扫描半径，所述S值为聚类最小包含样本数。

11、一种样本聚类设备，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-9中任一所述的样本聚类方法。

12、一种计算机可读存储介质，其上存储有计算机程序，其中，该程序被处理器执行时实现如权利要求1-9中任一所述的样本聚类方法。

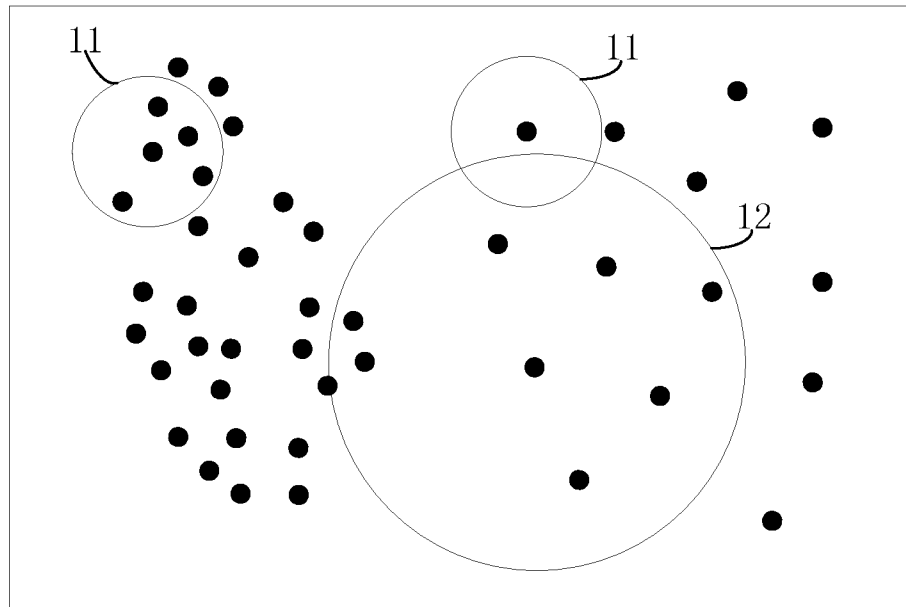


图 1

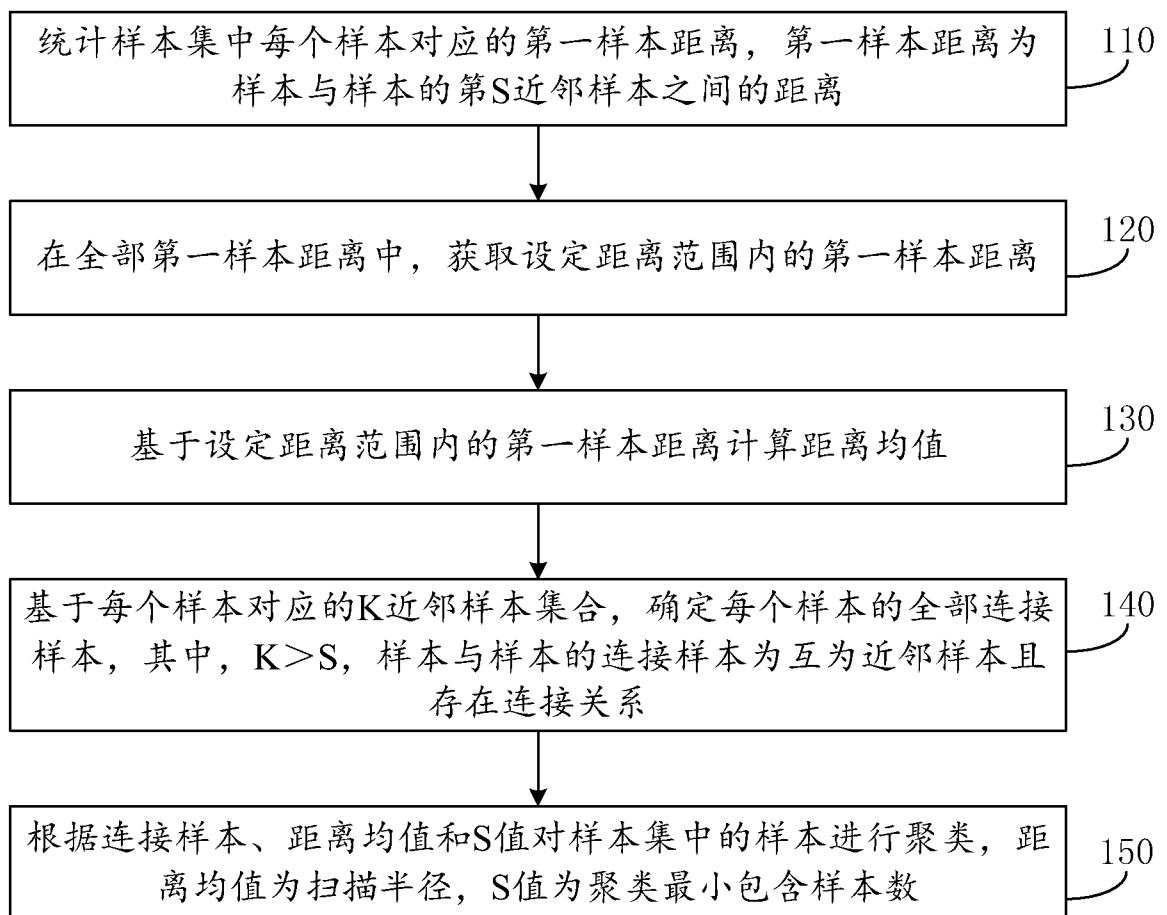


图 2

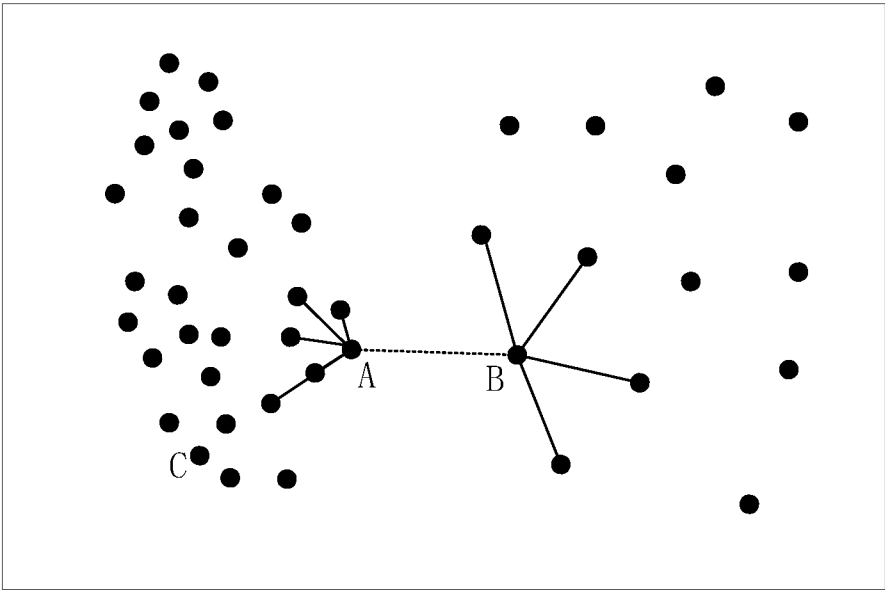


图 3

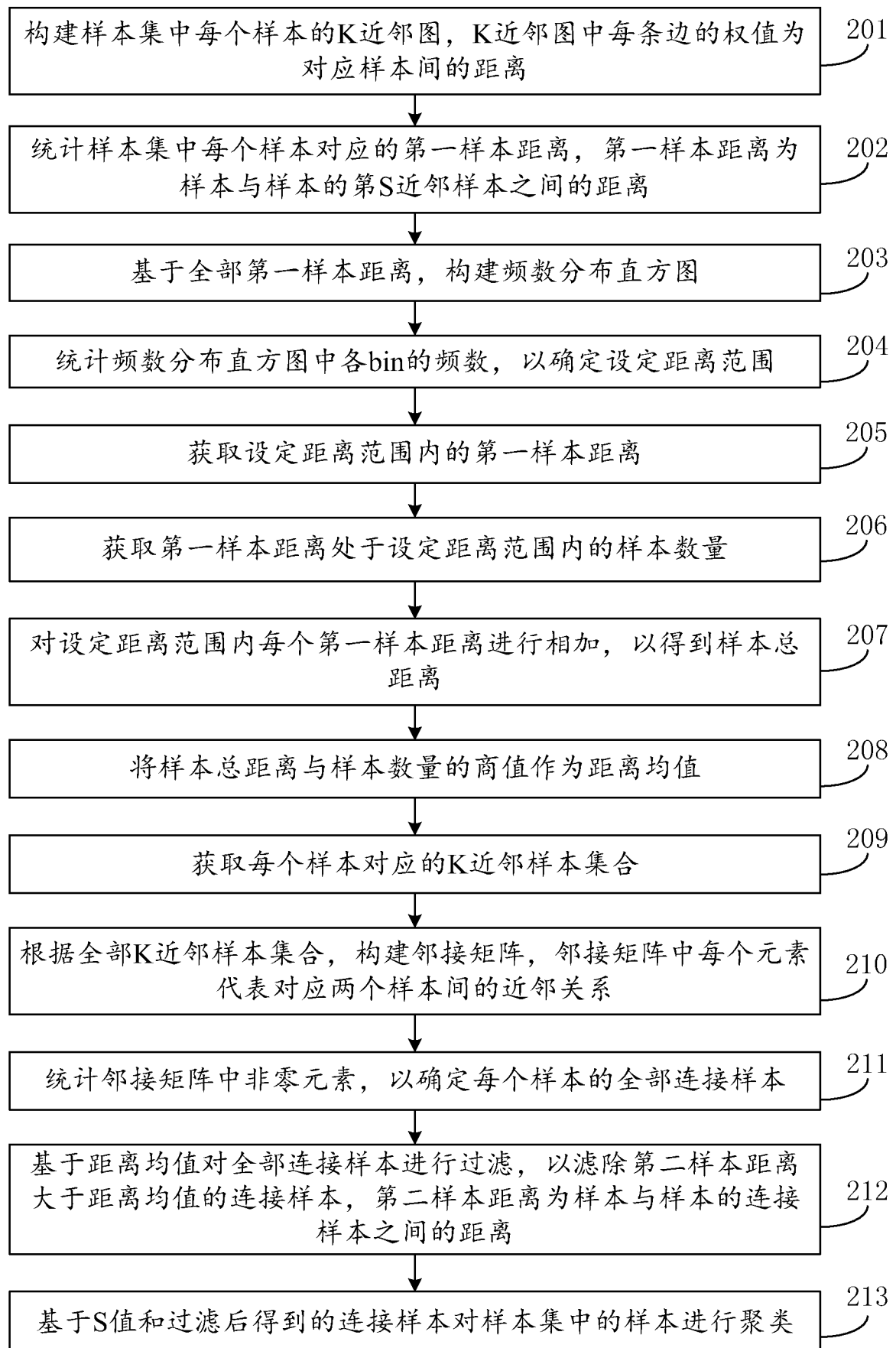


图 4

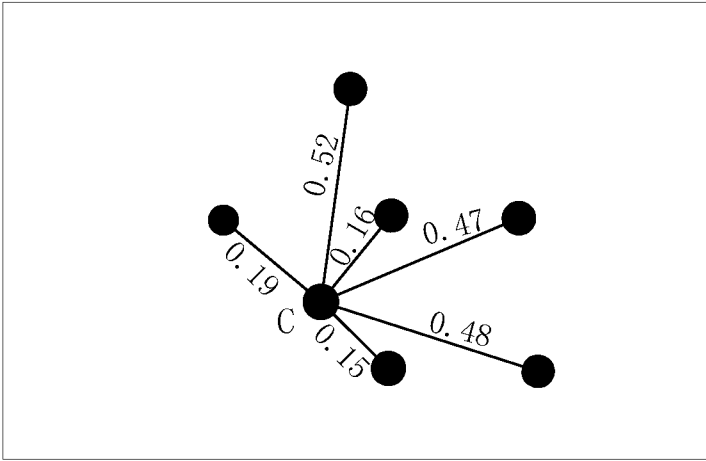


图 5

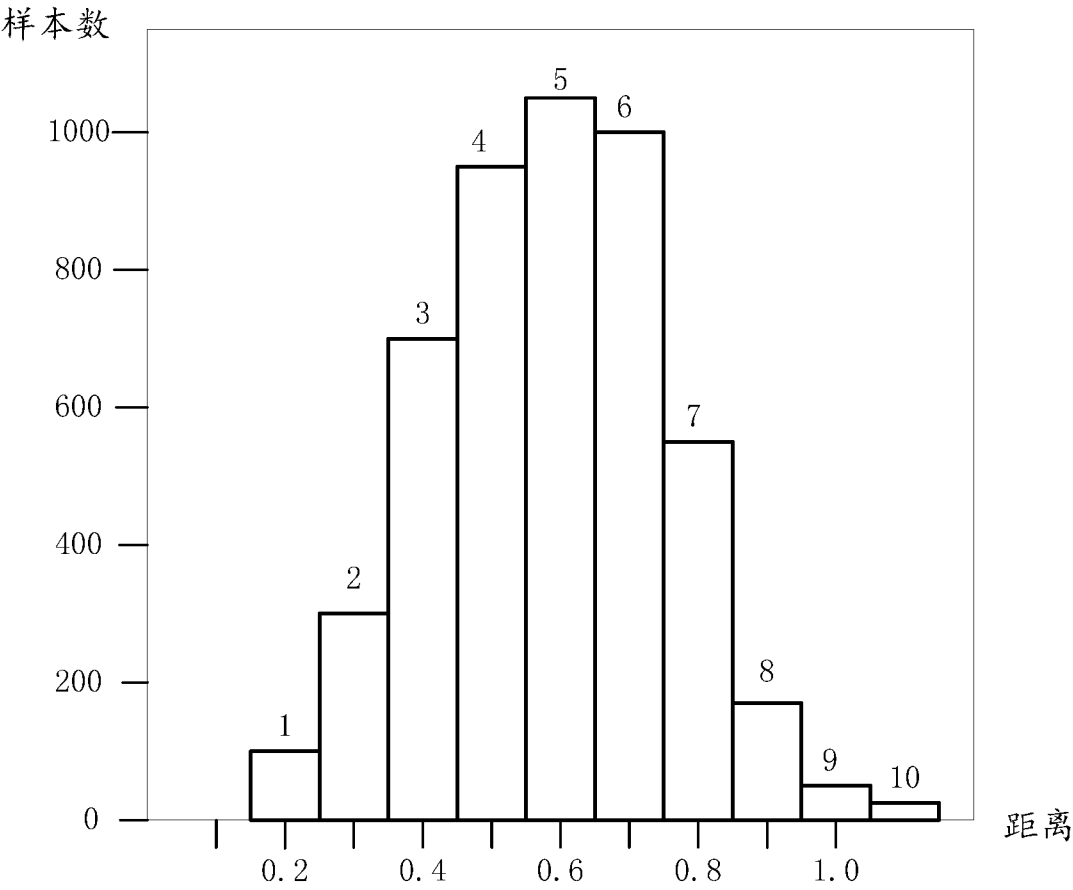


图 6

	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	0	1
3	1	1	1	1	1	0	0	1
4	1	1	0	1	1	1	0	1
5	1	1	1	0	1	0	1	1
6	1	0	0	1	1	1	1	0
7	0	1	1	0	1	0	1	1
8	0	0	1	0	1	1	0	1

图 7

	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	0	0
2	1	1	1	1	1	0	0	0
3	1	1	1	0	1	0	0	1
4	1	1	0	1	0	1	0	0
5	1	1	1	0	1	0	1	1
6	1	0	0	1	0	1	0	0
7	0	0	0	0	1	0	1	0
8	0	0	1	0	1	0	0	1

图 8

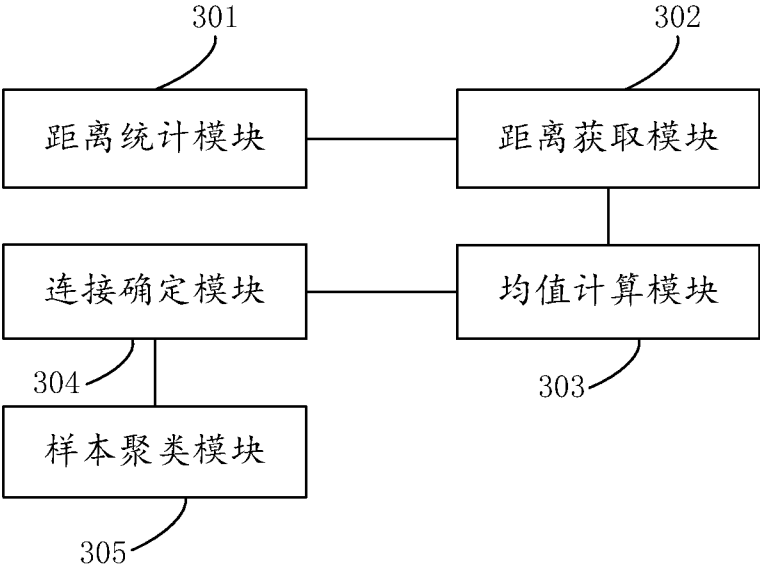


图 9

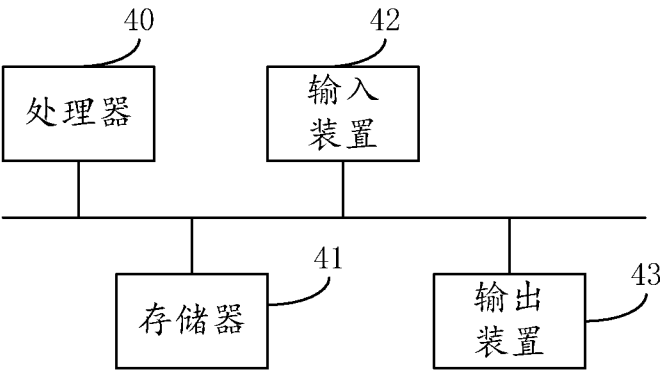


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/126716

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; VEN; USTXT; WOTXT; EPTXT; CNKI: 聚类, 距离, 近邻, 邻近, 互相, 互为, 连接, 半径, 密度, clustering, distance, neighbor, proximity, mutual, each other, connect, radius, density

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110276401 A (GUANGZHOU CVT ELECTRONICS TECHNOLOGY CO., LTD.) 24 September 2019 (2019-09-24) entire document	1-12
X	CN 103902655 A (XIAOMI TECHNOLOGY CO., LTD.) 02 July 2014 (2014-07-02) description, paragraphs 1-86, 100-238	1-12
A	CN 105930856 A (SHENZHEN YITONG TECHNOLOGY CO., LTD.) 07 September 2016 (2016-09-07) entire document	1-12
A	CN 108776806 A (HOHAI UNIVERSITY) 09 November 2018 (2018-11-09) entire document	1-12
A	US 2011055212 A1 (TSAI, Cheng-Fa et al.) 03 March 2011 (2011-03-03) entire document	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

28 February 2020

Date of mailing of the international search report

25 March 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/126716

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	110276401	A	24 September 2019	None		
CN	103902655	A	02 July 2014	EP	2913780 A2	02 September 2015
				KR	20150111343 A	05 October 2015
				EP	2913780 A3	30 September 2015
				KR	101646887 B1	08 August 2016
				BR	112015002059 A2	04 July 2017
				CN	103902655 B	04 January 2017
				MX	2015001410 A	27 April 2016
				WO	2015127801 A1	03 September 2015
				JP	6104454 B2	29 March 2017
				JP	2016520896 A	14 July 2016
				US	2015248472 A1	03 September 2015
CN	105930856	A	07 September 2016	None		
CN	108776806	A	09 November 2018	None		
US	2011055212	A1	03 March 2011	TW	201109949 A	16 March 2011
				TW	I385544 B	11 February 2013
				US	8171025 B2	01 May 2012

国际检索报告

国际申请号

PCT/CN2019/126716

A. 主题的分类

G06K 9/62 (2006.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;VEN;USTXT;WOTXT;EPTXT;CNKI: 聚类, 距离, 近邻, 邻近, 互相, 互为, 连接, 半径, 密度, clustering, distance, neighbor, proximity, mutual, each other, connect, radius, density

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 110276401 A (广州视源电子科技有限公司) 2019年 9月 24日 (2019 - 09 - 24) 全文	1-12
X	CN 103902655 A (小米科技有限责任公司) 2014年 7月 2日 (2014 - 07 - 02) 说明书第1-86、100-238段	1-12
A	CN 105930856 A (深圳市颐通科技有限公司) 2016年 9月 7日 (2016 - 09 - 07) 全文	1-12
A	CN 108776806 A (河海大学) 2018年 11月 9日 (2018 - 11 - 09) 全文	1-12
A	US 2011055212 A1 (TSAI, Cheng-Fa等) 2011年 3月 3日 (2011 - 03 - 03) 全文	1-12

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 2月 28日

国际检索报告邮寄日期

2020年 3月 25日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

邢丽超

电话号码 86-(20)-28950382

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/126716

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110276401	A	2019年 9月 24日	无			
CN	103902655	A	2014年 7月 2日	EP	2913780	A2	2015年 9月 2日
				KR	20150111343	A	2015年 10月 5日
				EP	2913780	A3	2015年 9月 30日
				KR	101646887	B1	2016年 8月 8日
				BR	112015002059	A2	2017年 7月 4日
				CN	103902655	B	2017年 1月 4日
				MX	2015001410	A	2016年 4月 27日
				WO	2015127801	A1	2015年 9月 3日
				JP	6104454	B2	2017年 3月 29日
				JP	2016520896	A	2016年 7月 14日
				US	2015248472	A1	2015年 9月 3日
CN	105930856	A	2016年 9月 7日	无			
CN	108776806	A	2018年 11月 9日	无			
US	2011055212	A1	2011年 3月 3日	TW	201109949	A	2011年 3月 16日
				TW	1385544	B	2013年 2月 11日
				US	8171025	B2	2012年 5月 1日