

(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
2018 年 1 月 25 日 (25.01.2018)



(10) 国际公布号
WO 2018/014717 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2017/091432

(22) 国际申请日: 2017 年 7 月 3 日 (03.07.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610586139.8 2016 年 7 月 22 日 (22.07.2016) CN

(71) 申请人: 北京市商汤科技开发有限公司(BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD) [CN/CN]; 中国北京市海淀区中关村东路 1 号院 3 号楼 7 层 710-712 房间程宝妹, Beijing 100084 (CN)。

(72) 发明人: 潘薪宇 (PAN, Xinyu); 中国北京市海淀区中关村东路 1 号院 3 号楼 7 层 710-712 房间潘薪宇, Beijing 100084 (CN)。

(74) 代理人: 北京思元知识产权代理事务所 (普通合伙) (BEIJING SIYUAN INTELLECTUAL PROPERTY AGENCY); 中国北京市海淀区永泰中路 25 号中关村永泰创新园 B 座 226 余光军, Beijing 100192 (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, 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,

(54) Title: METHOD AND DEVICE FOR CLUSTERING AND ELECTRONIC EQUIPMENT

(54) 发明名称: 聚类方法、装置及电子设备

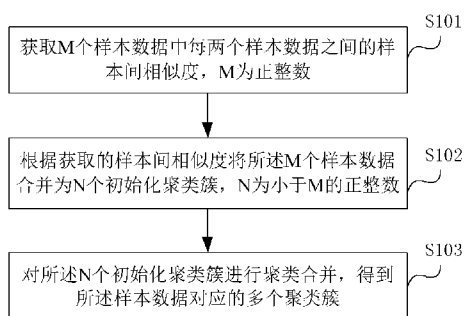


图 5

- S101 Acquire the inter-sample degree of similarity between every two pieces of sample data in M pieces of sample data, M being a positive integer
- S102 Merge the M pieces of sample data into N initialized clusters on the basis of the acquired inter-sample degree of similarity, N being a positive integer less than M
- S103 Cluster merge the N initialized clusters to produce multiple clusters corresponding to the sample data

(57) Abstract: A method and device for clustering and electronic equipment. The method comprises: acquiring the inter-sample degree of similarity between every two pieces of sample data in M pieces of sample data, M being a positive integer (S101); merging the M pieces of sample data into N initialized clusters on the basis of the acquired inter-sample degree of similarity, N being a positive integer less than M (S102); and cluster merging the N initialized clusters to produce multiple clusters corresponding to the M pieces of sample data (S103).

(57) 摘要: 一种聚类方法、装置及电子设备, 该方法包括: 获取M个样本数据中每两个样本数据之间的样本间相似度, M为正整数 (S101); 根据获取的样本间相似度将所述M个样本数据合并为N个初始化聚类簇, N为小于M的正整数 (S102); 对所述N个初始化聚类簇进行聚类合并, 得到所述M个样本数据对应的多个聚类簇 (S103)。

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))。

聚类方法、装置及电子设备

本公开要求在 2016 年 7 月 22 日提交中国专利局、申请号为 201610586139.8、发明名称为“聚类方法、装置及电子设备”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

5

技术领域

本发明涉及数据处理技术领域，特别涉及一种聚类方法、装置及电子设备。

背景技术

10 生活中常常存在大量错综复杂的数据，为了便于管理或者查找这些数据等因素，往往需要对这些数据进行聚类，将具有相同或相似特征的数据聚为一类，如：将包含花朵的图像聚为一类等，这样，用户就可以按照类别查找或管理需要的数据。如何快速准确的将具有相同或相似特征的数据聚为一类，是非常值得关注的。

15 发明内容

本公开公开一种聚类技术方案。

本公开公开了一种聚类方法，所述方法包括：获取 M 个样本数据中每两个样本数据之间的样本间相似度， M 为正整数；根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇， N 为小于 M 的正整数；对所述 N 个初始化聚类簇进行聚类合并，得到所述 M 个样本数据对
20 应的多个聚类簇。

在本公开的一种可选的实现方式中，所述根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，包括：所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；针对得到的多个疑似初始化聚类簇，将包含相同样本数据的个数不小于第二阈值的至少二个疑似初始化聚类簇合并为一初始化聚类簇。
25

在本公开的一种可选的实现方式中，所述根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，包括：所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；针对包含相同样本数据的个数不小于第二阈值的第一疑似初始化聚类簇和第二疑似初始化聚类簇，计算第一和第二疑似初始化聚类簇的交集的模，并计算第一初始化聚类簇的模与第二初始化聚类簇的模之和与第一常数的商，在所述交集的模与所述商之差不小于第二常数时，将所述
30 第一和第二疑似初始化聚类簇合并为一疑似初始化聚类簇。

在本公开的一种可选的实现方式中，所述根据获取的所述样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，还包括：针对得到的多个疑似初始化聚类簇，将不能与其他疑似
35 初始化聚类簇合并的各疑似初始化聚类簇分别作为一初始化聚类簇。

在本公开的一种可选的实现方式中，所述第二阈值根据所述至少二个疑似初始化聚类簇中的一个或者多个疑似初始化聚类簇所包括的样本数据的个数的线性函数确定。

在本公开的一种可选的实现方式中，所述方法还包括：对所述多个聚类簇中的至少一个聚类簇进行离群点分离，且所述离群点分离处理后获得的所有聚类簇被作为所述 M 个样本数据对应的多个聚类簇。

在本公开的一种可选的实现方式中，所述离群点分离包括：针对任一聚类簇，获得所述聚类簇内每一样本数据对应的待离群簇和非待离群簇，其中，每一样本数据对应的所述待离群簇中包括所述样本数据，所述非待离群簇中包括：所述聚类簇中除所述样本数据外的其它样本数据；获取每一样本数据对应的待离群簇与非待离群簇之间的簇间相似度；确定所述聚类簇中所有样本数据分别对应的多个簇间相似度中最小的簇间相似度是否小于第三阈值；响应于所述最小的簇间相似度小于所述第三阈值，将所述最小的簇间相似度对应的待离群簇和非待离群簇分别作为两个新的聚类簇。

在本公开的一种可选的实现方式中，所述对所述 N 个初始化聚类簇进行聚类合并，包括：将所述 N 个初始化聚类簇作为多个待聚类簇；获取每个待聚类簇与其他待聚类簇之间的簇间相似度；确定所述多个待聚类簇对应的所有簇间相似度中的最大簇间相似度是否大于第四阈值；响应于所述最大簇间相似度大于所述第四阈值，将所述最大簇间相似度对应的两个待聚类簇进行合并得到一新的待聚类簇；对所述新的待聚类簇与本次未合并的其它待聚类簇构成的新的多个待聚类簇继续进行聚类合并，直到没有可合并的待聚类簇。

在本公开的一种可选的实现方式中，所述 M 个样本数据为 M 个图像，且两个图像之间的样本间相似度包括：所述两个图像分别对应的两个特征向量之间的余弦距离。

本公开还提供了一种聚类装置，其特征不在于，所述装置包括：

获取单元，用于获取 M 个样本数据中每两个样本数据之间的样本间相似度，M 为正整数；合并单元，用于根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，N 为小于 M 的正整数；聚类单元，用于对所述 N 个初始化聚类簇进行聚类合并，得到所述 M 个样本数据对应的多个聚类簇。

在本公开的一种可选的实现方式中，所述合并单元包括：第一合并子单元，用于所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；第二合并子单元，用于针对得到的多个疑似初始化聚类簇，将包含相同样本数据的个数不小于第二阈值的至少二个疑似初始化聚类簇合并为一初始化聚类簇。

在本公开的一种可选的实现方式中，所述合并单元具体用于：所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；针对包含相同样本数据的个数不小于第二阈值的第一疑似初始化聚类簇和第二疑似初始化聚类簇，计算第一和第二疑似初始化聚类簇的交集的模，并计算第一初始化聚类簇的模与第二初始化聚类簇的模之和与第一常数的商，在所述交集的模与所述商之差不小于第二常数时，将所述第一和第二疑似初始化聚类簇合并为一疑似初始化聚类簇。

在本公开的一种可选的实现方式中，所述合并单元还包括：第一作为子单元，用于针对得到的多个疑似初始化聚类簇，将不能与其他疑似初始化聚类簇合并的各疑似初始化聚类簇分别作为一初始化聚类簇。

在本公开的一种可选的实现方式中，所述第二阈值根据所述至少二个疑似初始化聚类簇中的一个或者多个疑似初始化聚类簇所包括的样本数据的个数的线性函数确定。

在本公开的一种可选的实现方式中，所述装置还包括：

5 分离单元，用于对所述多个聚类簇中的至少一个聚类簇进行离群点分离，且所述离群点分离处理后获得的所有聚类簇被作为所述 M 个样本数据对应的多个聚类簇。

在本公开的一种可选的实现方式中，所述分离单元具体用于对一所述聚类簇进行离群点分离；所述分离单元包括：

10 获得子单元，针对任一聚类簇，用于获得所述聚类簇内每一样本数据对应的待离群簇和非待离群簇，其中，每一样本数据对应的所述待离群簇中包括所述样本数据，所述非待离群簇中包括：所述聚类簇中除所述样本数据外的其它样本数据；

第一获取子单元，用于获取每一样本数据对应的待离群簇与非待离群簇之间的簇间相似度；

第一确定子单元，用于确定所述聚类簇中所有样本数据分别对应的多个簇间相似度中最小的簇间相似度是否小于第三阈值；

15 第一响应子单元，用于响应于所述最小的簇间相似度小于所述第三阈值，将所述最小的簇间相似度对应的待离群簇和非待离群簇分别作为两个新的聚类簇。

在本公开的一种可选的实现方式中，所述聚类单元包括：

第二作为子单元，用于将所述 N 个初始化聚类簇作为多个待聚类簇；

第二获取子单元，用于获取每个待聚类簇与其他待聚类簇之间的簇间相似度；

20 第二确定子单元，用于确定所述多个待聚类簇对应的所有簇间相似度中的最大簇间相似度是否大于第四阈值；

第二响应子单元，用于响应于所述最大簇间相似度大于所述第四阈值，将所述最大簇间相似度对应的两个待聚类簇进行合并得到一新的待聚类簇，并触发所述第二获取子单元，对所述新的待聚类簇与本次未合并的其它待聚类簇构成的新的多个待聚类簇继续进行聚类合并，直到没有可合并的待聚类簇。

25 在本公开的一种可选的实现方式中，所述 M 个样本数据为 M 个图像，且两个图像之间的样本间相似度包括：所述两个图像分别对应的两个特征向量之间的余弦距离。

30 本公开还公开了一种电子设备，包括：壳体、处理器、存储器、电路板和电源电路，其中，所述电路板安置在所述壳体围成的空间内部，所述处理器和所述存储器设置在所述电路板上；所述电源电路，用于为终端的各个电路或器件供电；所述存储器用于存储可执行程序代码；所述处理器通过读取所述存储器中存储的可执行程序代码来运行与可执行程序代码对应的程序，以用于执行上述聚类方法对应的操作。

本公开还公开了一种非暂时性计算机存储介质，该介质可存储计算机可读指令，当这些指令被执行时可使处理器执行上述聚类方法对应的操作。

35 本公开中，对样本数据进行聚类时，先根据多个样本数据中每两个样本数据之间的样本间相似度对样本数据进行合并，获得初始化聚类簇，减少聚类时初始化聚类簇的数量，根据此时的初始化聚类簇进行聚类合并，得到所述多个样本数据对应的多个聚类簇，有利于提高聚类速度。

附图说明

为了更清楚地说明本公开或现有技术中的技术方案，下面将对本公开或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本公开的一些实施例，
5 对于本领域普通技术人员来讲，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的附图。

图 1 示出了本公开的一应用场景示意图；

图 2 示出了本公开的另一应用场景示意图；

图 3 示出了实现本公开的一示例性设备的框图；

10 图 4 示出了实现本公开的另一示例性设备的框图；

图 5 为本公开提供的聚类方法的流程示意图；

图 6 为本公开提供的初始聚类簇之间的簇间相似度示意图；

图 7 为本公开提供的初始化聚类簇聚类合并方法的流程示意图；

图 8 为本公开提供的簇间相似度的确定方法的流程示意图；

15 图 9 为本公开提供的聚类装置的结构示意图；

图 10 为本公开提供的聚类装置的结构示意图；

图 11 为本公开提供的电子设备的结构示意图。

具体实施例

20 下面将结合本公开中的附图，对本公开中的技术方案进行清楚、完整地描述，显然，所描述的技术方案仅仅是本公开一部分技术方案，而不是全部的技术方案。基于本公开中的技术方案，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

图 1 示意性地示出了根据本公开提供的聚类技术方案可以在其中实现的一应用场景。

25 图 1 中，用户的智能移动电话 1000 中存储有多张图片，例如，用户的智能移动电话 1000 中存储的图片包括：第一图片 1001、第二图片 1002、第三图片 1003、……以及第 M 图片 1004 等共 M 张图片，其中，M 的大小可以由智能移动电话 1000 中存储的图片数量决定，且在通常情况下，M 为不小于 2 的正整数。存储于智能移动电话 1000 中的图片可以是用户利用智能移动电话 1000 拍摄的图片，也可以是用户将其他终端设备中存储的图片传输并存储在智能移动电话
30 1000 中的图片，还可以是用户利用智能移动电话 1000 从网络中下载的图片等。

本公开提供的技术方案可以使智能移动电话 1000 中存储的第一图片 1001、第二图片 1002、第三图片 1003、……以及第 M 图片 1004 按照预定分类方式而被自动的划分在多个集合中，例如，在预定分类方式为根据图片中的人进行划分的情况下，由于智能移动电话 1000 中存储的第一图片 1001 和第 M 图片 1004 中均包含有第一用户的肖像，而智能移动电话 1000 中存储的第二图片
35 1002 和第三图片 1003 均包含有第 N 用户的肖像，因此，本公开提供的技术方案可以使智能移动电话 1000 中存储的第一图片 1001 和第 M 图片 1004 自动划分在第一用户的图片集合中，并使智能移动电话 1000 中存储的第二图片 1002 和第三图片 1003 自动划分在第 N 用户的图片集合中，

上述 N 为小于 M 的正整数。

本公开通过将智能移动电话 1000 中存储的第一图片 1001、第二图片 1002、第三图片 1003、……以及第 M 图片 1004 自动划分在多个集合中，有利于提高图片的可管理性，从而可以为用户带来操作上的便利，例如，用户可以方便的浏览智能移动电话 1000 中存储的包含有第一

5 用户的肖像的所有图片。

图 2 示意性地示出了根据本公开提供的聚类技术方案可以在其中实现的另一应用场景。

图 2 中，电商平台的设备 2000（如计算机或者服务器等）中存储有电商平台所能提供的所有商品信息或者部分商品信息，例如，电商平台的设备 2000 中存储的商品信息包括：第一商品信息 2001、第二商品信息 2002、第三商品信息 2003、……以及第 M 商品信息 2004 等共 M 个商品信息，其中，M 的大小可以由设备 2000 中存储的商品信息的数量决定，且在通常情况下，M

10 为不小于 2 的正整数。存储于设备 2000 中的商品信息可以包括：商品的图片以及商品的文字描述等。

本公开提供的技术方案可以使设备 2000 中存储的第一商品信息 2001、第二商品信息 2002、第三商品信息 2003、……以及第 M 商品信息 2004 按照预定分类方式而被自动划分在多个集合中，

15 例如，在预定分类方式为根据商品类别进行划分的情况下，由于设备 2000 中存储的第一商品信息 2001 所对应的商品和第 M 商品信息 2004 所对应的商品均属于第一类别（如女士旅游鞋类别），而设备 2000 中存储的第二商品信息 2002 所对应的商品和第三商品信息 2003 所对应的商品均属于第 N 类别（如婴儿奶粉类别），因此，本公开提供的技术方案可以使设备 2000 中存储的第一商品信息 2001 和第 M 商品信息 2004 自动划分在第一类别（如女士旅游鞋类别）集合中，并使

20 设备 2000 中存储的第二商品信息 2002 和第三商品信息 2003 自动划分在第 N 类别（如婴儿奶粉类别）集合中，上述 N 为小于 M 的正整数。

本公开通过将设备 2000 中存储的第一商品信息 2001、第二商品信息 2002、第三商品信息 2003、……以及第 M 商品信息 2004 划分在多个集合中，有利于提高商品信息的可管理性，从而可以为电商平台以及消费者带来操作上的便利，例如，在消费者查看设备 2000 中存储的第一商

25 品信息 2001 的情况下，电商平台可以将第一商品信息 2001 所在的第一类别集合中的其他商品信息（例如，第 M 商品信息）推荐展示给用户。

然而，本领域技术人员可以理解，本公开还可以适用于其他应用场景中，即本公开所能够适用的应用场景并不会受上述例举的两个应用场景的限制。

下面结合附图通过具体的实施例对本公开的聚类技术方案进行详细介绍。

30 图 3 示出了适于实现本公开的示例性设备 3000（例如，计算机系统/服务器）的框图。图 3 显示的设备 3000 仅仅是一个示例，不应对本公开的功能和使用范围带来任何限制。

如图 3 所示，设备 3000 可以以通用计算设备的形式表现。设备 3000 的组件可以包括但不限于：一个或者多个处理器或者处理单元 3001，系统存储器 3002，连接不同系统组件（包括系统存储器 3002 和处理单元 3001）的总线 3003。设备 3000 可以包括多种计算机系统可读介质。

35 这些介质可以是任何能够被设备 3000 访问的可用介质，包括易失性和非易失性介质，可移动的和不可移动的介质等。

系统存储器 3002 可以包括易失性存储器形式的计算机系统可读介质，例如，随机存取存储

器 (RAM) 3021 和/或高速缓存存储器 3022。设备 3000 可以进一步包括其他可移动的/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例, ROM 3023 可以用于读写不可移动的、非易失性磁介质 (图 3 中未显示, 通常称为“硬盘驱动器”)。尽管未在图 3 中示出, 可以提供用于对可移动非易失性磁盘 (例如“软盘”) 读写的磁盘驱动器, 以及对可移动非易失性光盘 (例如 CD-ROM, DVD-ROM 或者其他光介质) 读写的光盘驱动器。在这些情况下, 每个驱动器可以通过一个或者多个数据介质接口与总线 3003 相连。系统存储器 3002 中可以包括至少一个程序产品, 该程序产品具有一组 (例如至少一个) 程序模块, 这些程序模块被配置以执行本公开的功能。

具有一组 (至少一个) 程序模块 3024 的程序/实用工具 3025, 可以存储在例如系统存储器 3002 中, 这样的程序模块 3024 包括但不限于: 操作系统、一个或者多个应用程序、其他程序模块以及程序数据, 这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块 3024 通常执行本公开所描述的功能和/或方法。

设备 3000 也可以与一个或多个外部设备 3004 (如键盘、指向设备、显示器等) 通信。这种通信可以通过输入/输出 (I/O) 接口 3005 进行, 并且, 设备 3000 还可以通过网络适配器 3006 与一个或者多个网络 (例如局域网 (LAN), 广域网 (WAN) 和/或者公共网络, 例如因特网) 通信。如图 3 所示, 网络适配器 3006 通过总线 3003 与设备 3000 的其他模块 (如处理单元 3001 等) 通信。应当明白, 尽管图 3 中未示出, 可以结合设备 3000 使用其他硬件和/或软件模块。

处理单元 3001 (即处理器) 通过运行存储在系统存储器 3002 中的计算机程序, 从而执行各种功能应用以及数据处理, 例如, 执行用于实现上述方法中的各步骤的指令; 具体而言, 处理单元 3001 可以执行系统存储器 3002 中存储的计算机程序, 且该计算机程序被执行时, 下述步骤被实现: 获取 M 个样本数据中每两个样本数据之间的样本间相似度, M 为正整数; 根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇, N 为小于 M 的正整数; 对所述 N 个初始化聚类簇进行聚类合并, 得到所述 M 个样本数据对应的多个聚类簇。

图 4 示出了适于实现本公开的示例性设备 4000, 设备 4000 可以是移动终端、个人计算机 (PC)、平板电脑以及服务器等。图 4 中, 设备 4000 包括一个或者多个处理器、通信部等, 所述一个或者多个处理器可以为: 一个或者多个中央处理单元 (CPU) 4001, 和/或, 一个或者多个图像处理器 (GPU) 4130 等, 处理器可以根据存储在只读存储器 (ROM) 4002 中的可执行指令或者从存储部分 4008 加载到随机访问存储器 (RAM) 4003 中的可执行指令而执行各种适当的动作和处理。通信部 4120 可以包括但不限于网卡, 所述网卡可以包括但不限于 IB (Infiniband) 网卡。处理器可与只读存储器 4002 和/或随机访问存储器 4300 中通信以执行可执行指令, 通过总线 4004 与通信部 4120 相连、并经通信部 4120 与其他目标设备通信, 从而完成本公开中的相应操作。处理器所执行的操作包括: 获取 M 个样本数据中每两个样本数据之间的样本间相似度, M 为正整数; 根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇, N 为小于 M 的正整数; 对所述 N 个初始化聚类簇进行聚类合并, 得到所述 M 个样本数据对应的多个聚类簇。

此外, 在 RAM 4003 中, 还可以存储有装置操作所需的各种程序以及数据。CPU4001、ROM4002 以及 RAM4003 通过总线 4004 彼此相连。在有 RAM4003 的情况下, ROM4002 为可选模块。RAM4003 存储可执行指令, 或在运行时向 ROM4002 中写入可执行指令, 可执行指令使中央处理单元 4001

执行上述通信方法对应的操作。输入/输出 (I/O) 接口 4005 也连接至总线 4004。通信部 4120 可以集成设置,也可以设置为具有多个子模块(例如,多个 IB 网卡),并分别与总线连接。

以下部件连接至 I/O 接口 4005:包括键盘、鼠标等的输入部分 4006;包括诸如阴极射线管 (CRT)、液晶显示器 (LCD) 等以及扬声器等的输出部分 4007;包括硬盘等的存储部分 4008;
5 以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 4009。通信部分 4009 经由诸如因特网的网络执行通信处理。驱动器 4100 也根据需要连接至 I/O 接口 4005。可拆卸介质 4110,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器 4100 上,以便于从其上读出的计算机程序根据需要被安装入存储部分 4008。

需要说明的,如图 4 所示的架构仅为一种可选实现方式,在具体实践过程中,可根据实际
10 需要对上述图 4 的部件数量和类型进行选择、删减、增加或替换;在不同功能部件设置上,也可采用分离设置或集成设置等实现方式,例如,GPU 和 CPU 可分离设置,再如理,可将 GPU 集成在 CPU 上,通信部可分离设置,也可集成设置在 CPU 或 GPU 上等。这些可替换的实施方式均落入本公开的保护范围。

特别地,根据本公开的实施方式,下文参考流程图描述的过程可以被实现为计算机软件程序,例如,本公开的实施方式包括一种计算机程序产品,其包括有形地包含在机器可读介质上的计算机程序,计算机程序包含用于执行流程图所示的步骤的程序代码,程序代码可包括对应
15 执行本申请提供的步骤对应的可执行指令,例如,用于获取 M 个样本数据中每两个样本数据之间的样本间相似度的可执行指令,M 为正整数;用于根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇的可执行指令,N 为小于 M 的正整数;用于对所述 N 个初始化聚类簇
20 进行聚类合并,得到所述 M 个样本数据对应的多个聚类簇的可执行指令。

在这样的实施方式中,该计算机程序可以通过通信部分 4009 从网络上被下载及安装,和/或从可拆卸介质 4110 被安装。在该计算机程序被中央处理单元 (CPU) 4001 执行时,执行本公开中记载的上述可执行指令。

下面结合图 5-图 11 对本公开提供的聚类技术方案进行说明。本公开提供的任一种聚类技术方案可由软件或者硬件或者软硬结合的方式进行示例。例如,本公开提供的聚类技术方案可由
25 某一电子设备实施或者由某一处理器实施,本公开并不限制,所述电子设备可包括但不限于终端或服务器,所述处理器可包括但不限于 CPU 或 GPU。以下不再赘述。

图 5 中,S101:获取 M 个样本数据中每两个样本数据之间的样本间相似度,M 为正整数。通常情况下,M 大于 2,如 M 可为 100 或者 1001 或者 5107 等。

30 一种可选的实现方式中,步骤 S101 可以由处理器调用存储器存储的指令执行,或者,可以由被处理器运行的获取单元 501 执行。

本公开中的样本数据可以为图像(即图片)、语音、视频或者文本等。

一种可选的实现方式中,在处理器获得样本数据后,处理器可以采用卷积神经网络或其他传统的局部特征描述子对样本数据的特征进行提取,如:当样本数据为图像时,处理器可以采用
35 Sift(Scale-invariant feature transform,尺度不变特征转换)、HOG(Histogram of Oriented Gradien,方向梯度直方图)等对图像进行特征提取,获得每个样本数据的特征向量;处理器根据提取的样本数据的特征向量构建特征矩阵。假设,当前有 n 个样本数据,提取每一样本数据

的 m 类特征，根据提取到的特征构建 $n \times m$ 阶特征矩阵。

本公开中的两个样本数据之间的样本间相似度可通过但不限于采用样本数据特征向量的余弦距离、欧式距离等表示。

例如，处理器在建立特征矩阵后，可以计算样本数据间特征的余弦距离，生成相似度矩阵，
5 这里，一个样本数据的所有特征作为一个特征向量。根据上述假设，处理器生成了 $n \times m$ 阶特征矩阵，根据该特征矩阵，处理器可以生成 $n \times n$ 阶相似度矩阵，如下所示：

$$M = \begin{pmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \ddots & \vdots \\ & M_{ij} & \\ M_{n1} & \cdots & M_{nn} \end{pmatrix}。$$

其中， M_{ij} 为第 i 个样本数据和第 j 个样本数据之间的样本间相似度，

$$M_{ij} = \cos(X, Y) = \frac{\langle X, Y \rangle}{|X| * |Y|},$$

10 X 表示第 i 个样本数据的特征向量， Y 表示第 j 个样本数据的特征向量。

相似度矩阵表征的是样本数据间的相似度度量，当两个样本数据的样本间相似度非常高（如超过某一设定阈值，或者，相似度满足一定的预定条件等）时，可以近似地认为这两个样本数据属于同一个类。另外，该相似度矩阵为对称矩阵。在本公开的输入信息为 M 个样本数据的相似度矩阵的情况下，处理器可以通过接收输入信息而获得 M 个样本数据的相似度矩阵，然而，
15 在本公开的输入信息为 M 个样本数据的情况下，处理器可以采用上述描述的方案获得 M 个样本数据的相似度矩阵。本公开不限制处理器获得 M 个样本数据的相似度矩阵的具体实现方式。

S102：根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇， N 为小于 M 的正整数。

一种可选的实现方式中，步骤 S102 可以由处理器调用存储器存储的相应指令来执行，或者，
20 可以由被处理器运行的合并单元 502 执行。其中，每个初始化聚类簇包含至少一个样本数据。

本公开在对样本数据进行聚类前，一般需要确定样本数据的初始化聚类簇。通常的做法是将每一样本数据单独确定为一个聚类簇，这些聚类簇作为初始化聚类簇，如，当前有 4 个样本数据，分别为 a 、 b 、 c 和 d ，则可以确定出 4 个初始化聚类簇。这种方法，以每一个样本数据为基础，具有良好的鲁棒性，获得的最终的聚类结果具有良好的准确率。

25 上述将每一样本数据单独确定为一个初始化聚类簇，根据确定的初始化聚类簇获得最终的聚类结果的方法的计算复杂度为 $O(n^3 \log n)$ ，其中， n 为样本数据的数量，这种方法适用于对精度要求高、而对速度性能要求不高的场景。在对处理数据的速度性能要求较高的场景中，本公开的聚类技术方案通过对样本数据进行合并，减少了初始化聚类簇的个数，有利于提高聚类的速度性能。

一种可选的实现方式中，处理器根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，可以包括下述步骤：

S01、针对一个样本数据而言，查找与该样本数据之间的样本间相似度大于第一阈值的所有样本数据，并将查找到的所有样本数据分别与该样本数据合并为一个疑似初始化聚类簇，从而

5 处理器可以获得多个疑似初始化聚类簇。

一种可选的实现方式中，步骤 S01 可以由处理器调用存储器存储的相应指令来执行，或者，可以由被处理器运行的第一合并子单元执行。

假设，当前有 5 个样本数据，分别为 a、b、c、d 和 e，且此时处理器已经获得了 a 分别与 b、c、d、e 的样本间相似度、b 分别与 a、c、d、e 的样本间相似度、c 分别与 a、b、d、e 的样本间相似度、d 分别与 a、b、c、e 的样本间相似度、e 分别与 a、b、c、d 的样本间相似度。

根据上述假设，若 a 与 b、d 的样本间相似度大于第一阈值，且 b 和 c 的样本间相似度大于第一阈值，则处理器可以将 a 与 b 合并为一个疑似初始化聚类簇 C_1 ，将 a 与 d 合并为一个疑似初始化聚类簇 C_2 ，将 b 与 c 合并为一个疑似初始化聚类簇 C_3 ，由于 e 和其他所有样本数据的样本间相似度都不大于第一阈值，因此，处理器可以将 e 作为一个疑似初始化聚类簇 C_4 。

15 值得一提的是，第一阈值可选的可以为比较高的阈值，如，在样本间相似度的取值范围为 $[0, 1]$ 的情况下，第一阈值可以为 0.9、0.95 等，由此可以尽可能保证疑似初始化聚类簇分类的准确性。另外，第一阈值可以根据实际情况进行调整，例如，在每一个疑似初始化聚类簇均包括一个样本数据的情况下，处理器可以适当降低第一阈值；再例如，在包含有两个样本数据的疑似初始化聚类簇的数量与所有疑似初始化聚类簇的数量的比值低于预定比值（例如，1/100）的情况下，处理器可以适当降低第一阈值。

S02、针对得到的多个疑似初始化聚类簇，将包含相同样本数据的个数不小于第二阈值的至少二个疑似初始化聚类簇合并为一初始化聚类簇。

一种可选的实现方式中，步骤 S02 可以由处理器调用存储器存储的相应指令来执行，或者，可以由被处理器运行的第二合并子单元执行。

25 一种可选的实现方式中，上述第二阈值可以设定为固定的正整数，例如，假设第二阈值为 1，如果两个疑似初始化聚类簇中均包含有一个相同的样本数据，则处理器将这两个疑似初始化聚类簇进行合并；另外，不能与其他疑似初始化聚类簇合并的一疑似初始化聚类簇也确定为一初始化聚类簇，例如，被处理器运行的第一作为子单元将不能与其他疑似初始化聚类簇合并的一疑似初始化聚类簇确定为一初始化聚类簇。

30 样本间相似度具有传递性，因此处理器可以将包含相同样本数据的疑似初始化聚类簇合并为一个初始化聚类簇。

如步骤 S01 中的假设，疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_2 中都包含样本数据 a，疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_3 中都包含样本数据 b，则处理器可以将疑似初始化聚类簇 C_1 、疑似初始化聚类簇 C_2 和疑似初始化聚类簇 C_3 合并为一个初始化聚类簇 C_5 ；另外，疑似初始化聚类簇 C_4 与其他疑似初始化聚类簇不包含相同样本数据，则处理器不对疑似初始化聚类簇 C_4 进行合并处理，将疑似初始化聚类簇 C_4 确定为初始化聚类簇。

本公开的聚类技术方案的计算复杂度为 $O(n^2 \log n)$ ，与上述将每一样本数据单独确定为一

个初始化聚类簇，并根据确定的初始化聚类簇获得最终的聚类结果的技术方案的计算复杂度 $O(n^3 \log n)$ 相比，在较大程度上提高了对样本数据进行聚类获得最终的聚类结果的计算速度。

本公开合并疑似初始化聚类簇的一种方式，只要两个疑似初始化聚类簇所包含的相同样本数据的个数不小于第二阈值，就会将这两个疑似初始化聚类簇合并。

5 实际应用的某些情形中，两个疑似初始化聚类簇可能并不属于同一类，可这两个疑似初始化聚类簇中存在极少数的样本数据间的样本间相似度大于第一阈值，也就是，由于疑似初始化聚类簇中存在噪声点样本数据，导致两个疑似初始化聚类簇合并，这样在确定初始化聚类簇阶段可能出现误差。

10 如图 6 所示，图 6 中， X_i 为样本数据，样本数据间存在实线连接表示两样本数据间的样本间相似度大于第一阈值，疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_2 之间只有 X_{10} 和 X_{12} 相连，若采用上述疑似初始化聚类簇合并方法，那么会将疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_2 合并，这使得不应合并两个疑似初始化聚类簇合并为了一个，在确定初始化聚类簇阶段出现了误差。

15 为了减小甚至避免在确定初始化聚类簇阶段出现误差，本公开中的第二阈值可以根据所述至少二个疑似初始化聚类簇中一个或者多个疑似初始化聚类簇所包括的样本数据的个数的线性函数确定。如：假设斜率 k 为 0.5，当一个疑似初始化聚类簇 C_1 中包含 2 个样本数据时，若其想与疑似初始化聚类簇 C_2 合并，第二阈值为 1；当疑似初始化聚类簇 C_1 中包含 4 个样本数据时，若其想与疑似初始化聚类簇 C_2 合并，第二阈值为 2，依此类推。

20 本公开合并疑似初始化聚类簇的另一种可选的实现方式为：不仅两个疑似初始化聚类簇所包含的相同样本数据的个数不小于第二阈值，而且两个疑似初始化聚类簇还应满足下述公式 (1) 才能够被合并：

$$|S_1 \cap S_2| \geq (|S_1| + |S_2|) / \beta + \gamma; \quad (1)$$

25 其中， S_1 为一个疑似初始化聚类簇对应的样本数据集，在所述 S_1 中，一个样本数据与其他各样本数据的样本间相似度均大于第一阈值， S_2 为另一个疑似初始化聚类簇对应的样本数据集，在所述 S_2 中，一个样本数据与其他各样本数据之间的样本相似度均大于第一阈值， β 和 γ 为预设参数。其中，可选地， $2 \leq \beta \leq 3$ ， $-3 \leq \gamma \leq -1$ ，例如， $\beta = 2.5$ ， $\gamma = -1$ 。

合并两个疑似初始化聚类簇的过程可以具体为：

判断第一疑似初始化聚类簇和第二疑似初始化聚类簇是否满足公式 (1)，其中，第一疑似初始化聚类簇和第二疑似初始化聚类簇包含相同的样本数据；

30 若为是，将第一疑似初始化聚类簇和第二疑似初始化聚类簇合并为一个新的疑似初始化聚类簇，再次根据公式 (1) 将该新的疑似初始化聚类簇与其他疑似初始化聚类簇进行合并；

若为否，拒绝将第一疑似初始化聚类簇和第二疑似初始化聚类簇合并为一个新的疑似初始化聚类簇。

35 当所有疑似初始化聚类簇都不能再合并时，将所有疑似初始化聚类簇，都确定为初始化聚类簇。

如步骤 S01 中的假设，由于疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_2 中都包含样本数据

a, 因此, 处理器将疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_2 分别看作是一个样本数据集合, 若 C_1 和 C_2 的交集的模、 C_1 的模和 C_2 的模满足公式 (1), 也就是, 下述公式成立:

$$|C_1 \cap C_2| \geq (|C_1| + |C_2|) / \beta + \gamma,$$

那么处理器就将疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_2 合并为一个新的疑似初始化聚类簇 C_5 , 此时疑似初始化聚类簇 C_5 中包含样本数据 a 样本数据 b 和样本数据 d、而疑似初始化聚类簇 C_3 中包含样本数据 b 和样本数据 c, 疑似初始化聚类簇 C_5 和疑似初始化聚类簇 C_3 中都包含样本数据 b, 则处理器再次判断疑似初始化聚类簇 C_5 和疑似初始化聚类簇 C_3 是否满足公式 (1), 若不满足公式 (1), 则处理器拒绝合并疑似初始化聚类簇 C_5 和疑似初始化聚类簇 C_3 , 由于疑似初始聚类簇 C_4 包含样本数据 e, 疑似初始聚类簇 C_4 与疑似初始聚类簇 C_5 和疑似初始聚类簇 C_3 未包含有相同的样本数据, 因此, 处理器分别将疑似初始化聚类簇 C_5 、疑似初始化聚类簇 C_3 和疑似初始聚类簇 C_4 确定为初始化聚类簇;

若 C_1 和 C_2 的交集的模、 C_1 的模和 C_2 的模不满足公式 (1), 则处理器拒绝合并疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_2 , 由于疑似初始聚类簇 C_1 和疑似初始化聚类簇 C_3 都包含样本数据 b, 处理器再次判断疑似初始化聚类簇 C_1 和疑似初始化聚类簇 C_3 是否满足公式 (1), 假设还是不满足公式 (1), 另外, 疑似初始聚类簇 C_4 包含样本数据 e, 疑似初始聚类簇 C_4 与疑似初始聚类簇 C_1 、疑似初始聚类簇 C_2 和疑似初始化聚类簇 C_3 未包含有相同的样本数据, 因此, 处理器不对其进行合并处理, 处理器将疑似初始化聚类簇 C_1 、疑似初始化聚类簇 C_2 、疑似初始化聚类簇 C_3 和疑似初始化聚类簇 C_4 分别确定为初始化聚类簇。

本公开在两个疑似初始化聚类簇所包含的相同样数据的个数不小于第二阈值, 且在满足公式 (1) 的情况下, 才将两个疑似初始化聚类簇进行合并, 在两个疑似初始化聚类簇合并过程中考虑了 $(|S_1 \cap S_2|)$ 的情况, 因此有利于提高初始化聚类簇的准确性, 有利于提高整个聚类技术的鲁棒性。

S103: 对所述 N 个初始化聚类簇进行聚类合并, 得到所述 M 个样本数据对应的多个聚类簇。

一种可选的实现方式中, 步骤 S103 可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的聚类单元 503 执行。

一种可选的实现方式中, 参考图 7, 对所述 N 个初始化聚类簇进行聚类合并, 可以包括如下步骤:

S301: 将所述 N 个初始化聚类簇作为多个待聚类簇。

一种可选的实现方式中, 步骤 S301 可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的第二作为子单元执行。

S302: 获取每个待聚类簇与其他待聚类簇之间的簇间相似度。

一种可选的实现方式中, 步骤 S302 可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的第二获取子单元执行。

处理器 (如第二获取子单元) 在计算每两个待聚类簇之间的簇间相似度时, 可参考图 8, 图 8 为本公开提供的一种簇间相似度的确定方法的流程示意图, 该方法包括如下步骤:

S401、获得第一待聚类簇中的每一样本数据与第二待聚类簇中的每一样本数据之间的至少

一个第一样本间相似度。

假设，第一待聚类簇中包含的样本数据为：a、b；第二待聚类簇中包含的样本数据为：c、d；那么确定的第一样本间相似度包括：a和c间的样本间相似度、a和d间的样本间相似度、b和c间的样本间相似度、b和d间的样本间相似度。

5 S402、确定相似度范围。

其中，相似度范围的最小值大于第一样本间相似度中的最小的样本间相似度，相似度范围的最大值小于第一样本间相似度中的最大的样本间相似度。实际应用中，第一样本间相似度中的较大值和较小值对应的样本数据在一般情况下很可能为噪声点数据，因此，在确定相似度范围时，取第一样本间相似度中的中间范围为相似度范围，根据该相似度范围内的第一样本间相似度，计算两待聚类簇之间的簇间相似度，能够最大程度的去除噪声点数据对计算两待聚类簇之间的簇间相似度准确性的影响，进而提高最终聚类结果的准确率。

假设，处理器获得的第一样本间相似度为：0.2、0.32、0.4、0.3、0.7、0.5、0.75、0.8、0.9、0.92，此时处理器确定的相似度范围可以为：0.3-0.75；也可以为：0.4-0.7；当然也可以为其他满足上述条件的相似度范围，本公开对此不进行限定。

15 一种具体实现方式中，处理器可以通过以下方式确定相似度范围：

处理器对获得的第一样本间相似度进行排序，如， $e_1 \geq e_2 \geq e_3 \geq \dots \geq e_E$ ，其中，E为获得的第一样本间相似度的个数，e为第一样本间相似度；

处理器根据参数l、k确定相似度范围，这里，可选的，参数l可以取0.2E，参数k可以取0.8E。

20 根据上述假设，处理器对其获得的第一样本间相似度进行排序， $0.2 < 0.3 < 0.32 < 0.4 < 0.5 < 0.7 < 0.75 < 0.8 < 0.9 < 0.92$ ，处理器根据第一样本间相似度的个数E=10，确定l=2，k=8，此时处理器确定的相似度范围可以为0.3-0.9。

S403、根据相似度范围内的第一样本间相似度，计算第一待聚类簇和第二待聚类簇之间的簇间相似度。

25 具体地，处理器可以根据下述公式计算第一待聚类簇和第二待聚类簇之间的簇间相似度：

$$\sum_{i=l}^k \alpha^i e_i / \sum_{i=l}^k \alpha^i$$

其中， α 为预设参数，可选地， $0.85 < \alpha < 0.95$ ； e_i 为对获得的第一样本间相似度按照从小到大的顺序排序后，从最小侧起（或从最大侧起）的第i个第一样本间相似度，l和k为上述确定出的参数。

30 S303：确定所述多个待聚类簇对应的所有簇间相似度中的最大簇间相似度是否大于第四阈值。

一种可选的实现方式中，步骤S303可以由处理器调用存储器存储的相应指令来执行，或者，可以由被处理器运行的第二确定子单元执行。

35 可选地，所述第四阈值可以在0.75-0.95的范围内取值。若簇间相似度大于第四阈值，则说明该簇间相似度对应的两个待聚类簇非常相似，可以认为这两个待聚类簇为一类，合并这两

个待聚类簇。

S304: 响应于所述最大簇间相似度大于所述第四阈值, 将所述最大簇间相似度对应的两个待聚类簇进行合并得到一新的待聚类簇。

一种可选的实现方式中, 步骤 S304 可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的第二响应子单元执行。

S305: 对所述新的待聚类簇与本次未合并的其它待聚类簇构成的新的多个待聚类簇继续进行聚类合并, 直到没有可合并的待聚类簇, 即重复上述步骤 S302 至 S304, 直到没有可合并的待聚类簇。

假设, 当前计算得到的簇间相似度有 0.85、0.92、0.7、0.6、0.95, 从中获得的最大的簇间相似度为 0.95, 设定第四阈值为 0.75, 由于 $0.95 > 0.75$, 因此将 0.95 对应的两个待聚类簇合并为一个新的待聚类簇, 并返回步骤 S302, 继续获取每个待聚类簇与其他待聚类簇之间的簇间相似度, 直到最大的簇间相似度不大于第四阈值为止, 也就是, 直到没有可合并的待聚类簇为止。此时, 可以将当前多个待聚类簇作为所述多个样本数据对应的多个聚类簇。

由于样本数据本身存在一定的噪声, 并且聚类过程中存在一些可以影响最终的聚类结果鲁棒性的因素, 因此, 导致最终的聚类结果中可能会存在一些样本数据被错误的聚类。为了使最终呈现的聚类结果更加准确, 在得到所述多个样本数据对应的多个聚类簇之后, 处理器还可以对所述多个聚类簇中的每个聚类簇进行离群点分离, 得到所述多个样本数据优化后的聚类结果。一种可选的实现方式中, 对所述多个聚类簇中的每个聚类簇进行离群点分离处理的操作可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的分离单元 601 执行。

以一个聚类簇 C_1 进行离群点分离为例, 分离的过程包括如下步骤:

S31、获得聚类簇 C_1 内每一样本数据对应的待离群簇和非待离群簇。

一种可选的实现方式中, 步骤 S31 可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的获得子单元执行。

其中, 一样本数据 X 对应的待离群簇中包括: 样本数据 X , 所述样本数据 X 对应的非待离群簇中包括: 聚类簇 C_1 中除样本数据 X 外的其他样本数据。

假设, 聚类簇 C_1 包括的样本数据有: X_1 、 X_2 、 X_3 和 X_4 , 那么, X_1 对应的待离群簇为 $\{X_1\}$, X_1 对应的非待离群簇为 $\{X_2, X_3, X_4\}$; X_2 对应的待离群簇为 $\{X_2\}$, X_2 对应的非待离群簇为 $\{X_1, X_3, X_4\}$; X_3 对应的待离群簇为 $\{X_3\}$, X_3 对应的非待离群簇为 $\{X_1, X_2, X_4\}$; X_4 对应的待离群簇为 $\{X_4\}$, X_4 对应的非待离群簇为 $\{X_1, X_2, X_3\}$ 。

S32、获取每一样本数据对应的待离群簇与非待离群簇之间的簇间相似度。

一种可选的实现方式中, 步骤 S32 可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的第一获取子单元执行。

S33、确定聚类簇 C_1 中所有样本数据分别对应的多个簇间相似度中最小的簇间相似度是否小于第三阈值。

一种可选的实现方式中, 步骤 S33 可以由处理器调用存储器存储的相应指令来执行, 或者, 可以由被处理器运行的第一确定子单元执行。

这里, 第三阈值可以在 0.3-0.5 的范围内取值。若最小的簇间相似度小于第三阈值, 则说

明该簇间相似度对应的待离群簇与非待离群簇非常不相似，可以理解为该待离群簇与该非待离群簇不是一类，该待离群簇包含的样本数据相对于该非待离群簇来说为离群点，处理器需要将该待离群簇与该非待离群簇分离。

- 5 S34、响应于最小的簇间相似度小于第三阈值，将最小的簇间相似度对应的待离群簇和非待离群簇分别作为两个新的聚类簇。

一种可选的实现方式中，步骤 S34 可以由处理器调用存储器存储的相应指令来执行，或者，可以由被处理器运行的第一响应子单元执行。

S35、对非待离群簇对应的聚类簇继续进行离群点分离操作，如重复上述步骤 S32 至 S34，直到聚类簇中没有可分离的样本数据。

- 10 根据 S31 中的例子，假设，处理器计算得到的簇间相似度有 0.25、0.2、0.7、0.5，从中获得的最小的簇间相似度为 0.2（0.2 对应的待离群簇为 $\{X_2\}$ ），第三阈值为 0.3，由于 $0.2 < 0.3$ ，因此处理器可以确定待离群簇 $\{X_2\}$ 为离群点，处理器将 0.2 对应的待离群簇和非待离群簇分别作为两个新的聚类簇，即聚类簇 $\{X_2\}$ 和聚类簇 $\{X_1, X_3, X_4\}$ ；处理器对聚类簇 $\{X_1, X_3, X_4\}$ 继续进行离群点分离操作，直到最小的簇间相似度不小于第三阈值为止，也就是，直到没有可分
15 离的样本数据为止。

应用图 5-图 8 所示技术方案，处理器对样本数据进行聚类时，先根据多个样本数据中每两个样本数据之间的样本间相似度对样本数据进行合并，获得初始化聚类簇，减少了聚类时初始化聚类簇的数量，处理器根据此时的初始化聚类簇进行聚类合并，得到所述多个样本数据对应的多个聚类簇，有效地提高了聚类速度。

- 20 参考图 9 所示的装置包括：中，获取单元 501，用于获取 M 个样本数据中每两个样本数据之间的样本间相似度，M 为正整数；

合并单元 502，用于根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，N 为小于 M 的正整数；

- 25 聚类单元 503，用于对所述 N 个初始化聚类簇进行聚类合并，得到所述 M 个样本数据对应的多个聚类簇。

一种可选的实现方式中，所述合并单元 502 可以包括：

第一合并子单元（图 9 中未示出），用于所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；

- 30 第二合并子单元（图 9 中未示出），用于针对得到的多个疑似初始化聚类簇，将包含相同样本数据的个数不小于第二阈值的至少二个疑似初始化聚类簇合并为一初始化聚类簇。

一种可选的实现方式中，所述合并单元 502 还可以包括：

第一作为子单元（图 9 中未示出），用于针对得到的多个疑似初始化聚类簇，将不能与其他疑似初始化聚类簇合并的疑似初始化聚类簇作为一初始化聚类簇。

- 35 一种可选的实现方式中，所述第二阈值根据所述至少二个疑似初始化聚类簇中各疑似初始化聚类簇所包括的样本数据的个数总和的线性函数确定。

一种可选的实现方式中，可参考图 10，所述装置在图 9 所示装置的基础上还可以包括：

分离单元 601, 用于对所述多个聚类簇中的每个聚类簇进行离群点分离, 得到所述多个样本数据优化后的聚类结果。

一种可选的实现方式中, 所述分离单元 601 具体用于对一所述聚类簇进行离群点分离;

所述分离单元 601 可以包括:

- 5 获得子单元 (图 10 中未示出), 用于获得所述聚类簇内每一样本数据对应的待离群簇和非待离群簇, 其中, 每一样本数据对应的所述待离群簇中包括所述样本数据, 所述非待离群簇中包括: 所述聚类簇中除所述样本数据外的其它样本数据;

第一获取子单元 (图 10 中未示出), 用于获取每一样本数据对应的待离群簇与非待离群簇之间的簇间相似度;

- 10 第一确定子单元 (图 10 中未示出), 用于确定所述聚类簇中所有样本数据分别对应的多个簇间相似度中最小的簇间相似度是否小于第三阈值;

第一响应子单元 (图 10 中未示出), 用于响应于所述最小的簇间相似度小于所述第三阈值, 将所述最小的簇间相似度对应的待离群簇和非待离群簇分别作为两个新的聚类簇, 并触发所述获得子单元, 对所述非待离群簇对应的聚类簇继续进行离群点分离操作, 直到没有可分离的聚

- 15 类簇。

一种可选的实现方式中, 所述聚类单元 503 可以包括:

第二作为子单元 (图 9 中未示出), 用于将所述 N 个初始化聚类簇作为多个待聚类簇;

第二获取子单元 (图 9 中未示出), 用于获取每个待聚类簇与其他待聚类簇之间的簇间相似度;

- 20 第二确定子单元 (图 9 中未示出), 用于确定所述多个待聚类簇对应的所有簇间相似度中的最大簇间相似度是否大于第四阈值;

第二响应子单元 (图 9 中未示出), 用于响应于所述最大簇间相似度大于所述第四阈值, 将所述最大簇间相似度对应的两个待聚类簇进行合并得到一新的待聚类簇, 并触发所述第二获取子单元, 对所述新的待聚类簇与本次未合并的其它待聚类簇构成的新的多个待聚类簇继续进

- 25 行聚类合并, 直到没有可合并的待聚类簇。

一种可选的实现方式中, 所述 M 个样本数据为 M 个图像。

一种可选的实现方式中, 两个图像之间的样本间相似度包括: 所述两个图像分别对应的两个特征向量之间的余弦距离。

- 30 应用图 9 所示, 对样本数据进行聚类时, 先根据多个样本数据中每两个样本数据之间的样本间相似度对样本数据进行合并, 获得初始化聚类簇, 减少聚类时初始化聚类簇的数量, 根据此时的初始化聚类簇进行聚类合并, 得到所述多个样本数据对应的多个聚类簇, 有效地提高了聚类速度。

- 35 参考图 11, 图 11 为本公开提供的一种电子设备的结构示意图, 该电子设备包括: 壳体 701、处理器 702、存储器 703、电路板 704 和电源电路 705, 其中, 电路板 704 安置在壳体 701 围成的空间内部, 处理器 702 和存储器 703 设置在电路板 704 上; 电源电路 705, 用于为电子设备的各个电路或器件供电; 存储器 703 用于存储可执行程序代码; 处理器 702 通过读取存储器 703 中存储的可执行程序代码来运行与可执行程序代码对应的程序, 以用于执行以下步骤:

获取 M 个样本数据中每两个样本数据之间的样本间相似度，M 为正整数；

根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，N 为小于 M 的正整数；

对所述 N 个初始化聚类簇进行聚类合并，得到所述 M 个样本数据对应的多个聚类簇。

5 处理器 702 对上述步骤的具体执行过程以及处理器 702 通过运行可执行程序代码来进一步执行的步骤，可以参见本公开针对图 5-11 所示的描述，在此不再赘述。

本公开中，对样本数据进行聚类时，先根据多个样本数据中每两个样本数据之间的样本间相似度对样本数据进行合并，获得初始化聚类簇，减少聚类时初始化聚类簇的数量，根据此时的初始化聚类簇进行聚类合并，得到所述多个样本数据对应的多个聚类簇，有利于提高聚类速度。
10 度。

该电子设备以多种形式存在，包括但不限于：

(1) 移动通信设备：这类设备的特点是具备移动通信功能，并且以提供话音、数据通信为主要目标。这类终端包括：智能手机（例如 iPhone）、多媒体手机、功能性手机，以及低端手机等。

15 (2) 超移动个人计算机设备：这类设备属于个人计算机的范畴，有计算和处理功能，一般也具备移动上网特性。这类终端包括：PDA、MID 和 UMPC 设备等，例如 iPad。

(3) 便携式娱乐设备：这类设备可以显示和播放多媒体内容。该类设备包括：音频、视频播放器（例如 iPod），掌上游戏机，电子书，以及智能玩具和便携式车载导航设备。

(4) 服务器：提供计算服务的设备，服务器的构成包括处理器、硬盘、内存、系统总线等，
20 服务器和通用的计算机架构类似，但是由于需要提供高可靠的服务，因此在处理能力、稳定性、可靠性、安全性、可扩展性、可管理性等方面要求较高。

(5) 其他具有数据交互功能的电子装置。

对于装置、电子设备实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

25 需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在
30 没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

本领域普通技术人员可以理解实现上述方法实施方式中的全部或部分步骤是可以通程序来指令相关的硬件来完成，所述的程序可以存储于计算机可读取存储介质中，这里所称得的存储介质，如：ROM/RAM、磁碟、光盘等。

35 以上所述仅为本公开的较佳实施例而已，并非用于限定本公开的保护范围。凡在本公开的精神和原则之内所作的任何修改、等同替换、改进等，均包含在本公开的保护范围内。

权 利 要 求

1、一种聚类方法，其特征在于，所述方法包括：

获取 M 个样本数据中每两个样本数据之间的样本间相似度，M 为正整数；

5 根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，N 为小于 M 的正整数；

对所述 N 个初始化聚类簇进行聚类合并，得到所述 M 个样本数据对应的多个聚类簇。

2、根据权利要求 1 所述的方法，其特征在于，所述根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，包括：

10 所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；

针对得到的多个疑似初始化聚类簇，将包含相同样本数据的个数不小于第二阈值的至少二个疑似初始化聚类簇合并为一初始化聚类簇。

3、根据权利要求 1 所述的方法，其特征在于，所述根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，包括：

所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；

20 针对包含相同样本数据的个数不小于第二阈值的第一疑似初始化聚类簇和第二疑似初始化聚类簇，计算第一和第二疑似初始化聚类簇的交集的模，并计算第一初始化聚类簇的模与第二初始化聚类簇的模之和与第一常数的商，在所述交集的模与所述商之差不小于第二常数时，将所述第一和第二疑似初始化聚类簇合并为一疑似初始化聚类簇。

4、根据权利要求 2-3 任一所述的方法，其特征在于，所述根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，还包括：

25 针对得到的多个疑似初始化聚类簇，将不能与其他疑似初始化聚类簇合并的各疑似初始化聚类簇分别作为一初始化聚类簇。

5、根据权利要求 2-4 任一所述的方法，其特征在于，所述第二阈值根据所述至少二个疑似初始化聚类簇中的一个或者多个疑似初始化聚类簇所包括的样本数据的个数的线性函数确定。

30 6、根据权利要求 1-5 任一所述的方法，其特征在于，所述方法还包括：对所述多个聚类簇中的至少一个聚类簇进行离群点分离，且所述离群点分离处理后获得的所有聚类簇被作为所述 M 个样本数据对应的多个聚类簇。

7、根据权利要求 6 所述的方法，其特征在于，所述离群点分离包括：

针对任一聚类簇，获得所述聚类簇内每一样本数据对应的待离群簇和非待离群簇，其中，每一样本数据对应的所述待离群簇中包括所述样本数据，所述非待离群簇中包括：所述聚类簇中除所述样本数据外的其它样本数据；

35 获取每一样本数据对应的待离群簇与非待离群簇之间的簇间相似度；

确定所述聚类簇中所有样本数据分别对应的多个簇间相似度中最小的簇间相似度是否小于第三阈值；

响应于所述最小的簇间相似度小于所述第三阈值，将所述最小的簇间相似度对应的待离群簇和非待离群簇分别作为两个新的聚类簇。

8、根据权利要求 1-7 任一所述的方法，其特征在于，所述对所述 N 个初始化聚类簇进行聚类合并，包括：

5 将所述 N 个初始化聚类簇作为多个待聚类簇；

获取每个待聚类簇与其他待聚类簇之间的簇间相似度；

确定所述多个待聚类簇对应的所有簇间相似度中的最大簇间相似度是否大于第四阈值；

响应于所述最大簇间相似度大于所述第四阈值，将所述最大簇间相似度对应的两个待聚类簇进行合并得到一新的待聚类簇；

10 对所述新的待聚类簇与本次未合并的其它待聚类簇构成的新的多个待聚类簇继续进行聚类合并，直到没有可合并的待聚类簇。

9、根据权利要求 2-8 任一所述的方法，其特征在于，所述 M 个样本数据为 M 个图像，且两个图像之间的样本间相似度包括：所述两个图像分别对应的两个特征向量之间的余弦距离。

10、一种聚类装置，其特征在于，所述装置包括：

15 获取单元，用于获取 M 个样本数据中每两个样本数据之间的样本间相似度，M 为正整数；

合并单元，用于根据获取的样本间相似度将所述 M 个样本数据合并为 N 个初始化聚类簇，N 为小于 M 的正整数；

聚类单元，用于对所述 N 个初始化聚类簇进行聚类合并，得到所述 M 个样本数据对应的多个聚类簇。

20 11、根据权利要求 10 所述的装置，其特征在于，所述合并单元包括：

第一合并子单元，用于所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；

第二合并子单元，用于针对得到的多个疑似初始化聚类簇，将包含相同样本数据的个数不小于第二阈值的至少二个疑似初始化聚类簇合并为一初始化聚类簇。

25 12、根据权利要求 10 所述的装置，其特征在于，所述合并单元具体用于：

所述 M 个样本数据中，对于任意两个样本数据而言，如果这两个样本数据的样本间相似度大于第一阈值，则将这两个样本数据合并为一疑似初始化聚类簇；

30 针对包含相同样本数据的个数不小于第二阈值的第一疑似初始化聚类簇和第二疑似初始化聚类簇，计算第一和第二疑似初始化聚类簇的交集的模，并计算第一初始化聚类簇的模与第二初始化聚类簇的模之和与第一常数的商，在所述交集的模与所述商之差不小于第二常数时，将所述第一和第二疑似初始化聚类簇合并为一疑似初始化聚类簇。

13、根据权利要求 11-12 任一所述的装置，其特征在于，所述合并单元还包括：

第一作为子单元，用于针对得到的多个疑似初始化聚类簇，将不能与其他疑似初始化聚类簇合并的各疑似初始化聚类簇分别作为一初始化聚类簇。

35 14、根据权利要求 11-13 任一所述的装置，其特征在于，所述第二阈值根据所述至少二个疑似初始化聚类簇中的一个或者多个疑似初始化聚类簇所包括的样本数据的个数的线性函数确定。

15、根据权利要求 10-13 任一所述的装置，其特征在于，所述装置还包括：

分离单元，用于对所述多个聚类簇中的至少一个聚类簇进行离群点分离，且所述离群点分离处理后获得的所有聚类簇被作为所述 M 个样本数据对应的多个聚类簇。

5 16、根据权利要求 15 所述的装置，其特征在于，所述分离单元具体用于对一所述聚类簇进行离群点分离；

所述分离单元包括：

获得子单元，针对任一聚类簇，用于获得所述聚类簇内每一样本数据对应的待离群簇和非待离群簇，其中，每一样本数据对应的所述待离群簇中包括所述样本数据，所述非待离群簇中包括：所述聚类簇中除所述样本数据外的其它样本数据；

10 第一获取子单元，用于获取每一样本数据对应的待离群簇与非待离群簇之间的簇间相似度；

第一确定子单元，用于确定所述聚类簇中所有样本数据分别对应的多个簇间相似度中最小的簇间相似度是否小于第三阈值；

第一响应子单元，用于响应于所述最小的簇间相似度小于所述第三阈值，将所述最小的簇间相似度对应的待离群簇和非待离群簇分别作为两个新的聚类簇。

15 17、根据权利要求 10-16 任一所述的装置，其特征在于，所述聚类单元包括：

第二作为子单元，用于将所述 N 个初始化聚类簇作为多个待聚类簇；

第二获取子单元，用于获取每个待聚类簇与其他待聚类簇之间的簇间相似度；

第二确定子单元，用于确定所述多个待聚类簇对应的所有簇间相似度中的最大簇间相似度是否大于第四阈值；

20 第二响应子单元，用于响应于所述最大簇间相似度大于所述第四阈值，将所述最大簇间相似度对应的两个待聚类簇进行合并得到一新的待聚类簇，并触发所述第二获取子单元，对所述新的待聚类簇与本次未合并的其它待聚类簇构成的新的多个待聚类簇继续进行聚类合并，直到没有可合并的待聚类簇。

25 18、根据权利要求 10-17 任一所述的装置，其特征在于，所述 M 个样本数据为 M 个图像，且两个图像之间的样本间相似度包括：所述两个图像分别对应的两个特征向量之间的余弦距离。

30 19、一种电子设备，其特征在于，包括：壳体、处理器、存储器、电路板和电源电路，其中，所述电路板安置在所述壳体围成的空间内部，所述处理器和所述存储器设置在所述电路板上；所述电源电路，用于为终端的各个电路或器件供电；所述存储器用于存储可执行程序代码；所述处理器通过读取所述存储器中存储的可执行程序代码来运行与可执行程序代码对应的程序，以用于执行权利要求 1-9 任一项所述的聚类方法对应的操作。

20、一种计算机程序，包括计算机可读代码，当所述计算机可读代码在设备中运行时，所述设备中的处理器执行用于实现权利要求 1-9 中的任一权利要求所述的聚类方法中的步骤的可执行指令。

21、一种计算机可读介质，用于存储权利要求 20 所述的计算机程序。

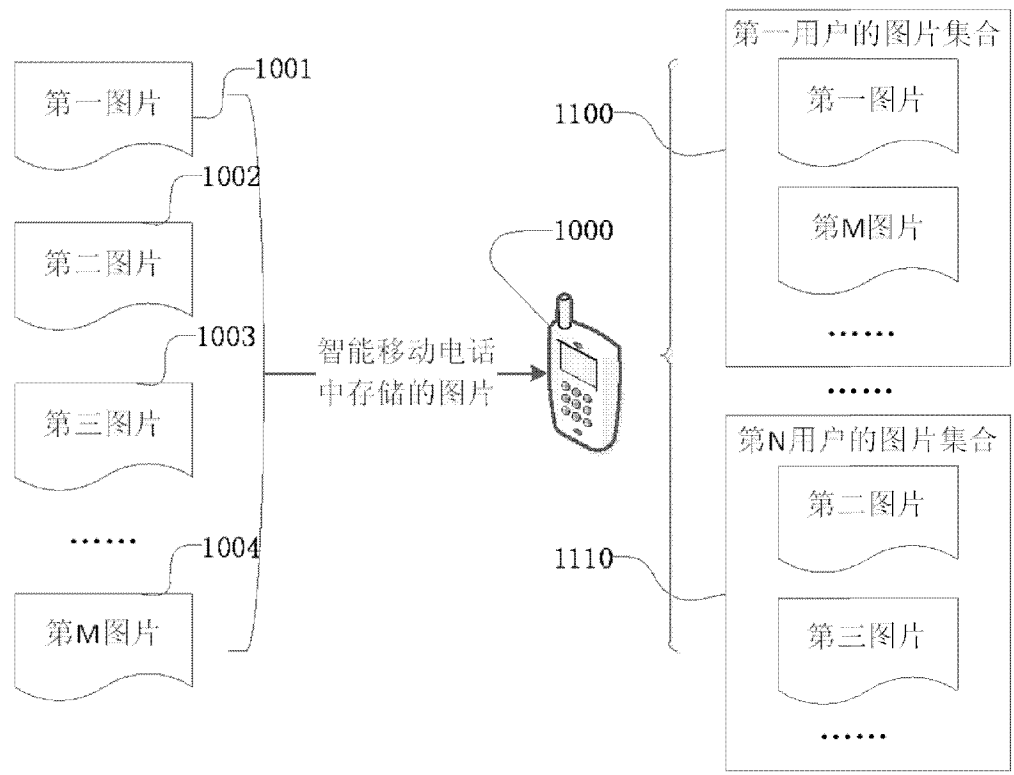


图 1

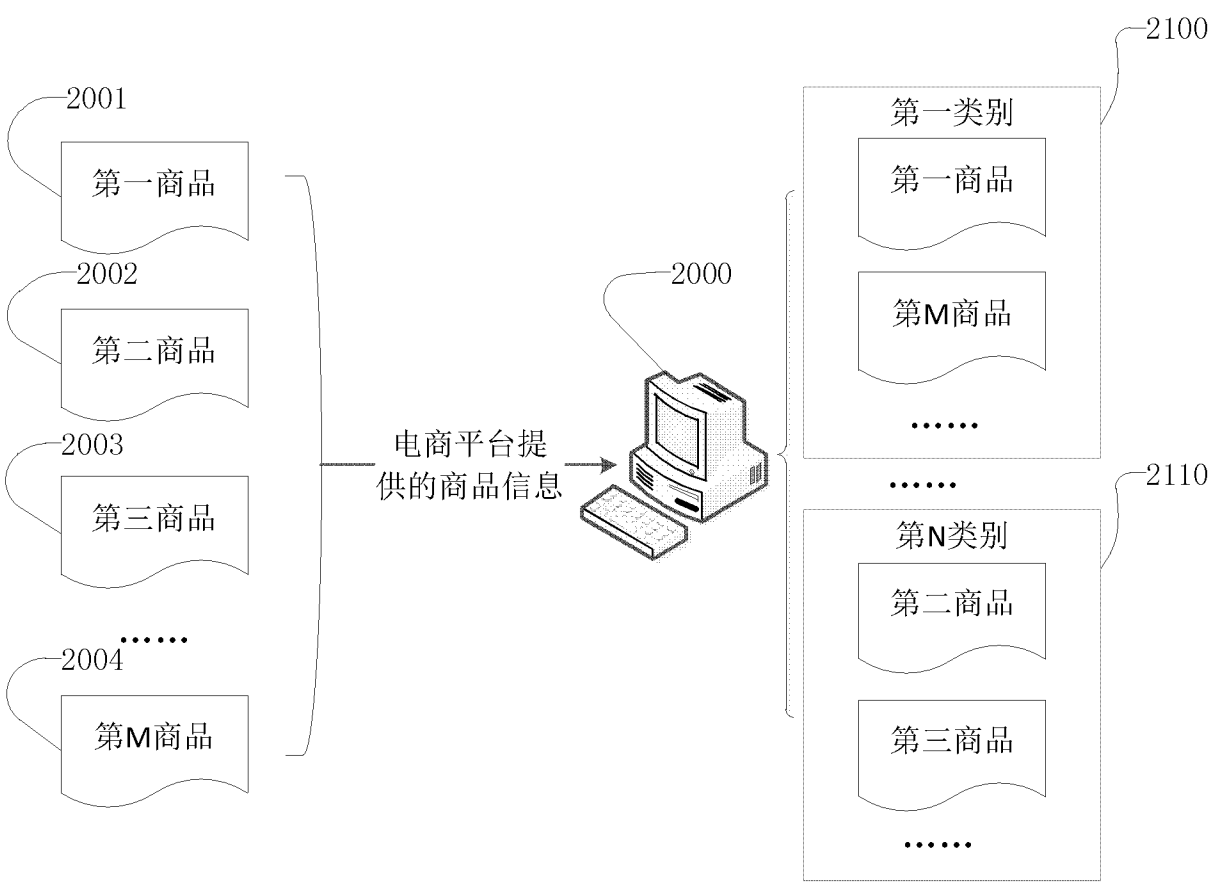


图 2

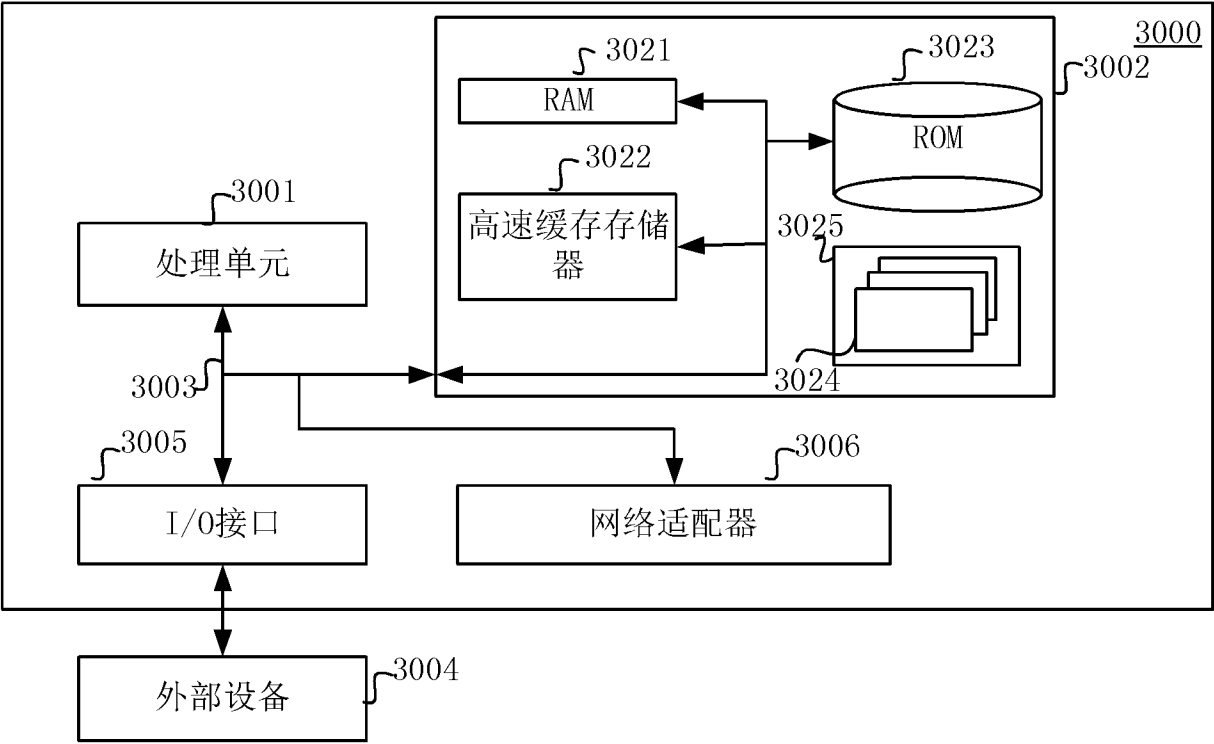


图 3

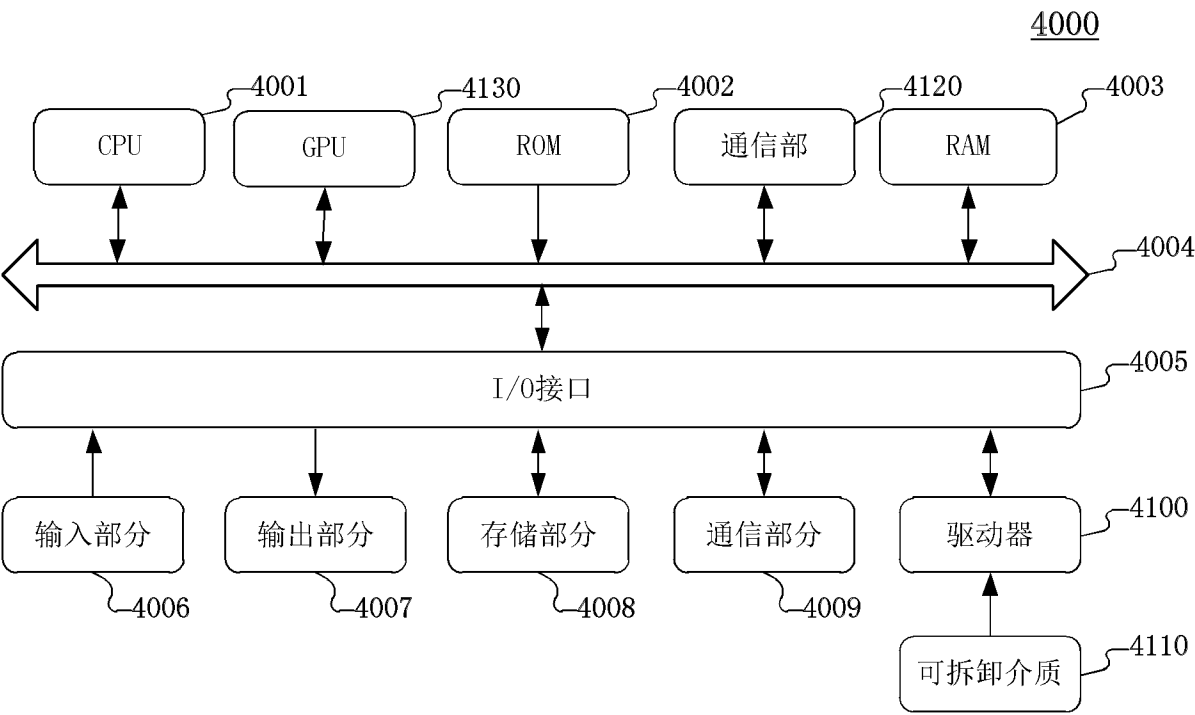


图 4

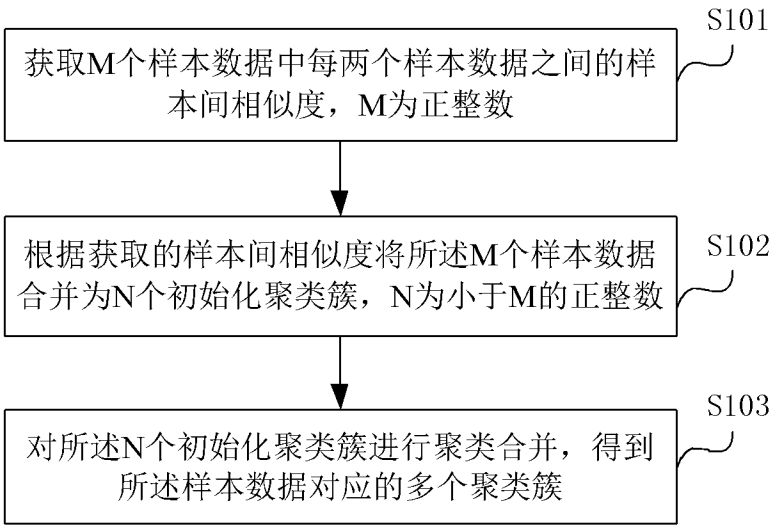


图 5

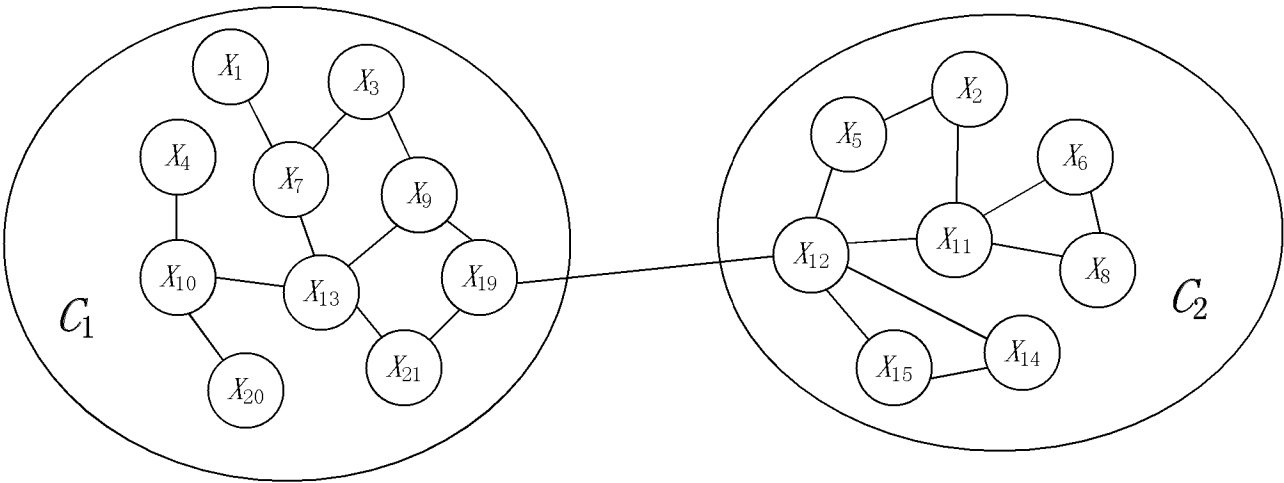


图6

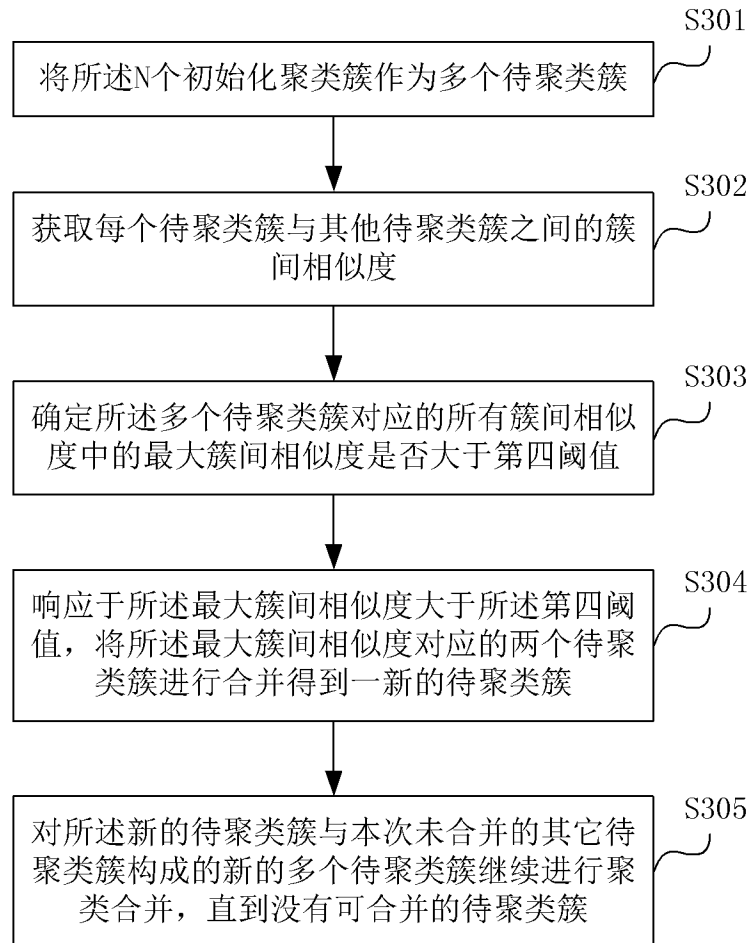


图7

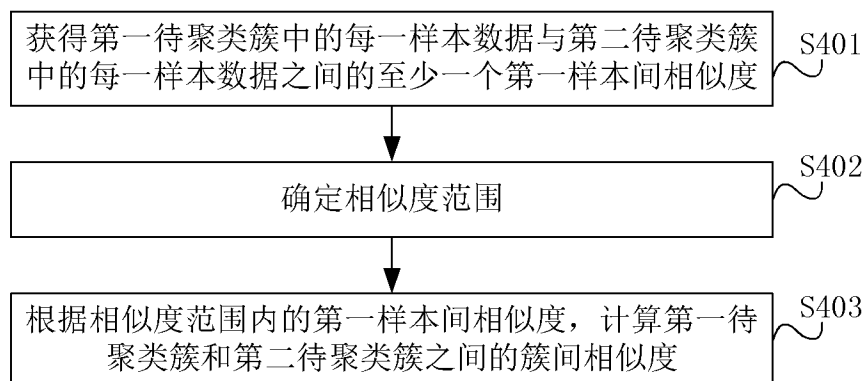


图8

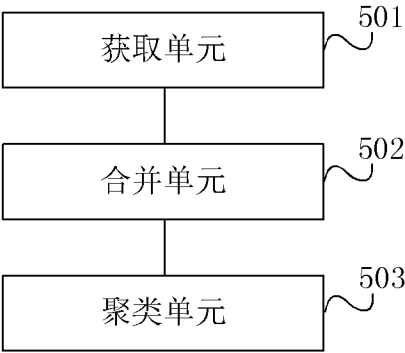


图9

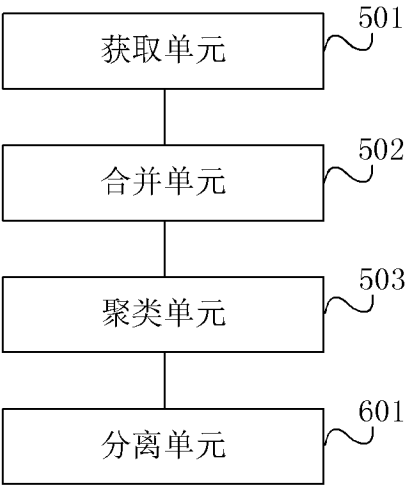


图10

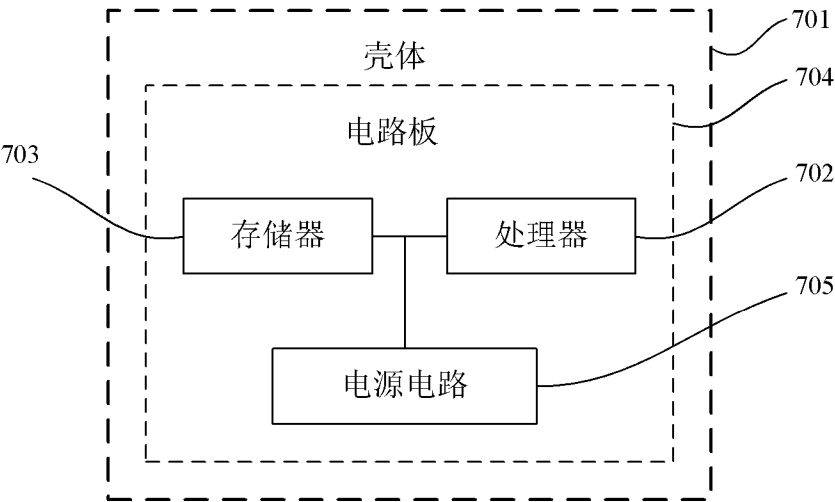


图11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/091432

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)

CNPAT, WPI, EPODOC, IEEE, CNKI: cluster, classification, similarity, threshold value, quotient, data, class, combination, set, sample, similar, preset value, model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106228188 A (BEIJING SENSETIME SCIENCE AND TECHNOLOGY DEVELOPMENT CO., LTD.), 14 December 2016 (14.12.2016), claims 1-10, and description, paragraphs [0068]-[0183]	1-21
X	CN 104252627 A (GUANGZHOU HUADUO INTERNET TECHNOLOGY CO., LTD.), 31 December 2014 (31.12.2014), description, paragraphs [0036]-[0058], figure 1	1-2, 4-6, 8-11, 13-15, 17-21
X	CN 103679190 A (FUJITSU LIMITED), 26 March 2014 (26.03.2014), description, paragraphs [0024]-[0030]	1-2, 4-6, 8-11, 13-15, 17-21
A	CN 103699653 A (SHENYANG AEROSPACE UNIVERSITY), 02 April 2014 (02.04.2014), the whole document	1-21
A	CN 104281569 A (FUJITSU LIMITED), 14 January 2015 (14.01.2015), the whole document	1-21
A	WO 2014067296 A1 (SHENZHEN EMPEROR ELECTRONIC TECHNOLOGY CO., LTD.), 08 May 2014 (08.05.2014), the whole document	1-21

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 18 September 2017 (18.09.2017)	Date of mailing of the international search report 11 October 2017 (11.10.2017)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer ZHANG, Qianzhen Telephone No.: (86-10) 62414072

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2017/091432

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106228188 A	14 December 2016	None	
CN 104252627 A	31 December 2014	None	
CN 103679190 A	26 March 2014	JP 2014063494 A	10 April 2014
		JP 6179305 B2	16 August 2017
CN 103699653 A	02 April 2014	None	
CN 104281569 A	14 January 2015	CN 104281569 B	01 August 2017
WO 2014067296 A1	08 May 2014	CN 102982314 A	20 March 2013
		CN 102982314 B	25 May 2016

国际检索报告

国际申请号

PCT/CN2017/091432

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)) CNPAT, WPI, EPODOC, IEEE, CNKI: 数据, 聚类, 分类, 合并, 样本, 相似度, 阈值, 商, 模, data, class, combination, set, sample, similar, preset value, model																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 106228188 A (北京市商汤科技开发有限公司) 2016年 12月 14日 (2016 - 12 - 14) 权利要求1-10, 说明书第[0068]-[0183]段</td> <td>1-21</td> </tr> <tr> <td>X</td> <td>CN 104252627 A (广州华多网络科技有限公司) 2014年 12月 31日 (2014 - 12 - 31) 说明书第[0036]-[0058]段, 图1</td> <td>1-2, 4-6, 8-11, 13-15, 17-21</td> </tr> <tr> <td>X</td> <td>CN 103679190 A (富士通株式会社) 2014年 3月 26日 (2014 - 03 - 26) 说明书第[0024]-[0030]段</td> <td>1-2, 4-6, 8-11, 13-15, 17-21</td> </tr> <tr> <td>A</td> <td>CN 103699653 A (沈阳航空航天大学) 2014年 4月 2日 (2014 - 04 - 02) 全文</td> <td>1-21</td> </tr> <tr> <td>A</td> <td>CN 104281569 A (富士通株式会社) 2015年 1月 14日 (2015 - 01 - 14) 全文</td> <td>1-21</td> </tr> <tr> <td>A</td> <td>WO 2014067296 A1 (深圳市恩普电子技术有限公司) 2014年 5月 8日 (2014 - 05 - 08) 全文</td> <td>1-21</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 106228188 A (北京市商汤科技开发有限公司) 2016年 12月 14日 (2016 - 12 - 14) 权利要求1-10, 说明书第[0068]-[0183]段	1-21	X	CN 104252627 A (广州华多网络科技有限公司) 2014年 12月 31日 (2014 - 12 - 31) 说明书第[0036]-[0058]段, 图1	1-2, 4-6, 8-11, 13-15, 17-21	X	CN 103679190 A (富士通株式会社) 2014年 3月 26日 (2014 - 03 - 26) 说明书第[0024]-[0030]段	1-2, 4-6, 8-11, 13-15, 17-21	A	CN 103699653 A (沈阳航空航天大学) 2014年 4月 2日 (2014 - 04 - 02) 全文	1-21	A	CN 104281569 A (富士通株式会社) 2015年 1月 14日 (2015 - 01 - 14) 全文	1-21	A	WO 2014067296 A1 (深圳市恩普电子技术有限公司) 2014年 5月 8日 (2014 - 05 - 08) 全文	1-21
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 106228188 A (北京市商汤科技开发有限公司) 2016年 12月 14日 (2016 - 12 - 14) 权利要求1-10, 说明书第[0068]-[0183]段	1-21																					
X	CN 104252627 A (广州华多网络科技有限公司) 2014年 12月 31日 (2014 - 12 - 31) 说明书第[0036]-[0058]段, 图1	1-2, 4-6, 8-11, 13-15, 17-21																					
X	CN 103679190 A (富士通株式会社) 2014年 3月 26日 (2014 - 03 - 26) 说明书第[0024]-[0030]段	1-2, 4-6, 8-11, 13-15, 17-21																					
A	CN 103699653 A (沈阳航空航天大学) 2014年 4月 2日 (2014 - 04 - 02) 全文	1-21																					
A	CN 104281569 A (富士通株式会社) 2015年 1月 14日 (2015 - 01 - 14) 全文	1-21																					
A	WO 2014067296 A1 (深圳市恩普电子技术有限公司) 2014年 5月 8日 (2014 - 05 - 08) 全文	1-21																					
☐ 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 2017年 9月 18日		国际检索报告邮寄日期 2017年 10月 11日																					
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (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 张乾桢 电话号码 (86-10)62414072																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/091432

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	106228188	A	2016年 12月 14日	无			
CN	104252627	A	2014年 12月 31日	无			
CN	103679190	A	2014年 3月 26日	JP	2014063494	A	2014年 4月 10日
				JP	6179305	B2	2017年 8月 16日
CN	103699653	A	2014年 4月 2日	无			
CN	104281569	A	2015年 1月 14日	CN	104281569	B	2017年 8月 1日
WO	2014067296	A1	2014年 5月 8日	CN	102982314	A	2013年 3月 20日
				CN	102982314	B	2016年 5月 25日