

基于多模态信息的聚类方法、服务器及存储介质

本申请要求于 2023 年 5 月 5 日提交中国专利局、申请号为 202310516381.8、发明名称为“基于多模态信息的聚类方法、服务器及存储介质”的中国专利申请的优先权，其全部
5 内容通过引用结合在本申请中。

技术领域

本申请涉及计算机技术，尤其涉及一种基于多模态信息的聚类方法、服务器及存储介
10 质。

背景技术

随着云计算和移动互联网的发展，各服务平台产生了海量的数据，这些数据是缺失多
模态的，不同的对象含有不同模态的数据，不同的模态具有不同的属性，利用多模态之间的
互补性，剔除模态间的冗余性，从而学习到更好的特征属性显得很重要。聚类是其中一
15 个重要的应用。

目前，主流的多模态聚类方法通常基于特征间的欧式距离或余弦相似度进行聚类，聚
类的准确度和召回率较低。

发明内容

20 本申请提供一种基于多模态信息的聚类方法、服务器及存储介质，用以解决主流的多
模态聚类方法聚类的准确度和召回率低问题。

第一方面，本申请提供一种基于多模态信息的聚类方法，包括：获取待聚类的目标对
象的多模态的特征信息；根据所述目标对象的多模态的特征信息，自学习生成各模态下的
后验匹配概率函数；根据所述各模态下的后验匹配概率函数，计算任意两个目标对象在多
25 模态下的后验匹配概率；根据任意两个目标对象在多模态下的后验匹配概率，对所述目标
对象进行聚类，生成所述目标对象的聚类结果。

第二方面，本申请提供一种基于多模态信息的聚类方法，包括：获取待聚类的图像的
多模态的特征信息，所述多模态的特征信息包括地理位置信息、所包含对象的至少一种模
态的特征信息；根据所述图像的多模态的特征信息，自学习生成各模态下的后验匹配概率
30 函数；根据所述各模态下的后验匹配概率函数，计算任意两个图像在多模态下的后验匹
配概率，其中两个图像在多模态下的后验匹配概率是两个图像包含同一对象的概率；根据任
意两个图像在多模态下的后验匹配概率，将所包含同一对象的图像聚为一类，得到所述图
像的聚类结果。

第三方面，本申请提供一种基于多模态信息的聚类方法，应用于端侧设备，包括：获
35 取待聚类的目标对象的多模态的特征信息；向服务器发送待聚类的目标对象的多模态的
特征信息；接收所述服务器发送的所述目标对象的聚类结果，所述目标对象的聚类结果是通
过如第一方面所述的方法生成的；根据所述聚类结果，执行如下至少一项处理逻辑：信息
推送、信息搜索、数据分类。

第四方面，本申请提供一种服务器，包括：处理器，以及与所述处理器通信连接的存储器；所述存储器存储计算机执行指令；所述处理器执行所述存储器存储的计算机执行指令，以实现如第一方面或第二方面所述的方法。

第五方面，本申请提供一种计算机可读存储介质，所述计算机可读存储介质中存储有计算机执行指令，所述计算机执行指令被处理器执行时用于实现如第一方面或第二方面所述的方法。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本申请的实施例，并与说明书一起用于解释本申请的原理。

- 图 1 为本申请所适用的一种示例性的基于多模态信息聚类的系统架构图；
- 图 2 为本申请一示例性实施例提供的基于多模态信息的聚类方法流程图；
- 图 3 为本申请另一示例性实施例提供的基于多模态信息的聚类方法流程图；
- 图 4 为本申请一示例性实施例提供的基于多模态信息的确定目标对象的多模态下的后验匹配概率的框架图；
- 图 5 为本申请另一示例性实施例提供的基于多模态信息的聚类方法流程图；
- 图 6 为本申请一示例性实施例提供的基于多模态信息的聚类方法的交互图；
- 图 7 为本申请一示例性实施例提供的基于多模态信息的聚类装置的结构示意图；
- 图 8 为本申请实施例提供的一种服务器的结构示意图。

通过上述附图，已示出本申请明确的实施例，后文中将有更详细的描述。这些附图和文字描述并不是为了通过任何方式限制本申请构思的范围，而是通过参考特定实施例为本领域技术人员说明本申请的概念。

具体实施方式

这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。

首先对本申请所涉及的名词进行解释：

模态：每一种信息的来源或者形式都可以称为一种模态。例如，人有触觉、听觉、视觉、嗅觉、味觉等，可以作为多种不同模态；信息的媒介有语音、视频、文字等，不同媒介可以作为多种不同模态；多种多样的传感器，如雷达、红外、加速度计等，不同的来源设备可以作为不同模态。例如，人员的多模态的特征信息，可以包含各类生物特征等视觉

信息, 基本资料信息, 时间空间等地理信息。在电商场景, 物品的多模态的特征信息, 可以包含图片特征、文字描述、物品类别、销量信息、评价信息等。

聚类: 将物理或抽象对象的集合分成由类似的对象组成的多个簇的过程被称为聚类。由聚类所生成的簇是一组对象的集合, 这些对象与同一个簇中的对象彼此相似, 与其他簇中的对象相异。

多模态聚类: 对拥有多元模态信息的对象进行聚类。在一示例场景中, 将在多个不同区域采集到的图像, 基于聚类对象的多模态信息, 将同一对象的图像汇集到同一个类别。在电商场景, 基于物品的多模态信息, 将不同店铺的物品按照物品唯一性维度汇集到同一类。

10 缺失多模态: 具有多模态信息的对象集合中部分对象的某个或者多个模态的信息缺失。例如, 部分对象不存在关联人员, 部分物品不存在关联的评价信息等。

欧式距离: 指在欧几里得空间中两点之间的距离。

余弦相似度: 通过计算两个向量的夹角余弦值来度量两个向量之间的相似度。

匹配概率: 两个对象属于同一个聚类类别(簇)的概率。

15 条件独立: 如果 $P(X,Y|Z)=P(X|Z)P(Y|Z)$, 或等价地 $P(X|Y,Z)=P(X|Z)$, 则称事件 X , Y 对于给定事件 Z 是条件独立的, 也就是说, 当 Z 发生时, X 发生与否与 Y 发生与否是无关的。在本申请中, 多个模态之间需要满足对给定类别标签和限定场景条件独立。例如, 不同生物特征相似度给定是否同一人满足条件独立, 生物特征相似度、空间距离给定是否同一人满足条件独立。

20 单一模态下的后验匹配概率: 是指基于对象的单一模态的特征信息确定的后验匹配概率。

多模态下的后验匹配概率: 是指基于对象的多模态的特征信息确定的对象间的后验匹配概率。

25 单一模态下的后验匹配概率函数: 是指基于单一模态的特征信息的后验匹配概率函数。

多模态下的后验匹配概率函数: 是指基于对象的多模态的特征信息的后验匹配概率函数, 包括两个模态下的交叉后验匹配概率函数, 以及三个及以上模态的混合后验匹配概率函数。

30 随着云计算和移动互联网的发展, 各服务平台产生了海量的数据。例如, 通过移动终端拍摄上传的大规模短视频数据, 通过获取视频的内容或图像、声音、文本等多模态信息进行在线聚类, 可用于同类视频聚类、检索等应用场景。再比如人们在社交网络上的表达方式变得越来越丰富, 可以通过图文、视频等表达自己的情绪和观点, 利用图片、视频、文本等多模态信息进行社交网络动态内容聚类和动态情绪聚类等, 向用户推荐同类动态内容。再比如越来越多的购物平台可提供搜索服务, 购物平台中不同商家对同一类物品有视
35 频、图片、文字、价格等不同的描述等多模态信息, 通过多模态信息聚类将不同商家的同

类物品聚类在一起，可以给用户提供视频、图片、文字相结合的搜索服务。

目前，主流的多模态聚类方法通常基于特征间的欧式距离或余弦相似度进行聚类，聚类的准确度和召回率较低。有的多模态聚类方法将关键特征的欧式距离或余弦相似度作为对象间的相似度，将其他模态的特征信息作为规则优化对象间的相似度，例如两张当图片的采集地址位于同一个街道时相似度提升 5%，位于同一个区域时相似度提升 2%等，规则设置不够灵活，影响聚类效果。

针对目前主流多模态聚类方法聚类的准确度和召回率较低的问题，本申请提供一种基于多模态信息的聚类方法，根据待聚类的目标对象的多模态的特征信息，自学习生成各单一模态下的后验匹配概率函数，无需使用标注信息即可自动学习到各模态下的后验匹配概率函数；根据各单一模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率；根据任意两个目标对象在多模态下的后验匹配概率，对目标对象进行聚类，生成目标对象的聚类结果，通过使用自学习生成的后验匹配概率函数，基于目标对象的多模态特征信息，计算目标对象属于同一类簇的后验匹配概率，作为目标对象间相似性的度量距离，相较于主流方法中使用特征间的欧式距离或余弦相似度，使用目标对象间的后验匹配概率能够更精准地度量目标对象间的相似性，基于目标对象间的后验匹配概率进行聚类，可以大大提升聚类的准确度和召回率。

本申请提供的基于多模态信息的聚类方法，可以自学习生成各单一模态下的后验匹配概率函数，从而可以无偏地快速应用到不同数据分布场景，例如，不同区域的图片库的聚集整理、同类商品的聚类等。

图 1 为本申请所适用的一种示例性的基于多模态信息聚类的系统架构图，如图 1 所示，该系统架构具体可包括服务器、端侧设备和数据服务设备。

其中，服务器具体可为设置在本地的服务器、也可以是设置在云端的服务器集群。服务器与各个端侧设备间具有可通信的通信链路，能够实现服务器与各个端侧设备间的通信连接。

端侧设备是指向服务器提供待聚类的目标对象的多模态特征信息的电子设备，具体可以是具有网络通信功能、运算功能以及信息显示功能的硬件设备，其包括但不限于用户使用的智能终端、平板电脑、个人电脑等终端设备，物联网设备，服务平台或机构的服务器等。

数据服务设备可以是能够向服务器提供多模态的信息的设备，数据服务设备具体可以是数据采集设备、或者可以收集数据采集设备所采集的一个或者多个模态的数据的服务器。服务器可以从一个或者多个数据服务设备获取目标对象的多模态的信息，并进行数据预处理及特征提取，得到目标对象的多模态的特征信息。

端侧设备向服务器发送对目标对象的聚类请求/指令，或者调用服务器提供的应用程序接口，以触发服务器基于目标对象的多模态信息进行聚类。服务器响应于对目标对象的聚类请求/指令、或应用程序接口被调用，从一个或者多个数据服务设备获取待聚类的目标对

象的多模态的信息，并生成目标对象的多模态的特征信息。服务器根据目标对象的多模态的特征信息，自学习生成各模态下的后验匹配概率函数；根据各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率；根据任意两个目标对象在多模态下的后验匹配概率，对目标对象进行聚类，生成目标对象的聚类结果。服务器向端侧设备发送聚类结果，端侧设备可以根据聚类结果执行如下至少一项处理逻辑：信息推送、信息搜索、数据分类。在另一实施方式中，服务器可以根据聚类结果执行如下至少一项处理逻辑：信息推送、信息搜索、数据分类，并将执行结果发送至端侧设备。

在其他实施方式中，可由端侧设备从数据服务设备获取目标对象的多模态数据，不同模态的数据可以来源于不同的设备，对目标对象的多模态数据进行预处理，生成目标对象的多模态的特征信息，并将目标对象的多模态的特征信息发送至服务器。

一种可能的应用场景为不同区域图像库的聚集整理场景，将在不同区域采集的图像，基于图像采集的地理位置信息、图像包含对象的至少一种模态的特征信息，将属于同一对象的图像汇聚到同一类簇中。例如，基于不同区域采集的图像，基于图像采集的地理位置信息，图像所包含的人的属性特征、生物特征，对图像进行多模态聚类，将同一人的图像汇聚到同一类簇中。例如，基于不同区域采集的机动车图像，基于机动车图像采集的地理位置信息、机动车图像所包含的机动车的车牌号、所属人身份、外观特征等，对机动车图像进行多模态聚类，将同一机动车的图像汇聚到同一类簇中。

该场景中，待聚类的目标对象为图像，服务器可以为具有信息库聚集整理需求并获得授权的系统平台的服务器，具体可以是部署在本地或云平台的服务器。服务器从该平台内部、或系统平台获得数据授权的图像采集设备，获取待聚类的图像及图像的地理位置信息，通过图像识别技术识别图像所包含的对象，并获取所包含对象的至少一种模态的特征信息，从而获取到待聚类的图像的多模态的特征信息。服务器根据图像的多模态的特征信息，自学习生成各模态下的后验匹配概率函数；根据各模态下的后验匹配概率函数，计算任意两个图像在多模态下的后验匹配概率，两个图像在多模态下的后验匹配概率是两个图像包含同一对象的概率；根据任意两个图像在多模态下的后验匹配概率，将所包含同一对象的图像聚为一类，得到图像的聚类结果。示例性地，服务器根据聚类结果，按照所属类簇，将同一类簇的图像与所包含的对象关联，并将图像进行分类存储，实现信息库聚集整理功能。其中，信息库中的待聚类的信息可以是图像、音频、或其他类型的数据信息，此处不做具体限定。

另一种可能的应用场景为电商场景，电商台中不同商家对同一类物品有视频、图片、介绍文字、价值信息等不同模态的信息，通过多模态聚类将不同商家的同类物品聚类到同一类簇中，实现基于视频、图片、介绍文字、价值信息等多模态信息的物品搜索、物品推荐等服务。

该场景中，待聚类的目标对象为物品，服务器可以为电商平台的服务器，具体可以是部署在本地或云平台的服务器。服务器从该电商平台内部获得不同商家对各物品的视频、

图片、介绍文字、价值信息等多模态信息，并进行数据预处理生成物品的多模态的特征信息。服务器根据物品的多模态的特征信息，自学习生成各模态下的后验匹配概率函数；根据各模态下的后验匹配概率函数，计算任意两个物品在多模态下的后验匹配概率，两个物品在多模态下的后验匹配概率是两个物品是同类物品的概率；根据任意两个物品在多模态下的后验匹配概率，将同类物品聚为一类，得到聚类结果。示例性地，基于聚类结果，服务器可以将属于同一类簇的物品进行关联，根据用户输入的查询信息或用户当前感兴趣（如点击浏览、加购、关注）的物品，搜索同类物品并推荐给用户。

本实施例的方法，具体可以用于各种应用领域中需要进行聚类的场景，包括但不限于信息库聚集整理、电商、教育、医疗等领域，可以对数据对象等抽象对象、物理对象等进行聚类，此处对于待聚类的目标对象的类型、聚类结果的具体应用场景不做具体限定。

下面以具体地实施例对本申请的技术方案以及本申请的技术方案如何解决上述技术问题进行详细说明。下面这几个具体的实施例可以相互结合，对于相同或相似的概念或过程可能在某些实施例中不再赘述。下面将结合附图，对本申请的实施例进行描述。

图2为本申请一示例性实施例提供的基于多模态信息的聚类方法流程图。本实施例的执行主体为上述基于多模态信息聚类的系统架构中的服务器，本实施例提供的方法用于基于目标对象的多模态的特征信息实现目标对象的多模态聚类。如图2所示，该方法具体步骤如下：

步骤S201、获取待聚类的目标对象的多模态的特征信息。

本实施例中，为了对待聚类的多个目标对象进行精准聚类，服务器获取目标对象的多个不同模态的特征信息。在不同的应用场景中，服务器获取多模态特征信息的来源和方式可以不同。

示例性地，服务器可以从所在系统平台内部的一个或多个数据服务设备获取目标对象的多模态的特征信息，不同模态的特征信息可以来源于相同或不同的数据服务设备。

示例性地，服务器还可以从所在系统平台内部的一个或多个数据服务设备获取目标对象的多模态数据，进行数据预处理和特征提取，生成目标对象的多模态的特征信息。

示例性地，服务器还可以接收端侧设备发送的目标对象的多模态的特征信息，端侧设备可以是向服务器发送聚类请求/指令，或者调用服务器提供的聚类应用程序接口的设备。端侧设备可以从一个或者多个数据服务设备提供的目标对象的多模态数据，进行数据预处理和特征提取，得到目标对象的多模态的特征信息。

其中，目标对象为以下任意一种：实体对象、包含实体对象的图像、包含实体对象的文本，目标对象的多模态的特征信息包含如下至少一种：图像特征、文本特征、属性特征、地理位置特征。

目标对象的多模态的特征信息包括但不限于：目标对象的一种或多种模态的属性信息（可以包括地理位置特征、类别特征等）、目标对象所包含内容的一种或多种模态的特征（可以包括图像特征、文本特征等）、目标对象的关联对象的一种或多种模态的特征（可

以包括图像特征、文本特征、属性特征、地理位置特征等)。

例如，在图像库聚集整理场景中，待聚类的目标对象可以是图像，目标对象的多模态的特征信息包括但不限于：图像采集的地理位置信息，图像包含对象的至少一种模态的特征信息，如图像所包含的人的生物特征，图像所包含的机动车的车牌号、所属人信息、外观特征等。

例如，在电商场景中，待聚类的目标对象可以是物品，目标对象的多模态的特征信息包括但不限于：视频、图片、介绍文字、价值信息。

步骤 S202、根据目标对象的多模态的特征信息，自学习生成各模态下的后验匹配概率函数。

本实施例中，通过自学习的方式，基于目标对象的多模态的特征信息，自动生成各模态下的后验匹配概率函数。任一模态下的后验匹配概率函数可以用于基于目标对象的该模态的特征信息，计算任意两个目标对象属于同一类簇的后验匹配概率，这一后验匹配概率体现了在当前模态的特征信息上目标对象间的相似性。

步骤 S203、根据各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率。

在自学习生成各模态下的后验匹配概率函数之后，基于各模态下的后验匹配概率函数，以及目标对象的多模态的特征信息，计算确定任意两个目标对象在多模态下的后验匹配概率，该后验匹配概率体现目标对象的多模态的特征信息的总体相似性。相较于基于单一模态的特征信息计算的后验匹配概率，使用不同目标对象在多模态下的后验匹配概率度量目标对象的相似性，更加精准。

步骤 S204、根据任意两个目标对象在多模态下的后验匹配概率，对目标对象进行聚类，生成目标对象的聚类结果。

在得到任意两个目标对象在多模态下的后验匹配概率之后，基于任意两个目标对象在多模态下的后验匹配概率，对目标对象进行聚类，将在多模态下的后验匹配概率较大的目标对象汇聚到同一类簇中，将在多模态下的后验匹配概率较小的目标对象分到不同的类簇中，实现目标对象的精准聚类。

该步骤中可以采用现有的任意一种聚类算法，使用目标对象在多模态下的后验匹配概率作为目标对象间相似度的度量距离，进行聚类，此处对于具体聚类流程不再赘述。

本实施例的方法，根据待聚类的目标对象的多模态的特征信息，自学习生成各单一模态下的后验匹配概率函数，无需使用标注信息即可自动学习到各模态下的后验匹配概率函数；根据各单一模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率；根据任意两个目标对象在多模态下的后验匹配概率，对目标对象进行聚类，生成目标对象的聚类结果，通过使用自学习生成的后验匹配概率函数，基于目标对象的多模态特征信息，计算目标对象属于同一类簇的后验匹配概率，作为目标对象间相似性的度量距离，相较于主流方法中使用特征间的欧式距离或余弦相似度，使用目标对象间的后验匹配

概率能够更精准地度量目标对象间的相似性，基于目标对象间的后验匹配概率进行聚类，可以大大提升聚类的准确度和召回率。

图3为本申请另一示例性实施例提供的基于多模态信息的聚类方法流程图。在前述方法实施例的基础上，本实施例中对基于多模态信息的聚类方法进行详细地说明，如图3所示，该方法具体步骤如下：

步骤S301、获取待聚类的目标对象的多模态的特征信息。

该步骤与前述步骤S201实现方式相同，此处不再赘述。

在一可选实施例中，服务器可以接收端侧设备发送的待聚类的目标对象的多模态的特征信息。进一步地，在得到目标对象的最终的聚类结果之后，服务器向端侧设备输出目标对象的聚类结果。

在获取到待聚类的目标对象的多模态的特征信息之后，通过步骤S302-S304，实现步骤S202，根据目标对象的多模态的特征信息，自学习生成各模态下的后验匹配概率函数。

步骤S302、根据目标对象的同一模态的特征信息，计算该模态下任意两个目标对象的相似度。

具体地，可以将目标对象的各模态的特征信息编码为特征向量，基于任一模态的特征向量，计算任意两个目标对象的该模态的特征向量间的欧式距离或余弦相似度，作为该模态下任意两个目标对象的相似度。

任一模态下的任意两个目标对象的相似度可以存储为矩阵，得到该模态下的相似度矩阵。用M表示模态的数量，各个模态下的相似度矩阵可以表示为 $w^{(1)}$ ， $w^{(2)}$ ，...， $w^{(M)}$ 。

步骤S303、根据各模态下任意两个目标对象的相似度，构建各模态下的多个样本对，样本对包括两个目标对象。

该步骤中，对于任一模态，基于该模态下任意两个目标对象的相似度，构建该模态下的多个样本对。基于相同的方式构建各个模态下的样本对，得到各模态下的多个样本对。

该步骤的一种可能的实现方式中，在构建任一模态下的多个样本对时，可以将任意两个目标对象组成样本对；或者，可以随机选择两个不同的目标对象组成样本对。

该步骤的另一种可能的实现方式中，在构建任一模态下的多个样本对时，还可以根据该模态下任意两个目标对象的相似度，确定该模态下各目标对象的K近邻对象；将该模态下任意目标对象分别与K近邻对象组成样本对，得到包含该目标对象的K个样本对，将目标对象的数量表示为N，可以得到该模态下的 $N \times K$ 个样本对。其中K和N均为正整数。通过这一方式构建的高质量样本对，并且减少样本对的数量，可以提升自学习生成各模态下的后验匹配概率函数的质量和效率。

其中，该模态下任一目标对象的K近邻对象，基于该模态下其他目标对象与该目标对象的相似度，采用K近邻（K Nearest Neighbor，简称KNN）算法确定与该目标对象的相似度较高的K个目标对象得到。K可以根据实际应用场景的需要进行设置和调整，此处不做具体限定。

步骤 S304、根据各模态下的多个样本对的相似度，采用优化算法，自学习生成各模态下的后验匹配概率函数。

在完成构建各模态下的多个样本对之后，可以基于所构建的样本对及样本对的相似度，使用优化算法，自学习生成各模态下的单一模态的后验匹配概率函数。具体可以通过如下步骤 S3041-S3044 实现：

步骤 S3041、将各模态下的后验匹配概率函数初始化为单调递增的分段函数，分段函数包含待求解的分段参数。

在实际应用中，随着目标对象间相似度的增大，目标对象间的后验匹配概率也随之增大。本实施例中，将各模态下的样本对按照相似度由小到大的顺序排列后分成 I 个分段集合，I 为正整数。用 T 表示单一模态下样本对的数量， $T=N \times K$ ，每个分段集合包含的样本对的数量为 $\frac{T}{I}$ 。将各模态下的单一模态的后验匹配概率函数，初始化为包含 I 个分段的单调递增的分段函数。用 M 表示模态的数量，可以将第 m 个模态下的后验匹配概率函数表示为如下分段函数：

$$\begin{aligned}
 & f^{(m)}(x): (x_{i_m}^{(m)}, f_{i_m}^{(m)}) \\
 15 \quad & \text{受约束于 (s.t.): } x_1^{(m)} < x_2^{(m)} < \dots < x_I^{(m)}, \\
 & f_1^{(m)} < f_2^{(m)} < \dots < f_I^{(m)}, \quad (1) \\
 & f_1^{(m)} = 0, \quad f_I^{(m)} = 1, \\
 & |z_{i_m}^{(m)}| = \text{length}(r_{i_m}^{(m)}) = \frac{T}{I}
 \end{aligned}$$

其中，式 (1) 中， $z_{i_m}^{(m)}$ 表示第 m 个模态下的第 i_m 个分段集合。 $|z_{i_m}^{(m)}|$ 表示第 m 个模态下的第 i_m 个分段集合包含的样本对的数量。 $r_{i_m}^{(m)}$ 表示第 m 个模态下的第 i_m 个分段集合 $z_{i_m}^{(m)}$ 中样本对的相似度的集合。 $\text{length}(r_{i_m}^{(m)})$ 表示 $r_{i_m}^{(m)}$ 中相似度值的数量。 $x_{i_m}^{(m)}$ 表示第 m 个模态下的第 i_m 个分段集合中样本对的相似度平均值，由于分段集合是将样本对按照相似度由小到大的顺序排列后分成的，因此满足 $x_1^{(m)} < x_2^{(m)} < \dots < x_I^{(m)}$ 。 $f_{i_m}^{(m)}$ 表示第 m 个模态下的后验匹配概率函数的第 i_m 个分段的函数值，是待求解的分段参数。该分段函数为单调递增函数，因此满足 $f_1^{(m)} < f_2^{(m)} < \dots < f_I^{(m)}$ 。后验匹配概率的最小值为 0，最大值为 1，因此满足 $f_1^{(m)} = 0, f_I^{(m)} = 1$ 。 $i_m \in [1, 2, \dots, I], m \in [1, 2, \dots, M]$ 。

通过式 (1)，可以将各模态下的后验匹配概率函数初始化为单调递增的分段函数，其中每个分段的自变量为对应分段集合中样本对相似度的平均值，函数值为待求解的分段参数，共有 $M \times I$ 个待求解的分段参数。

需要说明的是，分段函数的分段数量 I 可以根据实际应用场景进行设置，不同应用场景取值可以不同，此处不做具体限定。例如 I 可以取值为 1000、200 等。

步骤 S3042、基于各模态下的后验匹配概率函数，构建目标函数，目标函数包含待求解的分段参数。

本实施例中，各模态下的后验匹配概率函数即为单一模态的后验匹配概率函数，第 m 个模态下的单一模态的后验匹配概率函数，可以表示为 $Fsingle_{i_m}^{(m)}$ ：

$$Fsingle_{i_m}^{(m)} = f_{i_m}^{(m)} = f^{(m)}(x_{i_m}^{(m)}) \quad (2)$$

该步骤中，根据各模态下的单一模态的后验匹配概率函数，构建多模态下的后验匹配概率函数。

具体地，可以构建任意两个模态下的交叉后验匹配概率函数，第 m 和 b 模态下的交叉后验匹配概率函数，可以表示为 $Fcross_{i_m}^{(m)-(b)}$ ：

$$\begin{aligned} Fcross_{i_m}^{(m)-(b)} &= \frac{1}{|z_{i_m}^{(m)}|} \sum_{x \in r_{i_m}^{(m)}, x_{i_b}^{(b)} \in r_{i_m}^{(m)-(b)}} f^{(b)}(x_{i_b}^{(b)} | x) \\ &= \frac{1}{|z_{i_m}^{(m)}|} \sum_{x \in r_{i_m}^{(m)}, x_{i_b}^{(b)} \in r_{i_m}^{(m)-(b)}} f_{i_b}^{(b)} \end{aligned} \quad (3)$$

其中，式 (3) 中， $r_{i_m}^{(m)-(b)}$ 表示第 m 个模态下的第 i_m 个分段集合中的样本对在第 b 个模态下所属分段集合中样本对的相似度平均值的集合。 $x_{i_b}^{(b)}$ 为 $r_{i_m}^{(m)-(b)}$ 中的一个相似度平均值。 $i_m, i_b \in [1, 2, \dots, I]$ 。 $m, b \in [1, 2, \dots, M]$ 。其他各个字符的含义与前述内容中的含义一致，此处不再赘述。

进一步地，对于任意两个目标对象的给定 M 个模态的特征信息的相似度 $\{x_{i_1}^{(1)}, x_{i_2}^{(2)}, \dots, x_{i_M}^{(M)}\}$ ， M 个模态的联合概率分布函数可以表示为：

$$\begin{aligned} Fjoint(x_{i_1}^{(1)}, x_{i_2}^{(2)}, \dots, x_{i_M}^{(M)}) &= \frac{\prod_{m=1}^M f^{(m)}(x_{i_m}^{(m)})}{\prod_{m=1}^M f^{(m)}(x_{i_m}^{(m)}) + \prod_{m=1}^M (1 - f^{(m)}(x_{i_m}^{(m)}))} \\ &= \frac{\prod_{m=1}^M f_{i_m}^{(m)}}{\prod_{m=1}^M f_{i_m}^{(m)} + \prod_{m=1}^M (1 - f_{i_m}^{(m)})} \end{aligned} \quad (4)$$

进一步地，构建 M 个模态的后验匹配概率函数，可以表示为：

$$Fmulti_{i_m}^{(m)} = \frac{1}{|x_{i_m}^{(m)}|} \sum_{x \in x_{i_m}^{(m)}} Fjoint(x_{i_1}^{(1)}, x_{i_2}^{(2)}, \dots, x_{i_M}^{(M)} | x) \quad (5)$$

式 (4) 和 (5) 中各个字符的含义与前述内容中的含义一致, 此处不再赘述。

在实际应用中, 对于任意样本对的集合, 基于一个模态统计的样本对属于同一类别的概率的均值, 与基于任意多模态统计的样本对属于同一类别的概率的均值应该相等, 其期望值与真实的属于同一类别的样本度的占比相等。也即是对于同一样本对集合, 基于单一模态的后验匹配概率函数、基于两个模态的交叉后验匹配概率函数、基于三个及以上模态的后验匹配概率函数, 确定的后验匹配概率分布应当具有一致性。基于此, 根据各模态下的后验匹配概率函数和多模态下的后验匹配概率函数, 构建一致性损失函数。

示例性地, 可以构建如下所示的一致性损失函数:

$$L_{consistency} = \frac{1}{M} \sum_m \left(\frac{1}{I} \sum_{i_m} D \left(Fsingle_{i_m}^{(m)}, Fmulti_{i_m}^{(m)} \right) \right) + \frac{1}{M} \sum_m \left(\frac{1}{I} \sum_{i_m} \sum_{b \neq m} D \left(Fsingle_{i_m}^{(m)}, Fcross_{i_m}^{(m)-(b)} \right) \right) \quad (6)$$

其中, $D()$ 表示距离函数, 可以求欧式距离或余弦相似度。其他各个字符的含义与前述内容中的含义一致, 此处不再赘述。

鉴于式 (6) 中 $Fsingle_{i_m}^{(m)}$ 受到 $Fmulti_{i_m}^{(m)}$ 和 $Fcross_{i_m}^{(m)-(b)}$ 的约束, 式 (6) 中学习这三者的一致性损失较困难。基于 $Fsingle_{i_m}^{(m)}$ 、 $Fmulti_{i_m}^{(m)}$ 和 $Fcross_{i_m}^{(m)-(b)}$, 定义如下混合后验匹配概率函数:

$$Fmix_{i_m}^{(m)} = \sqrt{Fmulti_{i_m}^{(m)} \frac{1}{M} \left(Fsingle_{i_m}^{(m)} + \sum_{b \neq m} Fcross_{i_m}^{(m)-(b)} \right)} \quad (7)$$

式 (7) 的混合后验匹配概率函数用于约束 $Fsingle_{i_m}^{(m)}$ 和 $Fcross_{i_m}^{(m)-(b)}$ 的值。其他各个字符的含义与前述内容中的含义一致, 此处不再赘述。

进一步地, 基于式 (7), 式 (6) 的一致性损失函数可以使用如下式 (8) 的一致性损失函数代替:

$$L_{consistency} = \frac{1}{M} (L_{consistency1} + L_{consistency2}) \quad (8)$$

其中, $L_{consistency1} = \frac{1}{M} \sum_m \left(\frac{1}{I} \sum_{i_m} D \left(Fsingle_{i_m}^{(m)}, Fmix_{i_m}^{(m)} \right) \right)$,

$L_{consistency2} = \frac{1}{M} \sum_m \left(\frac{1}{I} \sum_{i_m} \sum_{b \neq m} D \left(Fcross_{i_m}^{(m)-(b)}, Fmix_{i_m}^{(m)} \right) \right)$ 。

其中, $L_{consistency1}$ 约束单一模态的后验匹配概率函数与混合后验匹配概率函数的一致性, $L_{consistency2}$ 约束两个模态的交叉后验匹配概率函数与混合后验匹配概率函数的一致性。

在实际应用中，无论是单一模态的后验匹配概率函数、两个模态的交叉后验匹配概率函数，还是混合后验匹配概率函数，其最小值是 0，最大值是 1，基于此构建约束损失函数。

示例性地，可以构建如下所示的约束损失函数：

$$\begin{aligned}
 5 \quad L_{\text{constraint}} = & \sum_m \left(\sum_{i \in r_i} D(F_{\text{multi}}^{(m)}_i, 0) + \sum_{j \in r_j} D(F_{\text{multi}}^{(m)}_j, 1) \right) \\
 & + \sum_m \left(\sum_{i \in r_i} D(F_{\text{single}}^{(m)}_i, 0) + \sum_{j \in r_j} D(F_{\text{single}}^{(m)}_j, 1) \right) \\
 & + \sum_m \left(\sum_{b \neq m} \left(\sum_{i \in r_i} D(F_{\text{cross}}^{(m)-(b)}_i, 0) + \sum_{j \in r_j} D(F_{\text{cross}}^{(m)-(b)}_j, 1) \right) \right) \quad (9)
 \end{aligned}$$

其中， $r_i \in [1, 2, \dots, \text{indi}]$ ，指代分段函数的前 indi 段。 $r_j \in [l - \text{indj}, \dots, l - 1, l]$ ，指代分段函数中后 $\text{indj} + 1$ 段。 indi 和 indj 的取值可以根据实际应用场景进行设置和调整，不同应用场景中的取值不同，此处不做具体限定。例如，在一些场景中 $\text{indi} = 10$ ， $\text{indj} = 4$ ，在另一些场景中 $\text{indi} = 10$ ， $\text{indj} = 2$ 。

进一步地，根据式 (8) 的一致性损失函数和式 (9) 的约束损失函数，构建目标函数如下：

$$L = \lambda L_{\text{consistency}} + L_{\text{constraint}} \quad (10)$$

15 其中，式 (10) 中 λ 为预设的权重参数，用于平衡一致性损失和约束损失。 λ 的值可以根据经验值进行设置，此处不做具体限定，不同应用场景中的取值不同，例如可以取值为 80、20、2 等。其他各个字符的含义与前述内容中的含义一致，此处不再赘述。

在一可选实施例中，可以仅基于前述的约束损失函数构建目标函数。

20 步骤 S3043、根据各模态下的多个样本对的相似度，采用优化算法最小化目标函数，求解得到分段参数。

在构建好目标函数之后，根据各模态下的多个样本对的相似度，采用优化算法求，解使得目标函数取得较小值时的分段参数。具体可以采用梯度下降法、牛顿法、拟牛顿法、动量优化法、共轭梯度法、启发式优化算法中的任意一种优化算法实现，此处不做具体限定。

25 步骤 S3044、将分段参数代入分段函数，得到各模态下的后验匹配概率函数。

在求解得到分段参数之后，将分段参数代入分段函数，即可得到各模态下的后验匹配概率函数。

30 在步骤 S304 的另一可选实施方式中，还可以将各模态下的后验匹配概率函数初始化为连续单调函数，该连续单调函数包含待求解的未知参数，基于该连续单调函数构建一致性损失函数和约束损失函数，并构建目标函数，通过优化算法最小化目标函数求解得到未知参数，从而得到各模态下的连续单调的后验匹配概率函数。

在学习得到各模态下的后验匹配概率函数之后，通过步骤 S305-S306，实现步骤 S203，根据各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配

概率。

步骤 S305、根据任一模态下任意两个目标对象的相似度，使用该模态下的后验匹配概率函数，确定任意两个目标对象在该模态下的后验匹配概率。

- 5 在得到各模态下的后验匹配概率函数之后，对于任一模态，以第 m 个模态为例，根据该模态下任意两个目标对象的相似度，在该模态下的后验匹配概率函数的 I 个分段的自变量 $\{x_1^{(m)}, x_2^{(m)}, \dots, x_I^{(m)}\}$ 中，确定与这两个目标对象的相似度的绝对差值最小的自变量，将该自变量对应的后验匹配概率值作为这两个目标对象在该模态下的后验匹配概率。同理可以确定任意两个目标对象在各模态下的后验匹配概率。

- 10 步骤 S306、根据任意两个目标对象在各模态下的后验匹配概率，计算任意两个目标对象在多模态下的后验匹配概率。

该步骤中，在计算得到任意两个目标对象在各模态下的后验匹配概率之后，可以通过如下式 (11)，计算得到任意两个目标对象 i 和 j 在多模态下的后验匹配概率 $P(i, j)$ ：

$$P(i, j) = \frac{\prod_{m=1}^M P(e_{ij}=1|w_{ij}^{(m)})}{\prod_{m=1}^M P(e_{ij}=1|w_{ij}^{(m)}) + \prod_{m=1}^M P(e_{ij}=0|w_{ij}^{(m)})} \quad (11)$$

- 其中， $P(e_{ij}=1|w_{ij}^{(m)})$ 表示目标对象 i 和 j 在第 m 模态下的后验匹配概率，
15 $P(e_{ij}=0|w_{ij}^{(m)})=1-P(e_{ij}=1|w_{ij}^{(m)})$ 。其他各个字符的含义与前述内容中的含义一致，此处不再赘述。

- 可选地，该步骤中还可以将任意两个目标对象在各模态下的后验匹配概率求均值，作为两个目标对象在多模态下的后验匹配概率。可选地，该步骤中还可以根据各个模态对应的权重系数，将任意两个目标对象在各模态下的后验匹配概率加权求均值，计算结果作为
20 两个目标对象在多模态下的后验匹配概率。

- 示例性地，图 4 为本实施例提供的基于多模态信息的确定目标对象的多模态下的后验匹配概率的框架图。如图 4 所示，首先获取到目标对象的 M 个模态的特征信息。对于任意模态，计算该模态下任意两个目标对象的特征信息间的相似度，得到各模态下的相似度矩阵 $w^{(1)}, \dots, w^{(M)}$ 。根据各模态下的相似度矩阵 $w^{(1)}, \dots, w^{(M)}$ ，采用 KNN 算法构建各
25 模态下的 T 个样本对 $(p_1, q_1), (p_2, q_2) \dots (p_T, q_T)$ ，并确定 T 个样本对的相似度，例如第 m 个模态下 (p_1, q_1) 的相似度可以表示为 $w_{p_1, q_1}^{(1)}$ 的形式。将各个模态下的 T 个样本对按照相似度排序后分成 I 个分段集合，用 $a_x, b_y \dots c_z$ 分别表示 I 个分段集合中样本对的数量，用 $w_{a_1}^{(1)}$ ，

$w_{a_2}^{(1)} \dots w_{a_x}^{(1)}$ 分别表示在第 1 个模态下的第一个分段集合中 a_x 个样本对的相似度, $w_{b_1}^{(1)}$,

$w_{b_2}^{(1)} \dots w_{b_y}^{(1)}$ 分别表示在第 1 个模态下的第二个分段集合中 b_y 个样本对的相似度, $w_{c_1}^{(1)}$,

$w_{c_2}^{(1)} \dots w_{c_z}^{(1)}$ 分别表示在第 1 个模态下的最后一个分段集合中 c_z 个样本对的相似度。其他模

态下各个分段集合中样本对的相似度的表示形式类似, 例如, $w_{a_1}^{(2)}, w_{a_2}^{(2)} \dots w_{a_x}^{(2)}$ 分别表示

- 5 在第 2 个模态下的第一个分段集合中 a_x 个样本对的相似度, 此处不再一一列举。将各单一模态的后验匹配概率函数初始为分段函数, 并基于分段函数构建两个模态 (即交叉模态) 的交叉后验匹配概率函数、M 模态下的后验匹配概率函数, 基于各类后验匹配概率函数的一致性, 自学习生成各模态下的后验匹配概率函数。基于各模态下的后验匹配概率函数, 以及各模态下任意两个目标对象的相似度, 可以计算得到任意两个目标对象在各单一模态
- 10 下的后验匹配概率, 进一步地, 任意两个目标对象在各单一模态下的后验匹配概率, 计算得到任意两个目标对象在多模态下的后验匹配概率, 作为两个目标对象最终的后验匹配概率。进一步地, 基于两个目标对象最终的后验匹配概率进行聚类, 即可得到目标对象的聚类结果。

- 15 在一可选实施例, 在根据各模态下的后验匹配概率函数, 计算任意两个目标对象在多模态下的后验匹配概率之后, 还可以通过如下步骤 S307-S308, 对任意两个目标对象在多模态下的后验匹配概率进行优化, 以提升任意两个目标对象在多模态下的后验匹配概率的精准度。

步骤 S307、对于任意两个目标对象, 确定两个目标对象在各模态下的 K 近邻对象, 并确定两个目标对象的共有近邻对象。

- 20 该步骤中, 确定任意模态下任一目标对象的 K 近邻对象时, 基于该模态下其他目标对象与该目标对象的相似度, 采用 KNN 算法确定与该目标对象的相似度较高的 K 个目标对象得到。K 可以根据实际应用场景的需要进行设置和调整, 此处不做具体限定。其中 K 为正整数。

- 25 进一步地, 在确定各模态下各目标对象的 K 近邻对象之后, 可以得到任意目标对象在各个模态下的 K 近邻对象集合的并集, 该并集作为目标对象的近邻对象集合, 包含目标对象在各个模态下的 K 近邻对象。对于任意两个目标对象, 确定这两个目标对象的近邻对象集合的交集, 在该交集集中的近邻对象即为这两个目标对象的共有近邻对象。

示例性地, 用 $knn_i^{(m)}$ 表示目标对象 i 在第 m 个模态下的 K 近邻对象的集合, 那么目标

对象*i*的近邻对象集合是 $knn_i^{(1)}, knn_i^{(2)} \dots knn_i^{(M)}$ 的并集, 可以表示为 $knn_i = \{\cup knn_i^{(m)}\}$ 。

同理, 用 $knn_j^{(m)}$ 表示目标对象*j*在第 *m* 个模态下的 *K* 近邻对象的集合, 那么目标对象*j*的近邻对象集合表示为 $knn_j = \{\cup knn_j^{(m)}\}$ 。目标对象*i*和目标对象*j*的共有近邻对象的集合为 $knn_{ij} = knn_i \cap knn_j$ 。

- 5 步骤 S308、根据两个目标对象与同一共有近邻对象在多模态下的后验匹配概率, 优化两个目标对象在多模态下的后验匹配概率。

本实施例中, 根据两个目标对象与同一共有近邻对象在多模态下的后验匹配概率, 基一致性和后验匹配概率的边界约束, 对任意两个目标对象在多模态下的后验匹配概率进行优化, 以提升任意两个目标对象在多模态下的后验匹配概率的精准度。

- 10 该步骤可使用的第一种优化方式为: 计算两个目标对象与同一共有近邻对象在多模态下的后验匹配概率的乘积, 根据乘积与两个目标对象在多模态下的后验匹配概率中的最大值, 优化两个目标对象在多模态下的后验匹配概率。

示例性地, 对于计算得到的任意两个目标对象*i*和*j*在多模态下的后验匹配概率 $P(i, j)$, 可以采用如下式 (12) 进行优化:

$$15 \quad P(i, j) = \max \{P(i, j), P(i, h) \times P(h, j)\} \quad (12)$$

其中, 式 (12) 中, *h*指代目标对象*i*和目标对象*j*的任意共有近邻对象, $h \in knn_{ij}$ 。

$knn_{ij} = knn_i \cap knn_j$, 表示目标对象*i*和目标对象*j*的共有近邻对象的集合。 $\max()$ 为求最大值函数。其他各个字符的含义与前述内容中的含义一致, 此处不再赘述。

- 20 该步骤可使用的第二种优化方式为: 将两个目标对象与同一共有近邻对象在多模态下的后验匹配概率之和, 作为该共有近邻对象对应的后验匹配概率和; 计算各共有近邻对象对应的后验匹配概率和的和, 得到第一求和结果; 对于两个目标对象中的第一对象和第二对象, 计算第一对象与各共有近邻对象在多模态下的后验匹配概率之和, 得到第二求和结果, 计算第二对象与各共有近邻对象在多模态下的后验匹配概率之和, 得到第三求和结果, 并计算第二求和结果与第三求和结果之和, 得到第四求和结果; 根据第一求和结果与第四求和结果的比值, 优化两个目标对象在多模态下的后验匹配概率。

25 示例性地, 对于计算得到的任意两个目标对象*i*和*j*在多模态下的后验匹配概率 $P(i, j)$, 可以采用如下式 (13) 进行优化:

$$P(i, j) = \frac{\sum_{h \in knn_{ij}} (P(i, h) + P(h, j))}{\sum_{h_i \in knn_i} P(i, h_i) + \sum_{h_j \in knn_j} P(j, h_j)} \quad (13)$$

其中, 式 (13) 中 h_i 指代目标对象*i*的近邻对象集合中任一近邻对象, $h_i \in knn_i$ 。 h_j 指

代目标对象 j 的近邻对象集合中任一近邻对象, $h_j \in knn_j$ 。其他各个字符的含义与前述内容中的含义一致, 此处不再赘述。

该步骤中, 还可以使用上述两种优化方式对任意两个目标对象在多模态下的后验匹配概率进行优化, 并且不限定基于两种优化方式进行优化先后顺序。例如, 可以先使用第一种优化方式对任意两个目标对象在多模态下的后验匹配概率进行优化, 再对第一种优化方式的优化结果使用第二种优化方式进行第二次优化; 或者, 可以先使用第二种优化方式对任意两个目标对象在多模态下的后验匹配概率进行优化, 再对第二种优化方式的优化结果使用第一种优化方式进行第二次优化。

步骤 S309、根据任意两个目标对象在多模态下的后验匹配概率, 对目标对象进行聚类, 生成目标对象的聚类结果。

在得到任意两个目标对象在多模态下的后验匹配概率之后, 基于任意两个目标对象在多模态下的后验匹配概率对目标对象进行聚类, 生成目标对象的聚类结果。

该步骤中可以采用现有的任意一种聚类算法, 使用目标对象在多模态下的后验匹配概率作为目标对象间相似度的度量距离, 进行聚类, 此处对于具体聚类流程不再赘述。

本实施例中, 对于任意两个目标对象, 其中目标对象的一个或者多个模态的特征信息缺失时, 可以忽略该模态的特征信息, 基于剩余不存在信息缺失的各模态的特征信息, 计算这两个目标对象在剩余多模态下的后验匹配概率, 从而在缺失部分模态的特征信息时, 在不补充缺失模态的特征信息的情况下, 仍然可以较精准地进行聚类。同时, 充分利用了多模态的特征信息进行聚类, 关注了不同模态之间的重要性, 提升了聚类的精准度和召回率。并且, 通过 KNN 算法构建各模态下的 T 个样本对, 基于各模态下 T 个样本对的相似度, 自学习生成各模态下的后验匹配概率函数, 可以缩短自学习生成各模态下的后验匹配概率函数的时间, 从而提升聚类的效率。

本申请提供的基于多模态信息的聚类方法, 可以自学习生成各单一模态下的后验匹配概率函数, 从而可以无偏地快速应用到不同数据分布场景, 例如, 不同区域的图像库的聚集整理、同类商品的聚类等。

图 5 为本申请另一示例性实施例提供的基于多模态信息的聚类方法流程图。本实施例提供的方法可以应用于不同区域图像库的聚集整理场景中, 将不同区域信息库中, 在不同区域采集的图像, 基于图像采集的地理位置信息、图像包含对象的至少一种模态的特征信息, 将属于同一对象的图像汇聚到同一类簇中。例如, 基于不同区域采集的图像, 基于图像采集的地理位置信息, 图像所包含的人的生物特征, 对图像进行多模态聚类, 将同一人的图像汇聚到同一类簇中。例如, 基于不同区域采集的机动车图像, 基于机动车图像采集的地理位置信息、机动车图像所包含的机动车的车牌号、所属人身份、外观特征等, 对机动车图像进行多模态聚类, 将同一机动车的图像汇聚到同一类簇中。

如图 5 所示, 方法具体步骤如下:

步骤 S501、获取待聚类的图像的多模态的特征信息，多模态的特征信息包括地理位置信息、所包含对象的至少一种模态的特征信息。

本实施例中待聚类的目标对象是图像，图像的多模态的特征信息包括图像采集的地理位置信息、以及图像所包含对象的至少一种模态的特征信息。

5 示例性地，在不同区域图像库聚集整理场景中，待聚类的图像可以是多个区域信息库中不同区域采集到的特定对象的图像，以人的图像为例，图像的多模态的特征信息可以包括图像采集的位置坐标，图像所包含的人的如下至少一种模态的特征：生物特征、步态特征、身份信息。

步骤 S502、根据图像的多模态的特征信息，自学习生成各模态下的后验匹配概率函数。

10 在获取到待聚类图像的多模态的特征信息之后，根据图像的多模态的特征信息，自学习生成各模态下的后验匹配概率函数，该步骤可以采用前述任一方法实施例中步骤 S202 类似的方式实现，该步骤中的目标对象为图像，具体参见前述实施例中的相关内容，此处不再赘述。

步骤 S503、根据各模态下的后验匹配概率函数，计算任意两个图像在多模态下的后验匹配概率，其中两个图像在多模态下的后验匹配概率是两个图像包含同一对象的概率。

该步骤的具体实现方式与前述步骤 S203 中根据各模态下的后验匹配概率函数计算任意两个目标对象在多模态下的后验匹配概率的实现方式类似，该步骤中的目标对象为图像，具体参见前述实施例中的相关内容，此处不再赘述。

20 步骤 S504、根据任意两个图像在多模态下的后验匹配概率，将所包含同一对象的图像聚为一类，得到图像的聚类结果。

在得到任意两个图像在多模态下的后验匹配概率之后，基于任意两个图像在多模态下的后验匹配概率，对图像进行聚类，将在多模态下的后验匹配概率较大的图像汇聚到同一类簇中，将在多模态下的后验匹配概率较小的图像分到不同的类簇中，实现基于图像所包含对象的唯一性进行精准第图像聚类。

25 该步骤中可以采用现有的任意一种聚类算法，使用图像在多模态下的后验匹配概率作为图像间相似度的度量距离进行聚类，此处对于具体聚类流程不再赘述。

进一步地，在得到图像聚类结果之后，服务器向图像聚类结果发送聚类结果，端侧设备可以根据图像聚类结果执行如下至少一项处理逻辑：信息推送、信息搜索、数据分类。在另一实施方式中，服务器可以根据图像聚类结果执行如下至少一项处理逻辑：信息推送、30 信息搜索、数据分类，并将执行结果发送至端侧设备。

示例性地，在得到图像聚类结果之后，服务器可以根据聚类结果，按照所属类簇，将同一类簇的图像与所包含的对象关联，并将图像进行分类存储，实现信息库聚集整理功能。

示例性地，在得到图像聚类结果之后，服务器还可以基于指定对象所关联的图像的地理位置信息，分析确定指定对象的活动范围、社交区域等信息。

35 本实施例中，在不同区域信息库的聚集整理场景中，获取待聚类的图像的地理位置信

息、及所包含对象的至少一种模态的特征信息，得到待聚类的图像的多模态的特征信息，通过自学习生成各单一模态下的后验匹配概率函数，无需使用标注信息即可自动学习到各模态下的后验匹配概率函数；根据各单一模态下的后验匹配概率函数，计算任意两个图像在多模态下的后验匹配概率；根据任意两个图像在多模态下的后验匹配概率，对图像进行

5 聚类，生成图像的聚类结果，通过使用自学习生成的后验匹配概率函数，基于图像的多模态特征信息，计算图像包含同一对象的后验匹配概率，作为图像间相似性的度量距离，相较于主流方法中使用图像特征间的欧式距离或余弦相似度，使用图像间的后验匹配概率能够更精准地度量图像所包含对象的相似性，基于图像间的后验匹配概率进行聚类，可以大大提升图像聚类的准确度和召回率。

10 图6为本申请一示例性实施例提供的基于多模态信息的聚类方法的交互图。如图6所示，基于多模态信息的聚类方法中服务器与端侧设备的交互流程如下：

步骤 S601、端侧设备获取待聚类的目标对象的多模态的特征信息。

本实施例中，端侧设备从数据服务设备获取目标对象的多模态数据，不同模态的数据可以来源于不同的设备，对目标对象的多模态数据进行预处理，生成目标对象的多模态的

15 特征信息，并将目标对象的多模态的特征信息发送至服务器。

其中，对目标对象的多模态数据进行预处理可以包括但不限于脏数据的清洗、特征提取等，在不同应用场景中所需获取的特征信息不同，可以根据实际应用场景的需要设置数据预处理和特征提取的方式，各类特征信息的提取方式可以采用提取同类特征常用的技术实现，此处不再赘述。

20 步骤 S602、端侧设备向服务器发送待聚类的目标对象的多模态的特征信息。

步骤 S603、服务器接收待聚类的目标对象的多模态的特征信息。

步骤 S604、服务器根据目标对象的多模态的特征信息，自学习生成各模态下的后验匹配概率函数。

该步骤的具体实现方式与前述步骤 S202 相同，具体参见前述实施例中的相关内容，

25 此处不再赘述。

步骤 S605、服务器根据各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率。

该步骤的具体实现方式与前述步骤 S203 相同，具体参见前述实施例中的相关内容，此处不再赘述。

30 步骤 S606、服务器根据任意两个目标对象在多模态下的后验匹配概率，对目标对象进行聚类，生成目标对象的聚类结果。

该步骤的具体实现方式与前述步骤 S204 相同，具体参见前述实施例中的相关内容，此处不再赘述。

步骤 S607、服务器向端侧设备发送目标对象的聚类结果。

35 步骤 S608、端侧设备接收服务器发送的目标对象的聚类结果。

步骤 S609、端侧设备根据聚类结果，执行如下至少一项处理逻辑：信息推送、信息搜索、数据分类。

5 示例性地，在电商场景下，通过多模态聚类将不同商家的同类物品聚类到同一类簇，得到物品聚类结果之后，可以基于视频、图片、介绍文字、价值信息等多模态信息，实现的物品搜索、物品推荐等服务。

例如，根据用户当前浏览的物品，搜索同类物品的信息，向用户展示/推送同类物品的信息，以便于用户选择所需的物品。或者，根据用户提交的图片，搜索图片中所包含物品及其同类物品的信息，并将图片中所包含物品及其同类物品的信息展示/推送给用户，以便于用户选择所需的物品。

10 例如，不同区域信息库的聚集整理场景中，将在不同区域采集的图像，基于图像采集的地理位置信息、图像包含对象的至少一种模态的特征信息，将属于同一对象的图像汇聚到同一类簇中。进一步地，根据聚类结果，按照所属类簇，将同一类簇的图像与所包含的对象关联，并将图像进行分类存储，实现信息库聚集整理功能。

本实施例提供了基于多模态信息的聚类方法中服务器与端侧设备的交互流程。

15 图 7 为本申请一示例性实施例提供的基于多模态信息的聚类装置的结构示意图。本申请实施例提供的基于多模态信息的聚类装置可以执行基于多模态信息的聚类方法实施例提供的处理流程。如图 7 所示，基于多模态信息的聚类装置 70 包括：多模态特征获取模块 71、自学习模块 72、后验匹配概率计算模块 73 和多模态聚类模块 74。

具体地，多模态特征获取模块 71 用于获取待聚类的目标对象的多模态的特征信息。

20 自学习模块 72 根据目标对象的多模态的特征信息，自学习生成各模态下的后验匹配概率函数。

后验匹配概率计算模块 73 根据各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率。

25 多模态聚类模块 74 根据任意两个目标对象在多模态下的后验匹配概率，对目标对象进行聚类，生成目标对象的聚类结果。

在一可选实施例中，在实现根据目标对象的多模态的特征信息，自学习生成各模态下的后验匹配概率函数时，自学习模块 72 还用于：

30 根据目标对象的同一模态的特征信息，计算该模态下任意两个目标对象的相似度；根据各模态下任意两个目标对象的相似度，构建各模态下的多个样本对，样本对包括两个目标对象；根据各模态下的多个样本对的相似度，采用优化算法，自学习生成各模态下的后验匹配概率函数。

在一可选实施例中，在实现根据各模态下任意两个目标对象的相似度，构建各模态下的多个样本对，样本对包括两个目标对象时，自学习模块 72 还用于：

35 对于任意模态，根据模态下任意两个目标对象的相似度，确定模态下各目标对象的 K 近邻对象，其中 K 为正整数；将模态下各目标对象分别与 K 近邻对象组成样本对，得到模

态下的多个样本对。

在一可选实施例中，在实现根据各模态下的多个样本对及样本对的相似度，采用优化算法，自学习生成各模态下的后验匹配概率函数时，自学习模块 72 还用于：

5 将各模态下的后验匹配概率函数初始化为单调递增的分段函数，分段函数包含待求解的分段参数；基于各模态下的后验匹配概率函数，构建目标函数，目标函数包含待求解的分段参数；根据各模态下的多个样本对的相似度，采用优化算法最小化目标函数，求解得到分段参数；将分段参数代入分段函数，得到各模态下的后验匹配概率函数。

在一可选实施例中，在实现基于各模态下的后验匹配概率函数，构建目标函数时，自学习模块 72 还用于：

10 根据各模态下的后验匹配概率函数，构建多模态下的后验匹配概率函数；根据各模态下的后验匹配概率函数和多模态下的后验匹配概率函数，构建一致性损失函数和约束损失函数；

根据一致性损失函数和约束损失函数，构建目标函数。

15 在一可选实施例中，在实现根据各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率时，后验匹配概率计算模块 73 还用于：

根据任一模态下任意两个目标对象的相似度，使用该模态下的后验匹配概率函数，确定任意两个目标对象在该模态下的后验匹配概率；根据任意两个目标对象在各模态下的后验匹配概率，计算任意两个目标对象在多模态下的后验匹配概率。

20 在一可选实施例中，在根据各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率之后，后验匹配概率计算模块 73 还用于：

对于任意两个目标对象，确定两个目标对象在各模态下的 K 近邻对象，并确定两个目标对象的共有近邻对象，其中 K 为正整数；根据两个目标对象与同一共有近邻对象在多模态下的后验匹配概率，优化两个目标对象在多模态下的后验匹配概率。

25 在一可选实施例中，在实现根据两个目标对象与同一共有近邻对象在多模态下的后验匹配概率，优化两个目标对象在多模态下的后验匹配概率时，后验匹配概率计算模块 73 还用于：

计算两个目标对象与同一共有近邻对象在多模态下的后验匹配概率的乘积，根据乘积与两个目标对象在多模态下的后验匹配概率中的最大值，优化两个目标对象在多模态下的后验匹配概率。

30 在一可选实施例中，在实现根据两个目标对象与同一共有近邻对象在多模态下的后验匹配概率，优化两个目标对象在多模态下的后验匹配概率时，后验匹配概率计算模块 73 还用于：

35 将两个目标对象与同一共有近邻对象在多模态下的后验匹配概率之和，作为该共有近邻对象对应的后验匹配概率和；计算各共有近邻对象对应的后验匹配概率和的和，得到第一求和结果；对于两个目标对象中的第一对象和第二对象，计算第一对象与各共有近邻对

象在多模态下的后验匹配概率之和，得到第二求和结果，计算第二对象与各共有近邻对象在多模态下的后验匹配概率之和，得到第三求和结果，并计算第二求和结果与第三求和结果之和，得到第四求和结果：根据第一求和结果与第四求和结果的比值，优化两个目标对象在多模态下的后验匹配概率。

5 在一可选实施例中，在实现获取待聚类的目标对象的多模态的特征信息时，模态特征获取模块 71 还用于：接收端侧设备发送的待聚类的目标对象的多模态的特征信息。

在生成目标对象的聚类结果之后，多模态聚类模块 74 还用于：向端侧设备输出目标对象的聚类结果。

10 本申请实施例提供的装置可以具体用于执行上述任一方法实施例中服务器所执行的流程，具体功能和所能实现的技术效果此处不再赘述。

图 8 为本申请实施例提供的一种服务器的结构示意图。如图 8 所示，该服务器包括：存储器 801 和处理器 802。存储器 801，用于存储计算机执行指令，并可被配置为存储其它各种数据以支持在服务器上的操作。处理器 802，与存储器 801 通信连接，用于执行存储器 801 存储的计算机执行指令，以实现上述任一方法实施例提供的技术方案，其具体功
15 能和所能实现的技术效果类似，此处不再赘述。图 8 中以服务器为云服务器为例进行示例性地说明，服务器还可以是本地部署的服务器。

可选的，如图 8 所示，该服务器还包括：防火墙 803、负载均衡器 804、通信组件 805、电源组件 806 等其它组件。图 8 中仅示意性给出部分组件，并不意味着服务器只包括图 8 所示组件。

20 本申请实施例还提供一种计算机可读存储介质，计算机可读存储介质中存储有计算机执行指令，计算机执行指令被处理器执行时用于实现上述任一方法实施例所提供的方案，具体功能和所能实现的技术效果此处不再赘述。

本申请实施例还提供了一种计算机程序产品，计算机程序产品包括：计算机程序，计算机程序存储在可读存储介质中，服务器的至少一个处理器可以从可读存储介质读取计算机程序，至少一个处理器执行计算机程序使得服务器执行上述任一方法实施例所提供的方案，具体功能和所能实现的技术效果此处不再赘述。本申请实施例提供一种芯片，包括：
25 处理模块与通信接口，该处理模块能执行前述方法实施例中服务器的技术方案。可选的，该芯片还包括存储模块（如，存储器），存储模块用于存储指令，处理模块用于执行存储模块存储的指令，并且对存储模块中存储的指令的执行使得处理模块执行前述任一方法实
30 施例提供的技术方案。

上述存储器可以是对象存储(Object Storage Service, OSS)。

上述存储器可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器 (SRAM)，电可擦除可编程只读存储器 (EEPROM)，可擦除可编程只读存储器 (EPROM)，可编程只读存储器 (PROM)，只读存储器 (ROM)，磁存储器，快
35 闪存储器，磁盘或光盘。

上述通信组件被配置为便于通信组件所在设备和其他设备之间有线或无线方式的通信。通信组件所在设备可以接入基于通信标准的无线网络，如移动热点（WiFi），第二代移动通信系统（2G）、第三代移动通信系统（3G）、第四代移动通信系统（4G）/长期演进（LTE）、第五代移动通信系统（5G）等移动通信网络，或它们的组合。在一个示例性实施
5 例中，通信组件经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，通信组件还包括近场通信（NFC）模块，以促进短程通信。例如，在 NFC 模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

上述电源组件，为电源组件所在设备的各种组件提供电力。电源组件可以包括电源管
10 理系统，一个或多个电源，及其他与为电源组件所在设备生成、管理和分配电力相关联的组件。

本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机
15 可用存储介质（包括但不限于磁盘存储器、只读光盘存储器（CD-ROM）、光学存储器等）上实施的计算机程序产品的形式。

本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机
20 程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方
25 式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他
30 可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

在一个典型的配置中，计算设备包括一个或多个处理器（CPU）、输入/输出接口、网络接口和内存。

内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器（RAM）和/或非
35 易失性内存等形式，如只读存储器（ROM）或闪存（flash RAM）。内存是计算机可读介质的

示例。

计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、
5 动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器 (CD-ROM)、数字多功能光盘 (DVD) 或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体 (transitory media)，如调制的
10 数据信号和载波。

需要说明的是，本申请所涉及的用户信息（包括但不限于用户设备信息、用户属性信息）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守
相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

15 另外，在上述实施例及附图中的描述的一些流程中，包含了按照特定顺序出现的多个操作，但是应该清楚了解，这些操作可以不按照其在本文中出现的顺序来执行或并行执行，仅仅是用于区分开各个不同的操作，序号本身不代表任何的执行顺序。另外，这些流程可以包括更多或更少的操作，并且这些操作可以按顺序执行或并行执行。需要说明的是，本文中的“第一”、“第二”等描述，是用于区分不同的消息、设备、模块等，不代表先后顺序，也不限定“第一”和“第二”是不同的类型。“多个”的含义是两个以上，除非另有明确具体的限定。
20

需要说明的是，本申请所涉及的用户信息（包括但不限于用户设备信息、用户个人信息）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守
25 相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或
30 惯用技术手段。说明书和实施例仅被视为示例性的，本申请的真正范围和精神由下面的权利要求书指出。

应当理解的是，本申请并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求书来限制。

权 利 要 求

- 1、一种基于多模态信息的聚类方法，其特征在于，包括：
获取待聚类的目标对象的多模态的特征信息；
根据所述目标对象的多模态的特征信息，自学习生成各模态下的后验匹配概率函
- 5 数；
根据所述各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后
验匹配概率；
根据任意两个目标对象在多模态下的后验匹配概率，对所述目标对象进行聚类，
生成所述目标对象的聚类结果。
- 10 2、根据权利要求1所述的方法，其特征在于，所述根据所述目标对象的多模态的
特征信息，自学习生成各模态下的后验匹配概率函数，包括：
根据所述目标对象的同一模态的特征信息，计算该模态下任意两个目标对象的相
似度；
根据各模态下任意两个目标对象的相似度，构建各模态下的多个样本对，所述样
- 15 本对包括两个目标对象；
根据各模态下的多个样本对的相似度，采用优化算法，自学习生成各模态下的后
验匹配概率函数。
- 3、根据权利要求2所述的方法，其特征在于，所述根据各模态下任意两个目标对
象的相似度，构建各模态下的多个样本对，所述样本对包括两个目标对象，包括：
- 20 对于任意模态，根据所述模态下任意两个目标对象的相似度，确定所述模态下各
所述目标对象的K近邻对象，其中K为正整数；
将所述模态下各所述目标对象分别与K近邻对象组成样本对，得到所述模态下的
多个样本对。
- 4、根据权利要求2所述的方法，其特征在于，所述根据各模态下的多个样本对及
- 25 所述样本对的相似度，采用优化算法，自学习生成各模态下的后验匹配概率函数，包
括：
将各模态下的后验匹配概率函数初始化为单调递增的分段函数，所述分段函数包
含待求解的分段参数；
基于各模态下的后验匹配概率函数，构建目标函数，所述目标函数包含所述待求
- 30 解的分段参数；
根据各模态下的多个样本对的相似度，采用优化算法最小化所述目标函数，求解
得到所述分段参数；
将所述分段参数代入所述分段函数，得到各模态下的后验匹配概率函数。
- 5、根据权利要求4所述的方法，其特征在于，所述基于各模态下的后验匹配概率
- 35 函数，构建目标函数，包括：

根据各模态下的后验匹配概率函数，构建多模态下的后验匹配概率函数；

根据各模态下的后验匹配概率函数和多模态下的后验匹配概率函数，构建一致性损失函数和约束损失函数；

根据所述一致性损失函数和约束损失函数，构建目标函数。

5 6、根据权利要求1所述的方法，其特征在于，所述根据所述各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率，包括：

根据任一模态下任意两个目标对象的相似度，使用该模态下的后验匹配概率函数，确定任意两个目标对象在该模态下的后验匹配概率；

10 根据任意两个目标对象在各模态下的后验匹配概率，计算任意两个目标对象在多模态下的后验匹配概率。

7、根据权利要求1-6中任一项所述的方法，其特征在于，所述根据所述各模态下的后验匹配概率函数，计算任意两个目标对象在多模态下的后验匹配概率之后，还包括：

15 对于任意两个目标对象，确定所述两个目标对象在各模态下的K近邻对象，并确定所述两个目标对象的共有近邻对象，其中K为正整数；

根据所述两个目标对象与同一共有近邻对象在多模态下的后验匹配概率，优化所述两个目标对象在多模态下的后验匹配概率。

20 8、根据权利要求7所述的方法，其特征在于，所述根据所述两个目标对象与同一共有近邻对象在多模态下的后验匹配概率，优化所述两个目标对象在多模态下的后验匹配概率，包括：

计算所述两个目标对象与同一共有近邻对象在多模态下的后验匹配概率的乘积，根据所述乘积与两个目标对象在多模态下的后验匹配概率中的最大值，优化所述两个目标对象在多模态下的后验匹配概率。

25 9、根据权利要求7所述的方法，其特征在于，所述根据所述两个目标对象与同一共有近邻对象在多模态下的后验匹配概率，优化所述两个目标对象在多模态下的后验匹配概率，包括：

将所述两个目标对象与同一共有近邻对象在多模态下的后验匹配概率之和，作为该共有近邻对象对应的后验匹配概率和；

计算各所述共有近邻对象对应的后验匹配概率和的和，得到第一求和结果；

30 对于所述两个目标对象中的第一对象和第二对象，计算所述第一对象与各所述共有近邻对象在多模态下的后验匹配概率之和，得到第二求和结果，计算所述第二对象与各所述共有近邻对象在多模态下的后验匹配概率之和，得到第三求和结果，并计算所述第二求和结果与所述第三求和结果之和，得到第四求和结果；

35 根据所述第一求和结果与第四求和结果的比值，优化所述两个目标对象在多模态下的后验匹配概率。

10、根据权利要求1-6中任一项所述的方法，其特征在于，所述获取待聚类的目标对象的多模态的特征信息，包括：

接收端侧设备发送的待聚类的目标对象的多模态的特征信息；

在生成所述目标对象的聚类结果之后，还包括：

5 向所述端侧设备输出所述目标对象的聚类结果。

11、一种基于多模态信息的聚类方法，其特征在于，包括：

获取待聚类的图像的多模态的特征信息，所述多模态的特征信息包括地理位置信息、所包含对象的至少一种模态的特征信息；

根据所述图像的多模态的特征信息，自学习生成各模态下的后验匹配概率函数；

10 根据所述各模态下的后验匹配概率函数，计算任意两个图像在多模态下的后验匹配概率，其中两个图像在多模态下的后验匹配概率是两个图像包含同一对象的概率；

根据任意两个图像在多模态下的后验匹配概率，将所包含同一对象的图像聚为一类，得到所述图像的聚类结果。

12、一种基于多模态信息的聚类方法，其特征在于，应用于端侧设备，包括：

15 获取待聚类的目标对象的多模态的特征信息；

向服务器发送待聚类的目标对象的多模态的特征信息；

接收所述服务器发送的所述目标对象的聚类结果，所述目标对象的聚类结果是通过如权利要求1-10中任一所述的方法生成的；

20 根据所述聚类结果，执行如下至少一项处理逻辑：信息推送、信息搜索、数据分类。

13、一种服务器，其特征在于，包括：

处理器，以及与所述处理器通信连接的存储器；

所述存储器存储计算机执行指令；

25 所述处理器执行所述存储器存储的计算机执行指令，以实现如权利要求1-11中任一项所述的方法。

14、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质中存储有计算机执行指令，所述计算机执行指令被处理器执行时用于实现如权利要求1-11中任一项所述的方法。