



(12)发明专利

(10)授权公告号 CN 106295673 B

(45)授权公告日 2019.12.13

(21)申请号 201510359225.0

(22)申请日 2015.06.25

(65)同一申请的已公布的文献号
申请公布号 CN 106295673 A

(43)申请公布日 2017.01.04

(73)专利权人 阿里巴巴集团控股有限公司
地址 开曼群岛大开曼岛

(72)发明人 蒋宇东

(74)专利代理机构 北京弘权知识产权代理事务
所(普通合伙) 11363
代理人 逯长明 许伟群

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

(56)对比文件

CN 101216830 A, 2008.07.09,
JP 特开2011-70422 A, 2011.04.07,
CN 103744966 A, 2014.04.23,
CN 103810208 A, 2014.05.21,
CN 102609523 A, 2012.07.25,

审查员 赵志洪

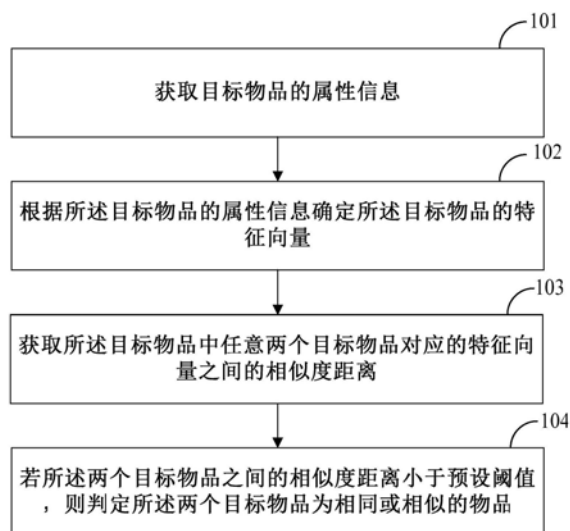
权利要求书2页 说明书7页 附图3页

(54)发明名称

物品信息处理方法及处理装置

(57)摘要

本发明的实施例中公开了一种物品信息处理方法和处理装置,所述方法包括:获取目标物品的属性信息;根据所述目标物品的属性信息确定所述目标物品的特征向量;获取所述目标物品中任意两个目标物品对应的特征向量之间的相似度距离;若所述两个目标物品之间的相似度距离小于预设阈值,则判定所述两个目标物品为相同或相似的物品。本发明的实施例能在物品信息处理时提高精确度,便于用户识别出相同或相似的物品,提高用户的使用体验。



1. 一种物品信息处理方法,其特征在于,所述方法包括:
获取目标物品的属性信息;
根据所述目标物品的属性信息确定所述目标物品的特征向量;
获取所述目标物品中任意两个目标物品对应的特征向量之间的相似度距离;
若所述两个目标物品之间的相似度距离小于预设阈值,则判定所述两个目标物品为相同或相似的物品;

其中,所述根据目标物品的属性信息确定所述目标物品的特征向量,所述方法包括:
根据所述目标物品的属性信息,从所述目标物品对应的属性字典中确定所述目标物品的特征向量,其中,所述属性字典包括属性及所述属性对应的特征向量;
所述根据目标物品的属性,从目标物品对应的属性字典中获取目标物品的特征向量之前,所述方法还包括:

获取目标物品的图像信息;
将图像信息相同的目标物品构成的集合作为第一参考集合;
获取目标物品的文本信息;
将所述文本信息相同的目标物品构成的集合作为第二参考集合;
获取所述第一参考集合和第二参考集合的并集作为第三参考集合;
获取第三参考集合中目标物品的属性名和对应的属性值;
对所述属性值进行编码生成所述属性值对应的子向量;
根据所述属性名、属性名对应的属性值、以及所述属性值对应的子向量,生成所述属性词典。

2. 如权利要求1所述的方法,其特征在于,所述将图像信息相同的目标物品构成的集合作为第一参考集合,包括:

获取所述图像信息中的图像特征数据;
根据所述图像特征数据生成指纹编码,所述相同图像信息的指纹编码相同;
将所述指纹编码相同的图像信息所对应的目标物品构成的集合作为第一参考集合。

3. 如权利要求1或2所述的方法,其特征在于,所述将所述文本信息相同的目标物品构成的集合作为第二参考集合,包括:

获取所述目标物品的文本信息;
获取所述文本信息中的文本特征数据;
将所述文本特征数据相同的文本信息对应的目标物品构成的集合作为第二参考集合。

4. 如权利要求1所述的方法,其特征在于,所述在判定两个目标物品为相同或相似的物品之后,所述方法还包括:

将所述相同或相似的物品作为第一同款物品集合中的元素;
将所述第一同款物品集合和所述第三参考集合的并集作为第二同款物品集合。

5. 如权利要求1所述的方法,其特征在于,所述获取目标物品的属性信息包括:
获取目标物品的属性名,以及所述属性名对应的属性值;
所述根据所述目标物品的属性信息,从所述目标物品对应的属性字典中获取所述目标物品的特征向量,包括:

根据所述目标物品的属性值,从所述目标物品对应的属性字典中获取所述属性值对应

的子向量；

根据所述属性值对应的子向量，获取所述属性名对应的子向量；

根据所述属性名对应的子向量，生成所述目标物品对应的特征向量。

6. 如权利要求5所述的方法，其特征在于，所述根据属性值对应的子向量，获取属性名对应的子向量，包括：

若所述属性名对应一个属性值，则将所述属性值对应的子向量作为所述属性名对应的子向量；

若所述属性名对应多个属性值，则将所述多个属性值对应的多个子向量加权平均后作为所述属性名对应的子向量。

7. 一种物品信息处理装置，其特征在于，所述装置包括：

属性信息获取单元，用于获取目标物品的属性信息；

特征向量获取单元，用于根据所述目标物品的属性信息确定所述目标物品的特征向量；

相似度距离获取单元，用于获取所述目标物品中任意两个目标物品对应的特征向量之间的相似度距离；

判定单元，用于若所述两个目标物品之间的相似度距离小于预设阈值，则判定所述两个目标物品为相同或相似的物品；

其中，所述判定单元还用于根据所述目标物品的属性信息，从所述目标物品对应的属性字典中确定所述目标物品的特征向量，其中，所述属性字典包括属性及所述属性对应的特征向量；

所述物品信息处理装置还包括：

第一获取单元，用于获取目标物品的图像信息；

第一处理单元，用于将图像信息相同的目标物品构成的集合作为第一参考集合；

第二获取单元，用于获取目标物品的文本信息；

第二处理单元，用于将所述文本信息相同的目标物品构成的集合作为第二参考集合；

第三获取单元，用于获取所述第一参考集合和第二参考集合的并集作为第三参考集合；

第四获取单元，用于获取第三参考集合中目标物品的属性名和对应的属性值；

编码单元，用于对所述属性值进行编码生成所述属性值对应的子向量；

生成单元，用于根据所述属性名、属性名对应的属性值、以及所述属性值对应的子向量，生成所述属性词典。

物品信息处理方法及处理装置

技术领域

[0001] 本发明涉及互联网领域,特别涉及一种物品信息处理方法及处理装置。

背景技术

[0002] 随着通信及网络技术的飞速发展,互联网成为了人们获取物品的重要平台。和传统的获取方式相比,通过互联网平台获取物品的弊端在与用户在拿到物品之前是无法亲眼见到物品,也无法触摸到物品,因此,通过互联网平台获取物品的首要问题是如何从互联网上呈现的众多的物品中筛选出自己需要的物品。在互联网平台上呈现的海量物品中,有很多物品是相同、相似或相关性较高的,导致用户筛选物品时冗余度很高,影响用户的选择,因此用户获取物品时通常需要判断多个物品是否相同或相似。

[0003] 判断多个物品是否相同或相似通常是通过分析物品信息来实现的,相同物品一般是指材质、视觉外观、规格、配置等属性均一致的物品。物品信息包括文本信息和图像信息。现有技术中,判断多个物品是否相似主要包括基于文本信息的识别方法和基于图像信息的识别方法。

[0004] 单纯的基于文本信息对物品进行识别的缺陷是如果物品的所有者提供的文本信息不准确,或不完备,或文字描述不规范时,会导致误判。单纯的基于物品图像信息的识别方法的识别方法,可能会因为图像的拍摄角度、图像的变形、物品的摆放位置、摆放方式以及背景干扰的因素造成误判。

[0005] 可见,现有技术中由于对物品信息的处理方法还不完善,使物品信息识别时精确度较低,影响用户的选择。

发明内容

[0006] 本发明实施例中提供了一种物品信息处理方法及处理装置,以解决物品信息识别时的精确度较低的问题。

[0007] 为了解决上述技术问题,本发明实施例公开了如下技术方案:

[0008] 一方面,提供了一种物品信息处理方法,所述方法包括:

[0009] 获取目标物品的属性信息;

[0010] 根据所述目标物品的属性信息确定所述目标物品的特征向量;

[0011] 获取所述目标物品中任意两个目标物品对应的特征向量之间的相似度距离;

[0012] 若所述两个目标物品之间的相似度距离小于预设阈值,则判定所述两个目标物品为相同或相似的物品。

[0013] 可选的,所述根据目标物品的属性信息确定所述目标物品的特征向量,所述方法包括:

[0014] 根据所述目标物品的属性信息,从所述目标物品对应的属性字典中确定所述目标物品的特征向量,所述属性字典包括属性及所述属性对应的特征向量。

[0015] 可选的,所述根据目标物品的属性,从目标物品对应的属性字典中获取目标物品

的特征向量之前,所述方法还包括:

- [0016] 获取目标物品的图像信息;
- [0017] 将图像信息相同的目标物品构成的集合作为第一参考集合;
- [0018] 获取目标物品的文本信息;
- [0019] 将所述文本信息相同的目标物品构成的集合作为第二参考集合;
- [0020] 获取所述第一参考集合和第二参考集合的并集作为第三参考集合;
- [0021] 获取第三参考集合中目标物品的属性名和对应的属性值;
- [0022] 对所述属性值进行编码生成所述属性值对应的子向量;
- [0023] 根据所述属性名、属性名对应的属性值、以及所述属性值对应的子向量,生成所述属性词典。
- [0024] 可选的,所述将图像信息相同的目标物品构成的集合作为第一参考集合,包括:
- [0025] 获取所述图像信息中的图像特征数据;
- [0026] 根据所述图像特征数据生成指纹编码,所述相同图像信息的指纹编码相同;
- [0027] 将所述指纹编码相同的图像信息所对应的目标物品构成的集合作为第一参考集合。
- [0028] 可选的,所述将所述文本信息相同的目标物品构成的集合作为第二参考集合,包括:
- [0029] 获取所述目标物品的文本信息;
- [0030] 获取所述文本信息中的文本特征数据;
- [0031] 将所述文本特征数据相同的文本信息对应的目标物品构成的集合作为第二参考集合。
- [0032] 可选的,所述在判断两个目标物品为相同或相似的物品之后,所述方法还包括:
- [0033] 将所述相同或相似的物品作为第一同款物品集合中的元素;
- [0034] 将所述第一同款物品集合和所述第三参考集合的并集作为第二同款物品集合。
- [0035] 可选的,所述获取目标物品的属性信息包括:
- [0036] 获取目标物品的属性名,以及所述属性名对应的属性值;
- [0037] 所述根据所述目标物品的属性信息,从所述目标物品对应的属性字典中获取所述目标物品的特征向量,包括:
- [0038] 根据所述目标物品的属性值,从所述目标物品对应的属性字典中获取所述属性值对应的子向量;
- [0039] 根据所述属性值对应的子向量,获取所述属性名对应的子向量;
- [0040] 根据所述属性名对应的子向量,生成所述目标物品对应的特征向量。
- [0041] 可选的,所述根据属性值对应的子向量,获取属性名对应的子向量,包括:
- [0042] 若所述属性名对应一个属性值,则将所述属性值对应的子向量作为所述属性名对应的子向量;
- [0043] 若所述属性名对应多个属性值,则将所述多个属性值对应的多个子向量加权平均后作为所述属性名对应的子向量。
- [0044] 第二方面,提供了一种物品信息处理装置,所述装置包括:
- [0045] 属性信息获取单元,用于获取目标物品的属性信息;

[0046] 特征向量获取单元,用于根据所述目标物品的属性信息确定所述目标物品的特征向量;

[0047] 相似度距离获取单元,用于获取所述目标物品中任意两个目标物品对应的特征向量之间的相似度距离;

[0048] 判定单元,用于若所述两个目标物品之间的相似度距离小于预设阈值,则判定所述两个目标物品为相同或相似的物品。

[0049] 可选的,所述判定单元还用于根据所述目标物品的属性信息,从所述目标物品对应的属性字典中确定所述目标物品的特征向量;

[0050] 所述属性字典包括属性及所述属性对应的特征向量。

[0051] 本发明的实施例中公开了一种物品信息处理方法,通过目标物品对应的属性信息获取目标物品的特征向量,获取目标物品中任意两个目标物品对应的特征向量之间的相似度距离和预设阈值的比较,来判断上述任意两个目标物品是否为相同或相似的物品。本发明实施例的方法可以将物品的属性信息转变为特征向量,通过对特征向量的数学运算,可以获得物品属性之间的相关度或相似度,能在物品信息处理时提高精确度,便于用户识别出相同或相似的物品,提高用户的使用体验。

附图说明

[0052] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例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发明的一些实施例,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。

[0053] 图1所示为本发明实施例的一种物品信息处理方法的流程图;

[0054] 图2所示为本发明实施例的一种物品信息处理方法的流程图;

[0055] 图3所示为本发明实施例的一种物品信息处理装置的示意图。

具体实施方式

[0056] 本发明如下实施例提供了一种物品信息处理方法,能提高物品信息处理的精确度。

[0057] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整的描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,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,都属于本发明保护的范围。

[0058] 图1所示为本发明实施例的一种物品信息处理方法的流程图,如图1所示,所述方法包括:

[0059] 步骤101,获取目标物品的属性信息。

[0060] 目标物品可以是提供给用户的各种物品,例如服装、首饰和箱包等。目标物品的属性信息包括属性名和属性名对应的属性值。

[0061] 属性名可以是颜色、样式、材料、尺码、风格等,属性值可以是属性名对应的具体内容,例如属性名为颜色,对应的属性值可以是红色、白色、黑白;属性名为样式,对应的属性

值可以是长裙、短裙、半裙等；属性名为材料，对应的属性值可以是纯棉、雪纺、欧根纱、毛呢、真皮等；属性名为风格，对应的属性值可以是森女、欧美、简约等。

[0062] 所述获取目标物品的属性信息包括：获取目标物品的属性名，以及所述属性名对应的属性值；

[0063] 步骤102，根据所述目标物品的属性信息确定所述目标物品的特征向量。

[0064] 本发明实施例中，根据所述目标物品的属性信息确定所述目标物品的特征向量，包括：

[0065] 根据所述目标物品的属性信息，从所述目标物品对应的属性字典中获取所述目标物品的特征向量；

[0066] 所述属性字典包括所述属性及所述属性对应的向量。

[0067] 步骤103，获取所述目标物品中任意两个目标物品对应的特征向量之间的相似度距离。

[0068] 本发明实施例中的两个特征向量之间的相似度距离及两个向量之间的距离。例如若两个目标物品A和B对应的特征向量分别为(a,b,c)和(x,y,z)，则相似度距离为

$$[0069] \quad d = \sqrt{(a-x)^2 + (b-y)^2 + (c-z)^2}$$

[0070] 步骤104，若所述两个目标物品之间的相似度距离小于预设阈值，则判定所述两个目标物品为相同或相似的物品。

[0071] 本发明的实施例中公开了一种物品信息处理方法，通过目标物品对应的属性信息获取目标物品的特征向量，获取目标物品中任意两个目标物品对应的特征向量之间的相似度距离和预设阈值的比较，判断上述任意两个目标物品是否为相同或相似的物品。本发明实施例的方法可以将物品的属性信息转变为特征向量，通过对特征向量的数学运算来获得物品属性之间的相关度或相似度，以识别物品是否相同或相似，能在物品信息处理时提高精确度，便于使用户识别出相同或相似的物品，提高用户的使用体验。

[0072] 图2所示为本发明实施例的一种物品信息处理方法的流程图，如图2所示，所述根据目标物品的属性，从目标物品对应的属性字典中获取目标物品的特征向量之前，所述方法还包括：

[0073] 步骤201，获取目标物品的图像信息。

[0074] 目标物品的图像信息可以是物品的照片，模特图等，也可以是视频信息。

[0075] 步骤202，将图像信息相同的目标物品构成的集合作为第一参考集合。

[0076] 所述将图像信息相同的目标物品构成的集合作为第一参考集合，包括：

[0077] 获取所述图像信息中的图像特征数据；

[0078] 根据所述图像特征数据生成指纹编码，所述相同图像信息的指纹编码相同；

[0079] 将所述指纹编码相同的图像信息所对应的目标物品构成的集合作为第一参考集合。

[0080] 根据图像特征数据生成指纹编码，可以采用现有技术中图像处理的方法，例如先提取重要特征，比如形状特征，颜色特征，轮廓等，然后进行编码，在此不再赘述。

[0081] 步骤203，获取目标物品的文本信息。

[0082] 目标物品的文本信息可以是以文字描写的物品的分类、颜色、风格、尺寸、货号 and

品牌等信息。

[0083] 步骤204,将所述文本信息相同的目标物品构成的集合作为第二参考集合。

[0084] 所述将所述文本信息相同的目标物品构成的集合作为第二参考集合,包括:

[0085] 获取所述目标物品的文本信息;

[0086] 获取所述文本信息中的文本特征数据;

[0087] 将所述文本特征数据相同的文本信息对应的目标物品构成的集合作为第二参考集合。

[0088] 步骤205,获取所述第一参考集合和第二参考集合的并集作为第三参考集合。

[0089] 获取第一参考集合和第二参考集合的并集时,先将第一参考集合和第二参考集合合并,然后去掉合集中的重复元素。例如第一参考集合为(A1,B1,C1),第二参考集合为(A1,B2,C2),则第三参考集合为(A1,B1,C1,B2,C2)。

[0090] 步骤206,获取第三参考集合中目标物品的属性名和对应的属性值。

[0091] 步骤207,对所述属性值进行编码生成所述属性值对应的子向量。

[0092] 对属性值进行编码时,越相近的属性值编码越相相近,例如红色可以是(0,0,1.1),那么酒红色可以是(0,0,1.2),而蓝色可以是(0,0,2.0)。编码得到的子向量可以是三维的,也可以是多维的。

[0093] 步骤208,根据所述属性名、属性名对应的属性值、以及所述属性值对应的子向量,生成所述属性词典。

[0094] 属性词典可以用表格来表示,还可以用其他形式来表示,表1所示为一个实施例的属性词典,表1所示的实施例中的子向量为多维向量。

[0095] 表1

[0096]

属性名	属性值	子向量
风格	荷叶边	0.268,0.074,⋯,0.732
风格	蕾丝	0.479,0.112,⋯,0.132
风格	镂空	0.876,0.201,⋯,0.001
年份季节	2014年秋季	0.321,0.101,⋯,0.009
年份季节	秋季	0.318,0.099,⋯,0.011
年份季节	2014夏季	0.891,0.321,⋯,0.033
腰型	中腰	0.732,0.001,⋯,0.029
腰型	宽腰	0.098,0.341,⋯,0.145
腰型	窄腰	0.098,0.124,⋯,0.121
颜色分类	酒红色	0.546,0.782,⋯,0.003
颜色分类	红色	0.442,0.231,⋯,0.000
颜色分类	白色	0.000,0.413,⋯,0.817
裙型	百褶裙	0.417,0.282,⋯,0.811
裙型	灯笼裙	0.001,0.293,⋯,0.000
裙型	A字裙	0.374,0.784,⋯,0.387

[0097] 本发明实施例中,可选的,本发明实施例中,在步骤104之后,即在判断两个目标物

品为相同或相似的物品之后,所述方法还包括:

[0098] 将所述相同或相似的物品作为第一同款物品集合中的元素;

[0099] 将所述第一同款物品集合和所述第三参考集合的并集作为第二同款物品集合。

[0100] 用户可以设置预设阈值,以调整精确度。预设阈值越小,第一同款集合中的元素就越少,即相同的物品越少,预设阈值越大,第一同款集合中的元素就越多,即相同的物品越多。

[0101] 将第一同款物品集合和第三参考集合的并集作为第二同款物品集合,可以提高识别的精确度,避免遗漏物品。

[0102] 本发明实施例中,可选的,根据所述目标物品的属性信息,从所述目标物品对应的属性字典中获取所述目标物品的特征向量,包括:

[0103] 根据所述目标物品的属性值,从所述目标物品对应的属性字典中获取所述属性值对应的子向量;

[0104] 根据所述属性值对应的子向量,获取所述属性名对应的子向量;

[0105] 根据所述属性名对应的子向量,生成所述目标物品对应的特征向量。

[0106] 可选的,所述根据属性值对应的子向量,获取属性名对应的子向量,包括:

[0107] 若所述属性名对应一个属性值,则将所述属性值对应的子向量作为所述属性名对应的子向量;

[0108] 若所述属性名对应多个属性值,则将所述多个属性值对应的多个子向量加权平均后作为所述属性名对应的子向量。

[0109] 本发明一个实施例中,采用如表1所述的属性字典,如果一件物品的风格为:荷叶边、蕾丝,即风格属性名对应了两个属性值,则风格属性名的子向量为“荷叶边”对应的子向量以及“蕾丝”对应的子向量的加权平均值,加权值可以根据经验设置。

[0110] 获得子向量以后,可以将子向量顺序排列作为目标物品对应的特征向量,例如一件物品对应的子向量为A1,A2,A3,那么该物品对应的特征向量为(A1,A2,A3)。

[0111] 通过属性字典,可以将物品的属性信息转变为特征向量,通过对特征向量的数学运算,可以获得物品属性之间的相关度或相似度,来判断目标物品是否为相同物品或相似物品,能在物品信息处理时提高精确度,便于使用户识别出相同或相似的物品,提高用户的使用体验。

[0112] 图3所示为本发明实施例的一种物品信息处理装置的示意图,如图3所示,所述装置包括:

[0113] 属性信息获取单元301,用于获取目标物品的属性信息;

[0114] 特征向量获取单元302,用于根据所述目标物品的属性信息确定所述目标物品的特征向量;

[0115] 相似度距离获取单元303,用于获取所述目标物品中任意两个目标物品对应的特征向量之间的相似度距离;

[0116] 判定单元304,用于若所述两个目标物品之间的相似度距离小于预设阈值,则判定所述两个目标物品为相同或相似的物品。

[0117] 可选的,所述判定单元还用于根据所述目标物品的属性信息,从所述目标物品对应的属性字典中确定所述目标物品的特征向量;

[0118] 所述属性字典包括属性及所述属性对应的特征向量。

[0119] 本发明实施例的物品信息处理装置,可以将物品的属性信息转变为特征向量,通过对特征向量的数学运算,可以获得物品属性之间的相关度或相似度,来判断目标物品是否为相同物品或相似物品,能在物品信息处理时提高精确度,便于用户识别出相同或相似的物品,提高用户的使用体验。

[0120] 本发明实施例提供了一种物品信息处理方法及处理装置,通过目标物品对应的属性字典获取目标物品的特征向量,获取目标物品中任意两个目标物品对应的特征向量之间的相似度距离和预设阈值的比较,判断上述任意两个目标物品是否为相同或相似的物品。本发明实施例的方法提供了一种物品信息处理方法,可以将物品的属性信息转变为特征向量,通过对特征向量的数学运算,可以获得物品属性之间的相关度或相似度,能在物品信息处理时提高精确度,便于用户识别出相同或相似的物品,提高用户的使用体验。

[0121] 本领域的技术人员可以清楚地了解到本发明实施例中的技术可借助软件加必需的通用硬件的方式来实现,通用硬件包括通用集成电路、通用CPU、通用存储器、通用元器件等,当然也可以通过专用硬件包括专用集成电路、专用CPU、专用存储器、专用元器件等来实现,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明实施例中的技术方案本质上或者说对现有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存储介质中,如只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟、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)执行本发明各个实施例或者实施例的某些部分所述的方法。

[0122] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,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。

[0123] 以上所述的本发明实施方式,并不构成对本发明保护范围的限定。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明的保护范围之内。

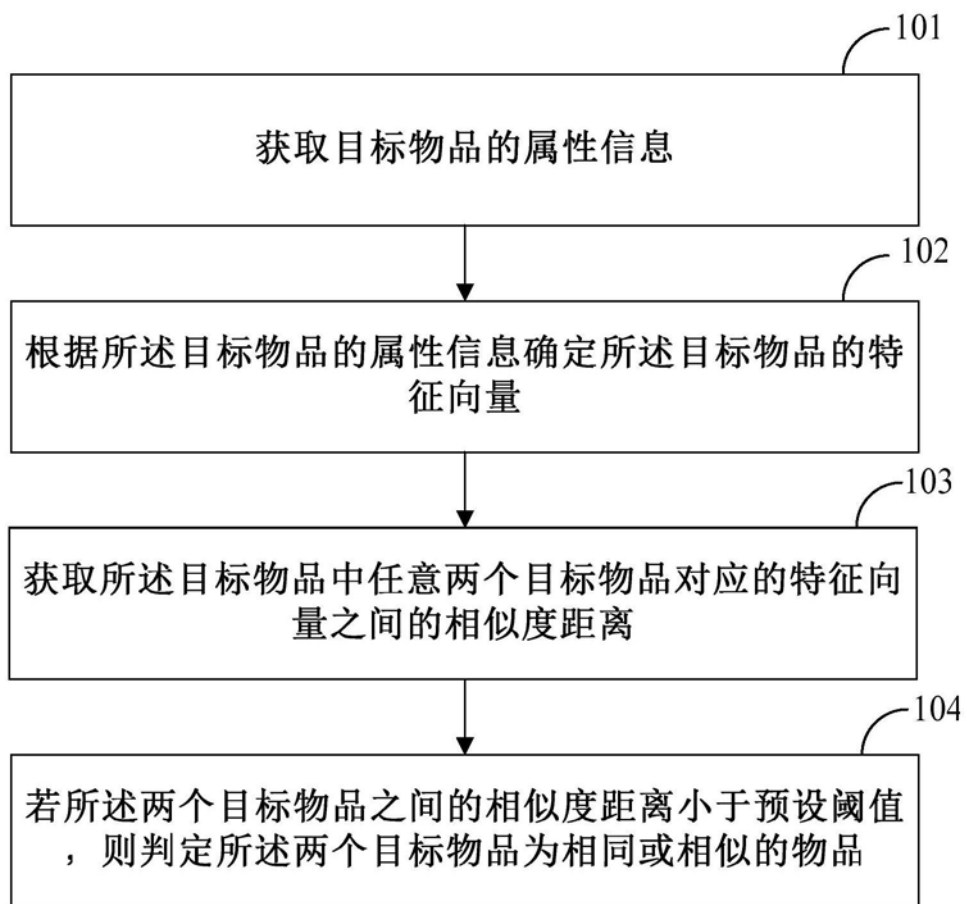


图1

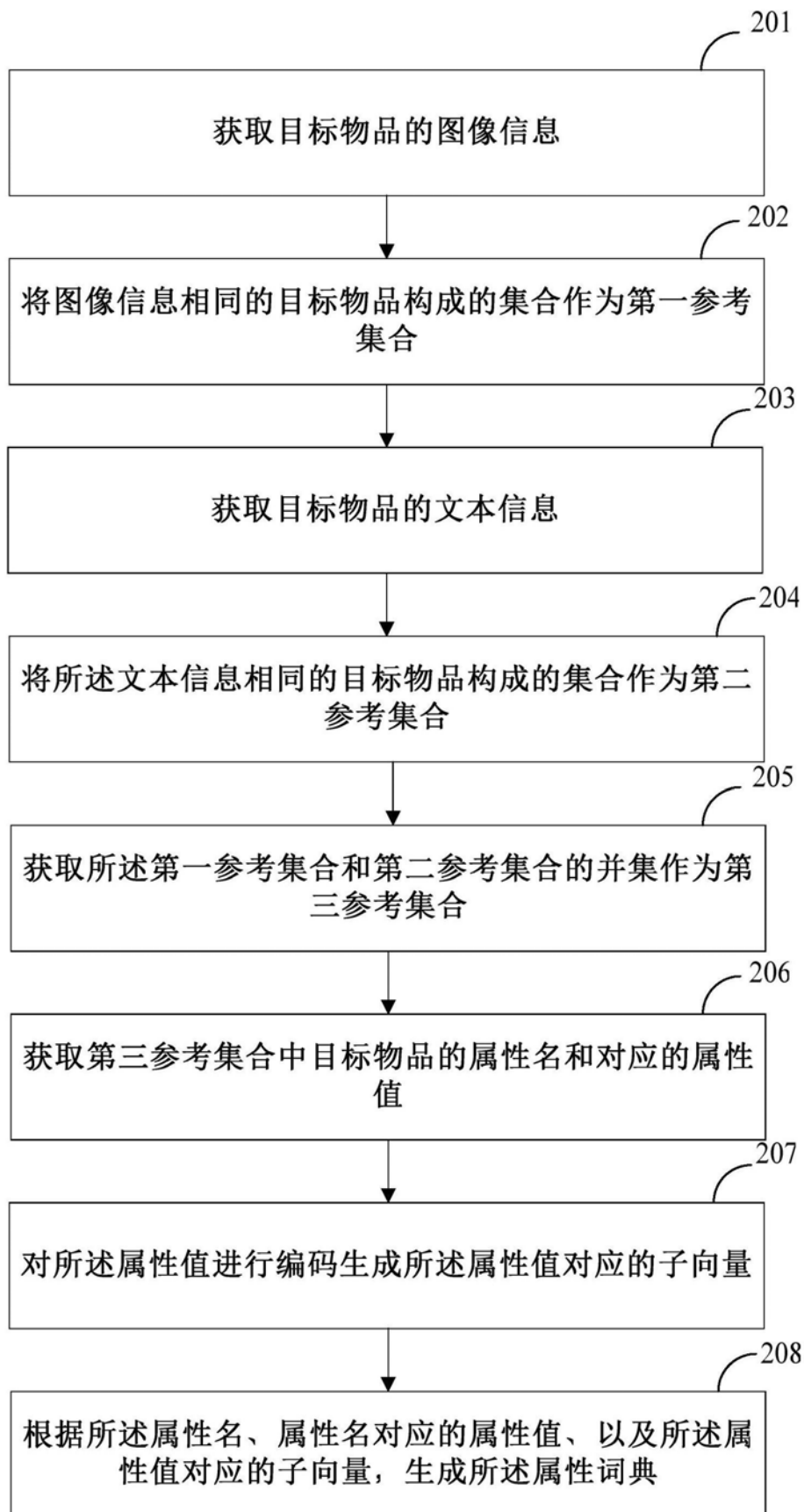


图2

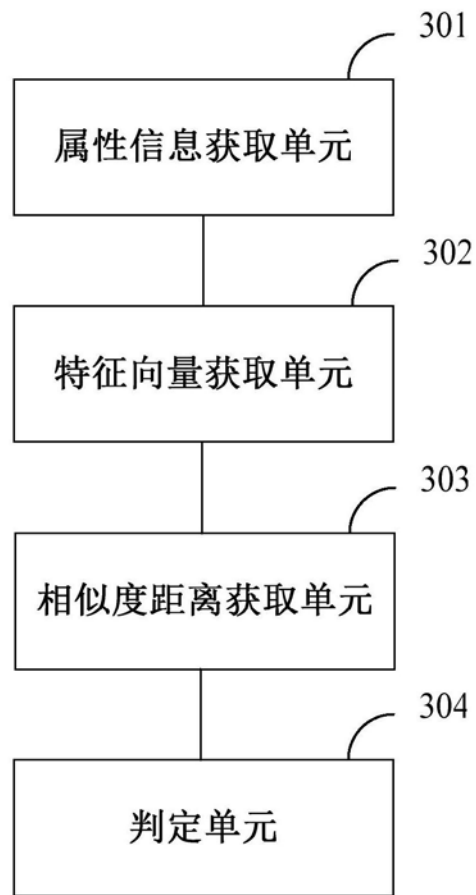


图3