



(12)发明专利

(10)授权公告号 CN 104504055 B

(45)授权公告日 2017.12.26

(21)申请号 201410798968.3

(22)申请日 2014.12.19

(65)同一申请的已公布的文献号
申请公布号 CN 104504055 A

(43)申请公布日 2015.04.08

(73)专利权人 常州飞寻视讯信息科技有限公司
地址 213164 江苏省常州市常武中路801号
常州科教城天润科技大厦A座3楼

(72)发明人 姚志强

(74)专利代理机构 北京科亿知识产权代理事务
所(普通合伙) 11350

代理人 汤东风

(51)Int.Cl.

G06F 17/30(2006.01)

G06Q 30/02(2012.01)

(56)对比文件

CN 103324650 A,2013.09.25,

CN 103345645 A,2013.10.09,

JP 2006331460 A,2006.12.07,

CN 102436483 A,2012.05.02,

审查员 朱琦

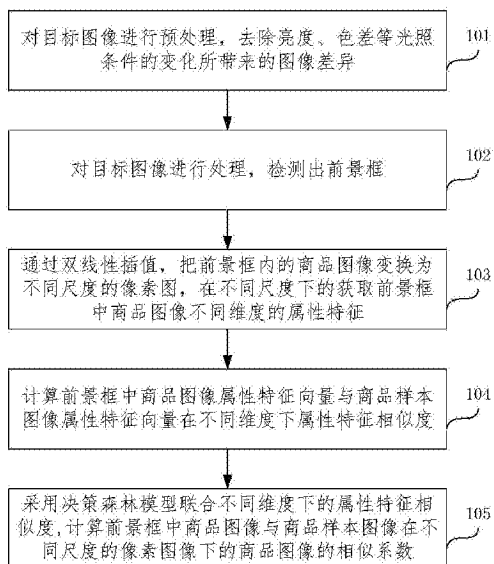
权利要求书3页 说明书8页 附图4页

(54)发明名称

基于图像相似度的商品相似计算方法及商品推荐系统

(57)摘要

本发明涉及互联网电子商务领域,特别涉及基于图像相似度的商品相似计算方法及商品推荐系统,所述方法包括:对目标图像进行预处理,去除亮度、色差等光照条件的变化所带来的图像差异;对目标图像进行处理检测出前景框;通过双线性插值,把前景框内的商品图像变换为不同尺度的像素图,在不同尺度下的获取前景框中商品图像不同维度的属性特征;计算前景框中商品图像属性特征向量与商品样本图像属性特征向量在不同维度下属性特征相似度;采用决策森林模型联合不同维度下的属性特征相似度,计算前景框中商品图像与商品样本图像在不同尺度的像素图像下的商品图像的相似度;以商品图像作为商品在不同商务平台的统一标识,极大地增加了推荐系统的可靠性。



1. 基于图像相似度的商品相似计算方法,其特征在於:包括:

步骤101、对目标图像进行预处理,去除亮度、色差光照条件的变化所带来的图像差异;

步骤102、对目标图像进行处理检测出前景框;

步骤103、通过双线性插值,把前景框内的商品图像变换为不同尺度的像素图,在不同尺度下的获取前景框中商品图像不同维度的属性特征;

步骤104、计算前景框中商品图像属性特征向量与商品样本图像属性特征向量在不同维度下属性特征相似度;

步骤105、采用决策森林模型联合不同尺度下的属性特征相似度,计算前景框中商品图像与商品样本图像在不同尺度的像素图像下的商品图像的相似度。

2. 根据权利要求1所述基于图像相似度的商品相似计算方法,其特征在於:所述对目标图像进行处理检测出前景框包括:

102-1、通过图像分割算法从目标图像生成若干候选框;

102-2、对每个候选框采用卷积神经网络提取属性特征,即通过卷积神经网络的卷积核对候选框中的图像像素值进行卷积运算得到卷积值;

102-3、采用支持向量机SVM分类器将候选框分成候选前景框或非前景框;

102-4、从候选前景框中选择置信度最大的作为商品图像的前景框。

3. 根据权利要求2所述基于图像相似度的商品相似计算方法,其特征在於:

所述卷积神经网络通过以下方式建立:

A) 建立样本库,包含NX种不同的商品,NX是样本数,取值范围为5000—20000;

B) 在3种光照条件下,从3种拍摄角度拍摄商品图片;

C) 以步骤A得到的不同商品作为不同的类别,建立分类器,以卷积神经网络的输出为该分类器所采用的分类特征,以步骤B得到的商品图片为评估样本;通过“后向反馈”的方法,逐层调整卷积神经网络中的卷积层,使得以卷积神经网络的输出作为分类特征的分类器,在评估样本上达到最好的分类精度。

4. 根据权利要求1所述基于图像相似度的商品相似计算方法,其特征在於:

所述属性特征相似度通过以下方式计算:

$$\text{属性特征相似度} = \frac{\overline{feature}_i \cdot \overline{feature}_i'}{\|\overline{feature}_i\| \times \|\overline{feature}_i'\|}$$

其中, $\overline{feature}_i$ 是前景框中商品图像在一个尺度下属性i的特征向量, $\overline{feature}_i'$ 是商品样本图像在相同尺度下属性i的特征向量,i取自然数。

5. 根据权利要求1所述基于图像相似度的商品相似计算方法,其特征在於:

所述商品图像的相似度采用以下方式计算:

$$\text{商品图像的相似度} = \sum_n \text{TreeWweight}_n \times \text{Scroe}_n$$

其中,TreeWweight_n是第n个决策树的权重;Scroe_n是从第n个决策树的决策路径上获得的相似度分值。

6. 根据权利要求1所述基于图像相似度的商品相似计算方法,其特征在於:

所述采用的决策森林模型通过“成对排序学习”迭代构建,具体包括:

每次迭代时,选择进行分支扩展的节点,分支扩展所需的属性特征及相似度阈值,以及扩展分支的相似度分值;

每次迭代的选择,使得商品样本中任意两个商品的相似度分值差距,与其来自决策森林的商品图像相似度差距,尽可能一致,即最大化以下公式:

$$\sum_{Sample_m} \sum_{mi,mj} (1 - e^{FSimilarity_{mj} - FSimilarity_{mi}}) \times (LSimilarity_{mi} - LSimilarity_{mj})$$

其中,Sample_m,m=1,2,3...代表第m个商品样本;FSimilarity_{mj},j=1,2,3...代表第m个商品样本中第j个商品从决策森林获得的商品图像的相似度;LSimilarity_{mi}代表第m个商品样本中第i个商品被标注的相似度分值。

7. 基于图像相似度的商品推荐系统,其特征在于:包括:

历史购买商品收集模块(100),用于从当前用户正在浏览的电子商务系统的购买记录中,收集用户在当前电子商务系统的历史购买商品,收集当前电子商务系统中其他用户在当前电子商务系统的历史购买商品,提供给相似商品合并模块;

跨系统历史购买商品收集模块(200),用于解析其他电子商务系统的购买链接,收集用户在其他电子商务系统的历史购买商品图像,收集当前电子商务系统中其他用户在其他电子商务系统的历史购买商品图像;

线下购买商品收集模块(300),用于上传在线下商户购买的商品图片,收集用户在线下的历史购买商品图像,收集当前电子商务系统中其他用户在线下的历史购买商品图像;

相似商品合并模块(500),用于通过上述权利要求1-6任一基于图像相似度的商品相似计算方法计算商品图像的相似度,合并来自不同商家的同一款商品,把图像相似度>0.95的商品赋予相同的商品标识,建立用户-商品矩阵;

推荐商品计算模块(400),用于根据相似商品合并模块(500)的用户-商品矩阵计算推荐商品。

8. 根据权利要求7所述基于图像相似度的商品推荐系统,其特征在于:所述相似商品合并模块(500)还用于:

获取所述当前电子商务系统中其他用户在其他电子商务系统的历史购买商品图像,通过上述权利要求1-6任一基于图像相似度的商品相似计算方法计算商品图像的相似度,把这些历史购买商品合并到用户-商品矩阵中,即把图像相似度<0.95的商品赋予新的商品标识,添加为新的商品,更新用户-商品矩阵;

获取所述当前电子商务系统中其他用户在线下的历史购买商品图像;通过上述权利要求1-6任一基于图像相似度的商品相似计算方法计算商品图像相似度,把这些历史购买商品合并到用户-商品矩阵中,即把图像相似度<0.95的商品赋予新的商品标识,添加为新的商品,更新用户-商品矩阵。

9. 根据权利要求7所述基于图像相似度的商品推荐系统,其特征在于:

所述推荐商品计算模块(400)包括相似用户计算模块(401A)和推荐商品获取模块(401B);

所述相似用户计算模块(401A)用于根据所述用户-商品矩阵计算该用户行与其他用户行的商品重合数,基于商品重合数计算用户之间的相似度,选择用户相似度>0.75的用户作为该用户的相似用户;

所述基于商品重合数计算用户之间的相似度为：

$$\text{用户相似度} = \text{Max}\left(\frac{\text{AmountSITEM}_{\text{user1vs.user2}}}{\text{AmountITEM}_{\text{user1}}}, \frac{\text{AmountSITEM}_{\text{user1vs.user2}}}{\text{AmountITEM}_{\text{user2}}}\right)$$

其中, AmountITEM_{user1}是用户1的历史购买商品数, AmountITEM_{user2}是用户2的历史购买商品数; AmountSITEM_{user1vs.user2}是用户1和用户2重合的商品数;

所述推荐商品获取模块(401B)用于从用户正在浏览的电子商务系统中,选择和相似用户的历史购买商品相似度>0.95的商品,计算所选商品的商品推荐指数,将商品推荐指数>0.8的商品作为推荐商品;

所述计算所选商品的商品推荐指数为：

$$\text{商品推荐指数} = \left(0.8 + \frac{\text{Min}(\text{Amount}_{\text{user}}, 5)}{10}\right) \times \text{ImageSimilarity}$$

其中, Amount_{user}是相似用户的数量, ImageSimilarity是该商品与相似用户购买商品的图像相似度。

10. 根据权利要求7所述基于图像相似度的商品推荐系统,其特征在于：

所述推荐商品计算模块(400)包括相似商品计算模块(402A)和推荐商品获取模块(402B)；

所述相似商品计算模块(402A)用于根据所述用户-商品矩阵计算该用户正在浏览的商品列与其他商品列的用户重合数,基于用户重合数计算商品之间的相关度,将商品相关度>0.75的商品作为相关商品；

优选地,所述基于用户重合数计算商品之间的相关度为：

$$\text{商品相关度} = \text{Max}\left(\frac{\text{AmountSUSER}_{\text{item1vs.item2}}}{\text{AmountUSER}_{\text{item1}}}, \frac{\text{AmountSUSER}_{\text{item1vs.item2}}}{\text{AmountUSER}_{\text{item2}}}\right)$$

其中, AmountUSER_{item1}是购买商品1的用户数, AmountUSER_{item2}是购买商品2的用户数; AmountSUSER_{item1vs.item2}是商品1和2的用户重合数;

所述推荐商品获取模块(402B)用于从用户正在浏览的电子商务系统中,选择和相关商品相似度>0.95的商品作为推荐商品。

基于图像相似度的商品相似计算方法及商品推荐系统

技术领域

[0001] 本发明涉及互联网电子商务技术领域,特别涉及基于图像相似度的商品相似计算方法及商品推荐系统。

技术背景

[0002] 协同过滤是当前用户推荐中广泛使用的技术。协同过滤分析用户兴趣,在用户群中找到指定用户的相似(兴趣)用户,综合相似用户对某一信息的评价,形成对该指定用户对此信息的喜好程度预测。

[0003] 协同过滤通过用户的购买行为,建立用户-商品矩阵,以此矩阵为基础,进行“相似用户的协同过滤”以及“相关商品的协同过滤”。

[0004] “相似用户的协同过滤”是从用户-商品矩阵的“用户行”为出发点,通过比较用户和当前浏览平台中其它用户的历史购买数据,获得用户购买行为的相似度(譬如喜好,关注点),以相似度高的用户曾发生的购买行为作为推荐依据。

[0005] 表1相似用户的协同过滤

[0006]

用户/商品	商品A	商品B	商品C	商品D
用户A	√		√	推荐
用户B		√		
用户C	√		√	√

[0007] 如表1中,用户C是用户A的相似用户,用户A从用户C处获得商品C作为推荐(商品C为用户C已购买过的但用户A未曾购买的商品)。

[0008] “相关商品的协同过滤”是从用户-商品矩阵的“商品列”为出发点,通过累计不同商品的购买行为发生在同一用户上的次数,获得商品之间的相关性,以相关度高的商品作为推荐依据。

[0009] 表2相关商品的协同过滤

[0010]

用户/商品	商品A	商品B	商品C
用户A	√		√
用户B	√	√	√
用户C	√		推荐

[0011] 如表2中,从用户A和用户B的购买行为中,发现商品A和商品C具有相关性,因此当用户C购买商品A时,商品C作为商品A的相关商品,作为提供给用户C的推荐。

[0012] 现有以协同过滤为基础的商品推荐技术,存在以下问题:

[0013] 1) 用户的历史购买数据只来自于一个电子商务系统的购买行为,有限的历史购买数据导致用户-商品矩阵的稀疏性;

[0014] 就矩阵的“用户行”角度而言,它限制了用户购买倾向的描述能力,进而限制了相

似用户的判断准确性,影响推荐的质量;特别是对于平台的新用户,现有的系统无法做出推荐。

[0015] 就矩阵的“商品列”而言,它降低了相关商品发生在多个不同用户购买行为中的概率,降低了相关商品的判断准确性,影响推荐的质量。

[0016] 2) 相同的商品可能具有多个不同的标识,来自不同商家的同一商品在协同过滤的用户-商品矩阵被标识为不同的商品,这增加了用户-商品矩阵的稀疏性。

[0017] 3) 无论是“相似用户的协同过滤”,还是“相关商品的协同过滤”,所依据的推荐商品,必须来自用户当前浏览的电子商务系统,在其他电子商务系统或是线下的购买商品无法作为推荐的候选。

发明内容

[0018] 为解决以上技术问题,本发明提供基于图像相似度的商品相似计算方法及以此相似度为基础的商品推荐系统。

[0019] 本发明基于图像相似度的商品相似计算方法,包括:

[0020] 步骤101、对目标图像进行预处理,去除亮度、色差等光照条件的变化所带来的图像差异;

[0021] 步骤102、对目标图像进行处理检测出前景框;

[0022] 步骤103、通过双线性插值,把前景框内的商品图像变换为不同尺度的像素图,在不同尺度下的获取前景框中商品图像不同维度的属性特征;

[0023] 步骤104、计算前景框中商品图像属性特征向量与商品样本图像属性特征向量在不同尺度下属性特征相似度;

[0024] 步骤105、采用决策森林模型联合不同尺度下的属性特征相似度,计算前景框中商品图像与商品样本图像在不同尺度的像素图像下的商品图像的相似度。

[0025] 本发明基于图像相似度的商品推荐系统,包括:

[0026] 历史购买商品收集模块100,用于从当前用户正在浏览的电子商务系统的购买记录中,收集用户在当前电子商务系统的历史购买商品,收集当前电子商务系统中其他用户在当前电子商务系统的历史购买商品,提供给相似商品合并模块;

[0027] 跨系统历史购买商品收集模块200,用于解析其他电子商务系统的购买链接,收集用户在其他电子商务系统的历史购买商品图像,收集当前电子商务系统中其他用户在其他电子商务系统的历史购买商品图像;

[0028] 线下购买商品收集模块300,用于上传在线下商户购买的商品图片,收集用户在线下的历史购买商品图像,收集当前电子商务系统中其他用户在线下的历史购买商品图像;

[0029] 相似商品合并模块500,用于通过上述权利要求1-6任一基于图像相似度的商品相似计算方法计算得到的商品图像的相似度,合并来自不同商家的同一款商品,把图像相似度 >0.95 的商赋予相同的商品标识,建立用户-商品矩阵;

[0030] 推荐商品计算模块400,用于根据相似商品合并模块500的用户-商品矩阵计算推荐商品。

[0031] 优选地,所述相似商品合并模块500还用于:

[0032] 获取所述当前电子商务系统中其他用户在其他电子商务系统的历史购买商品图

像,通过上述权利要求1—6任一基于图像相似度的商品相似计算方法计算商品图像的相似度,把这些历史购买商品合并到用户-商品矩阵中,即把图像相似度 <0.95 的商品赋予新的商品标识,添加为新的商品,更新用户-商品矩阵;

[0033] 获取所述当前电子商务系统中其他用户在线下的历史购买商品图像;通过上述权利要求1—6任一基于图像相似度的商品相似计算方法计算商品图像相似度,把这些历史购买商品合并到用户-商品矩阵中,即把图像相似度 <0.95 的商品赋予新的商品标识,添加为新的商品,更新用户-商品矩阵。

[0034] 本发明以构建“商品图像作为统一标识”的基础,利用多个深度卷积神经网络提取商品图像不同尺度/不同属性(包括:形状,颜色,纹理,风格等)的特征,并通过图像相似度的对比,联合不同尺度/不同属性特征的决策森林模型(该模型反映了不同尺度/不同属性特征在计算图像相似度时的权重),以此计算商品图像的相似度。避免了用户的历史购买数据只来自于一个电子商务系统的购买行为,有限的历史购买数据导致用户-商品矩阵的稀疏性的问题,并且有效解决了相同的商品可能具有多个不同的标识、来自不同商家的同一商品在协同过滤的用户-商品矩阵被标识为不同的商品的问题,在其他电子商务系统或是线下的购买商品均可以作为推荐的候选,极大地增加了推荐商品来源的数据可靠性。

附图说明

[0035] 图1为本发明基于图像相似度的商品相似计算方法优选实施例流程示意图;

[0036] 图2为本发明基于图像相似度的商品相似计算方法对目标图像进行预处理后的示例示意图;

[0037] 图3为本发明基于图像相似度的商品相似计算方法对目标图像进行处理检测出前景框优选实施例流程示意图;

[0038] 图4为本发明基于图像相似度的商品相似计算方法对目标图像进行处理检测出前景框中通过图像分割算法从目标图像生成若干候选框示例示意图;

[0039] 图5为本发明基于图像相似度的商品相似计算方法联合不同尺度/不同属性特征的决策森林优选实施例流程示意图;

[0040] 图6为本发明基于图像相似度的商品推荐系统优选实施例结构示意图;

[0041] 图7为本发明基于图像相似度的商品推荐系统推荐商品计算模块优选实施例结构示意图;

[0042] 图8为本发明基于图像相似度的商品推荐系统推荐商品计算模块另一优选实施例结构示意图。

具体实施方式

[0043] 下面结合附图,对本发明实施例中的技术方案进行清楚、完整地描述。

[0044] 本发明中公式中若无特别说明,Max()表示取最大值运算,Min()表示取最小值运算。

[0045] 本发明基于图像相似度的商品相似计算方法优选实施例,如图1所示,包括:

[0046] 步骤101、对目标图像进行预处理,去除亮度、色差等光照条件的变化所带来的图像差异;

[0047] 所述目标图像来自于互联网以及现实交易,既可以通过网络获取,也可以通过对实现进行现场拍摄获取。

[0048] 优选地,使用Retinex方法对目标图像进行预处理,去除亮度,色差等光照条件的变化所带来的图像差异,如图2所示的(a)(b)两图为经过预处理的目标图像。

[0049] 步骤102、对目标图像进行处理,检测出前景框。

[0050] 作为一种可实现方式,所述对目标图像进行处理,检测出前景框可以为:人为地将目标图像用特定大小方形框进行选择,例如提供一个前景选择界面给用户,将包含有商品图像的方形框确定为前景框。该方式需要通过用户自身的手动选择标定前景框,前景框选择不准确,操作也不方便。

[0051] 优选地,本发明提供另一种可实现方式,所述对目标图像进行处理,检测出前景框为:使用基于区域的卷积神经网络检测出商品图像的前景框,如图3所示,具体包括:

[0052] 102-1、通过图像分割算法从目标图像生成若干候选框

[0053] 所述图像分割算法根据基于像素值相似度的图像分割信息从一定数量的滑动框中选择出候选框,可以采用Selective Search图像分割算法。

[0054] 所述Selective Search图像分割算法为本领域公知技术手段,不再赘述。

[0055] 如图4(a)所示的服装图像,图像中的方框为用图像分割算法得到的部分候选框(完整的候选框数目为1292个)。

[0056] 102-2、对每个候选框采用卷积神经网络提取属性特征,即通过卷积神经网络的卷积核对待选框中的图像像素值进行卷积运算得到卷积值,输出是属性特征向量。

[0057] 所述卷积神经网络是这近年来提出的技术,在特征表达上具有最好的效果,它的卷积核都是通过学习样本或商品样本和所连接的分类器训练得到的。一般而言包含5个卷积层,每个卷积层包含多个卷积核,可以理解这些核是空间变换的基本基函数,可以通过卷积核进行空间映射,譬如非线性到线性。

[0058] 所述卷积神经网络可以采用本领域公知技术手段,优选地,本发明采用以下方式建立卷积神经网络:

[0059] A) 建立样本库,包含NX种不同的商品,NX是样本数,取值范围为5000-20000,优选10000

[0060] B) 在3种光照条件下,从3种拍摄角度拍摄商品图片;

[0061] C) 以步骤A得到的不同商品作为不同的类别,建立分类器,以卷积神经网络的输出为该分类器所采用的分类特征,以步骤B得到的商品图片为评估样本;通过“后向反馈”的方法,逐层调整卷积神经网络中的卷积层,使得以卷积神经网络的输出作为分类特征的分类器,在评估样本上达到最好的分类精度。

[0062] 所述分类器可以采用SoftMax回归算法,所述“SoftMax回归算法”,“后向反馈”为本领域公知技术手段,不再赘述。

[0063] 102-3、采用SVM分类器将候选框分成候选前景框或非前景框

[0064] 采用的支持向量机SVM分类器把候选框分为前景框或非前景框,从正、负两类商品样本中建立得到的,正样本为商品图像中的前景区域(即包含有商品的最小方形区域),负样本为从商品图像的非前景区域中随机选择的任意方形区域

[0065] 如图4(b)所示的服装图像,图像中的方框为用该方法得到的候选前景框。

[0066] 所述SVM分类器本身不是本发明内容,其为本领域公知技术手段,不再赘述。

[0067] 102-4、从候选前景框中选择置信度最大的作为商品图像的前景框。

[0068] 所述置信度是从SVM分类器得到的类别归属概率。如图4(c)所示的服装图像,图像中的方框为最终的前景框。

[0069] 所述置信度是SVM的输出,例如属于前景的概率为0.8,属于非前景的概率为0.2,则该候选框被SVM分为前景,其置信度为0.8。

[0070] 步骤103、通过双线性插值,把前景框内的商品图像变换为不同尺度的像素图,在不同尺度下的获取前景框中商品图像不同维度的属性特征。

[0071] 所述双线性插值,又称为双线性内插,在数学上,其是有两个变量的插值函数的线性插值扩展,其核心思想是在横轴和纵轴两个方向分别进行一次线性插值。具体算法为本领域公知技术,不再赘述。

[0072] 本发明图像变换的尺度可以进行灵活设置,本发明实施例优选使用3个尺度,128像素×128像素、160像素×160像素、224像素×224像素。

[0073] 而维度的选择取决于商品所在的大类,如服装,鞋,包等等;.

[0074] 以服装类商品举例,其维度包含“服装类型”(毛衣,针织衫,短外套,卫衣,风衣,T恤,衬衫,大衣,羽绒服等)，“风格”(淑女,韩版,民族,欧美,通勤,波西米亚,复古,性感,运动等等)，“领型”(V领,POLO领,圆领,方领,立领,一字领,娃娃领等)，“袖型”(喇叭袖,灯笼袖,泡泡袖,蝙蝠袖,荷叶袖,等)，“材质”(PU,棉,麻,羊毛,羊绒,蚕丝,莫代尔等)，“图案”(格子,碎花,植物花卉,花色,纯色等)，“颜色”(花色,黑色,白色,透明,浅灰色等)，“版型”(宽松,修身,直筒,高腰型等)，“裙型”(低腰,中腰,高腰,超低腰等),等等。

[0075] 在每个尺度每个维度,通过有监督学习建立一个卷积神经网络,(譬如对于服装类商品,在“服装类型”这个维度,通过优化“服装类型”的分类精度,得到能够表述“服装类别”的最优的卷积核),以此来抽取每个尺度每个维度的属性特征(该特征抽取方式与步骤102-2相同)。

[0076] 步骤104、计算前景框中商品图像属性特征向量与商品样本图像属性特征向量在不同维度下属性特征相似度

[0077] 优选地,所述属性特征相似度可以通过以下方式计算:

$$[0078] \quad \text{属性特征相似度} = \frac{\overline{feature_i} \cdot \overline{feature_i'}}{\| \overline{feature_i} \| \times \| \overline{feature_i'} \|}$$

[0079] 其中, $\overline{feature_i}$ 是前景框中商品图像在一个尺度下属性i的特征向量, $\overline{feature_i'}$ 是商品样本图像在相同尺度下属性i的特征向量,i取自自然数; \cdot 表示矢量点乘;双竖线为矢量的模。

[0080] 所述商品样本图像为预先建立的商品样本图像库,图像库里的图像是对商品在3种光照条件下进行拍摄,并变换为不同尺度的像素图像,在每个尺度下提取了属性i的特征向量 $\overline{feature_i'}$ 。

[0081] 步骤105、采用决策森林模型联合不同尺度下的属性特征相似度,计算前景框中商品图像与商品样本图像在不同尺度的像素图像下的商品图像的相似度。

[0082] 本发明采用决策森林模型联合不同尺度下商品图像的属性特征的相似度,如图5

所示联合不同尺度/不同属性特征的决策森林,决策森林由一系列决策树构成,每个决策树代表一个决策子空间(即相似度子空间),由一系列决策路径所构成,每条决策路径代表一个决策模式(即相似度模式),由决策树上的“根结点”(图5中的“黑色实心圆”),“子节点”(图5中的“空心圆”)和“叶子结点”(图5中的“灰色实心圆”)组合而成,以“根结点”作为起始结点,以“叶子结点”为终止结点。“根结点”和“子节点”是决策路径上的分支结点,它包含一个尺度下商品图像的一个属性特征(图5中的 $feature_i, i=1,2,3,\dots$),以及与该属性特征相似度(由步骤104计算得到)对应的相似度阈值(图5中的 $th_j, j=1,2,3,\dots$),以此作为决策路径扩展的依据。

[0083] 具体的,当该属性特征相似度大于阈值时,决策路径选择左边的分支进行扩展,反之,选择右边的分支。“叶子结点”作为决策路径上的终止结点,包含一个相似度分值(图5中的 $score_k, k=1,2,3,\dots$),描述该路径所代表的相似度模式。相似度分值的获得以“根结点”作为起始结点,以“叶子结点”为终止结点。

[0084] 所述“分支”是决策树的基本术语,根据每个分支所用的特征可能可以具体解释它的意义。分支是多路径存在的前提,直观的可以理解为把特征空间进行分段线性表示。

[0085] 联合决策树上的决策路径,计算商品图像的相似度,

[0086] 商品图像的相似度 = $\sum_n TreeWeight_n \times Score_n$

[0087] 其中, $TreeWeight_n$ 是第n个决策树的权重; $Score_n$ 是从第n个决策树的决策路径上获得的相似度分值。

[0088] 所述采用的决策森林模型是通过“成对排序学习”迭代构建的:

[0089] 每次迭代时,选择进行分支扩展的节点,分支扩展所需的属性特征及相似度阈值,以及扩展分支的相似度分值。

[0090] 而每次迭代的选择,使得商品样本中任意两个商品的相似度分值差距,与其来自决策森林的商品图像相似度差距,尽可能一致,即最大化以下公式:

[0091]
$$\sum_{Sample_m} \sum_{mi,mj} (1 - e^{FSimilarity_{mj} - FSimilarity_{mi}}) \times (LSimilarity_{mi} - LSimilarity_{mj})$$

[0092] 其中, $Sample_m, m=1,2,3,\dots$ 代表第m个商品样本; $FSimilarity_{mj}, j=1,2,3,\dots$ 代表第m个商品样本中第j个商品从决策森林获得的商品图像的相似度; $LSimilarity_{mi}$ 代表第m个商品样本中第i个商品被标注的相似度分值。

[0093] 这是决策森林训练时所需要优化的目标函数,和最终的相似度计算没有关系,每次迭代生成的决策森林都可以给出一个相似度的中间结果,该目标函数简单说来就是检查该中间结果是否和人工标定的结果的“一致性”。

[0094] 得到的商品图像的相似度是一个概然系数,表示与商品样本的相似度。每个商品样本都由一个商品及与该商品相似或不相似的若干商品构成,并对这若干商品根据它们的相似程度赋予不同的商品相似阈值,商品相似阈值取值范围为0—1,例如,可以设定1.0代表完全相似;0.75代表基本相似,在以商品内容相似度为基础进行推荐的系统中,该商品可以作为相似内容的商品进行推荐;0代表不相似。当然,可以根据需要设定不同的商品相似阈值。

[0095] 本发明提出基于图像相似度的商品相似计算方法,以此构建“商品图像作为统一

标识”的基础,利用多个深度卷积神经网络提取商品图像不同尺度/不同属性(包括:形状,颜色,纹理,风格等)的特征,并通过图像相似度的对比,联合不同尺度/不同属性特征的决策森林模型(该模型反映了不同尺度/不同属性特征在计算图像相似度时的权重),以此计算商品图像的相似度。避免了用户的历史购买数据只来自于一个电子商务系统的购买行为,有限的历史购买数据导致用户-商品矩阵的稀疏性的问题,并且有效解决了相同的商品可能具有多个不同的标识、来自不同商家的同一商品在协同过滤的用户-商品矩阵被标识为不同的商品的问题,在其他电子商务系统或是线下的购买商品均可以作为推荐的候选,极大地增加了推荐商品来源的数据可靠性。

[0096] 本发明基于图像相似度的商品推荐系统优选实施例,如图6所示,包括:

[0097] 历史购买商品收集模块100,用于从当前用户正在浏览的电子商务系统的购买记录中,收集用户在当前电子商务系统的历史购买商品,收集当前电子商务系统中其他用户当前电子商务系统的历史购买商品,提供给相似商品合并模块500;

[0098] 相似商品合并模块500,用于通过上述任一基于图像相似度的商品相似计算方法计算商品图像的相似度,合并来自不同商家的同一款商品,把图像相似度 >0.95 的商品(譬如不同商家的同一款商品)赋予相同的商品标识,建立用户-商品矩阵。

[0099] 优选地,相似商品合并模块500还用于:

[0100] 获取当前电子商务系统中其他用户在其他电子商务系统的历史购买商品图像,通过上述任一基于图像相似度的商品相似计算方法计算商品图像的相似度,把这些历史购买商品合并到相似商品合并模块100的用户-商品矩阵中,即把图像相似度 <0.95 的商品赋予新的商品标识,添加为新的商品,更新用户-商品矩阵。

[0101] 获取当前电子商务系统中其他用户在线下的历史购买商品图像;通过上述任一基于图像相似度的商品相似计算方法计算商品图像相似度,把这些历史购买商品合并到跨系统历史购买商品收集模块200的用户-商品矩阵中,即把图像相似度 <0.95 的商品赋予新的商品标识,添加为新的商品,更新用户-商品矩阵。

[0102] 跨系统历史购买商品收集模块200,用于解析其他电子商务系统的购买链接,收集用户在其他电子商务系统的历史购买商品图像,收集当前电子商务系统中其他用户在其他电子商务系统的历史购买商品图像,提供给相似商品合并模块500;

[0103] 线下购买商品收集模块300,用于上传在线下商户购买的商品图片,收集用户在线下的历史购买商品图像,收集当前电子商务系统中其他用户在线下的历史购买商品图像,提供给相似商品合并模块500;

[0104] 推荐商品计算模块400,用于根据相似商品合并模块500的用户-商品矩阵计算推荐商品;

[0105] 作为一种可实现方式,如图7所示,所述推荐商品计算模块400包括相似用户计算模块401A和推荐商品获取模块401B;

[0106] 所述相似用户计算模块401A用于根据所述用户-商品矩阵计算该用户行与其他用户行的商品重合数,基于商品重合数计算用户之间的相似度,选择用户相似度 >0.75 的用户作为该用户的相似用户;

[0107] 优选地,所述基于商品重合数计算用户之间的相似度为:

$$[0108] \quad \text{用户相似度} = \text{Max}\left(\frac{\text{AmountSITEM}_{\text{user1vs.user2}}}{\text{AmountITEM}_{\text{user1}}}, \frac{\text{AmountSITEM}_{\text{user1vs.user2}}}{\text{AmountITEM}_{\text{user2}}}\right)$$

[0109] 其中, AmountITEM_{user1}是用户1的历史购买商品数, AmountITEM_{user2}是用户2的历史购买商品数; AmountITEM_{user1vs.user2}是用户1和用户2重合的商品数。

[0110] 所述推荐商品获取模块401B用于通过权利所述的计算商品图像相似度的方法, 从用户正在浏览的电子商务系统中, 选择和相似用户的历史购买商品相似度>0.95的商品, 计算所选商品的商品推荐指数, 将商品推荐指数>0.8的商品作为推荐商品;

[0111] 优选地, 所述计算所选商品的商品推荐指数为:

$$[0112] \quad \text{商品推荐指数} = \left(0.8 + \frac{\text{Min}(\text{Amount}_{\text{suser}}, 5)}{10}\right) \times \text{ImageSimilarity}$$

[0113] 其中, Amount_{suser}是相似用户的数量, ImageSimilarity是该商品与相似用户购买商品的图像相似度。

[0114] 作为一种另可实现方式, 如图8所示, 所述推荐商品计算模块400包括相似商品计算模块402A和推荐商品获取模块402B

[0115] 所述相似商品计算模块402A用于根据所述用户-商品矩阵计算该用户正在浏览的商品列与其他商品列的用户重合数, 基于用户重合数计算商品之间的相关度, 将商品相关度>0.75的商品作为相关商品。

[0116] 优选地, 所述基于用户重合数计算商品之间的相关度为:

$$[0117] \quad \text{商品相关度} = \text{Max}\left(\frac{\text{AmountSUSER}_{\text{item1vs.item2}}}{\text{AmountUSER}_{\text{item1}}}, \frac{\text{AmountSUSER}_{\text{item1vs.item2}}}{\text{AmountUSER}_{\text{item2}}}\right)$$

[0118] 其中, AmountUSER_{item1}是购买商品1的用户数, AmountUSER_{item2}是购买商品2的用户数; AmountSUSER_{item1vs.item2}是商品1和2的用户重合数。

[0119] 所述推荐商品获取模块402B用于通过权利所述的计算商品图像相似度的方法, 从用户正在浏览的电子商务系统中, 选择和商品相似度>0.95的商品作为推荐商品。

[0120] 本发明基于图像相似度的商品推荐系统, 提出了以商品图像作为商品在不同商务平台的统一标识, 以此联合用户在不同平台的购买数据(不同商家/不同电子商务系统/甚至线下的商店), 建立用户-商品矩阵的协同过滤推荐系统。本发明解决了用户的历史购买数据只来自于一个电子商务系统的购买行为, 有限的历史购买数据导致用户-商品矩阵的稀疏性, 使得推荐不准确及新用户冷启动问题; 并且有效解决了相同的商品可能具有多个不同的标识、来自不同商家的同一商品在协同过滤的用户-商品矩阵被标识为不同的商品的问题, 在其他电子商务系统或是线下的购买商品均可以作为推荐的候选, 极大地增加了推荐商品来源的数据可靠性。

[0121] 显然, 描述的实施例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。

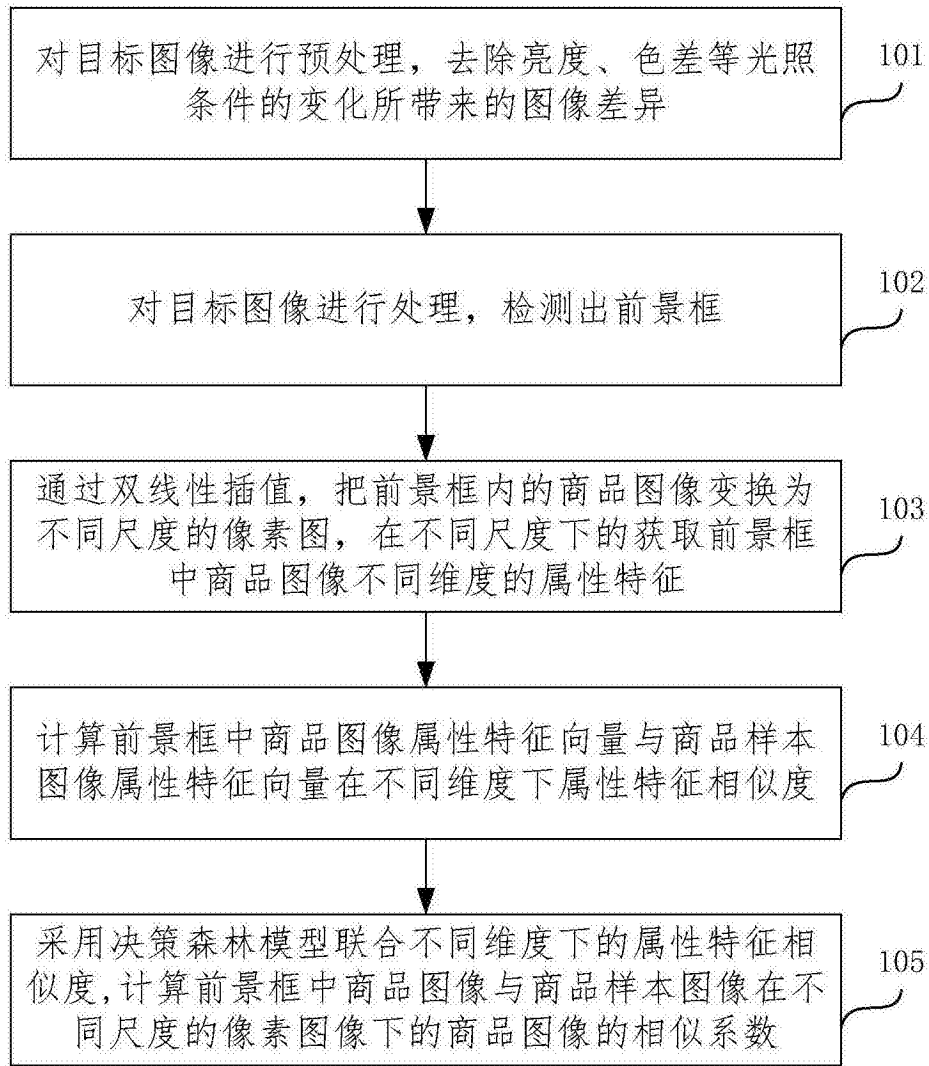


图1

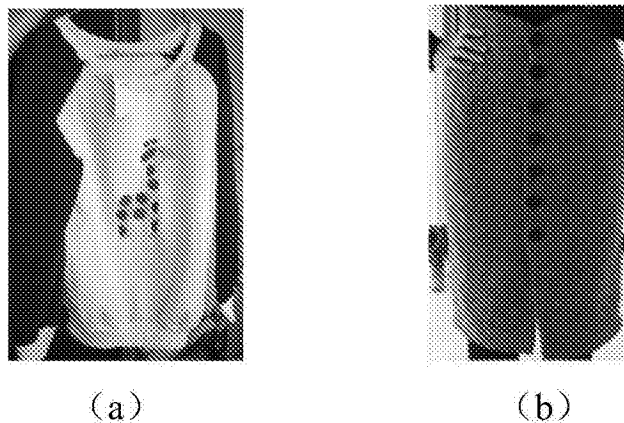


图2

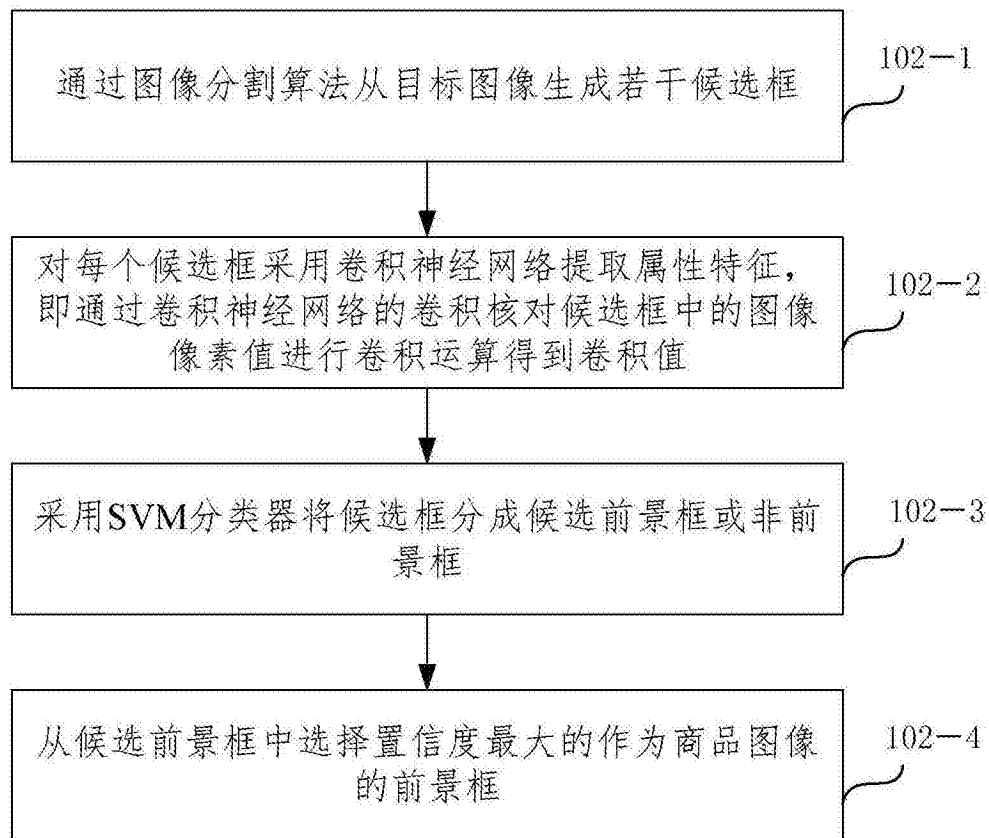


图3

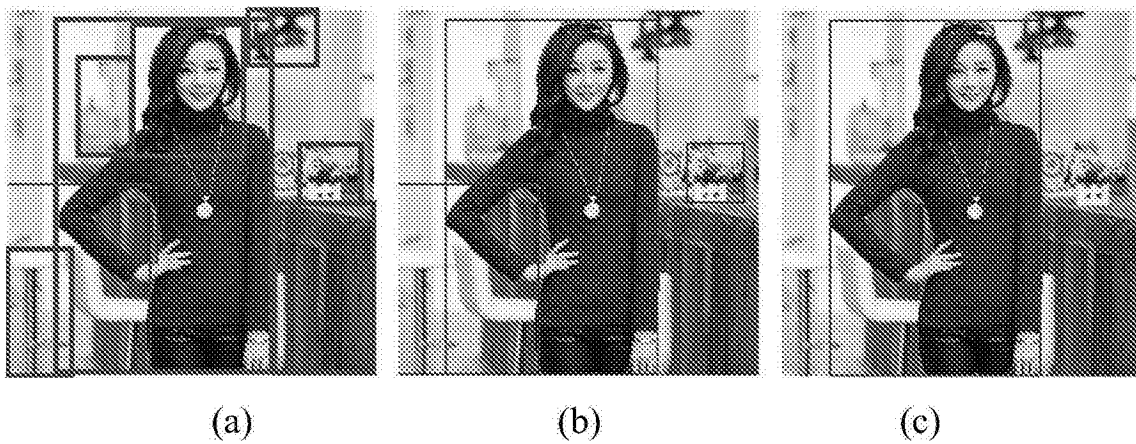


图4

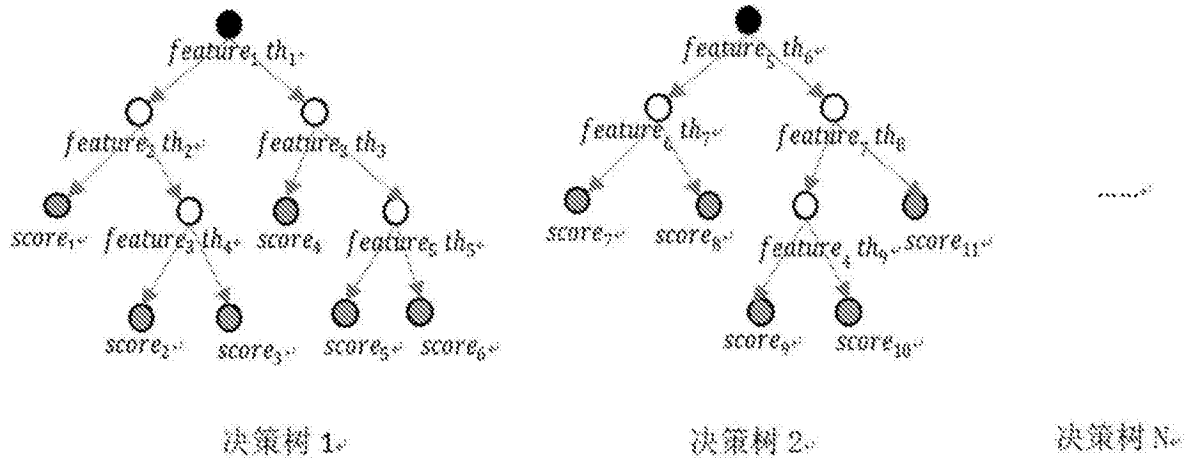


图5

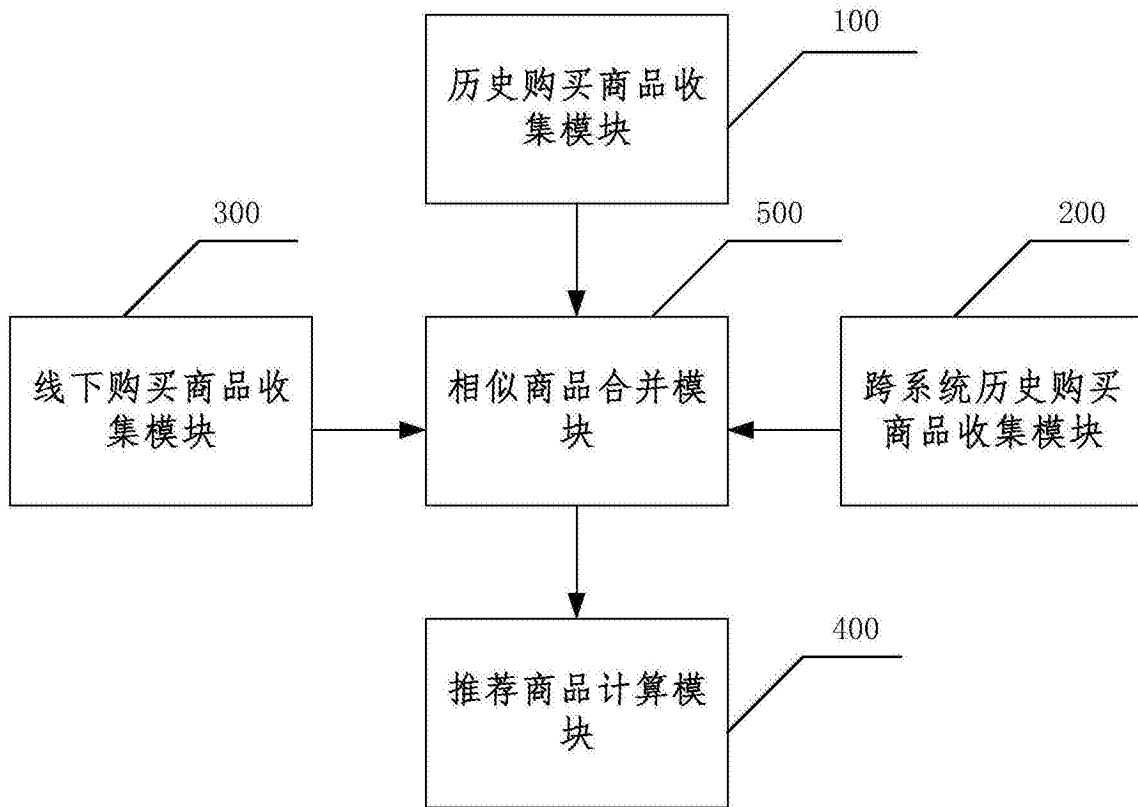


图6

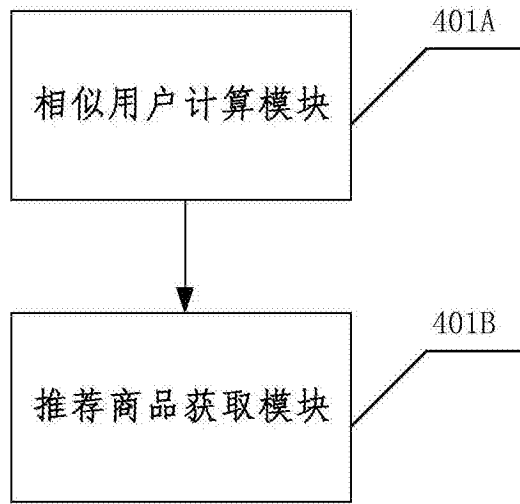


图7

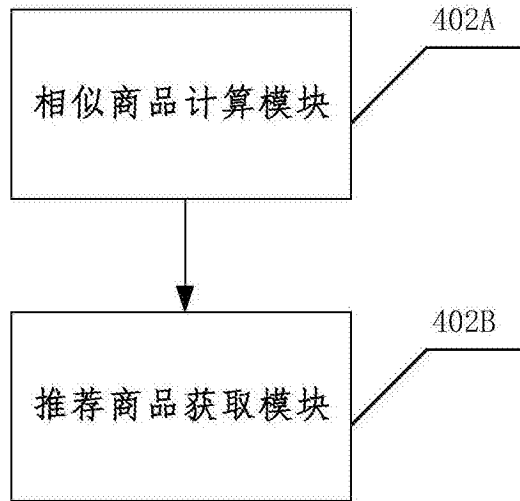


图8