



(12) 发明专利

(10) 授权公告号 CN 108665329 B

(45) 授权公告日 2024. 09. 24

(21) 申请号 201710200470.6

(22) 申请日 2017.03.29

(65) 同一申请的已公布的文献号  
申请公布号 CN 108665329 A

(43) 申请公布日 2018.10.16

(73) 专利权人 北京京东尚科信息技术有限公司  
地址 100195 北京市海淀区杏石口路65号  
西杉创意园四区11号楼东段1-4层西  
段1-4层  
专利权人 北京京东世纪贸易有限公司

(72) 发明人 杨俊

(74) 专利代理机构 中原信达知识产权代理有限  
责任公司 11219  
专利代理师 张一军 姜劲

(51) Int.Cl.

G06Q 30/0601 (2023.01)

G06F 16/2458 (2019.01)

G06F 16/28 (2019.01)

(56) 对比文件

CN 105469263 A, 2016.04.06

CN 103489117 A, 2014.01.01

CN 102156932 A, 2011.08.17

CN 105787770 A, 2016.07.20

审查员 李芳

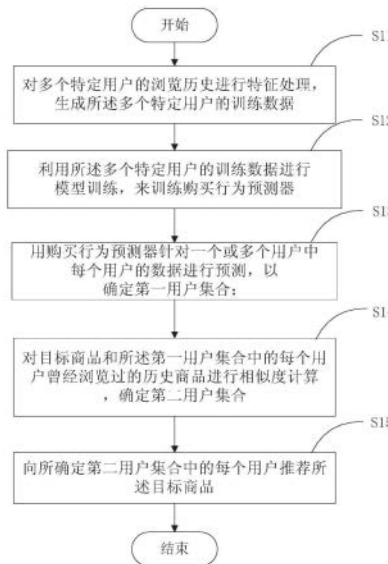
权利要求书3页 说明书11页 附图4页

(54) 发明名称

一种基于用户浏览行为的商品推荐方法

(57) 摘要

本发明提供一种基于用户浏览行为的商品推荐的方法,包括:由购买行为预测器分析一个或多个用户的商品浏览历史数据,确定在一未来时段较有可能进行购买的第一用户集合;针对所确定的第一用户集合中的每个用户:从该用户的商品浏览历史数据中获取该用户浏览过的一个或多个商品,将所述一个或多个商品与目标商品进行商品相似度计算,基于所述商品相似度计算的结果,确定所述用户在所述未来时段较有可能购买目标商品,以及将所述用户添加至第二用户集合;以及向所确定的第二用户集合中的每个用户推荐所述目标商品。



1. 一种基于用户浏览行为的商品推荐方法,其特征在于,所述方法包括:
  - 由购买行为预测器分析一个或多个用户的商品浏览历史数据,确定在一未来时段较有可能进行购买的第一用户集合;
  - 针对所确定的第一用户集合中的每个用户:
    - 从该用户的商品浏览历史数据中获取该用户浏览过的一个或多个商品,
    - 将所述一个或多个商品与目标商品进行商品相似度计算,包括:
      - 针对所述目标商品的类别,基于预先定义的商品属性特征模型,构建所述目标商品的目标商品向量;
    - 对于所述第一用户集合中的每个用户:
      - 基于所述预先定义的商品属性特征模型,构建该用户浏览过的商品中与所述目标商品同类别的商品的对比商品向量,其中所述目标商品向量与所述对比商品向量包含相同的向量维度;
      - 对目标商品向量和所述对比商品向量进行相似度计算;
      - 基于所述商品相似度计算的结果,确定所述用户在所述未来时段较有可能购买目标商品,以及
      - 将所述用户添加至第二用户集合;以及
      - 向所确定的第二用户集合中的每个用户推荐所述目标商品。
2. 根据权利要求1所述的方法,其特征在于,其中由所述购买行为预测器分析所述一个或多个用户的所述商品浏览历史数据,确定在所述未来时段较有可能进行购买的所述第一用户集合包括:
  - 针对所述一个或多个用户中的每个用户:
    - 基于该用户的商品浏览历史数据,由所述购买行为预测器确定该用户在所述未来时段进行购买行为的可能性;
    - 响应于确定所述可能性大于阈值,将该用户添加到所述第一用户集合。
3. 根据权利要求1所述的方法,还包括,对所述购买行为预测器进行训练,其中所述训练包括:
  - 接收多个特定用户的样本商品浏览历史数据;
  - 基于所接收的样本商品浏览历史数据生成与所述多个特定用户相关联的训练数据;
  - 利用所述训练数据训练所述购买行为预测器。
4. 根据权利要求3所述的方法,其特征在于,生成与所述多个特定用户相关联的训练数据包括:
  - 针对所述多个特定用户中的每个特定用户,基于预先定义的用户行为特征,提取该特定用户的样本基础历史数据的特征并添加标签,生成与该特定用户相关联的训练数据;
  - 聚集与每个特定用户相关联的训练数据,对所聚集的训练数据进行标准化处理,生成与所述多个特定用户相关联的所述训练数据。
5. 根据权利要求3-4中任一项所述的方法,其特征在于,其中所述特定用户相关联的训练数据是以向量来表示的。
6. 根据权利要求3-4中任一项所述的方法,其特征在于,其中所述训练的方法是二分类算法。

7. 根据权利要求1-4中任一项所述的方法,其特征在於,其中所述相似度计算是带有修正系数的余弦相似度计算。

8. 一种基于用户浏览行为的商品推荐的装置,其特征在於,包括:

模型计算模块,用于:

由购买行为预测器分析一个或多个用户的商品浏览历史数据,确定在一未来时段较有可能进行购买的第一用户集合;

针对所确定的第一用户集合中的每个用户:

从该用户的商品浏览历史数据中获取该用户浏览过的一个或多个商品,

将所述一个或多个商品与目标商品进行商品相似度计算,包括:

针对所述目标商品的类别,基于预先定义的商品属性特征模型,构建所述目标商品的目标商品向量;

对于所述第一用户集合中的每个用户:

基于所述预先定义的商品属性特征模型,构建该用户浏览过的商品中与所述目标商品同类别的商品的对比商品向量,其中所述目标商品向量与所述对比商品向量包含相同的向量维度;

对目标商品向量和所述对比商品向量进行相似度计算;

基于所述商品相似度计算的结果,确定所述用户在所述未来时段较有可能购买目标商品,以及

将所述用户添加至第二用户集合;以及

向所确定的第二用户集合中的每个用户推荐所述目标商品。

9. 根据权利要求8所述的装置,其特征在於,所述模型计算模块进一步用于:

针对所述一个或多个用户中的每个用户:

基于该用户的商品浏览历史数据,由所述购买行为预测器确定该用户在所述未来时段进行购买行为的可能性;

响应于确定所述可能性大于阈值,将该用户添加到所述第一用户集合。

10. 根据权利要求8所述的装置,其特征在於,所述模型计算模块进一步用于对所述购买行为预测器进行训练,其中所述训练包括:

接收多个特定用户的样本商品浏览历史数据;

基于所接收的样本商品浏览历史数据生成与所述多个特定用户相关联的训练数据;

利用所述训练数据训练所述购买行为预测器。

11. 根据权利要求10所述的装置,其特征在於,其中所述模型计算模块进一步用于:

针对所述多个特定用户中的每个特定用户,基于预先定义的用户行为特征,提取该特定用户的样本基础历史数据的特征并添加标签,生成与该特定用户相关联的训练数据;

聚集与每个特定用户相关联的训练数据,对所聚集的训练数据进行标准化处理,生成与所述多个特定用户相关联的所述训练数据。

12. 根据权利要求10-11中任一项所述的装置,其特征在於,其中所述特定用户相关联的训练数据是以向量来表示的。

13. 根据权利要求10-11中任一项所述的装置,其特征在於,其中所述训练的方法是二分类算法。

14. 根据权利要求8-11中任一项所述的装置,其特征在于,其中所述相似度计算是带有修正系数的余弦相似度计算。

15. 一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-7中任一所述的方法。

16. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行时实现如权利要求1-7中任一所述的方法。

## 一种基于用户浏览行为的商品推荐方法

### 技术领域

[0001] 本发明涉及一种基于用户浏览行为的商品推荐方法、系统、电子设备和可读存储介质。

### 背景技术

[0002] 随着互联网的兴起,越来越多的人到网上购物。电商行业的竞争也越来越激烈。各种营销手段也层出不穷。为了促进商品的销售,市场营销部门会进行商品的宣传,通常都会通过email、短信、移动电话推送消息等向用户发送各种营销信息。挖掘最有可能购买商品的潜在用户已经成为市场营销部门不可或缺的一环。

[0003] 向用户发送营销信息之前,需要分析潜在购买所推销的目标商品的用户。目前一般有三种常见的商品推荐方法:

[0004] 基于人群标签的推荐:分析用户的各种标签,如性别标签、收入标签、消费力标签、职业标签等等,针对具有相应标签的人群进行营销。基于人群标签方法,实际效果并不好,标签的确认是一个复杂的过程,并且很多标签是一个动态变化的过程,导致标签的准确率不高。

[0005] 基于商品的推荐:找出与目标营销商品相似的商品,然后提取购买相似商品的用户进行营销。基于商品的方法,完全从商品属性方面考虑,忽略了用户本身的特征,比如两款商品虽然很相似,但是品牌不同,如果用户的品牌忠诚度很高,这样的营销甚至会导致用户的反感。

[0006] 基于模型训练的推荐:目前使用的基于模型训练的推荐主要是以用户评价和对商品的评分作为基础数据的。但是用户评价和评分受很多主观因素影响,并且在用户对评分或评价不感兴趣而未进行相应行为的情况下,数据的收集并不全面。因此,以评价或评分进行模型训练,无法获取到用户最真实的购买习惯,也无法精确分析潜在购买目标商品的人群。

### 发明内容

[0007] 有鉴于此,本发明实施例提供一种基于用户浏览行为的商品推荐方法,能够利用模型计算结合相似度计算的方法,精确分析潜在购买目标商品的人群,提高了用户的实际订单转化率。利用模型计算来针对每个用户的历史数据进行分析,确定出每个用户在未来时段进行购买的可能性,从而得到较有可能进行购买行为的用户集合;然后针对目标商品和有可能进行购买行为的用户曾浏览过的商品进行相似度计算,最终得到在未来时段有可能购买该目标商品的用户。从而能够向有可能购买该目标商品的用户推荐该目标商品。

[0008] 为实现上述目的,根据本发明的技术方案的一个方面,提供了一种基于用户浏览行为的商品推荐方法。

[0009] 根据本发明的技术方案的一个方面的基于用户浏览行为的商品推荐方法,包括:由购买行为预测器分析一个或多个用户的商品浏览历史数据,确定在一未来时段较有可能

进行购买的第一用户集合;针对所确定的第一用户集合中的每个用户:从该用户的商品浏览历史数据中获取该用户浏览过的一个或多个商品,将所述一个或多个商品与目标商品进行商品相似度计算,基于所述商品相似度计算的结果,确定所述用户在所述未来时段较有可能购买目标商品,以及将所述用户添加至第二用户集合;以及向所确定的第二用户集合中的每个用户推荐所述目标商品。

[0010] 可选地,其中由所述购买行为预测器分析所述一个或多个用户的所述商品浏览历史数据,确定在所述未来时段较有可能进行购买的所述第一用户集合包括:针对所述一个或多个用户中的每个用户:基于该用户的商品浏览历史数据,由所述购买行为预测器确定该用户在所述未来时段进行购买行为的可能性;响应于确定所述可能性大于阈值,将该用户添加到所述第一用户集合。

[0011] 可选地,对所述购买行为预测器进行训练,其中所述训练包括:接收多个特定用户的样本商品浏览历史数据;基于所接收的样本商品浏览历史数据生成与所述多个特定用户相关联的训练数据;利用所述训练数据训练所述购买行为预测器。

[0012] 可选地,生成与所述多个特定用户相关联的训练数据包括:针对所述多个特定用户中的每个特定用户,基于预先定义的用户行为特征,提取该特定用户的样本基础历史数据的特征并添加标签,生成与该特定用户相关联的训练数据;聚集与每个特定用户相关联的训练数据,对所聚集的训练数据进行标准化处理,生成与所述多个特定用户相关联的所述训练数据。

[0013] 可选地,其中将所述一个或多个商品与目标商品进行商品相似度计算包括:针对所述目标商品的类别,基于预先定义的商品属性特征模型,构建所述目标商品的目标商品向量;对于所述第一用户集合中的每个用户:基于所述预先定义的商品属性特征模型,构建该用户浏览过的商品中与所述目标商品同类别的商品的对比商品向量,其中所述目标商品向量与所述对比商品向量包含相同的向量维度;对目标商品向量和所述对比商品向量进行相似度计算。

[0014] 可选地,其中所述用户的训练数据是以向量来表示的。

[0015] 可选地,其中所述模型训练的方法是二分类算法。

[0016] 可选地,其中所述相似度计算是带有修正系数的余弦相似度计算。

[0017] 根据本发明的技术方案的另一个方面,提供了一种基于用户浏览行为的商品推荐装置。

[0018] 根据本发明的技术方案的另一个方面的基于用户浏览行为的商品推荐的装置,包括:模型计算模块,用于:由购买行为预测器分析一个或多个用户的商品浏览历史数据,确定在一未来时段较有可能进行购买的第一用户集合;针对所确定的第一用户集合中的每个用户:从该用户的商品浏览历史数据中获取该用户浏览过的一个或多个商品,将所述一个或多个商品与目标商品进行商品相似度计算,基于所述商品相似度计算的结果,确定所述用户在所述未来时段较有可能购买目标商品,以及将所述用户添加至第二用户集合;以及向所确定的第二用户集合中的每个用户推荐所述目标商品。

[0019] 可选地,所述模型计算模块进一步用于:针对所述一个或多个用户中的每个用户:基于该用户的商品浏览历史数据,由所述购买行为预测器确定该用户在所述未来时段进行购买行为的可能性;响应于确定所述可能性大于阈值,将该用户添加到所述第一用户集合。

[0020] 可选地,所述模型计算模块进一步用于对所述购买行为预测器进行训练,其中所述训练包括:接收多个特定用户的样本商品浏览历史数据;基于所接收的样本商品浏览历史数据生成与所述多个特定用户相关联的训练数据;利用所述训练数据训练所述购买行为预测器。

[0021] 可选地,其中所述模型计算模块进一步用于:针对所述多个特定用户中的每个特定用户,基于预先定义的用户行为特征,提取该特定用户的样本基础历史数据的特征并添加标签,生成与该特定用户相关联的训练数据;聚集与每个特定用户相关联的训练数据,对所聚集的训练数据进行标准化处理,生成与所述多个特定用户相关联的所述训练数据。

[0022] 可选地,所述相似度计算模块进一步用于:针对所述目标商品的类别,基于预先定义的商品属性特征模型,构建所述目标商品的目标商品向量;对于所述第一用户集合中的每个用户:基于所述预先定义的商品属性特征模型,构建该用户浏览过的商品中与所述目标商品同类别的商品的对比商品向量,其中所述目标商品向量与所述对比商品向量包含相同的向量维度;对目标商品向量和所述对比商品向量进行相似度计算。

[0023] 可选地,其中所述用户的训练数据是以向量来表示的。

[0024] 可选地,其中所述模型训练的方法是二分类算法。

[0025] 可选地,其中所述相似度计算是带有修正系数的余弦相似度计算。

[0026] 为实现上述目的,根据本发明的再一方面,提供了一种电子设备。

[0027] 本发明的一种电子设备包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现本发明所提供的基于用户浏览行为的商品推荐的方法。

[0028] 为实现上述目的,根据本发明的再一方面,提供了一种计算机可读存储介质。

[0029] 本发明的一种计算机可读存储介质,其上存储有计算机程序,所述程序被处理器执行时实现本发明所提供的基于用户浏览行为的商品推荐的方法。

[0030] 上述发明中的一个实施例具有如下优点或有益效果:因为采用模型计算结合相似度计算的技术手段,所以克服了现有技术中无法精确定位目标用户的问题,进而达到精确分析购买目标商品的潜在人群的技术效果。

[0031] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

## 附图说明

[0032] 附图用于更好地理解本发明,不构成对本发明的不当限定。其中:

[0033] 图1是根据本发明实施例的一种基于用户浏览行为的商品推荐方法的总体流程图;

[0034] 图2是根据本发明实施例的特征处理以生成用户的样本的流程图;

[0035] 图3是根据本发明实施例的模型训练以生成模型文件的流程图;

[0036] 图4是根据本发明实施例的相似度计算的流程图;

[0037] 图5是根据本发明实施例的一种基于用户浏览行为的商品推荐系统的系统模块图;

[0038] 图6是用于实现本发明实施例的一种基于用户浏览行为的商品推荐方法的电子设

备的硬件结构示意图。

### 具体实施方式

[0039] 以下结合附图对本发明的示范性实施例做出说明,其中包括本发明实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本发明的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0040] 本发明基于经特征处理的用户的样本,利用模型训练来对用户进行第一次筛选,预测有可能进行购买的用户;然后通过将目标商品与有可能进行购买的用户所浏览过的历史商品进行比较,对用户进行第二次筛选,预测有可能购买目标商品的用户。

[0041] 用户的样本:将用户浏览历史数据进行特征处理并打标签,生成用户的样本。多个用户的样本可用于作为的训练数据,以用于模型训练。在示例实施例中,用户Amy在3月1日浏览过2件商品并将其中一件商品加入购物车,在3月2日浏览过4件商品,在3月4日购买了2件商品。这些动作的集合构成用户Amy的基础历史数据。经由特征处理步骤对Amy的基础历史数据进行处理,最终形成具有特征和标签的样本。在一种实施方式中,标签表示用户是否购买商品。例如0表示没有购买商品,1表示购买了商品。在替选实施方式中,标签可以表示其他信息,诸如0表示购买了不超过预定数目的商品,1表示购买了超过预定数目的商品。

[0042] 特征:我们通过用户浏览的历史记录来挖掘各种用户行为的特征。特征是根据预定的特征定义规则来定义的。在一种实施方式中,特征值以数字来表示。

[0043] 模型:模型是对成千上万的样本进行模型训练而生成。例如,对成千上万个用户的样本进行训练,计算出一个模型。在一种实施方式中,模型可以是函数。在替选实施方式中,模型可以是算法。可以将模型应用于每个用户的样本,来预测该用户在未来时段的行为。

[0044] 模型文件:模型文件保持的是模型的参数。将模型文件中的参数传入模型对用户样本进行数学运算,最后输出诸如0-1之间的概率值。

[0045] 图1是根据本发明实施例的一种基于用户浏览行为的商品推荐方法的总体流程图。如图1所示,图1包括步骤S11-S15。

[0046] 在图1的步骤S11中,对多个用户中的每个用户的浏览历史进行特征处理,生成所述多个用户的样本,作为训练数据。详细步骤参见图2。

[0047] 跳转至图2,图2根据本发明实施例的生成样本的流程图,示出了基于用户的历史数据进行特征处理并生成样本的过程。在图2的步骤S11-1中,收集一个或多个用户浏览网站的基础历史数据。在一种实施方式中,基础历史数据来自诸如分布式文件系统的存储单元中所存储的流量表、点击表、订单表、预定表等业务数据存储库。可以采用诸如mapreduce+hive等技术来对商品特征进行采集。在一种实施方式中,可以按照用户维度来处理数据,用户浏览的粒度可以按商品库存量单位(SKU)编号进行采集。在一个实施例中,采用定时器来设置数据的采集时间,实现自动采集数据。例如,夜间系统负荷小,可以将定时器设置为夜间自动启动数据采集而避免影响主要业务。

[0048] 在图2的步骤S11-2中,对每个用户的基础数据进行特征抽取,并生成用户样本向量。下表1示例性列出了用户行为的示例特征定义规则。特征抽取过程即按照预先定义的特征规则对诸如表1中列出的维度和特征进行筛选。在示例实施例中,例如,用户Amy最近一周



关注了4个商品,但是没有加入购物车的动作,则特征抽取过程将“关注商品个数”特征设置为4,“加入购物车”特征设置为0。用户购买商品的欲望是随着时间逐渐减弱的。经过分析用户的历史购买记录发现,浏览时间越近对用户购买商品的影响最大。在一种实施例,仅出于示例目的,可以将特征提取时间维度定为最近7天,特征抽取的示例维度和相对应的示例特征定义如表1所示。在其他实施例中可以定义不同的特征和提取时间。

[0049] 在图2的步骤S11-3中,特征清洗去掉异常数据并对特征的各个维度进行归一化处理。由于每个特征的量级不一致,因此必须对用户的样本向量的所有维度都进行归一化处理。在示例实施例中,用户Amy浏览商品个数为5、最近一次购买的时间距今的天数为200,为了方便算法进行处理,对“5”和“200”进行归一化处理。在一种实施方式中,可以采用Min-Max Normalization归一化算法将数值归一化为区间表1示例特征定义规则

维度	特征	特征类型
购买维度	关注商品个数（最近 7 天）	整数
	收藏商品个数（最近 7 天）	整数
	商品加入购物车个数（最近 7 天）	整数
	预定商品个数（最近 7 天）	整数
历史订单	最近一次购买商品的时间距今的天数	整数
	历史订单中购买商品的平均价格	浮点数
	历史订单中购买商品的平均间隔周期（单位月）	整数
PC 端	每天浏览商品的比例（最近 1-7 天）	浮点数
	每天浏览商品库存存量（SKU）的个数（最近 1-7 天）	整数
	当天第一次浏览商品时间，小时维度（最近 1-7 天）	整数
	当天浏览商品次数最多的前三个时间，小时维度（最近 1-7 天）	整数
	点击商品商品详情页的记录点次数（最近 1-7 天）	浮点数
APP 端	每天浏览商品的比例（最近 1-7 天）	浮点数
	每天浏览商品 SKU 的个数（最近 1-7 天）	整数
	当天第一次商品浏览时间，小时维度（最近 1-7 天）	整数
	当天浏览商品次数最多的前三个时间段，小时维度（最近 1-7 天）	整数
	点击商品商品详情页的记录点次数（最近 1-7 天）	浮点数
即时聊天工具端	每天浏览商品的比例（最近 1-7 天）	浮点数
	每天浏览商品 SKU 的个数（最近 1-7 天）	整数
	当天第一次商品浏览的时间，小时维度（最近 1-7 天）	整数
	当天浏览商品次数最多的前三个时间段，小时维度（最近 1-7 天）	整数
	点击商品商品详情页的记录点次数（最近 1-7 天）	浮点数
商品属性	统计用户与目标商品同一个品类的商品浏览记录，然后对商品的各个属性进行值，组成一个特征向量	向量

[0050]

[0051]

[0052] [0,1]之间。替选地,也可以采用Z-score等各种其他归一化算法。在示例实施例中,用户Amy、Bob、Cindy、Dora的样本分别具有“关注商品个数”和“收藏商品个数”两个维度的特征:

[0053] 表2示例用户样本

用户名	关注商品个数	收藏商品个数
Amy	2	4
Bob	8	3
Cindy	10	3
Dora	10000	2

[0055] 在该示例实施例中,用户Dora的样本“关注商品个数”值为10000,远远大于其它样本的值,作为异常值去掉Dora样本。

[0056] 在一种实施方式中,以Min-Max Normalization归一化算法为例,用每一维度的特征值除以该维度特征的最大值。用户的样本转化为如下形式:

[0057] 表3经特征清洗的示例用户样本

用户名	关注商品个数	收藏商品个数
Amy	0.2	1
Bob	0.8	0.75
Cindy	1	0.75

[0059] 在图2的步骤S11-4中,输出经特征处理的多个用户的样本,以供下面步骤使用。经特征处理之后,为样本打上标签。在示例实施例中,将样本的标签定义为例如“正负样本”,通过参数设置来定义正负样本的选择标准。例如,正样本可以定义为购买商品的用户,负样本可以定义为有浏览行为,但是没有购买商品的用户。输出的用户样本,即为具有特征+标签的样本。经特征处理并打标签的用户样本是本发明后续操作的基础,在后面的步骤中都是基于经特征处理并打标签的用户样本进行操作。

[0060] 跳转回图1,在图1的步骤S12中,根据预先定义的参数设置利用多个特定用户的样本进行模型训练,生成购买行为预测器,并输出模型文件。预先定义的参数可以设置样本的起止时间、模型的说明、样本划分标准等等。详细步骤参见图3。

[0061] 图3是根据本发明实施例的流程图。在图3的步骤S12-1中,划分训练集和测试集。训练集就是用于模型训练的样本集合,测试集就是用来验证模型准确率的样本集合。在模型训练中通常采用8:2原则。例如,10万个特定用户样本,其中8万个作为训练集,2万个作为测试集。

[0062] 在图3的步骤S12-2中,利用训练集中的特定用户样本进行模型训练,被模型训练的模型称之为购买行为预测器。在一种实施方式中,可以采用有监督学习的二分类算法来训练模型(诸如逻辑回归,决策树等算法)。模型训练的输入数据是用户样本的二维矩阵,其中每一行是一个用户样本的向量,每一列代表一个特征,表示该样本对应于此特征的值。对于每一个样本,都对应有一个经模型训练的目标值。在采用有监督学习的二分类算法来训练模型的实施方式中,用于模型训练的用户样本仅有两种目标值。

[0063] 在对训练测试集进行模型训练之后,利用之前划分的测试集来验证模型的准确性。

[0064] 在图3的步骤S12-3中,输出模型文件。模型文件是购买行为预测器的参数。针对不同场景,可以生成不同的模型文件。将不同的模型文件存储在存储单元中以供将来选择,用于不同的预测场景。

[0065] 跳转回图1,在图1的步骤S13中,用所输出的模型文件对每个用户的样本进行预测,以确定在给定的未来时段有可能进行购买的第一用户集合。对目标用户进行预测时,获取该用户的样本,然后利用所选择的模型文件里面的参数通过购买行为预测器进行分类算法的数学运算,得到购买概率。在一种实施方式中,如果结果概率为0.8,代表80%的概率有可能进行购买,20%的概率不会购买。将每个用户的购买概率与阈值进行比较,选择购买概率大于阈值的用户作为有可能进行购买的第一用户集合。

[0066] 在图1的步骤S14中,对目标商品和较有可能进行购买的第一用户集合中的每个用户曾经浏览过的商品进行比较,确定有可能购买目标商品的第二用户集合。详细步骤参见图4。

[0067] 图4是根据本发明实施例的相似度计算的流程图。在图4的S14-1中,基于预先定义的商品的属性特征模型,针对目标商品的类别,构建目标商品的目标商品向量。商品的属性特征诸如价格、品牌、产地等。

[0068] 在图4的S14-2中,针对每个用户曾经浏览过的商品中与目标商品同类别的商品,基于预先定义的商品的属性特征模型,构建该用户的对比商品向量,其中所述目标商品向量与所述对比商品向量包含相同的向量维度,向量维度即为属性特征。

[0069] 在图4的S14-3中,对目标商品向量和该用户的对比商品向量进行相似度计算。在一种实施方式中,可以采用修正系数结合余弦相似度来计算。

[0070] 在采用修正系数结合余弦相似度的实施例中,相似度计算公式如下:

$$[0071] \quad simi(v_1, v_2) = 1 - \sigma * \frac{\sum_{i=1}^n v_{1i} * v_{2i} * w_i}{\sqrt{\sum_{i=1}^n (v_{1i} * w_i)^2} * \sqrt{\sum_{i=1}^n (v_{2i} * w_i)^2}} \quad \text{公式 1}$$

[0072] 其中 $\sigma$ 为修正系数, $v_1$ 为目标商品属性特征向量, $v_2$ 为用户浏览过的商品属性特征向量, $w_i$ 为每个属性对应的权重向量。其中, $v_1$ 和 $v_2$ 包含的是相同的向量维度。并且用户浏览过的商品与目标商品同类别。向量维度表示商品的属性特征,相同的向量维度表示相同的商品属性。

[0073]  $V_1 = (v_{11}, v_{12}, v_{13}, \dots)$   $V_2 = (v_{21}, v_{22}, v_{23}, \dots)$   $W = (w_1, w_2, w_3, \dots)$

[0074] 在该实施例中,

[0075] 目标商品向量预设为 $V_1 = (1, 1, 1, 1, \dots)$

[0076] 目标商品的属性特征值都设为1。替选地,价格属性的特征值可以用实际价格数值除以归一化系数(如10000)来获得。

[0077] 用户浏览的对比商品向量 $V_2$ 的计算方式如下:

[0078] 用户浏览商品特征向量中的每一个属性特征值是根据下面的公式计算得到的。其中价格属性是特殊的属性,因此以单独的“价格属性”公式来计算,诸如品牌等其他属性都用“其他属性”公式来计算。

[0079] 以电脑为例,属性与该属性的权重可以例如:

[0080] 电脑:价格-1.0,屏幕尺寸-0.3,热点-0.3,品牌-0.6,机身颜色-0.3,系统-0.6,核数-0.3,内存-0.5。

[0081] 1. 价格属性的特征值计算公式:

$$[0082] \quad v = \frac{\sum_{i=1}^n (a_i b_i)}{\sum_{i=1}^n (b_i)}$$

[0083] 公式 2

[0084]  $a_i$  为所浏览的商品价格,  $b_i$  为该商品的浏览次数,  $n$  为浏览的商品个数。

[0085] 例如, 用户浏览了4款商品, 价格分别为5000、4500、6000、3500, 浏览次数分别为1、3、1、1那么平均价格为

$$[0086] \quad (5000*1+4500*3+6000*1+3500*1)/6 = 4667,$$

[0087] 然后除以归一化系数(如10000), 得到价格特征值0.4667。

[0088] 2. 其它属性的特征值计算公式:

$$[0089] \quad v = \frac{\sum_{i=1}^n (f(i, attr))}{n} \quad \text{公式 3}$$

$$[0090] \quad f(n, attr) = \begin{cases} 0 & \text{第 } n \text{ 次浏览的商品具有属性 } attr \\ 1 & \text{第 } n \text{ 次浏览的商品没有属性 } attr \end{cases}$$

[0091] 例如, 用户浏览了4款商品, 品牌分别A、B、C、D, 总的浏览次数为4次, 品牌C出现2次, 则品牌C的特征值为2/4=0.5。

[0092] 然后采用基于权重的余弦相似度分别计算目标商品的各个特征与用户浏览过的同类别商品的各个特征之间的相似度。

[0093] 上面计算余弦相似度的时候, 是从商品属性的角度来考虑, 需要用修正系数增加用户自身的行为。例如, 最近一周用户浏览商品详情页的次数、商品详情页的点击记录点次数等等, 且时间最近的行为其权重较大。本文对余弦相似度进行改进, 将余弦相似度的结果乘以修正系数。

[0094] 修正系数计算公式 = (浏览系数+记录点系数)/2

[0095] 1. 浏览系数 $a_1$ 的计算方法:

$$[0096] \quad a_1 = \left( \sum_{i=1}^m n_i * \frac{8-i}{7} \right) / \sum_{i=1}^m n_i \quad (m=7) \quad \text{公式 4}$$

[0097]  $n_i$  为每天浏览的次数

[0098] 2. 记录点系数 $a_2$ :

$$[0099] \quad a_2 = \left( \sum_{i=1}^x \left( \sum_{j=1}^y n_{ij} * w_j \right) * \frac{8-i}{7} \right) / \sum_{i=1}^x n_i \quad (x=7, y=8) \quad \text{公式 5}$$

[0100]  $n_{ij}$  为记录点的点击次数,  $w_j$  为该记录点权重(取权重排名前8个记录点)。

[0101] 在图4的S14-4中将所计算的相似度与阈值进行比较, 确定与大于阈值的相似度相对应的用户, 生成有可能购买目标商品的潜在用户集合。在示例实施例中, 可以设定诸如0.8的阈值, 提取最终相似度计算结果大于0.8的用户作为有可能购买目标商品的用户。

[0102] 返回图1, 在图1的步骤S15中, 向最终确定的较有可能购买目标商品的第二用户集合中的用户推荐目标商品。

[0103] 图5是根据本发明实施例的一种基于用户浏览行为的商品推荐系统的系统模块图。本系统分为模型计算模块和相似度计算模块两部分。其中,模型计算模块包括特征处理子模块、模型训练子模块、以及模型预测子模块。模型计算模块针对每个用户的历史数据进行特征处理得到用户的样本、对多个用户的样本进行模型训练生成模型文件、用目标用户的样本和所选用的模型文件来预测该用户未来时段的购买行为,确定有可能进行购买行为的第一用户集合;相似度计算模块针对第一用户集合中的每个用户,计算目标商品和该用户曾经浏览过的历史商品之间的相似度,从而确定有可能购买目标商品的第二用户集合。

[0104] 根据本发明的实施例,本发明还提供了一种电子设备和一种可读存储介质。

[0105] 本发明的电子设备包括:至少一个处理器;以及,与所述至少一个处理器通信连接的存储器;其中,所述存储器存储有可被所述一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器执行本发明所提供的基于用户浏览行为的商品推荐的方法。

[0106] 本发明的非暂态计算机可读存储介质,所述非暂态计算机可读存储介质存储计算机指令,所述计算机指令用于使所述计算机执行本发明所提供的基于用户浏览行为的商品推荐的方法。

[0107] 如图6所示,是用于实现本发明实施例的基于用户浏览行为的商品推荐的方法的电子设备的硬件结构示意图。如图6,该电子设备包括:一个或多个处理器61以及存储器62,图6中以一个处理器61为例。其中,存储器62即为本发明所提供的非暂态计算机可读存储介质。

[0108] 基于用户浏览行为的商品推荐的方法的电子设备还可以包括:输入装置63和输出装置64。

[0109] 处理器61、存储器62、输入装置63和输出装置64可以通过总线或者其他方式连接,图6中以通过总线连接为例。

[0110] 存储器62作为一种非暂态计算机可读存储介质,可用于存储非暂态软件程序、非暂态计算机可执行程序以及模块,如本发明实施例中的基于用户浏览行为的商品推荐的方法对应的程序指令/模块(例如,附图5所示的模型计算模块51和相似度计算模块52)。处理器61通过运行存储在存储器62中的非暂态软件程序、指令以及模块,从而执行服务器的各种功能应用以及数据处理,即实现上述方法实施例中的基于用户浏览行为的商品推荐的方法。

[0111] 存储器62可以包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需要的应用程序;存储数据区可存储根据基于用户浏览行为的商品推荐的装置的使用所创建的数据等。此外,存储器62可以包括高速随机存取存储器,还可以包括非暂态存储器,例如至少一个磁盘存储器件、闪存器件、或其他非暂态固态存储器件。在一些实施例中,存储器62可选包括相对于处理器61远程设置的存储器,这些远程存储器可以通过网络连接至基于用户浏览行为的商品推荐的装置。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0112] 输入装置63可接收输入的数字或字符信息,以及产生与基于用户浏览行为的商品推荐的装置的用户设置以及功能控制有关的键信号输入。输出装置64可包括显示屏等显示设备。

[0113] 所述一个或者多个模块存储在所述存储器62中,当被所述一个或者多个处理器61执行时,执行上述任意方法实施例中的基于用户浏览行为的商品推荐的方法。

[0114] 上述产品可执行本发明实施例所提供的方法,具备执行方法相应的功能模块和有益效果。未在本实施例中详尽描述的技术细节,可参见本发明实施例所提供的方法。

[0115] 根据本发明实施例的技术方案,精确地挖掘购买目标商品的潜在用户,对所筛选出的潜在用户推荐目标商品,最终达到提升用户订单转化率的目的。

[0116] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,取决于设计要求和因素,可以发生各种各样的修改、组合、子组合和替代。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,均应包含在本发明保护范围之内。

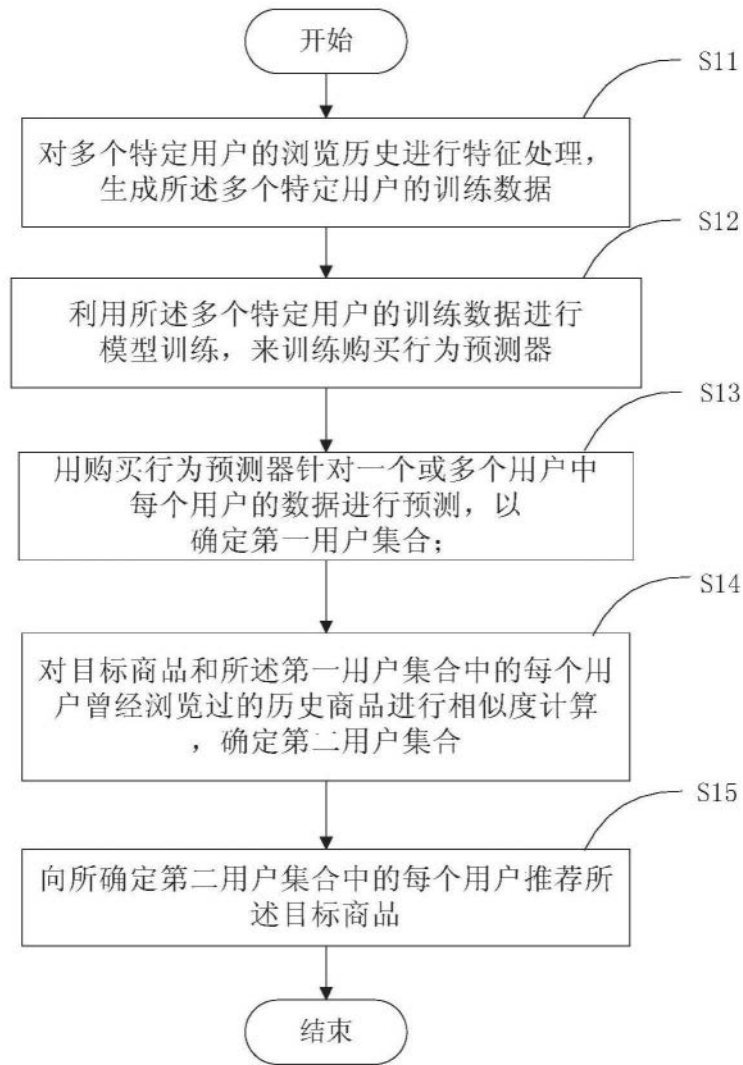


图1



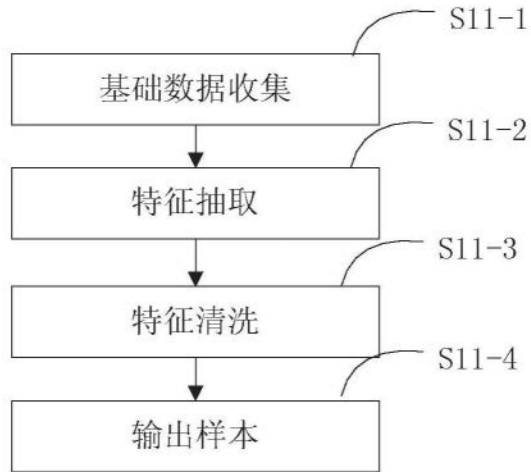


图2

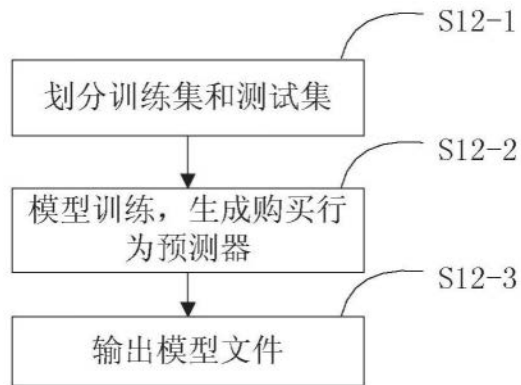


图3

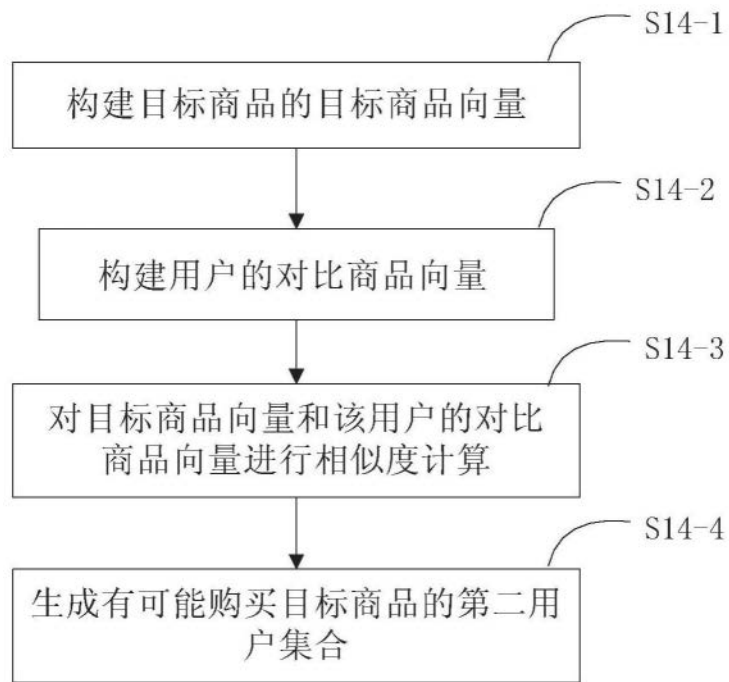


图4

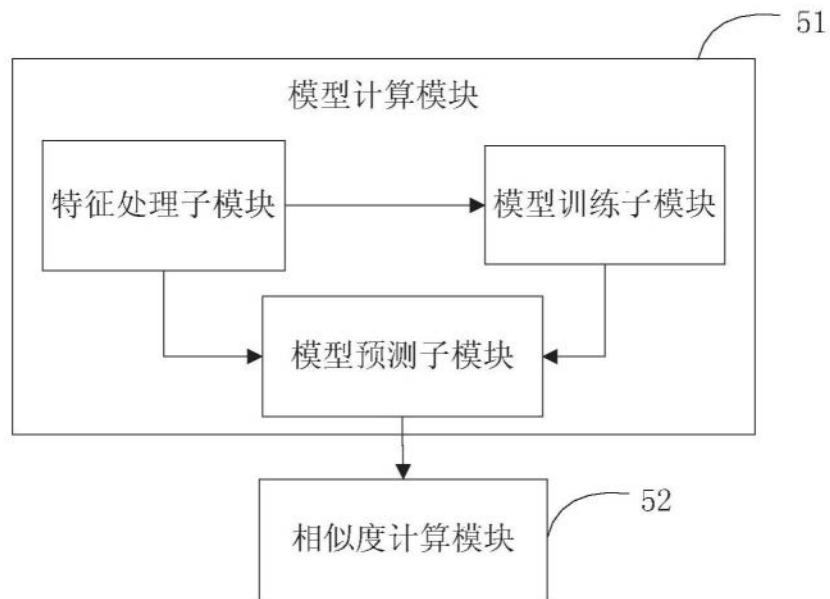


图5

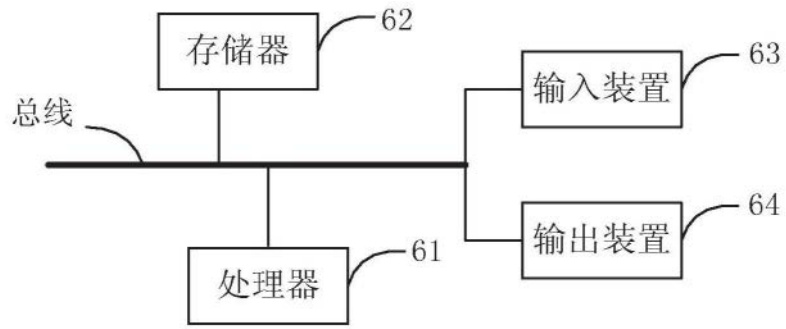


图6