

(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
2020 年 10 月 1 日 (01.10.2020)



(10) 国际公布号
WO 2020/192013 A1

(51) 国际专利分类号:
G06Q 30/02 (2012.01)

(21) 国际申请号: PCT/CN2019/103608

(22) 国际申请日: 2019 年 8 月 30 日 (30.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910238703.0 2019 年 3 月 27 日 (27.03.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(72) 发明人: 刘继宇(LIU, Jiyu); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518033 (CN)。

(74) 代理人: 北京鸿元知识产权代理有限公司 (GRANDER IP LAW FIRM); 中国北京市朝阳区光华路 7 号汉威大厦东区 18A6 室, Beijing 100004 (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,

(54) Title: DIRECTIONAL ADVERTISEMENT DELIVERY METHOD AND APPARATUS, AND DEVICE AND STORAGE MEDIUM

(54) 发明名称: 定向广告投放方法和装置、设备及存储介质

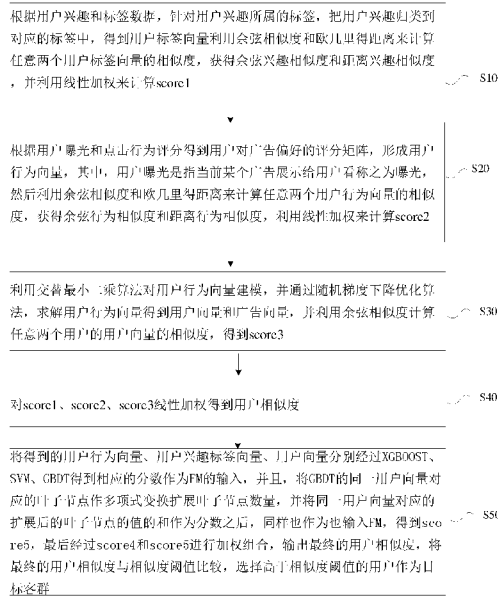


图 1

S10 According to user interests and label data, with regard to labels to which the user interests belong, classify the user interests into corresponding labels to obtain user label vectors, calculate the similarity between any two user label vectors by means of a cosine similarity and Euclidean distance to obtain a cosine interest similarity and a distance interest similarity, and calculate a score 1 by means of linear weighting

S20 Obtain a user scoring matrix for advertisement preferences according to user exposure and click behavior scoring to form user behavior vectors, wherein user exposure refers to the current display of a certain piece of advertisement to the user being referred to as exposure, then calculate the similarity between any two user behavior vectors by means of the cosine similarity and Euclidean distance to obtain the cosine behavior similarity and the distance behavior similarity, and calculate a score 2 by means of linear weighting

S30 Model the user behavior vectors by means of an alternating least squares algorithm, solve the user behavior vectors by means of a stochastic gradient descent optimization algorithm to obtain user vectors and advertisement vectors, and calculate the similarity between user vectors of any two users by means of the cosine similarity to obtain a score 3

S40 Carry out linear weighting on the score 1, the score 2 and the score 3 to obtain a user similarity

S50 Respectively subject the user behavior vectors, user interest label vectors and the user vectors to XGBOOST, SVM and GBDT to obtain corresponding scores as an input of an FM, carry out polynomial transformation on leaf nodes corresponding to the same user vector of the GBDT to expand the number of leaf nodes, take the sum of values of the expanded leaf nodes corresponding to the same user vector as the score, also take same as an input of the FM to obtain a score 5, then carry out weighted combination on the score 4 and the score 5 to output a final user similarity, compare the final user similarity with a similarity threshold, and select users with the similarity therebetween being higher than the similarity threshold as a target customer group

(57) Abstract: Provided are a directional advertisement delivery method and apparatus, and a device and a storage medium. The method comprises: calculating the similarity between label vectors of two users to obtain a score 1; calculating the similarity between two user behavior vectors to obtain a score 2; modeling user vectors using an alternating least squares algorithm, obtaining user vectors and advertisement vectors by means of a stochastic gradient descent optimization algorithm, and calculating the similarity between the user vectors of the two users to obtain a score 3; carrying out linear weighting on the score 1 to the score 3 to obtain a score 4; and

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,
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))。

respectively subjecting the user behavior vectors, the user label vectors and the user vectors to XGBOOST, SVM and GBDT to obtain corresponding scores as an input of an FM, so as to obtain a score 5, then carrying out weighted combination on the score 4 and the score 5 to output a final user similarity, comparing the final user similarity with a similarity threshold, and selecting users with the similarity therebetween being higher than the similarity threshold as a target customer group.

(57) 摘要: 一种定向广告投放方法和装置、设备及存储介质, 方法包括: 计算两个用户的标签向量的相似度, 获得分值1; 计算两个用户行为向量的相似度, 得到分值2; 利用交替最小二乘算法对用户向量建模, 并通过随机梯度下降优化算法, 得到用户向量和广告向量, 并计算两个用户的用户向量的相似度, 得到分值3, 对分值1至分值3线性加权得到分值4。将用户行为向量、用户标签向量、用户向量分别经过XGBOOST、SVM、GBDT得到相应的分数作为FM的输入, 得到分值5, 再对分值4和分值5进行加权组合, 输出最终的用户相似度, 将最终的用户相似度与相似度阈值比较, 选择高于相似度阈值的用户作为目标客群。

定向广告投放方法和装置、设备及存储介质

本申请要求于 2019 年 3 月 27 日提交的中国专利申请号 201910238703.0 的优先权益，上述案件全部内容以引用的方式并入本文中。

技术领域

本申请涉及人工智能，具体地说，涉及一种定向广告投放方法和装置、设备及存储介质。

背景技术

随着网络技术的发展，社交网络成为了重要的交流平台之一。消息推送得到了越来越多的应用。推送技术是指根据一定的发送规则，依据相关的标准和协议通过互联网向用户发送信息的技术。电子商务厂商纷纷向用户的智能移动终端上推送各类信息。然而目前定向广告基本上是根据用户的兴趣或者标签来进行人群定向；而目标客群是属于群体中绝大部分人都具备的行为，这会使得目标客群的覆盖率非常低。由于金融产品用户相对互联网用户，用户行为数据丰富度较低，从而导致挖掘用户的兴趣和标签的方法存在一定的挑战性。真实场景下，目标客群往往没有非常清晰的界限和区分性，很难利用用户的单个兴趣和标签圈定目标客群。

发明内容

为解决以上技术问题，本申请提供一种定向广告投放方法，应用于电子设备，包括：根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score（分值）1；

根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户向量的

相似度，得到 score3；

对 score1、score2、score3 线性加权得到用户相似度score4，

$$score4 = W1 * score1 + W2 * score2 + W3 * score3$$

其中，W1、W2、W3是权重系数；

将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，从而提升模型的精度，得到 score5，计算公式如下：

$$score5 = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$score6 = M1 * score4 + M2 * score5$$

其中，M1、M2是系数；

将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

本申请还提供一种定向广告投放装置，其特征在于，包括：

用户兴趣相似度获取模块，用于根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

用户行为向量相似度获取模块，用于根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

用户向量的相似度获取模块，用于利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似

度计算任意两个用户的用户向量的相似度，得到 score3；

用户相似度获取模块，用于对 score1、score2、score3 线性加权得到用户相似度 score4，

$$\text{score4} = W1 * \text{score1} + W2 * \text{score2} + W3 * \text{score3}$$

其中，W1、W2、W3 是权重系数；

用户相似度获取模块，将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，得到 score5，计算公式如下：

$$\text{score5} = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$\text{score6} = M1 * \text{score4} + M2 * \text{score5}$$

其中，M1、M2 是系数；

相似度阈值比较模块，用于将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

本申请还提供一种电子设备，该电子设备包括：存储器和处理器，所述存储器中存储有定向广告投放程序，所述定向广告投放程序被所述处理器执行时实现如下步骤：

(1) 根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

(2) 根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

(3) 利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户

向量的相似度，得到 score3；

(4) 对 score1、score2、score3 线性加权得到用户相似度score4，

$$score4 = W1 * score1 + W2 * score2 + W3 * score3$$

其中，W1、W2、W3是权重系数；

(5) 将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，得到 score5，计算公式如下：

$$score5 = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$score6 = M1 * score4 + M2 * score5$$

其中，M1、M2是系数；

将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

本申请还提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令被处理器执行时，实现以上所述的定向广告投放方法。

本申请的定向广告投放方法和装置、设备及存储介质提升了目标客群的覆盖率，能带来更多有价值的目标客群。本申请不仅利用用户兴趣标签数据和用户行为数据，还能够挖掘用户和广告潜在的关联关系，从而提升了金融用户数据的丰富度，提高算法的精确度。采用多模型融合更加具有通用性。

附图说明

通过结合下面附图对其实施例进行描述，本申请的上述特征和技术优点将会变得更加清楚和容易理解。

图 1 是本申请实施例的定向广告投放方法的流程示意图；

图 2 是本申请实施例的电子设备的硬件架构示意图；

图 3 是本申请实施例的定向广告投放程序的模块构成图。

具体实施方式

下面将参考附图来描述本申请所述的定向广告投放方法和装置、设备及存储介质的实施例。本领域的普通技术人员可以认识到，在不偏离本申请的精神和范围的情况下，可以用各种不同的方式或其组合对所描述的实施例进行修正。因此，附图和描述在本质上是说明性的，而不是用于限制权利要求的保护范围。此外，在本说明书中，附图未按比例画出，并且相同的附图标记表示相同的部分。

图 1 为本申请实施例提供的定向广告投放方法的流程示意图。该方法包括以下步骤：

步骤 S10，根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量。其中用户兴趣可以是基于用户的基本信息获得的用户兴趣，例如根据用户注册时填写的职业、爱好、居住地等信息推算出的用户兴趣，当然还可以包括是针对用户历史浏览、点击等行为获得的用户兴趣。而针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量。例如，定投、保底、货币型基金对应的标签是稳健型。股票、外汇、期货对应着高风险型。那么针对每个用户，都可以根据用户兴趣和对应的标签数据，得到用户标签向量。

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1。

步骤 S20，根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光。所述偏好就是指用户喜不喜欢该广告（即点击或不点击），并根据点击情况给出分值，通过这样评分得到用户对所有广告的评分矩阵。然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

步骤 S30，利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，从而挖掘用户、广告及其之间的潜在关联关系。其中，用户向量表示与广告特征相关的用户信息。例如，广告特征是电影，用户信息可能是喜欢言情、经常看电影等。大大提升了刻画用户和广告数据维度的丰富度。并利用余弦相似度计算任意两个用户的用户向量的相似度，得到 score3，其公式如下：

$$score3 = \frac{\sum_{i=1}^n userA_i userB_i}{\sqrt{\sum_{i=1}^n userA_i^2} \sqrt{\sum_{i=1}^n userB_i^2}}$$

其中, $userA_i$ 表示得到用户 A 的用户向量的第 i 个分量; $userB_i$ 表示算法求解得到用户 B 的用户向量的第 i 个分量; n 是用户向量的维度。

步骤 S40, 对 score1、score2、score3 线性加权得到用户相似度score4,

$$score4 = W1 * score1 + W2 * score2 + W3 * score3$$

其中, W1、W2、W3是权重系数;

步骤 S50, 将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST (极端梯度上升)、SVM (支持向量机)、GBDT (梯度提升树) 得到相应的分数作为 FM (因式分解机) 的输入。其中, 将 GBDT 的叶子节点, 并作多项式变换之后, 同样也作为 FM 的输入, FM 输出 score5, 从而提升模型的精度, 计算公式如下:

$$score5 = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中, T 是激活函数;

X 是特征;

$X_i * X_j$ 是二阶的特征交互;

$X_i * X_j * X_k$ 是三阶的特征交互;

$X_i * X_j * X_k * X_l$ 是四阶的特征交互;

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数;

最后经过 score4 和 score5 进行加权组合, 输出最终的用户相似度,

$$score6 = M1 * score4 + M2 * score5$$

其中, M1、M2是系数;

将最终的用户相似度与相似度阈值比较, 选择高于相似度阈值的用户作为目标客群。

进一步地, 欧几里得距离计算公式如下:

$$\begin{aligned} d(userA, userB) &= \sqrt{(userA_1 - userB_1)^2 + (userA_2 - userB_2)^2 + \dots + (userA_n - userB_n)^2} \\ &= \sqrt{\sum_{i=1}^n (userA_i - userB_i)^2} \end{aligned}$$

其中, $userA$ 表示用户 A 的用户向量;

$userB$ 表示用户 B 的用户向量;

n 表示用户向量的维度;

i 表示第 i 个维度;

$d(userA, userB)$ 表示用户 A 与用户 B 之间的距离。

进一步地, 各用户行为向量经过XGBOOST得到相应的分数的过程是:

根据预设的用户行为特征将各用户行为向量划分到多颗回归树的叶子节点, 组成XGBOOST模型 $\hat{y}$, 其中预设的用户行为特征可以是例如用户浏览推送信息的特征, 用户检索信息的特征, 用户收藏信息的特征等。根据这些特征分别构建回归树, 则多颗回归树组成XGBOOST模型 $\hat{y}$ 。

$\hat{y} = \sum_{k=1}^K f_k(x_b)$, $f_k \in F$, F 对应所有回归树组成的函数空间, x_b 表示一棵树中与第 b 个用户行为向量对应的所有节点的分值,

构建目标函数Obj, 利用梯度下降法使得目标函数最小化, 获得XGBOOST模型,

$$Obj = \sum_{m=1}^P l(y_m, \hat{y}_m) + \sum_{k=1}^K \Omega(f_k)$$

其中, $l()$ 表示训练误差, y_m 为第 m 个样本的实际分数, $\hat{y}_m$ 为第 m 个样本输入XGBOOST模型 $\hat{y}$ 的预测分数, K 表示共 K 棵树, k 表示第 k 棵树, p 表示训练样本数量, m 表示第 m 个样本, $\sum_{k=1}^K \Omega(f_k)$ 表示正则化项, 利用XGBOOST模型获得各用户行为向量所在的各分类树的叶子节点的值和分别作为该用户行为向量对应的分数。

进一步地, 各用户标签向量经过GBDT得到相应的分数的过程是:

构建分类树以及叶子节点, 其中, 把用户标签向量相近的作为同一叶子节点, 并以叶子节点的平均值作为预测值, 计算各叶子节点的实际分数与预测分数的残差, 按照各残差的相近程度划分下一棵分类树的叶子节点, 持续进行分类树的划分, 直至残差小于预设的阈值, 将各用户标签向量所在的各分类树的叶子节点的平均值的和分别作为该用户标签向量对应的分数。

进一步地, 使用preprocessing库的PolynomialFeatures类对叶子节点进行多项式变换, 并将各用户向量所在的各分类树的叶子节点的平均值的和分别作为该用户向量对应的分数。例如, 如果一个用户向量对应叶子节点Q1和叶子节点Q2, 这两个叶子节点对应的平均值分别为0.4和1.2, 则进行多项式变换后, 得到Q12、Q22、Q1Q2、Q1、Q2五个叶子节点, 其对应的值分别为0.16, 1.44, 0.48, 0.4和1.2, 则将这五个叶子节点的值相加, 得到3.68, 则以该3.68作为这个用户向量的分数。

在一个可选实施例中, 还根据用户点击广告后浏览时间来确定用户对广告偏好的评分矩阵, 形成用户行为向量, 其中, 先设定分值区间, 所述分值区间包括0至分值上限, 用

户没有点击为0分，还结合用户点开页面以及用户浏览页面时间的长短区分出直至分值上限的分值。

在一个可选实施例中，还设置有修正系数，所述修正系数根据用户点击广告后是否转发，是否删除来确定，其中，用户转发广告则加1分，用户删除广告则减1分。

在一个可选实施例中，还检测用户的位置坐标，根据广告委托者所在的位置坐标与用户的位置坐标的距离远近来推送广告，以距离用户的广告委托者的距离由近及远的方式向用户推送广告。这里的广告委托者是指所有能够提供产品展示的供应商。例如，商场，虽然商场不生产沙发，但是商场却会做广告推销沙发，所以商场就是以上所述的广告委托者，当然沙发的生产商如果也提供产品展示，那么其也可以添加到广告委托者中。当然，具体说，广告可能涉及一些虚拟产品，例如理财基金、股票推荐等，也可能是一些实体产品，例如沙发、电视等。以商场为例，商场会通过推送平台向用户推送广告，然而，有些用户可能距离较远，并不会特意去这个商场。而用户却有可能经过这个商场并不知道商场的推荐广告。而通过定位捕捉到用户的位置坐标，并与电子地图中的各个商场、工厂等广告所关联的广告委托者的坐标进行比较，并依据距离用户的位置坐标由近及远的距离来推送广告，这使得用户能够更快的阅读到距离其最近的生产商的广告。由此，可以使得用户无论走到哪里，都有对应的最近的产品的广告推送。甚至在商场中也可以及时推送，例如，用户经常可能在商场中查找某些品牌的位置，而通过以上的推送方式，则可以在用户行走的过程中，不断的推送给距离其最近的产品。

更进一步地，还可以设置距离阈值，当距离超过距离阈值，则不再推送超出该距离阈值的产品广告信息。例如，距离阈值为20米，则在20米内的广告推广按照距离远近逐个推送，20米外的则不推送。随着人员的行走，距离远近在变动，则推送的广告信息也逐渐变化。

进一步地，采用GPS的方式来定位用户的位置坐标，并与标记有广告委托者的位置坐标的电子地图进行比较，来确定广告委托者所在的位置坐标与用户的位置坐标的距离。

参阅图2所示，是本申请电子设备的实施例的硬件架构示意图。本实施例中，所述电子设备2是一种能够按照事先设定或者存储的指令，自动进行数值计算和/或信息处理的设备。例如，可以是智能手机、平板电脑、笔记本电脑、台式计算机、机架式服务器、刀片式服务器、塔式服务器或机柜式服务器(包括独立的服务器，或者多个服务器所组成的服务器集群)等。如图2所示，所述电子设备2至少包括，但不限于，可通过系统总线相互通信连接的存储器21、处理器22、网络接口23。其中：所述存储器21至少包括一种类型的计算机可读存储介质，所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如，SD或DX存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器

(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中,所述存储器 21 可以是所述电子设备 2 的内部存储单元,例如该电子设备 2 的硬盘或内存。在另一些实施例中,所述存储器 21 也可以是电子设备 2 的外部存储设备,例如该电子设备 2 上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。当然,所述存储器 21 还可以既包括所述电子设备 2 的内部存储单元也包括其外部存储设备。本实施例中,所述存储器 21 通常用于存储安装于所述电子设备 2 的操作系统和各类应用软件,例如所述定向广告投放程序代码等。此外,所述存储器 21 还可以用于暂时地存储已经输出或者将要输出的各类数据。

所述处理器 22 在一些实施例中可以是中央处理器(Central Processing Unit, CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器 22 通常用于控制所述电子设备 2 的总体操作,例如执行与所述电子设备 2 进行数据交互或者通信相关的控制和处理等。本实施例中,所述处理器 22 用于运行所述存储器 21 中存储的程序代码或者处理数据,例如运行所述的定向广告投放程序等。

所述网络接口 23 可包括无线网络接口或有线网络接口,该网络接口 23 通常用于在所述电子设备 2 与其他电子设备之间建立通信连接。例如,所述网络接口 23 用于通过网络将所述电子设备 2 与推送平台相连,在所述电子设备 2 与推送平台之间建立数据传输通道和通信连接等。所述网络可以是企业内部网(Intranet)、互联网(Internet)、全球移动通讯系统(Global System of Mobile communication, GSM)、宽带码分多址(Wideband CodeDivision Multiple Access, WCDMA)、4G 网络、5G 网络、蓝牙(Bluetooth)、Wi-Fi 等无线或有线网络。

可选地,该电子设备 2 还可以包括显示器,显示器也可以称为显示屏或显示单元。在一些实施例中可以是 LED 显示器、液晶显示器、触控式液晶显示器以及有机发光二极管(Organic Light-Emitting Diode, OLED)显示器等。显示器用于显示在电子设备 2 中处理的信息以及用于显示可视化的用户界面。

需要指出的是,图 2 仅示出了具有组件 21-23 的电子设备 2,但是应理解的是,并不要求实施所有示出的组件,可以替代的实施更多或者更少的组件。

包含可读存储介质的存储器 21 中可以包括操作系统、定向广告投放程序 50 等。处理器 22 执行存储器 21 中定向广告投放程序 50 时实现如下步骤:

步骤 S10,根据用户兴趣和标签数据,针对用户兴趣所属的标签,把用户兴趣归类到对应的标签中,得到用户标签向量,

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度,获得余弦兴

趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

步骤 S20，根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

步骤 S30，利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户向量的相似度，得到 score3，其公式如下：

$$score3 = \frac{\sum_{i=1}^n userA_i userB_i}{\sqrt{\sum_{i=1}^n userA_i^2} \sqrt{\sum_{i=1}^n userB_i^2}}$$

其中， $userA_i$ 表示得到用户 A 的用户向量的第 i 个分量； $userB_i$ 表示算法求解得到用户 B 的用户向量的第 i 个分量；n 是用户向量的维度。

步骤 S40，对 score1、score2、score3 线性加权得到用户相似度 score4，

$$score4 = W1 * score1 + W2 * score2 + W3 * score3$$

其中，W1、W2、W3是权重系数；

步骤 S50，将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，从而提升模型的精度，得到 score5，计算公式如下：

$$score5 = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$score6 = M1 * score4 + M2 * score5$$

其中，M1、M2是系数；

将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

在本实施例中，存储于存储器 21 中的所述定向广告投放程序可以被分割为一个或者多个程序模块，所述一个或者多个程序模块被存储于存储器 21 中，并可由一个或多个处理器(本实施例为处理器 22)所执行，以完成本申请。例如，图 3 示出了所述定向广告投放程序的程序模块示意图，该实施例中，所述定向广告投放程序 50 可以被分割为用户兴趣相似度获取模块 501、用户行为向量相似度获取模块 502、用户向量的相似度获取模块 503、用户相似度获取模块 504、相似度阈值比较模块 505。其中，本申请所称的程序模块是指能够完成特定功能的一系列计算机程序指令段，比程序更适合于描述所述定向广告投放程序在所述电子设备 2 中的执行过程。以下描述将具体介绍所述程序模块的具体功能。

其中，用户兴趣相似度获取模块 501 用于根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

用户行为向量相似度获取模块 502 用于根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，

然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

用户向量的相似度获取模块 503 用于利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户向量的相似度，得到 score3，其公式如下：

$$score3 = \frac{\sum_{i=1}^n userA_i userB_i}{\sqrt{\sum_{i=1}^n userA_i^2} \sqrt{\sum_{i=1}^n userB_i^2}}$$

其中， $userA_i$ 表示得到用户 A 的用户向量的第 i 个分量； $userB_i$ 表示算法求解得到用户 B 的用户向量的第 i 个分量；n 是用户向量的维度。

用户相似度获取模块 504 用于对 score1、score2、score3 线性加权得到用户相似度 score4，

$$score4 = W1 * score1 + W2 * score2 + W3 * score3$$

其中，W1、W2、W3是权重系数；

用户相似度获取模块 504 还将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用

户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，得到 score5，计算公式如下：

$$\text{score5} = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$\text{score6} = M1 * \text{score4} + M2 * \text{score5}$$

其中，M1、M2是系数；

相似度阈值比较模块 505 用于将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

另外，本申请还提供一种定向广告投放装置，包括：

用户兴趣相似度获取模块 501，用于根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

用户行为向量相似度获取模块 502，用于根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

用户向量的相似度获取模块 503，用于利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户向量的相似度，得到 score3；

用户相似度获取模块 504，用于对 score1、score2、score3 线性加权得到用户相似度 score4，

$$\text{score4} = W1 * \text{score1} + W2 * \text{score2} + W3 * \text{score3}$$

其中，W1、W2、W3是权重系数；

用户相似度获取模块 505，将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，从而提升模型的精度，得到 score5，计算公式如下：

$$\text{score5} = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$\text{score6} = M1 * \text{score4} + M2 * \text{score5}$$

其中，M1、M2 是系数；

相似度阈值比较模块，用于将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

此外，本申请实施例还提出一种计算机可读存储介质，所述计算机可读存储介质可以是硬盘、多媒体卡、SD 卡、闪存卡、SMC、只读存储器(ROM)、可擦除可编程只读存储器(EPROM)、便携式紧致盘只读存储器(CD-ROM)、USB 存储器等等中的任意一种或者几种的任意组合。所述计算机可读存储介质中包括定向广告投放程序等，所述定向广告投放程序 50 被处理器 22 执行时实现如下操作：

步骤 S10，根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

步骤 S20，根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

步骤 S30，利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算

法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户向量的相似度，得到 score3，其公式如下：

$$\text{score3} = \frac{\sum_{i=1}^n \text{userA}_i \text{userB}_i}{\sqrt{\sum_{i=1}^n \text{userA}_i^2} \sqrt{\sum_{i=1}^n \text{userB}_i^2}}$$

其中， userA_i 表示得到用户 A 的用户向量的第 i 个分量； userB_i 表示算法求解得到用户 B 的用户向量的第 i 个分量；n 是用户向量的维度。

步骤 S40，对 score1、score2、score3 线性加权得到用户相似度 score4，

$$\text{score4} = W1 * \text{score1} + W2 * \text{score2} + W3 * \text{score3}$$

其中，W1、W2、W3是权重系数；

步骤 S50，将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，从而提升模型的精度，得到 score5，计算公式如下：

$$\text{score5} = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$\text{score6} = M1 * \text{score4} + M2 * \text{score5}$$

其中，M1、M2是系数；

将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

本申请之计算机可读存储介质的具体实施方式与上述定向广告投放方法以及电子设备 2 的具体实施方式大致相同，在此不再赘述。

以上所述仅为本申请的优选实施例，并不用于限制本申请，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权 利 要 求 书

1. 一种定向广告投放方法，应用于电子设备，其特征在于，包括：

根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

根据用户曝光和点击行为评分得到用户对广告偏好的评分，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户向量的相似度，得到 score3；

对 score1、score2、score3 线性加权得到用户相似度score4，

$$\text{score4} = W1 * \text{score1} + W2 * \text{score2} + W3 * \text{score3}$$

其中，W1、W2、W3是权重系数；

将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，得到 score5，计算公式如下：

$$\text{score5} = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$\text{score6} = M1 * \text{score4} + M2 * \text{score5}$$

其中，M1、M2是系数；

将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

2. 根据权利要求1所述的定向广告投放方法，其特征在于：所述score3通过公式

$$\text{score3} = \frac{\sum_{i=1}^n \text{userA}_i \text{userB}_i}{\sqrt{\sum_{i=1}^n \text{userA}_i^2} \sqrt{\sum_{i=1}^n \text{userB}_i^2}}$$

获得，其中， userA_i 表示得到用户A的用户向量的第i个分量； userB_i 表示算法求解得到用户B的用户向量的第i个分量；n是用户向量的维度。

3. 根据权利要求1所述的定向广告投放方法，其特征在于：

欧几里得距离计算公式如下：

$$\begin{aligned} d(\text{userA}, \text{userB}) &= \sqrt{(\text{userA}_1 - \text{userB}_1)^2 + (\text{userA}_2 - \text{userB}_2)^2 + \dots + (\text{userA}_n - \text{userB}_n)^2} \\ &= \sqrt{\sum_{i=1}^n (\text{userA}_i - \text{userB}_i)^2} \end{aligned}$$

其中， userA 表示用户A的用户向量；

userB 表示用户B的用户向量；

n表示用户向量的维度；

i表示用户向量的第i个维度；

$d(\text{userA}, \text{userB})$ 表示用户A与用户B之间的距离。

4. 根据权利要求1所述的定向广告投放方法，其特征在于：

还根据用户点击广告后浏览时间来确定用户对广告偏好的评分矩阵，形成用户行为向量，其中，先设定分值区间，所述分值区间包括0至分值上限，用户没有点击为0分，还结合用户点开页面以及用户浏览页面时间的长短区分出直至分值上限的分值。

5. 根据权利要求4所述的定向广告投放方法，其特征在于：还设置有评分矩阵修正系数，所述修正系数根据用户点击广告后是否转发，是否删除来确定，其中，用户转发广告则加1分，用户删除广告则减1分。

6. 根据权利要求4所述的定向广告投放方法，其特征在于：

还检测用户的位置坐标，根据广告委托者所在的位置坐标与用户的位置坐标的距离远近来推送广告，以距离用户的所述广告委托者的距离由近及远的方式向用户推送广告。

7. 根据权利要求6所述的定向广告投放方法，其特征在于：

采用GPS的方式来定位用户的位置坐标，并与标记有广告委托者的位置坐标的电子地

图进行比较, 来确定广告委托者所在的位置坐标与用户的位置坐标的距离。

8. 根据权利要求6所述的定向广告投放方法, 其特征在于:

还设置有距离阈值, 当距离超过距离阈值, 则不再推送超出该距离阈值的产品广告信息。

9. 根据权利要求1所述的定向广告投放方法, 其特征在于:

各用户行为向量经过 XGBOOST 得到相应的分数的过程是:

将各用户行为向量分别划分到多颗回归树的叶子节点, 构成 XGBOOST 模型 $\hat{y}$,

$\hat{y} = \sum_{k=1}^K f_k(x_b)$, $f_k \in F$, F 对应所有回归树组成的函数空间, x_b 表示一棵树中与第**b**个用户行为向量对应的所有节点的分值,

构建目标函数Obj, 利用梯度下降法使得目标函数最小化, 获得 XGBOOST 模型,

$$Obj = \sum_{m=1}^P l(y_m, \hat{y}_m) + \sum_{k=1}^K \Omega(f_k)$$

其中, $l()$ 表示训练误差, y_m 为第 m 个样本的实际分数, $\hat{y}_m$ 为第 m 个样本输入 XGBOOST 模型 $\hat{y}$ 预测分数, K 表示共 K 棵树, k 表示第 k 棵树, p 表示训练样本数量, m 表示第 m 个样本, $\sum_{k=1}^K \Omega(f_k)$ 表示正则化项, 利用 XGBOOST 模型获得各用户行为向量所在的各分类树的叶子节点的值和分别作为该用户行为向量对应的分数。

10. 根据权利要求1所述的定向广告投放方法, 其特征在于:

各用户标签向量经过 GBDT 得到相应的分数的过程是:

构建分类树以及叶子节点, 其中, 把用户标签向量相近的作为同一叶子节点, 并以叶子节点的平均值作为预测值, 计算各叶子节点的实际分数与预测分数的残差, 按照各残差的相近程度划分下一棵分类树的叶子节点, 持续进行分类树的划分, 直至残差小于预设的阈值, 将各用户标签向量所在的各分类树的叶子节点的平均值的和分别作为该用户标签向量对应的分数。

11. 根据权利要求1所述的定向广告投放方法, 其特征在于:

根据用户注册填写的职业、爱好、居住地信息以及用户历史浏览、点击行为获得用户兴趣。

12. 根据权利要求1所述的定向广告投放方法, 其特征在于:

利用交替最小二乘算法对任一用户行为向量建模为 $R_{u*v} \approx X_{Z*u}^T Y_{Z*v}$,

其中, R_{u*v} 为用户行为向量;

u 表示第 u 个广告;

v 表示第 v 个点击行为评分;

X_{z*u}^T 为广告向量;

Y_{z*v} 为用户向量;

z 为对用户推送的的特征,

通过随机梯度下降优化算法, 求解用户行为向量得到用户向量和广告向量。

13. 根据权利要求 10 所述的定向广告投放方法, 其特征在于:

使用 preprocessing 库的 PolynomialFeatures 类对叶子节点进行多项式变换。

14. 一种定向广告投放装置, 其特征在于, 包括:

用户兴趣相似度获取模块, 用于根据用户兴趣和标签数据, 针对用户兴趣所属的标签, 把用户兴趣归类到对应的标签中, 得到用户标签向量,

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度, 获得余弦兴趣相似度和距离兴趣相似度, 并利用线性加权来计算 score1;

用户行为向量相似度获取模块, 用于根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵, 形成用户行为向量, 其中, 用户曝光是指当前某个广告展示给用户看称之为曝光, 然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度, 获得余弦行为相似度和距离行为相似度, 利用线性加权来计算 score2;

用户向量的相似度获取模块, 用于利用交替最小二乘算法对用户行为向量建模, 并通过随机梯度下降优化算法, 求解用户行为向量得到用户向量和广告向量, 并利用余弦相似度计算任意两个用户的用户向量的相似度, 得到 score3;

用户相似度获取模块, 用于对 score1、score2、score3 线性加权得到用户相似度 score4,

$$\text{score4} = W1 * \text{score1} + W2 * \text{score2} + W3 * \text{score3}$$

其中, $W1$ 、 $W2$ 、 $W3$ 是权重系数;

用户相似度获取模块, 将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入, 并且, 将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量, 并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM, 得到 score5, 计算公式如下:

$$\text{score5} = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中, T 是激活函数;

X 是特征;

$X_i * X_j$ 是二阶的特征交互;

$X_i * X_j * X_k$ 是三阶的特征交互;

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$\text{score6} = M1 * \text{score4} + M2 * \text{score5}$$

其中，M1、M2是系数；

相似度阈值比较模块，用于将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

15. 根据权利要求 14 所述的定向广告投放装置，其特征在于：

还包括用户行为向量修正模块，根据用户点击广告后浏览时间来确定用户对广告偏好的评分矩阵，形成用户行为向量，其中，先设定分值区间，所述分值区间包括 0 至分值上限，用户没有点击为 0 分，还结合用户点开页面以及用户浏览页面时间的长短区分出直至分值上限的分值。

16. 根据权利要求 14 所述的定向广告投放装置，其特征在于：还设置有评分矩阵修正系数，所述修正系数根据用户点击广告后是否转发，是否删除来确定，其中，用户转发广告则加 1 分，用户删除广告则减 1 分。

17. 根据权利要求 14 所述的定向广告投放装置，其特征在于：

还设置有距离判断推送模块，用于检测用户的位置坐标，根据广告委托者所在的位置坐标与用户的位置坐标的距离远近来推送广告，以距离用户的所述广告委托者的距离由近及远的方式向用户推送广告。

18. 一种电子设备，其特征在于，该电子设备包括：存储器和处理器，所述存储器中存储有定向广告投放程序，所述定向广告投放程序被所述处理器执行时实现如下步骤：

根据用户兴趣和标签数据，针对用户兴趣所属的标签，把用户兴趣归类到对应的标签中，得到用户标签向量，

利用余弦相似度和欧几里得距离来计算任意两个用户标签向量的相似度，获得余弦兴趣相似度和距离兴趣相似度，并利用线性加权来计算 score1；

根据用户曝光和点击行为评分得到用户对广告偏好的评分矩阵，形成用户行为向量，其中，用户曝光是指当前某个广告展示给用户看称之为曝光，然后利用余弦相似度和欧几里得距离来计算任意两个用户行为向量的相似度，获得余弦行为相似度和距离行为相似度，利用线性加权来计算 score2；

利用交替最小二乘算法对用户行为向量建模，并通过随机梯度下降优化算法，求解用

户行为向量得到用户向量和广告向量，并利用余弦相似度计算任意两个用户的用户向量的相似度，得到 score3；

对 score1、score2、score3 线性加权得到用户相似度score4，

$$\text{score4} = W1 * \text{score1} + W2 * \text{score2} + W3 * \text{score3}$$

其中，W1、W2、W3是权重系数；

将得到的用户行为向量、用户标签向量、用户向量分别经过 XGBOOST、SVM、GBDT 得到相应的分数作为 FM 的输入，并且，将 GBDT 的同一用户向量对应的叶子节点作多项式变换扩展叶子节点数量，并将同一用户向量对应的扩展后的叶子节点的值的和作为分数也输入 FM，得到 score5，计算公式如下：

$$\text{score5} = T(V_1 * X + V_{ij} * X_i * X_j + V_{ijk} * X_i * X_j * X_k + V_{ijkl} * X_i * X_j * X_k * X_l)$$

其中，T 是激活函数；

X 是特征；

$X_i * X_j$ 是二阶的特征交互；

$X_i * X_j * X_k$ 是三阶的特征交互；

$X_i * X_j * X_k * X_l$ 是四阶的特征交互；

$V_1, V_{ij}, V_{ijk}, V_{ijkl}$ 是权重系数；

最后经过 score4 和 score5 进行加权组合，输出最终的用户相似度，

$$\text{score6} = M1 * \text{score4} + M2 * \text{score5}$$

其中，M1、M2是系数；

将最终的用户相似度与相似度阈值比较，选择高于相似度阈值的用户作为目标客群。

19. 根据权利要求 18 所述的电子设备，其特征在于：

还根据用户点击广告后浏览时间来确定用户对广告偏好的评分矩阵，形成用户行为向量，其中，先设定分值区间，所述分值区间包括 0 至分值上限，用户没有点击为 0 分，还结合用户点开页面以及用户浏览页面时间的长短区分出直至分值上限的分值。

20. 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存储有计算机程序，所述计算机程序包括程序指令，所述程序指令被处理器执行时，实现权利要求 1 至 13 中任一项所述的定向广告投放方法。

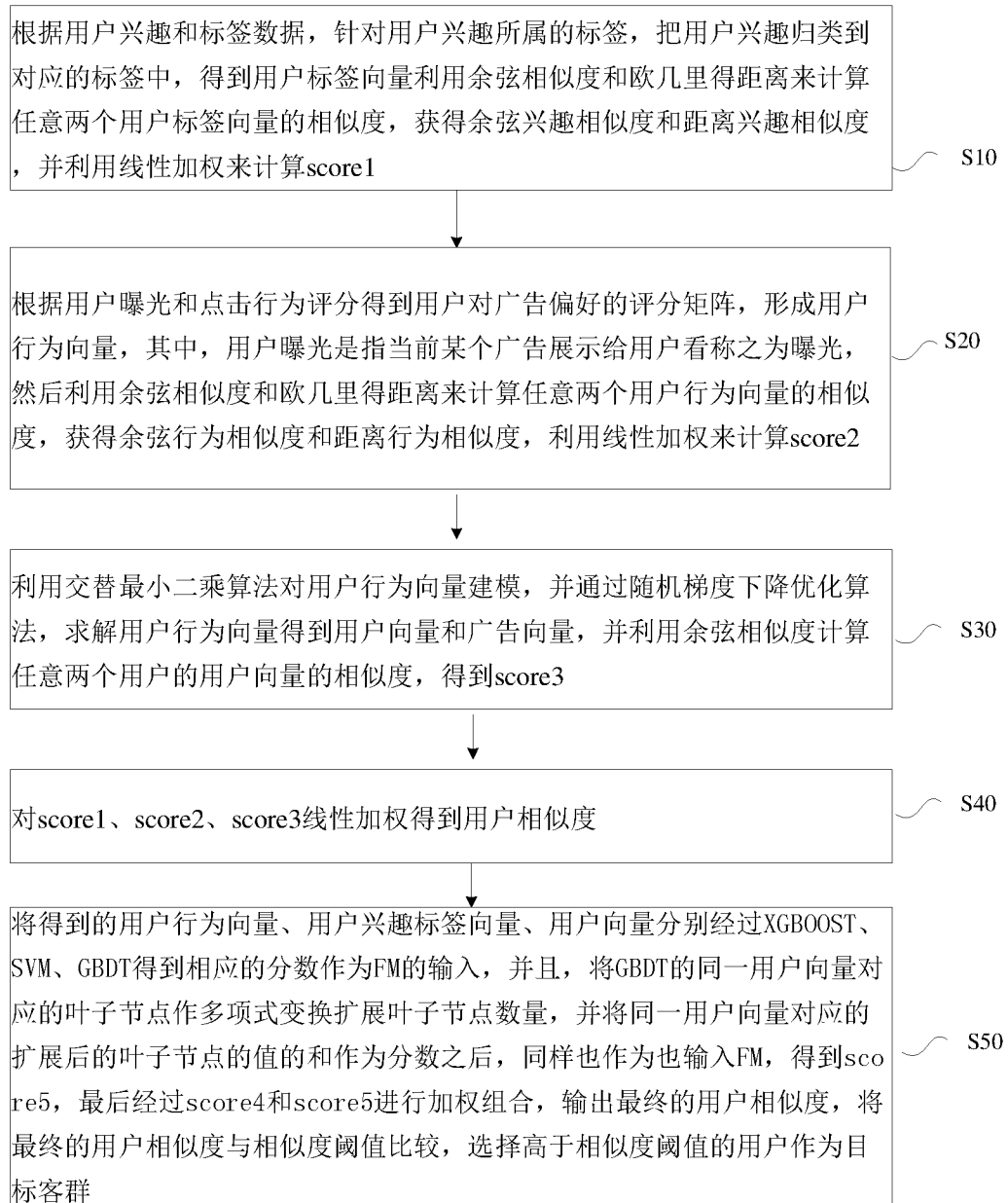


图 1

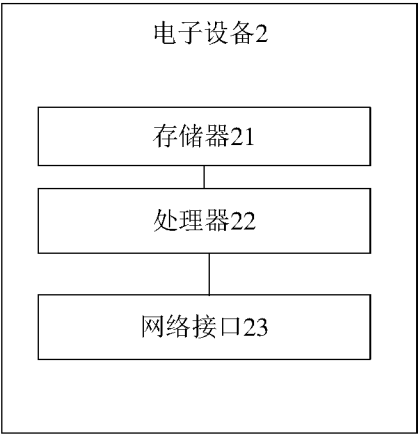


图 2

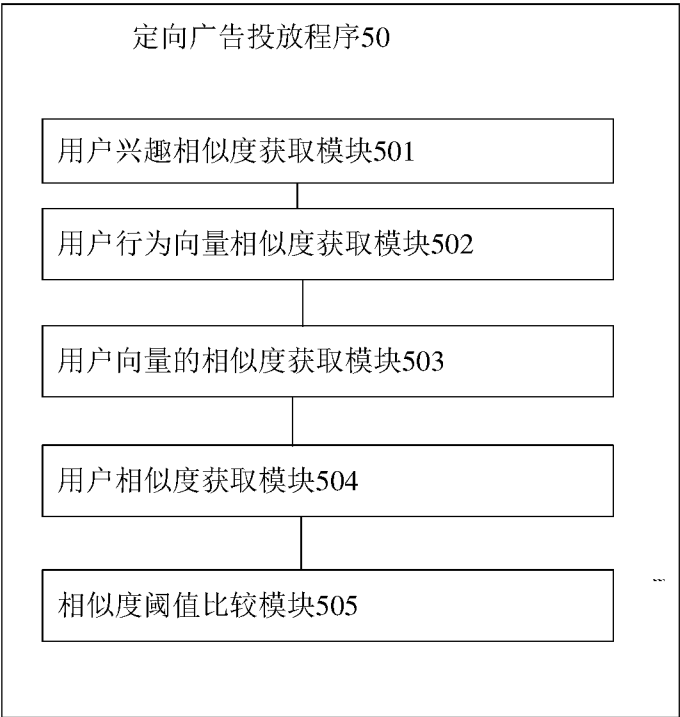


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/103608

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 30/02(2012.01)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)

G06Q

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)

WPI, EPODOC, CNKI, CNPAT, IEEE: 定投, 定向, 投放, 标签, 广告, 用户, 兴趣, 爱好, 相似度, 向量, 阈值, directional, put, advertisement, user, interest, preference, tag, label, vector, similarity, threshold

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110097395 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 06 August 2019 (2019-08-06) description, paragraphs [0066]-[0169]	1-20
A	CN 108415913 A (ZHOU, Meng) 17 August 2018 (2018-08-17) description, paragraphs [0063]-[0116]	1-20
A	US 2013073569 A1 (LEE, S.K. et al.) 21 March 2013 (2013-03-21) entire document	1-20
A	CN 103778145 A (CHINA MOBILE GROUP GUANGDONG COMPANY LIMITED) 07 May 2014 (2014-05-07) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“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)

“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

“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

“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

“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

“&” document member of the same patent family

Date of the actual completion of the international search

02 December 2019

Date of mailing of the international search report

30 December 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/103608

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110097395	A	06 August 2019	None			
CN	108415913	A	17 August 2018	None			
US	2013073569	A1	21 March 2013	WO	2011096690	A2	11 August 2011
				KR	20110090354	A	10 August 2011
				EP	2533430	A2	12 December 2012
CN	103778145	A	07 May 2014	None			

国际检索报告

国际申请号

PCT/CN2019/103608

A. 主题的分类

G06Q 30/02 (2012.01) 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)

G06Q

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))

WPI, EPDOC, CNKI, CNPAT, IEEE: 定投, 定向, 投放, 标签, 广告, 用户, 兴趣, 爱好, 相似度, 向量, 阈值, directional, put, advertisement, user, interest, preference, tag, label, vector, similarity, threshold

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 110097395 A (平安科技深圳有限公司) 2019年 8月 6日 (2019 - 08 - 06) 说明书第[0066]-[0169]段	1-20
A	CN 108415913 A (周孟) 2018年 8月 17日 (2018 - 08 - 17) 说明书第[0063]-[0116]段	1-20
A	US 2013073569 A1 (LEE, Sang Keun 等) 2013年 3月 21日 (2013 - 03 - 21) 全文	1-20
A	CN 103778145 A (中国移动通信集团广东有限公司) 2014年 5月 7日 (2014 - 05 - 07) 全文	1-20

☐ 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

2019年 12月 2日

国际检索报告邮寄日期

2019年 12月 30日

ISA/CN的名称和邮寄地址

中国国家知识产权局(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

授权官员

马春黎

电话号码 86-(10)-53961336

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/103608

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	110097395	A	2019年 8月 6日	无	
CN	108415913	A	2018年 8月 17日	无	
US	2013073569	A1	2013年 3月 21日	W0 2011096690 A2	2011年 8月 11日
				KR 20110090354 A	2011年 8月 10日
				EP 2533430 A2	2012年 12月 12日
CN	103778145	A	2014年 5月 7日	无	