(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110458360 B
(45)授权公告日 2023.07.18

(21)申请号 201910745843.7

(22)申请日 2019.08.13

(65)同一申请的已公布的文献号
申请公布号 CN 110458360 A

(43)申请公布日 2019.11.15

(73)专利权人 腾讯科技(深圳)有限公司
地址 518057 广东省深圳市南山区高新区科技中一路腾讯大厦35层

(72)发明人 谢思发 程序 张涵宇 江小琴 刘文强

(74)专利代理机构 北京三高永信知识产权代理有限责任公司 11138
专利代理师 祝亚男

(51)Int.Cl.
G06Q 10/04(2023.01)
G06N 20/00(2019.01)
G06F 18/2431(2023.01)

(56)对比文件
CN 103258248 A,2013.08.21
CN 105786962 A,2016.07.20
CN 106257449 A,2016.12.28
CN 107301471 A,2017.10.27
CN 107463660 A,2017.12.12
CN 107798027 A,2018.03.13
CN 108205536 A,2018.06.26
CN 108320171 A,2018.07.24
CN 108573408 A,2018.09.25
CN 108665148 A,2018.10.16
CN 108733631 A,2018.11.02
CN 109275031 A,2019.01.25
CN 109361932 A,2019.02.19
CN 109492827 A,2019.03.19
CN 109840541 A,2019.06.04
CN 109858562 A,2019.06.07
CN 109993566 A,2019.07.09 (续)

审查员 马贺

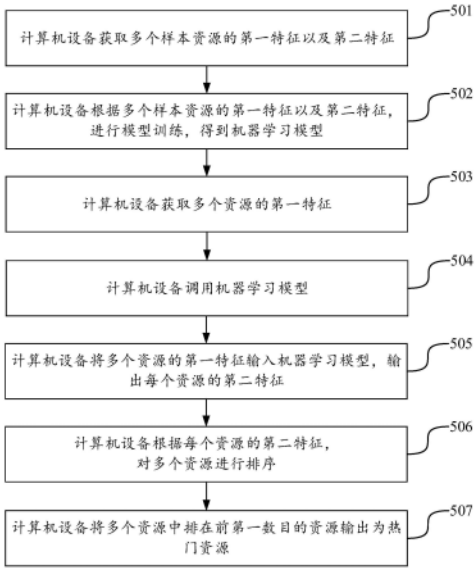
权利要求书5页 说明书24页 附图10页

(54)发明名称

热门资源的预测方法、装置、设备及存储介质

(57)摘要

本发明公开了一种热门资源的预测方法、装置、设备及存储介质，属于网络技术领域。该方法包括：获取多个资源的第一特征，第一特征包括资源的标签数量、价格、发行时长、评测数量、访问数量中的一项或多项；调用机器学习模型，机器学习模型用于根据资源的第一特征对资源的第二特征进行预测，第二特征能够表示资源的热门程度；将多个资源的第一特征输入机器学习模型，输出每个资源的第二特征；根据每个资源的第二特征，对多个资源进行排序；将多个资源中排在前第一数目的资源输出为热门资源。本发明提供了一种自动化的预测出热门资源的方法，节省了预测热门资源的时间，提高了预测热门资源的效率。此外，能够提高预测的准确性。



CN 110458360 B

[转续页]

[接上页]

(56)对比文件

CN 109993645 A,2019.07.09

US 2018005253 A1,2018.01.04

US 2018014038 A1,2018.01.11

WO 02069210 A1,2002.09.06

WO 2019109797 A1,2019.06.13

1.一种热门资源的预测方法,其特征在于,所述方法包括:

针对于发行时间与当前时间之间的时间间隔小于第一时间阈值的资源,获取多个资源的第一特征,所述第一特征包括资源的标签数量、价格、发行时长、评测数量、访问数量中的一项或多项;调用机器学习模型,所述机器学习模型用于根据资源的第一特征对资源的第二特征进行预测,所述第二特征能够表示资源的热门程度;将所述多个资源的第一特征输入所述机器学习模型,输出每个资源的第二特征;根据每个资源的第二特征,对所述多个资源进行排序;将所述多个资源中排在前第一数目的资源输出为热门资源;

针对于发行时间与当前时间之间的时间间隔大于第二时间阈值且在上一个统计周期具有所述第二特征的资源,对所述资源在当前统计周期的第二特征的平均值以及上一个统计周期的第二特征的平均值进行对比,得到所述资源的环比增长率,所述第二时间阈值不小于所述第一时间阈值;按照环比增长率从大到小的顺序,对多个资源进行排序;将所述多个资源中排在前第二数目的资源输出为热门资源;

针对于发行时间与当前时间之间的时间间隔大于所述第二时间阈值且在上一个统计周期不具有所述第二特征的资源,选取发行时间点位于当前统计周期的多个资源,并获取每个资源在当前统计周期的第二特征的平均值;对所述多个资源按照平均值从大到小的顺序排序;将所述多个资源中排在前第三数目的资源输出为热门资源。

2.根据权利要求1所述的方法,其特征在于,所述机器学习模型包括多个回归树,所述将所述多个资源的第一特征输入所述机器学习模型,输出每个资源的第二特征,包括:

对于所述多个资源中的每个资源,将所述资源的第一特征输入所述多个回归树,得到每个回归树的预测结果;

获取所述多个回归树的预测结果之和,作为所述资源的第二特征。

3.根据权利要求2所述的方法,其特征在于,每个回归树包括一个或多个非叶子节点、多个分支以及多个叶子节点,所述非叶子节点为第一节点、第二节点、第三节点、第四节点或第五节点;

所述第一节点用于根据标签数量进行分类,所述第一节点的每个分支表示标签数量的一个取值范围;

所述第二节点用于根据价格进行分类,所述第二节点的每个分支表示价格的一个取值范围;

所述第三节点用于根据发行时长进行分类,所述第三节点的每个分支表示发行时长的一个取值范围;

所述第四节点用于根据评测数量进行分类,所述第四节点的每个分支表示评测数量的一个取值范围;

所述第五节点用于根据访问数量进行分类,所述第五节点的每个分支表示访问数量的一个取值范围;

所述将所述资源的第一特征输入所述多个回归树,得到每个回归树的预测结果,包括:

对于所述多个回归树中的每个回归树,根据所述资源的第一特征在所述回归树中经过的目标非叶子节点以及目标分支,确定所述资源对应的目标叶子节点;

将所述目标叶子节点表示的预测结果,获取为所述回归树的预测结果。

4.根据权利要求2或3所述的方法,其特征在于,

所述第二特征为所述资源的销量，所述多个回归树中第一个回归树的预测结果表示销量的预测值，所述多个回归树中第n个回归树的预测结果表示销量的预测误差，所述预测误差为第1个回归树至第n-1个回归树的误差之和，所述n为大于1的正整数；或者，

所述第二特征为所述资源的在线用户数量，所述多个回归树中第一个回归树的预测结果表示在线用户数量的预测值，所述多个回归树中第n个回归树的预测结果表示在线用户数量的预测误差。

5.根据权利要求1所述的方法，其特征在于，所述方法还包括：

获取多个资源类型的第三特征以及第四特征，所述第三特征能够表示厂商对资源类型的偏好程度，所述第四特征能够表示用户对资源类型的偏好程度；

根据每个资源类型的第三特征以及第四特征，生成目标图像，所述目标图像的第一坐标轴表示所述第三特征，所述目标图像的第二坐标轴表示所述第四特征；

输出所述目标图像。

6.根据权利要求5所述的方法，其特征在于，所述获取多个资源类型的第三特征以及第四特征，包括：

根据每个资源类型的资源在第一统计周期新增的数量和在第二统计周期新增的数量，获取每个资源类型的资源数量的增长率，作为每个资源类型的第三特征，所述第一统计周期晚于所述第二统计周期；

根据每个资源类型的资源在第三统计周期的评论数量和在第四统计周期的评论数量，获取每个资源类型的评论数量的增长率，作为每个资源类型的第四特征，所述第三统计周期晚于所述第四统计周期；或者，

根据每个资源类型的资源在第五统计周期的在线用户数量和在第六统计周期的在线用户数量，获取每个资源类型的在线用户数量的增长率，作为每个资源类型的第四特征，所述第五统计周期晚于所述第六统计周期；或者，

根据每个资源类型的资源在第七统计周期的销量和在第八统计周期的销量，获取每个资源类型的销量的增长率，作为每个资源类型的第四特征，所述第七统计周期晚于所述第八统计周期。

7.根据权利要求1所述的方法，其特征在于，所述方法还包括：

获取多个榜单；

为所述多个榜单中的每个榜单分配对应的权重；

根据每个榜单的权重以及每个榜单中资源的排名，获取每个资源的权重，榜单的权重越大或排名越靠前，资源的权重越大；

根据每个资源的权重，对多个资源进行排序；

根据排序结果，生成目标榜单。

8.根据权利要求1所述的方法，其特征在于，所述获取多个资源的第一特征，包括：

爬取目标网站的网页内容，所述目标网站包括资源发布网站、视频网站、直播网站、资讯网站中的一项或多项；

解析所述网页内容，得到所述多个资源的数据；

根据所述多个资源的数据，获取所述多个资源的第一特征。

9.一种热门资源的预测装置，其特征在于，所述装置包括：

获取模块，用于针对于发行时间与当前时间之间的时间间隔小于第一时间阈值的资源，获取多个资源的第一特征，所述第一特征包括资源的标签数量、价格、发行时长、评测数量、访问数量中的一项或多项；

调用模块，用于调用机器学习模型，所述机器学习模型用于根据资源的第一特征对资源的第二特征进行预测，所述第二特征能够表示资源的热门程度；

模型运算模块，用于将所述多个资源的第一特征输入所述机器学习模型，输出每个资源的第二特征；

排序模块，用于根据每个资源的第二特征，对所述多个资源进行排序；

输出模块，用于将所述多个资源中排在前第一数目的资源输出为热门资源；

对比模块，用于针对于发行时间与当前时间之间的时间间隔大于第二时间阈值且在上一个统计周期具有所述第二特征的资源，对所述资源在当前统计周期的第二特征的平均值以及上一个统计周期的第二特征的平均值进行对比，得到所述资源的环比增长率，所述第二时间阈值不小于所述第一时间阈值；

所述排序模块，还用于按照环比增长率从大到小的顺序，对多个资源进行排序；

所述输出模块，还用于将所述多个资源中排在前第二数目的资源输出为热门资源；

用于执行以下步骤的模块：针对于发行时间与当前时间之间的时间间隔大于所述第二时间阈值且在上一个统计周期不具有所述第二特征的资源，选取发行时间点位于当前统计周期的多个资源，并获取每个资源在当前统计周期的第二特征的平均值；对所述多个资源按照平均值从大到小的顺序排序；将所述多个资源中排在前第三数目的资源输出为热门资源。

10.根据权利要求9所述的装置，其特征在于，所述机器学习模型包括多个回归树，所述模型运算模块包括：

回归树运算子模块，用于对于所述多个资源中的每个资源，将所述资源的第一特征输入所述多个回归树，得到每个回归树的预测结果；

求和子模块，用于获取所述多个回归树的预测结果之和，作为所述资源的第二特征。

11.根据权利要求10所述的装置，其特征在于，每个回归树包括一个或多个非叶子节点、多个分支以及多个叶子节点，所述非叶子节点为第一节点、第二节点、第三节点、第四节点或第五节点；

所述第一节点用于根据标签数量进行分类，所述第一节点的每个分支表示标签数量的一个取值范围；

所述第二节点用于根据价格进行分类，所述第二节点的每个分支表示价格的一个取值范围；

所述第三节点用于根据发行时长进行分类，所述第三节点的每个分支表示发行时长的一个取值范围；

所述第四节点用于根据评测数量进行分类，所述第四节点的每个分支表示评测数量的一个取值范围；

所述第五节点用于根据访问数量进行分类，所述第五节点的每个分支表示访问数量的一个取值范围；

所述回归树运算子模块，用于对于所述多个回归树中的每个回归树，根据所述资源的

第一特征在所述回归树中经过的目标非叶子节点以及目标分支，确定所述资源对应的目标叶子节点；将所述目标叶子节点表示的预测结果，获取为所述回归树的预测结果。

12.根据权利要求10或11所述的装置，其特征在于，

所述第二特征为所述资源的销量，所述多个回归树中第一个回归树的预测结果表示销量的预测值，所述多个回归树中第n个回归树的预测结果表示销量的预测误差，所述预测误差为第1个回归树至第n-1个回归树的误差之和，所述n为大于1的正整数；或者，

所述第二特征为所述资源的在线用户数量，所述多个回归树中第一个回归树的预测结果表示在线用户数量的预测值，所述多个回归树中第n个回归树的预测结果表示在线用户数量的预测误差。

13.根据权利要求9所述的装置，其特征在于，所述装置还包括：

所述获取模块，还用于获取多个资源类型的第三特征以及第四特征，所述第三特征能够表示厂商对资源类型的偏好程度，所述第四特征能够表示用户对资源类型的偏好程度；

生成模块，用于根据每个资源类型的第三特征以及第四特征，生成目标图像，所述目标图像的第一坐标轴表示所述第三特征，所述目标图像的第二坐标轴表示所述第四特征；

所述输出模块，还用于输出所述目标图像。

14.根据权利要求13所述的装置，其特征在于，所述获取模块用于执行：

根据每个资源类型的资源在第一统计周期新增的数量和在第二统计周期新增的数量，获取每个资源类型的资源数量的增长率，作为每个资源类型的第三特征，所述第一统计周期晚于所述第二统计周期；

根据每个资源类型的资源在第三统计周期的评论数量和在第四统计周期的评论数量，获取每个资源类型的评论数量的增长率，作为每个资源类型的第四特征，所述第三统计周期晚于所述第四统计周期；或者，

根据每个资源类型的资源在第五统计周期的在线用户数量和在第六统计周期的在线用户数量，获取每个资源类型的在线用户数量的增长率，作为每个资源类型的第四特征，所述第五统计周期晚于所述第六统计周期；或者，

根据每个资源类型的资源在第七统计周期的销量和在第八统计周期的销量，获取每个资源类型的销量的增长率，作为每个资源类型的第四特征，所述第七统计周期晚于所述第八统计周期。

15.根据权利要求9所述的装置，其特征在于，所述获取模块，还用于获取多个榜单；

分配模块，用于为所述多个榜单中的每个榜单分配对应的权重；

所述获取模块，还用于根据每个榜单的权重以及每个榜单中资源的排名，获取每个资源的权重，榜单的权重越大或排名越靠前，资源的权重越大；

所述排序模块，还用于根据每个资源的权重，对多个资源进行排序；

生成模块，用于根据排序结果，生成目标榜单。

16.根据权利要求9所述的装置，其特征在于，所述获取模块，包括：

爬取子模块，用于爬取目标网站的网页内容，所述目标网站包括资源发布网站、视频网站、直播网站、资讯网站中的一项或多项；

解析子模块，用于解析所述网页内容，得到所述多个资源的数据；

获取子模块，用于根据所述多个资源的数据，获取所述多个资源的第一特征。

17.一种计算机设备,其特征在于,所述计算机设备包括一个或多个处理器和一个或多个存储器,所述一个或多个存储器中存储有至少一条指令,所述指令由所述一个或多个处理器加载并执行以实现如权利要求1至权利要求8任一项所述的热门资源的预测方法。

18.一种计算机可读存储介质,其特征在于,所述存储介质中存储有至少一条指令,所述指令由处理器加载并执行以实现如权利要求1至权利要求8任一项所述的热门资源的预测方法。

热门资源的预测方法、装置、设备及存储介质

技术领域

[0001] 本发明涉及网络技术领域，特别涉及一种热门资源的预测方法、装置、设备及存储介质。

背景技术

[0002] 时下，可以通过互联网来发布、传播、销售或使用各种各样的资源。例如，游戏的厂商可以在游戏平台上发布游戏，用户可以在游戏平台上购买游戏，从游戏平台上下载游戏。为了帮助资源的开发，需要从海量的资源中，预测出哪个资源将成为市场中的热门资源，以便资源的厂商了解资源的流行趋势，制作更受欢迎的资源。

[0003] 以资源为游戏为例，热门游戏的预测过程目前通常是：人工在游戏平台上搜索近期发布的每个游戏的数据，根据每个游戏的数据，结合个人经验，从近期发布的各个游戏中，预测出哪个游戏将要成为爆款游戏，即热门游戏。

[0004] 从上述例子可以看出，人工预测热门资源的方式费时费力，效率较低，且存在主观性，无法有统一而客观的标准，导致预测的准确性较差。

发明内容

[0005] 本发明实施例提供了一种热门资源的预测方法、装置、设备及存储介质，能够解决相关技术中预测热门资源的效率较低、准确性差的技术问题。所述技术方案如下：

[0006] 一方面，提供了一种热门资源的预测方法，所述方法包括：

[0007] 获取多个资源的第一特征，所述第一特征包括资源的标签数量、价格、发行时长、评测数量、访问数量中的一项或多项；

[0008] 调用机器学习模型，所述机器学习模型用于根据资源的第一特征对资源的第二特征进行预测，所述第二特征能够表示资源的热门程度；

[0009] 将所述多个资源的第一特征输入所述机器学习模型，输出每个资源的第二特征；

[0010] 根据每个资源的第二特征，对所述多个资源进行排序；

[0011] 将所述多个资源中排在前第一数目的资源输出为热门资源。

[0012] 可选地，所述方法还包括：

[0013] 获取多个榜单；

[0014] 为所述多个榜单中的每个榜单分配对应的权重；

[0015] 根据每个榜单的权重以及每个榜单中资源的排名，获取每个资源的权重，榜单的权重越大或排名越靠前，资源的权重越大；

[0016] 根据每个资源的权重，对所述多个资源进行排序；

[0017] 根据排序结果，生成目标榜单。

[0018] 可选地，所述获取多个资源的第一特征，包括：

[0019] 爬取目标网站的网页内容，所述目标网站包括资源发布网站、视频网站、直播网站、资讯网站中的一项或多项；

[0020] 解析所述网页内容，得到多个资源的数据；

[0021] 根据所述多个资源的数据，获取所述多个资源的第一特征。

[0022] 另一方面，提供了一种热门资源的预测装置，所述装置包括：

[0023] 获取模块，用于获取多个资源的第一特征，所述第一特征包括资源的标签数量、价格、发行时长、评测数量、访问数量中的一项或多项；

[0024] 调用模块，用于调用机器学习模型，所述机器学习模型用于根据资源的第一特征对资源的第二特征进行预测，所述第二特征能够表示资源的热门程度；

[0025] 模型运算模块，用于将所述多个资源的第一特征输入所述机器学习模型，输出每个资源的第二特征；

[0026] 排序模块，用于根据每个资源的第二特征，对所述多个资源进行排序；

[0027] 输出模块，用于将所述多个资源中排在前第一数目的资源输出为热门资源。

[0028] 可选地，所述机器学习模型包括多个回归树，所述模型运算模块，包括：

[0029] 回归树运算子模块，用于对于所述多个资源中的每个资源，将所述资源的第一特征输入所述多个回归树，得到每个回归树的预测结果；

[0030] 求和子模块，用于获取所述多个回归树的预测结果之和，作为所述资源的第二特征。

[0031] 可选地，每个回归树包括一个或多个非叶子节点、多个分支以及多个叶子节点，所述非叶子节点为第一节点、第二节点、第三节点、第四节点或第五节点；

[0032] 所述第一节点用于根据标签数量进行分类，所述第一节点的每个分支表示标签数量的一个取值范围；

[0033] 所述第二节点用于根据价格进行分类，所述第二节点的每个分支表示价格的一个取值范围；

[0034] 所述第三节点用于根据发行时长进行分类，所述第三节点的每个分支表示发行时长的一个取值范围；

[0035] 所述第四节点用于根据评测数量进行分类，所述第四节点的每个分支表示评测数量的一个取值范围；

[0036] 所述第五节点用于根据访问数量进行分类，所述第五节点的每个分支表示访问数量的一个取值范围；

[0037] 所述回归树运算子模块，用于：对于所述多个回归树中的每个回归树，根据所述资源的第一特征在所述回归树中经过的目标非叶子节点以及目标分支，确定所述资源对应的目标叶子节点；将所述目标叶子节点表示的预测结果，获取为所述回归树的预测结果。

[0038] 可选地，所述第二特征为所述资源的销量，所述多个回归树中第一个回归树的预测结果表示销量的预测值，所述多个回归树中第n个回归树的预测结果表示销量的预测误差，所述预测误差为第1个回归树至第n-1个回归树的误差之和，所述n为大于1的正整数；或者，

[0039] 所述第二特征为所述资源的在线用户数量，所述多个回归树中第一个回归树的预测结果表示在线用户数量的预测值，所述多个回归树中第n个回归树的预测结果表示在线用户数量的预测误差。

[0040] 可选地，所述获取模块，还用于获取多个资源类型的第三特征以及第四特征，所述

第三特征能够表示厂商对资源类型的偏好程度，所述第四特征能够表示用户对资源类型的偏好程度；

[0041] 所述装置还包括：生成模块，用于根据每个资源类型的第三特征以及第四特征，生成目标图像，所述目标图像的第一坐标轴表示第三特征，所述目标图像的第二坐标轴表示第四特征；

[0042] 所述输出模块，还用于输出所述目标图像。

[0043] 可选地，所述获取模块，用于执行下述至少一项：

[0044] 根据每个资源类型的资源在第一统计周期新增的数量和在第二统计周期新增的数量，获取每个资源类型的资源数量的增长率，作为每个资源类型的第三特征，所述第一统计周期晚于所述第二统计周期；

[0045] 根据每个资源类型的资源在第三统计周期的评论数量和在第四统计周期的评论数量，获取每个资源类型的评论数量的增长率，作为每个资源类型的第四特征，所述第三统计周期晚于所述第四统计周期；

[0046] 根据每个资源类型的资源在第五统计周期的在线用户数量和在第六统计周期的在线用户数量，获取每个资源类型的在线用户数量的增长率，作为每个资源类型的第四特征，所述第五统计周期晚于所述第六统计周期；

[0047] 根据每个资源类型的资源在第七统计周期的销量和在第八统计周期的销量，获取每个资源类型的销量的增长率，作为每个资源类型的第四特征，所述第七统计周期晚于所述第八统计周期。

[0048] 可选地，所述装置还包括：

[0049] 对比模块，用于对于所述多个资源中的每个资源，对所述资源在当前统计周期的第二特征以及上一个统计周期的第二特征进行对比，得到所述资源的环比增长率；

[0050] 所述排序模块，还用于根据每个资源的环比增长率，对所述多个资源进行排序；

[0051] 所述输出模块，还用于将所述多个资源中排在前第二数目的资源输出为热门资源。

[0052] 可选地，所述获取模块，还用于获取多个榜单；

[0053] 所述装置还包括：分配模块，用于为所述多个榜单中的每个榜单分配对应的权重；

[0054] 所述获取模块，还用于根据每个榜单的权重以及每个榜单中资源的排名，获取每个资源的权重，榜单的权重越大或排名越靠前，资源的权重越大；

[0055] 排序模块，用于根据每个资源的权重，对所述多个资源进行排序；

[0056] 生成模块，用于根据排序结果，生成目标榜单。

[0057] 可选地，所述获取模块，包括：

[0058] 爬取子模块，用于爬取目标网站的网页内容，所述目标网站包括资源发布网站、视频网站、直播网站、资讯网站中的一项或多项；

[0059] 解析子模块，用于解析所述网页内容，得到多个资源的数据；

[0060] 获取子模块，用于根据所述多个资源的数据，获取所述多个资源的第一特征。

[0061] 另一方面，提供了一种计算机设备，所述计算机设备包括一个或多个处理器和一个或多个存储器，所述一个或多个存储器中存储有至少一条指令，所述指令由所述一个或多个处理器加载并执行以实现上述热门资源的预测方法。

[0062] 另一方面，提供了一种计算机可读存储介质，所述存储介质中存储有至少一条指令，所述指令由处理器加载并执行以实现上述热门资源的预测方法。

[0063] 本发明实施例提供的技术方案带来的有益效果至少包括：

[0064] 本发明实施例提供了一种自动化的预测出热门资源的方法，根据资源的标签数量、价格、发行时长、评测数量、访问数量等特征，利用机器学习模型来预测出资源的热门程度，从而发现出哪个资源将要成为市场中的热门资源，节省了预测热门资源的时间，提高了预测热门资源的效率。此外，通过机器学习模型，从大量的样本资源中，学习出资源的特征与热门程度之间的规律，使得通过机器学习模型进行预测时，能够提高预测的准确性。

附图说明

[0065] 为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述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。

[0066] 图1是本发明实施例提供的一种实施环境的示意图；

[0067] 图2是本发明实施例提供的一种资源监控系统的功能架构图；

[0068] 图3是本发明实施例提供的一种资源数据的采集方法的流程图；

[0069] 图4是本发明实施例提供的一种采集游戏数据的示意图；

[0070] 图5是本发明实施例提供的一种热门资源的预测方法的流程图；

[0071] 图6是本发明实施例提供的一种热门资源的预测方法的流程图；

[0072] 图7是本发明实施例提供的一种资源类型的流行趋势的预测方法的流程图；

[0073] 图8是本发明实施例提供的一种目标图像的示意图；

[0074] 图9是本发明实施例提供的一种榜单整合方法的流程图；

[0075] 图10是本发明实施例提供的一种热门资源的预测装置的框图；

[0076] 图11是本发明实施例提供的一种终端的结构示意图；

[0077] 图12是本发明实施例提供的一种服务器的结构示意图。

具体实施方式

[0078] 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、完整地描述，显然，所描述的实施例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做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。

[0079] 本发明实施例提供了一套自动化的资源监控系统，该资源监控系统能够监控与预测行业的趋势，具体地，会监控资源相关的各种数据源，采集到市场上每个资源的数据，预测出哪个资源可能成为市场中的热门资源，还能够预测出资源类型的流行趋势，此外还能够对各大平台提供的榜单进行整合和汇总，提供综合的榜单，将得出的监控结果自动通知给资源的开发者。

[0080] 资源可以而不限于应用程序、商品、音频、视频或文件等。示意性地，资源可以是游戏，该资源监控系统可以提供为游戏监控系统。该游戏监控系统通过监控游戏平台、游戏视

频、直播平台、新闻咨询等各个数据源，采集到市场上发行的每个游戏的数据，发现潜在的爆款游戏，预测游戏玩法的流行趋势，此外能够整合游戏热销榜、预约榜、优惠榜等各个榜单，提供爆款游戏榜。

[0081] 以下，通过下述各个实施例对该资源监控系统的各个方面分别进行描述。

[0082] 以下，示例性描述该资源监控系统的实施环境。

[0083] 图1是本发明实施例提供的一种实施环境的示意图。该实施环境包括：终端110和资源监控平台120。

[0084] 终端110通过无线网络或有线网络与资源监控平台120相连。终端110可以是智能手机、游戏主机、台式计算机、平板电脑、电子书阅读器、MP3(Moving Picture Experts Group Audio Layer III，动态影像专家压缩标准音频层面3)播放器或MP4(Moving Picture Experts Group Audio Layer IV，动态影像专家压缩标准音频层面4)播放器和膝上型便携计算机中的至少一种。可选地，终端110安装和运行应用程序，资源监控平台120可以向该应用程序发送信息，该应用程序可以向用户推送信息，该信息可以为热门资源、资源类型的流行趋势或目标榜单中的任一种，例如可以爆款游戏、游戏玩法的流行趋势、爆款游戏榜单等。示例性的，终端110是用户使用的终端，终端110中运行的应用程序内登录有用户账号。

[0085] 终端110通过无线网络或有线网络与资源监控平台120相连。

[0086] 可选地，资源监控平台120包括：服务器和数据库。服务器可以是一台服务器、多台服务器、云计算平台和虚拟化中心中的至少一种。服务器可以用于提供资源数据采集、热门资源预测、资源类型的流行趋势预测、榜单生成中任一项或多项有关的后台服务，终端可以通过应用程序，请求服务器提供相应的后台服务。服务器可以是一台或多台。当服务器是多台时，存在至少两台服务器用于提供不同的服务，和/或，存在至少两台服务器用于提供相同的服务，比如以负载均衡方式提供同一种服务，本发明实施例对此不加以限定。服务器可以设置有机器学习模型，服务器可以调用该机器学习模型，以预测热门资源。数据库可以用于存储服务器计算时需要的数据，或者存储服务器的计算结果。比如说，数据库可以存储采集到的资源数据、模型训练时需要使用的样本资源的第一特征以及第二特征，模型预测时需要使用的资源的第一特征、预测出的热门资源、表示资源类型的流行趋势的目标图像、生成的目标榜单中的一项或多项。

[0087] 以下，示例性描述该资源监控系统的功能架构。

[0088] 图2是本发明实施例提供的一种资源监控系统的功能架构图，该资源监控系统的功能主要包括数据爬取、数据清洗、数据存储、预测分析以及通知服务。

[0089] 数据爬取：能够基于爬虫技术，从资源发布网站、视频网站、直播网站、资讯网站爬取资源的数据。例如，如果资源是游戏，可以从游戏的发行平台、游戏视频、直播平台和游戏资讯分别爬取游戏数据。

[0090] 数据清洗：能够对爬取的数据进行清洗，具体可以包括实体抽取、缺失值补全、格式化、去重等功能。

[0091] 数据存储：能够将清洗后的数据存储至数据库。

[0092] 预测分析：可以包括热门资源预测、环比监控、资源类型的流行趋势、榜单整合等功能。其中，热门资源的预测方法还请参见下述图5实施例，环比监控的方法还请参见下述

图6实施例，资源类型的流行趋势的方法还请参见下述图7实施例，榜单整合的方法还请参见下述图9实施例。

[0093] 通知服务：用于将预测分析的结果通知给用户，该用户可以是订阅了预测分析功能的用户，例如可以为游戏开发者、厂商或玩家等。示意性地，可以将预测分析的结果携带在邮件中，发送给用户，或者将监控结果发布至网站，在网站中展示监控结果，方便用户在网站中查询监控结果。

[0094] 本发明实施例中，该资源监控系统能够通过图3实施例，监控能够提供资源数据的各个数据源，从而采集到市场中出现的每个资源的数据，通过采集的数据，来执行图5实施例、图6实施例、图7实施例以及图9实施例的方法。需要说明的一点是，图3实施例可以在图5实施例、图6实施例、图7实施例以及图9实施例之前执行，也可以不执行。图5实施例、图6实施例、图7实施例以及图9实施例可以全部执行，也可以只执行部分实施例，不同实施例可以按照时间的先后顺序串行执行，也可以同时执行，本实施例对图5实施例、图6实施例、图7实施例以及图9实施例的时序不做限定。

[0095] 以下，通过图3实施例，对资源数据采集的流程进行描述。

[0096] 图3是本发明实施例提供的一种资源数据的采集方法的流程图，该方法可以应用于上述实施环境中的资源监控平台，该资源监控平台可以通过终端或者服务器来执行采集数据的流程，而终端和服务器均可以视为一种计算机设备，因此，本发明实施例基于计算机设备作为执行主体来进行介绍。参见图3，该实施例具体可以包括以下步骤：

[0097] 301、计算机设备爬取目标网站的网页内容。

[0098] 目标网站用于提供资源的数据。具体地，目标网站包括资源发布网站、视频网站、直播网站、资讯网站中的一项或多项。

[0099] 资源发布网站用于发布资源，用户可以在资源发布网站上访问资源、下载资源、购买资源或使用资源。例如，如果资源是应用程序，资源发布网站可以是应用商店，如果资源是游戏，资源发布网站可以是游戏的发行平台，该游戏的发行平台可以用于发布端游、手游、主机游戏中的一项或多项。

[0100] 视频网站用于提供资源相关的视频，用户可以在视频网站上观看资源相关的视频。例如，如果资源是游戏，视频网站可以是专门提供游戏视频的网站，也可以是具有游戏频道的综合视频网站。

[0101] 直播网站用于提供直播服务，主播可以对资源的使用过程进行实时录制，得到直播，将直播上传至直播网站，用户可以在直播网站观看资源的使用过程。例如，如果资源是游戏，直播网站可以是电竞直播平台。

[0102] 资讯网站用于提供资源相关的资讯新闻，用户可以在资讯网站上浏览资源相关的资讯新闻。例如，如果资源是游戏，资讯网站可以是游戏资讯平台、发布游戏资讯的公众号、博客、论坛等。

[0103] 示意性地，参见图4，可以从平台、流以及新闻这三个方向来监控市场上出现的新游戏，以便采集新游戏的数据。其中，平台可以细分为端游、手游以及主机三个方向，流可以细分为直播平台、视频弹幕网站、视频分享网站以及游戏视频平台。

[0104] 关于爬取数据的具体过程，可以利用爬虫技术，对目标网站发布的网页进行爬取，得到目标网站的网页内容。具体来说，计算机设备可以通过爬虫程序，生成超文本传输协议

(Hypertext transfer protocol,HTTP)请求,向目标网站发送HTTP请求,该HTTP请求用于请求网页内容。目标网站的服务器接收到HTTP请求后,会生成HTTP响应,向计算机设备发送HTTP响应。计算机设备可以接收HTTP响应,从HTTP响应中,得到目标网站的网页内容。

[0105] 在一些可能的实施例中,爬取网页内容的过程具体可以包括以下步骤一至步骤二:

[0106] 步骤一、计算机设备从目标网站爬取每个资源的标识,根据每个资源的标识,获取每个资源的统一资源定位符(Uniform Resource Locator,URL);或者,从目标网站爬取每个资源的URL。

[0107] 其中,资源的标识用于标识资源,例如可以是资源的身份标识号(Identity,ID)、名称等。以资源为游戏为例,可以从游戏的发行平台,爬取每个游戏的ID(game id),根据每个游戏的ID,获取每个游戏的URL。又如,可以访问视频网站的游戏频道,从游戏频道,爬取每个游戏的URL。

[0108] 步骤二、根据每个资源的URL,从目标网站爬取每个资源对应的网页内容。

[0109] 对于任一资源而言,可以根据该资源的URL,生成HTTP请求,向目标网站的服务器发送HTTP请求,该目标网站的服务器会根据HTTP请求中的URL,确定该资源对应的网页,将网页内容通过HTTP响应返回给计算机设备。

[0110] 302、计算机设备解析网页内容,得到多个资源的数据。

[0111] 资源的数据可以包括资源的标签数量、价格、发行时间点、评论数量、测试数量、访问数量、资源类型、销量、榜单中的一项或多项。

[0112] 资源的标签为资源的用户为资源提交的词语。具体地,用户可以根据对资源的使用体验,输入形容该资源的词语,作为给资源打上的标签,或者,可以从其他用户提交的词语中,选择词语,作为给资源打上的标签。以资源为游戏为例,资源的标签可以是“角色扮演”、“高画质”、“和风”等。资源的标签数量可以为资源打上标签的用户数量,通常来讲,如果资源的标签数量越多,则表明存在很多用户为资源打上标签,能够反映资源越热门。

[0113] 访问数量用于表示资源的流量大小,可以为观看人数、阅读量、浏览量等。例如,如果资源是游戏,访问数量可以是该游戏的直播视频或宣传视频的观看人数、该游戏的新闻的阅读量、该游戏的介绍网页的浏览量等。

[0114] 资源类型可以根据资源的内容划分,例如,如果资源为游戏,资源类型可以是剧情、模拟经营、养成、穿越、策略、乙女、恋爱、宫斗、换装、放置、收集、卡牌、休闲、益智、对战、逃生或二次元。

[0115] 榜单可以为热门榜、预约榜、热销榜、新品榜、最近更新榜、好评榜、热玩榜等,每个榜单可以包括多个资源的标识以及每个资源的排名。不同榜单的排名方式可以不同,示意性地,热门榜可以按照热度排名,热销榜可以按照销量排名,预约榜可以按照预约用户数量排名,优惠榜可以按照优惠比例排名,最近更新榜可以按照最近一次更新的日期距离当前时间点的距离排名。

[0116] 303、计算机设备对多个资源的数据进行清洗。

[0117] 对数据进行清洗的过程可以包括缺失值补全、格式化处理、实体抽取、去重中的一项或多项,详见以下(1)至(4)。

[0118] (1)缺失值补全是指将存在缺失部分的数据补全为完整数据。具体地,由于数据源

本身的问题或者爬虫程序不稳定，可能会导致缺失部分数据，则可以根据实际情况，将数据的缺失部分置为预设数值，或者将数据的缺失部分置为上一个统计周期的均值。其中，该预设数值可以为0或空。

[0119] (2)格式化处理是指将不同格式或不同单位的数据统一为格式相同且单位相同的数据。举例来说，针对直播人数而言，直播网站A记录直播人数时，单位可以是人，比如11000人，直播网站B记录直播人数时，单位可以是万人，比如1.1万人，这个例子中直播网站A和直播网站B记录的直播人数实质上相同，但由于单位不同，导致两个网站记录的数据不同。为此，可以将以万人为单位的直播人数换算为以人为单位，从而统一各个直播网站提供的直播人数。又例如，针对日期而言，游戏平台A记录游戏的发行日期时，格式可以是2019年8月8日，游戏平台B记录游戏的发行日期时，格式可以是2019-8-8，游戏平台C记录游戏的发行日期时，格式可以是2019/8/8，为此，可以将爬取到的每个日期统一转换为xxxx-xx-xx的格式，从而方便数据库存储格式相同的各个日期。

[0120] (3)实体通常是指文本中具有特定意义或者指代性强的实体，通常包括人名、地名、组织机构名、日期时间、专有名词等，命名实体识别(Named Entity Recognition,NER)是自然语言处理的关键技术。本实施例中，实体是指资源的名称，例如为游戏名，实体抽取是指从文本形式的数据中，提取资源的名称。

[0121] 在一种可能的实现中，可以识别文本中的书名号，将位于书名号中的内容作为实体。可选地，可以根据过滤规则，对书名号中的内容进行过滤，将过滤后的内容作为实体。其中，过滤规则可以包括：过滤掉长度超过预设长度的内容、过滤掉以预设词汇结尾的内容中的一项或多项。

[0122] 其中，预设长度可以根据实验、经验或需求设置，具体可以根据资源的形式来确定，比如如果资源是游戏，结合游戏名称的通常长度，预设长度可以设置为10，从而过滤掉长度超过10的词汇，避免将长度超过10的词汇误作为游戏名称。预设词汇也可以根据实验、经验或需求设置，例如如果资源是游戏，从游戏资讯中提取游戏名称时，有时会匹配到一些报告、白皮书等非游戏名称，因此，可以将“报告”、“白皮书”作为预设词汇，从而过滤掉以报告或白皮书作为结尾的名字。经实验，这种实现方式召回率和准确率较高。

[0123] 可选地，可以对文本形式的数据进行分词，得到多个词语，对每个词语与游戏名库中的游戏名匹配，将匹配的词语作为实体。其中，该游戏名库包括多个游戏名，可以预先存储游戏名库，也可以调用某一平台发布的游戏名库，也可以由计算机设备构建游戏名库。具体地，从资源发布网站、视频网站以及直播网站中的一项或多项爬取到的数据后，可以从爬取的数据中提取资源的名称，将资源的名称存入游戏名库。

[0124] (4)去重是指删除重复的数据。

[0125] 304、计算机设备将清洗后的数据存储至数据库。

[0126] 数据库可以是mysql(一种关系型数据库管理系统)、neo4j(一种将结构化数据存储在网络的数据库)或HBASE(一种分布式的面向列的开源数据库)，当然也可以是其他数据库。

[0127] 相关技术中，为了分析时下的热门资源以及流行趋势等，需要依靠人工操作，手动地在各个网站中进行检索，下载检索到的数据，对检索的数据人工进行汇总和整理，导致数据采集的过程费时费力、过程繁琐，并且，收集一轮数据的统计周期长，导致信息更新不及

时。

[0128] 而本实施例提供了一种自动化的数据采集方案，通过爬虫技术，自动地从目标网站中爬取到资源的数据，从而免去了人工检索数据以及下载数据的繁琐操作，此外，通过数据清洗技术，能够实现缺失值补全、格式化处理、实体抽取、去重等各个功能，从而自动地对数据进行整理，免去了人工整理数据的繁琐操作，因此节约了数据采集的时间，提高了数据采集的效率。

[0129] 上述实施例示出了资源数据自动采集的方法，本发明实施例中，可以自动化的预测出热门资源，例如预测出市场上的爆款游戏。具体参见下述图5实施例。需要说明的一点是，图5实施例的执行主体和图3实施例的执行主体可以为同一计算机设备或不同的计算机设备，本实施例对此不做限定。

[0130] 图5是本发明实施例提供的一种热门资源的预测方法的流程图，参见图5，该实施例具体可以包括以下步骤：

[0131] 501、计算机设备获取多个样本资源的第一特征以及第二特征。

[0132] 本实施例中，可以采用机器学习的手段，通过大量的样本资源，学习出资源的第一特征与第二特征之间的映射关系，得到机器学习模型。当要从待预测的大量资源中，发现热门资源时，可以通过该机器学习模型来预测出每个资源的第二特征，以便根据每个资源的第二特征衡量每个资源的热门程度。

[0133] 在一些可能的实施例中，样本资源可以为发布时间点与当前时间点之间的时间间隔大于第一时长的资源。例如，可以查询数据库中记录的每个资源的发布时间点，从已记录的每个资源中，选择发布时间点与当前时间点之间的时间间隔大于第一时长的资源，作为样本资源，以便利用较早之前发布的老资源来进行模型训练。以第一时长为30天为例，可以将30天之前发布的游戏，比如将近180天至近30天发布的游戏作为样本游戏。当确定样本资源后，可以从数据库中读取样本资源的第一特征以及第二特征，或者根据数据库存储的样本资源的数据进行统计，得到样本资源的第一特征以及第二特征。

[0134] 第一特征包括资源的标签数量、价格、发行时长、评测数量、访问数量中的一项或多项。其中，可选地，可以对资源的标签数量进行归一化，得到归一化后的标签数量，将归一化的标签数量作为第一特征，来执行模型预测的步骤或者模型训练的步骤。具体地，可以获取资源的每个标签的数量与总标签的数量之间的比值，作为归一化后的标签数量。该归一化的标签数量的取值范围可以为0至1。资源的发行时长可以是资源的发行时间点与当前时间点之间的时间差。资源的发行时长可以以天为单位，例如，资源的发行时长可以为从资源的发行日到模型训练当天之间经过的天数。评测数量可以是评论数量或测试数量。

[0135] 第二特征能够表示资源的热门程度。可选地，第二特征可以为销量，资源的销量越大，表示资源越热门，资源成为爆款资源的概率越高，相应地，机器学习模型可以是销量预测模型。可选地，第二特征也可以为在线用户数量。资源的在线用户数量可以是登录或使用资源的数量，在线用户数量可以以天为单位，例如，在线用户数量可以是日活跃用户数量(Daily Active User，DAU)，资源的在线用户数量越大，表示资源越热门，资源成为爆款资源的概率越高，相应地，机器学习模型可以是在线用户数量预测模型。

[0136] 需要说明的一点是，销量以及在线用户数量仅是对第二特征的举例说明，第二特征也可以是销量以及在线用户数之外的其他数据，任意能够表示资源的热门程度的特征均

可以作为第二特征，比如资源的评论数量、观看人数等，本实施例对第二特征的具体类型并不做限定。

[0137] 502、计算机设备根据多个样本资源的第一特征以及第二特征，进行模型训练，得到机器学习模型，该机器学习模型用于根据资源的第一特征对资源的第二特征进行预测。

[0138] 可以将第一特征作为模型训练的特征，将第二特征作为模型训练的标签(label)，构建训练集，采用机器学习算法，对多个样本资源的第一特征以及第二特征进行模型训练。

[0139] 机器学习算法可以为集成学习算法，例如可以是提高(boosting)算法、引导聚合(bagging)算法、堆叠(stacking)算法等。相应地，训练得出的机器学习模型可以是集成学习模型，该集成学习模型可以包括多个子模型，每个子模型均可以根据资源的第一特征进行预测，通过结合多个子模型的预测结果，可以让机器学习模型最终输出的第二特征更加准确。

[0140] 举例来说，机器学习算法可以是极端梯度提升(eXtreme Gradient Boosting，XGBoost)算法，XGBoost算法中，会不断地生成回归树，通过将许多弱监督的回归树集成在一起，形成一个强监督的机器学习模型，以避免单个回归树预测时容易出现过拟合的情况。具体地，会参考第1个回归树进行预测时产生的误差，生成第2个回归树，参考第1个回归树以及第2个回归树的组合进行预测时产生的误差，生成第3个回归树，依次类推，参考n-1个回归树组合而成的模型进行预测时产生的误差，生成第n个回归树，从而得到n个回归树，其中n为正整数。在生成回归树的过程中，每当新增一个回归树，模型整体的损失会不断降低。其中，误差具体可以为模型的预测结果与实际结果之间的残差，例如，误差可以为销量的预测值与实际的销量之间的残差，又如，误差可以为在线用户数量的预测值与实际的在线用户数量之间的残差。

[0141] 机器学习模型可以包括多个回归树。每个回归树可以包括一个或多个非叶子节点、多个分支以及多个叶子节点。

[0142] 每个非叶子节点可以视为一个分类器，每个非叶子节点用于根据一种第一特征进行分类。每个非叶子节点连接多个分支，每个分支可以视为一个分类结果，该分类结果可以是第一特征的一种取值范围。具体来说，每个非叶子节点可以为第一节点、第二节点、第三节点、第四节点或第五节点。

[0143] 第一节点用于根据标签数量进行分类，第一节点的每个分支表示标签数量的一个取值范围。例如，第一节点可以连接2个分支，第一个分支表示标签数量＜k1，第二个分支表示标签数量≥k1，那么对于任一个资源来说，如果该资源的标签数量为a，且a＜k1，则将该资源的第一特征输入至第一节点后，第一特征会输出至第一个分支，落入左子树。其中，k1和a均≥0，如果对标签数量进行了归一化，则k1和a均≥0，且k1和a均≤1。

[0144] 第二节点用于根据价格进行分类，第二节点的每个分支表示价格的一个取值范围。例如，第二节点可以连接2个分支，第一个分支表示价格＜k2，第二个分支表示价格≥k2，那么对于任一个资源来说，如果该资源的价格为b，且b≥k2，则将该资源的第一特征输入至第二节点后，第一特征会输出至第二个分支，落入右子树。其中，b以及k2均≥0。

[0145] 第三节点用于根据发行时长进行分类，第三节点的每个分支表示发行时长的一个取值范围。例如，第三节点可以连接2个分支，第一个分支表示发行时长＜k3，第二个分支表示发行时长≥k3，那么对于任一个资源来说，如果该资源的发行时长为c，且c≥k3，则将该

资源的第一特征输入至第三节点后，第一特征会输出至第二个分支。其中，c以及k3均≥0。

[0146] 第四节点用于根据评测数量进行分类，第四节点的每个分支表示评测数量的一个取值范围。例如，第四节点可以连接2个分支，第一个分支表示评测数量＜k4，第二个分支表示评测数量≥k4，那么对于任一个资源来说，如果该资源的评测数量为d，且d＜k4，则将该资源的第一特征输入至第四节点后，第一特征会输出至第一个分支。其中，d以及k4均≥0。

[0147] 第五节点用于根据访问数量进行分类，第五节点的每个分支表示访问数量的一个取值范围。例如，第五节点可以连接2个分支，第一个分支表示访问数量＜k5，第二个分支表示访问数量≥k5，那么对于任一个资源来说，如果该资源的访问数量为e，且e≥k5，则将该资源的第一特征输入至第五节点后，第一特征会输出至第二个分支。其中，e以及k5均≥0。

[0148] 需要说明的一点是，第一节点、第二节点、第三节点、第四节点或第五节点具有两个分支仅是举例说明，在一些可能的实施例中，第一节点、第二节点、第三节点、第四节点或第五节点可以具有三个或三个以上的分支，本实施例对第一节点、第二节点、第三节点、第四节点或第五节点具有的分支数量并不做限定。例如，第一节点可以连接3个分支，第一个分支表示标签数量＜k1，第二个分支表示k1≤标签数量≤k2，第三个分支表示标签数量＞k2，则第一节点会根据第一特征中的标签数量，将第一特征输入至三个分支中的一个分支。

[0149] 需要说明的另一点是，上述每个分支对应的取值范围仅是举例说明，每个分支对应的取值范围可以是开区间、闭区间或半开半闭区间中的任一种，本实施例对分支对应的取值范围是否包含端点值不做限定。

[0150] 每个叶子节点表示一种预测结果，将资源的第一特征输入至任一个回归树后，该第一特征会经过非叶子节点和分支，落入至多个叶子节点中的一个叶子节点，该叶子节点表示的预测结果即为这个回归树的预测结果。其中，如果机器学习模型是XGBoost算法训练得到的模型，第一个回归树的每个叶子节点可以表示第二特征的预测值，例如可以为销量的预测值。第二个回归树的每个叶子节点可以表示第一个回归树的误差，例如可以表示销量的预测值与销量的真实值之间的偏差。第三个回归树的每个叶子节点可以表示第一个回归树与第二个回归树的误差之和，第四个回归树的每个叶子节点可以表示第一个回归树、第二个回归树至第三个回归树的误差之和，依次类推，第n个回归树的叶子节点可以表示第二特征的预测误差，预测误差为第1个回归树至第n-1个回归树的误差之和，n为大于1的正整数。

[0151] 需要说明的一点是，模型训练时采用XGBoost算法，机器学习模型包括多个回归树仅是举例说明，在另一些可能的实施例中，也可以采用XGBoost算法之外的其他集成学习算法来进行模型训练，相应地，机器学习模型可以是其他集成学习模型，比如可以采用随机森林算法进行模型训练，则机器学习模型可以是随机森林。应理解，也可以采用集成学习算法之外的机器学习算法进行模型训练，比如采用神经网络算法、支持向量机算法、贝叶斯算法、逻辑回归算法进行模型训练，相应地，机器学习模型可以是深度神经网络、支持向量机、贝叶斯网络、逻辑回归模型等，本实施例对机器学习算法以及机器学习模型的具体类型均不做具体限定。

[0152] 503、计算机设备获取多个资源的第一特征。

[0153] 在一些可能的实施例中，可以从数据库中记录的各个资源中，选择发布时间点与当前时间点之间的时间间隔小于第二时长的资源，这些资源可以视为新发布的资源，可以

将这些资源作为待预测的资源，获取这些资源的第一特征。例如，如果第二时长为30天，资源为游戏，可以将最近30天发布的游戏作为待预测的游戏。

[0154] 504、计算机设备调用机器学习模型。

[0155] 505、计算机设备将多个资源的第一特征输入机器学习模型，输出每个资源的第二特征。

[0156] 对于多个资源中的每个资源，计算机设备可以将该资源的第一特征输入机器学习模型，通过机器学习模型，根据该资源的第一特征预测资源的第二特征。具体地，如果机器学习模型包括多个回归树，通过机器学习模型进行预测的过程可以包括下述步骤(1)至步骤(2)：

[0157] 步骤(1)对于多个资源中的每个资源，计算机设备将资源的第一特征输入多个回归树，得到每个回归树的预测结果。

[0158] 在一些可能的实施例中，对于多个回归树中的每个回归树，将该资源的第一特征输入该回归树后，可以通过该回归树，根据资源的第一特征得出预测结果。具体地，针对任一个回归树而言，该回归树计算预测结果的过程可以包括以下步骤(1.1)至步骤(1.2)。

[0159] 步骤(1.1)计算机设备根据资源的第一特征在回归树中经过的目标非叶子节点以及目标分支，确定资源对应的目标叶子节点。

[0160] 目标非叶子节点是指回归树的所有非叶子节点中第一特征经过的非叶子节点，目标非叶子节点可以是为第一节点、第二节点、第三节点、第四节点或第五节点中的任意一种。目标分支是指回归树的所有分支中第一特征经过的分支，目标分支可以是目标非叶子节点的分支。目标叶子节点是指回归树的所有叶子节点中第一特征到达的叶子节点。

[0161] 具体地，将资源的第一特征输入回归树后，第一个目标非叶子节点为回归树的根节点，根节点会对第一特征进行分类，判断第一特征落入第一个目标非叶子节点的哪个分支的取值范围，得出第一特征所属的目标分支，将该第一特征输入该目标分支连接的子节点，则该目标分支连接的子节点为第二个目标非叶子节点；第二个目标非叶子节点也会对第一特征进行分类，判断第一特征落入第二个目标非叶子节点的哪个分支的取值范围，将该第一特征输入该目标分支连接的子节点，则该目标分支连接的子节点为第三个目标非叶子节点，依次类推，资源的第一特征会经过一个或多个目标非叶子节点以及一个或多个目标分支，到达目标叶子节点。

[0162] 结合目标非叶子节点的具体类型，具体地，如果目标非叶子节点为第一节点，则该目标非叶子节点会根据标签数量，对第一特征进行分类，判断第一特征中的标签数量属于哪个分支对应的标签数量的取值范围，得出第一特征所属的目标分支。如果该目标非叶子节点为第二节点，则该目标非叶子节点会根据价格，对第一特征进行分类，判断第一特征中的价格属于哪个分支对应的价格的取值范围，得出第一特征所属的目标分支。如果该目标非叶子节点为第三节点，则该目标非叶子节点会根据发行时长，对第一特征进行分类，判断第一特征中的发行时长属于哪个分支对应的发行时长的取值范围，得出第一特征所属的目标分支。如果该目标非叶子节点为第四节点，则该目标非叶子节点会根据评测数量，对第一特征进行分类，判断第一特征中的评测数量属于哪个分支对应的评测数量的取值范围，得出第一特征所属的目标分支。如果该目标非叶子节点为第五节点，则该目标非叶子节点会根据访问数量，对第一特征进行分类，判断第一特征中的访问数量属于哪个分支对应的访

问数量的取值范围，得出第一特征所属的目标分支。

[0163] 作为示意，如果资源的第一特征为：标签数量(归一化后的)为0.3，价格为50，发行时长为10天，评测数量为500，目标非叶子节点为第二节点，该目标非叶子节点具有3个分支，第一个分支表示价格≤10，第二个分支表示10＜价格≤80，第三个分支表示价格＞80，那么则目标非叶子节点对第一特征进行分类时，由于10＜50≤80，则该资源的价格属于第二个分支对应的取值范围，因此会得出第二个分支为目标分支。

[0164] 步骤(1.2)计算机设备将目标叶子节点表示的预测结果，获取为回归树的预测结果。

[0165] 在一种可能的实现中，机器学习模型可以包括回归树、叶子节点与预测结果之间的对应关系，每个回归树的每个叶子节点可以对应一个预测结果。计算机设备可以从该对应关系中，获取该回归树以及该目标非叶子节点均对应的预测结果，得到目标叶子节点表示的预测结果。

[0166] 步骤(2)计算机设备获取多个回归树的预测结果之和，作为资源的第二特征。

[0167] 具体地，如果机器学习模型包括n个回归树，则n个回归树中的每个回归树可以得出一个预测结果，共计得到n个预测结果，可以将n个预测结果累加起来，将得到的和值作为资源的第二特征。

[0168] 其中，如果机器学习模型用于预测销量，且机器学习模型为XGBoost训练的模型，则多个回归树中第一个回归树的预测结果可以表示销量的预测值，第二个回归树的预测结果可以表示第一个回归树的误差，第三个回归树的预测结果可以表示第一个回归树和第二个回归树的误差之和，以此类推，第n个回归树的预测结果可以表示销量的预测误差，预测误差为第1个回归树至第n-1个回归树的误差之和，n为大于1的正整数。

[0169] 例如，第一个回归树得出1000，表示销量的预测值为1000；第二个回归树得出100，表示第一个回归树少预测了100，则将第一个回归树和第二个回归树的预测结果相加后，预测结果之和为1100，从而修正了第一个回归树的误差；第三个回归树得出-30，表示第一个回归树和第二个回归树总计多预测了30，则第一个回归树和第二个回归树的预测结果相加后，预测结果之和为1070，从而修正了第一个回归树和第二个回归树总计的误差；以此类推，每当加上一个回归树的预测结果，可以修正该回归树之前的所有回归树的总计的预测结果，从而让预测结果通过累加得以不断优化。

[0170] 同理地，如果机器学习模型用于预测在线用户数量，且机器学习模型为XGBoost算法训练的模型，则多个回归树中第一个回归树的预测结果可以表示在线用户数量的预测值，第二个回归树的预测结果可以表示第一个回归树的误差，第三个回归树的预测结果可以表示第一个回归树和第二个回归树的误差之和，以此类推，第n个回归树的预测结果可以表示在线用户数量的预测误差，预测误差为第1个回归树至第n-1个回归树的误差之和。

[0171] 506、计算机设备根据每个资源的第二特征，对多个资源进行排序。

[0172] 例如，可以根据每个资源的销量，对多个资源按照销量从大到小的顺序进行排序。又如，可以根据每个资源的在线用户数量，对多个资源按照在线用户数量从大到小的顺序进行排序。

[0173] 507、计算机设备将多个资源中排在前第一数目的资源输出为热门资源。

[0174] 第一数目可以根据多个资源的数量以及预设百分比确定，例如，如果多个资源共

计100个，预设百分比为5%，可以输出第二特征排在前5位的资源，从而将前5%的资源提供给用户。其中，输出热门资源的方式可以而不限于是：计算机设备可以将热门资源的标识发布至网页，用户通过访问网页可以浏览热门资源；或者，计算机设备可以在屏幕中显示热门资源的标识；或者，计算机设备可以将热门资源的标识存储至数据库；或者，计算机设备可以将热门资源的标识携带在邮件中，向用户发送邮件。

[0175] 在一些可能的实施例中，还可以从排在前第一数目的资源中，选择发行时长小于第三时长的资源，将选择的资源输出为潜力热门资源。其中，该第三时长可以为5天。作为示例，可以将排名前5%，且发行时间在5天以内的游戏，作为潜力爆款游戏，输出潜力爆款游戏的标识。

[0176] 在一些可能的实施例中，还可以将资源在排序结果中的位置，作为资源的排名，输出资源的排名。比如说，如果游戏A在排序结果中排在第一位，可以输出：游戏A将要成为爆款游戏中的第一名。

[0177] 本实施例提供了一种自动化的预测出热门资源的方法，根据资源的标签数量、价格、发行时长、评测数量、访问数量等特征，利用机器学习模型来预测出资源的热门程度，从而发现出哪个资源将要成为市场中的热门资源，节省了预测热门资源的时间，提高了预测热门资源的效率。此外，通过机器学习模型，从大量的样本资源中，学习出资源的特征与热门程度之间的规律，使得通过机器学习模型进行预测时，能够提高预测的准确性。

[0178] 上述实施例示出了采用机器学习的手段，来预测出热门资源的方法。本发明实施例中，也可以采用计算环比增长率的手段，来预测出热门资源，具体参见下述图6实施例。可选地，图6实施例可以作为图5实施例的补充。在一个示例性场景中，可以应用图5实施例，从市场中近期发行的新游戏中，发现爆款游戏。而对于发行时间较早的老游戏而言，受到主播直播或其他因素的影响，可能会重新受到玩家的跟风游玩，则可以通过下述图6实施例，发现这种将要成为爆款游戏的老游戏。需要说明的一点是，图6实施例的执行主体和图2实施例或者图5实施例的执行主体可以为同一计算机设备或不同的计算机设备，本实施例对此不做限定。

[0179] 图6是本发明实施例提供的一种热门资源的预测方法的流程图，参见图6，该实施例具体可以包括以下步骤：

[0180] 601、对于多个资源中的每个资源，计算机设备对资源在当前统计周期的第二特征以及上一个统计周期的第二特征进行对比，得到该资源的环比增长率。

[0181] 环比是一种时间对比的方式，表示连续2个统计周期内的量的变化比，例如，环比可以是本周和上周进行对比、本月和上个月进行对比等。环比增长率是指当前统计周期与上一个统计周期相比的增长率。其中，在计算环比增长率时，可以将评论数量作为第二特征，也可以将观看人数作为第二特征，本实施例对此不做限定。

[0182] 在一些可能的实施例中，可以采用下述公式来计算环比增长率：

[0183] 环比增长率＝(当前统计周期的第二特征-上一个统计周期的第二特征)/上一个统计周期的第二特征，/表示相除。

[0184] 在一些可能的实施例中，也可以采用下述公式来计算环比增长率：

[0185] 环比增长率＝(当前统计周期的第二特征的平均值-上一个统计周期的第二特征的平均值)/上一个统计周期的第二特征的平均值，/表示相除。

[0186] 考虑到如果某一天爬虫程序出现不稳定的情况，造成爬虫程序没有爬取到个别资源的数据，那么如果在这一天计算环比增长率，会由于该资源的数据缺失，导致遗漏该资源，没能发现该资源。而通过这种方式，使用第二特征在统计周期的均值来进行预测，由于整个统计周期的第二特征的均值相对于单独一天的第二特征来说，数据的可靠性更高，因此能够降低遗漏的概率，提高预测的准确性。

[0187] 示例性地，如果资源为游戏，统计周期的长度为7天，第二特征为评论数量，可以获取游戏最近7天的评论数量的均值1以及最近8天至最近14天的评论数量的均值2，对均值1和均值2进行对比，得到评论数量的周环比＝(均值1-均值2)/均值2。

[0188] 602、计算机设备根据每个资源的环比增长率，对多个资源进行排序。

[0189] 可以按照环比增长率从大到小的顺序，对多个资源进行排序。

[0190] 603、计算机设备将多个资源中排在前第二数目的资源输出为热门资源。

[0191] 第二数目可以根据多个资源的数量以及预设百分比确定，例如，如果多个资源共计100个，预设百分比为5％，可以输出环比增长率排在前5位的资源，从而将前5％的资源提供给用户。

[0192] 需要说明的一点是，图6实施例是针对资源在上一个统计周期具有数据的情况，在一些可能的实施例中，可以存在上一个统计周期中没有第二特征的资源，比如对于当前统计周期新近发布的多个资源而言，这些资源在上一个统计周期没有数据，那么可以通过下述步骤一至步骤四来从这些资源中预测出热门资源。

[0193] 步骤一、选取发布时间点位于当前统计周期的多个资源。

[0194] 步骤二、获取每个资源在当前统计周期的第二特征的平均值。

[0195] 步骤三、对多个资源按照平均值从大到小的顺序排序。

[0196] 步骤四、将多个资源中排在前第三数目的资源输出为热门资源。

[0197] 示意性地，以资源为游戏为例，可以筛选最近7天发行的每个新游戏，可以计算每个新游戏在最近7天的评论数量的均值，对每个新游戏按照平均值从大到小的顺序排序，选择评论数量的均值为前5％的游戏，作为爆款游戏。

[0198] 本实施例提供了另一种自动化的预测出热门资源的方法，设计了环比监控的机制，通过对同一资源在当前统计周期的第二特征以及上一个统计周期的第二特征进行对比，得到资源的环比增长率，根据每个资源的环比增长率的大小，发现出哪个资源将要成为市场中的热门资源，同样可以节省了预测热门资源的时间，提高了预测热门资源的效率。

[0199] 上述图5实施例以及图6实施例示出了预测热门资源的方法，本发明实施例中，还可以预测出资源类型的流行趋势，例如游戏玩法的流行趋势，具体参见图7实施例。需要说明的一点是，图7实施例的执行主体和图5实施例或图6实施例的执行主体可以为同一计算机设备或不同的计算机设备，本实施例对此不做限定。

[0200] 图7是本发明实施例提供的一种资源类型的流行趋势的预测方法的流程图，参见图7，该实施例具体可以包括以下步骤：

[0201] 701、计算机设备获取多个资源类型的第三特征以及第四特征。

[0202] 第三特征能够表示厂商对资源类型的偏好程度，例如，资源类型的第三特征越大，表示厂商越偏好开发该资源类型的资源。可选地，第三特征可以是资源类型的资源数量的增长率。具体地，如果厂商越偏好某一资源类型，则会开发出更多的该资源类型的资源，那

么市场上该资源类型的资源的数量会迅猛增长，使得资源数量的增长率会越高，因此通过资源数量的增长率，能够表示出厂商对该资源类型的偏好程度的增加趋势。以资源类型为游戏玩法为例，如果厂商看好模拟经营玩法的游戏，则会开发出很多模拟经营玩法的游戏，那么模拟经营玩法的游戏的数量的增长率会高于其他游戏玩法的游戏的数量的增长率，因此通过游戏数量的增长率，可以表示出厂商对模拟经营玩法的游戏的偏好程度的增加趋势。

[0203] 关于资源数量的增长率的获取方式，可以根据每个资源类型的资源在第一统计周期新增的数量和在第二统计周期新增的数量，获取每个资源类型的资源数量的增长率，作为每个资源类型的第三特征。

[0204] 其中，第一统计周期晚于第二统计周期。可选地，第一统计周期可以是当前统计周期，第二统计周期可以与当前统计周期之间相距预设时长，例如，如果统计周期为K天，预设时长为M天，第一统计周期可以是最近K天，第二统计周期可以是M天前的最近K天。其中，K和M为正整数。

[0205] 可选地，预设时长可以根据资源数量的趋势变化的反馈周期正相关。以资源为游戏为例，游戏的开发周期较长，导致游戏数量的趋势变化的反馈周期较长，则预设时长可以是180天，第一统计周期可以是近30天，第二统计周期可以是180天前的近30天。

[0206] 示意性地，可以采用下述公式计算资源数量的增长率：

[0207] RN＝(N1-N2)/N2；

[0208] 其中，N1表示资源类型的资源在第一统计周期新增的数量，N2表示资源类型的资源在第二统计周期新增的数量，RN表示资源数量的增长率，/表示相除。以资源类型为游戏玩法为例，N1具体可以表示最近30天新增的游戏个数，N2具体可以表示180天前该游戏玩法最近30天新增的游戏个数，RN具体可以表示游戏玩法的游戏增长率。

[0209] 需要说明的一点是，第二统计周期与当前统计周期之间相距预设时长仅是举例说明，第二统计周期也可以是当前统计周期的上一个统计周期，例如，第一统计周期可以是近30天，第二统计周期可以是近60天至近30天，本实施例对此不做限定。

[0210] 需要说明的另一点是，资源数量的增长率仅是第三特征的举例说明，第三特征也可以是资源数量的增长率之外的其他数据，任意能够表示厂商对资源类型的偏好程度均可以作为第三特征，比如厂商对资源类型的资源的研发费用、广告营销预算等，本实施例对第三特征并不做限定。

[0211] 第四特征能够表示用户对资源类型的偏好程度，例如，资源程度的第四特征越大，表示用户越偏好使用该资源类型的资源。可选地，第四特征可以是评论数量的增长率、在线用户数量的增长率、销量的增长率中的任一项或多项。以评论数量的增长率为例，如果用户越偏好某一资源类型，通常会对该资源类型的资源发表更多的评论，那么该资源类型的资源的评论数量会迅猛增长，使得评论数量的增长率会越高，因此通过评论数量的增长率，能够表示出用户对该资源类型的偏好程度的增加趋势。以资源类型为游戏玩法为例，如果用户看好养成玩法的游戏，则会对养成玩法的很多游戏发表评论，那么养成玩法的游戏的数量的增长率会高于其他游戏玩法的游戏的数量的增长率，因此通过游戏数量的增长率，可以表示出用户对养成玩法的游戏的偏好程度的增加趋势。

[0212] 评论数量的增长率的获取方式可以包括：计算机设备根据每个资源类型的资源在

第三统计周期的评论数量和在第四统计周期的评论数量，获取每个资源类型的评论数量的增长率。其中，第三统计周期晚于第四统计周期。可选地，第三统计周期可以是当前统计周期，第四统计周期可以与当前统计周期之间相距预设时长。

[0213] 其中，预设时长可以根据评论数量的趋势变化的反馈周期正相关。以资源为游戏为例，游戏的评论数量的变化速度较快，导致评论数量的趋势变化的反馈周期较段，则预设时长可以是30天，第三统计周期可以是当天，第四统计周期可以是30天前的当天。

[0214] 示意性地，可以采用下述公式计算评论数量的增长率：

[0215] RC＝(C1-C2)/C2；

[0216] 其中，C1表示资源类型在第三统计周期的评论数量，C2表示资源类型在第四统计周期的评论数量，RC表示评论数量的增长率，/表示相除。以资源类型为游戏玩法为例，C1具体可以表示游戏玩法当天的评论数量，C2具体可以表示30天前该游戏玩法的评论数量，RC具体可以表示游戏玩法的评论数量的增长率。

[0217] 在线用户数量的增长率的获取方式可以包括：计算机设备根据每个资源类型的资源在第五统计周期的在线用户数量和在第六统计周期的在线用户数量，获取每个资源类型的在线用户数量的增长率，作为每个资源类型的第四特征，第五统计周期晚于第六统计周期。其中，在线用户数量的增长率的获取方式与评论数量的增长率的获取方式同理，具体细节不做赘述。

[0218] 销量的增长率的获取方式可以包括：计算机设备根据每个资源类型的资源在第七统计周期的销量和在第八统计周期的销量，获取每个资源类型的销量的增长率，作为每个资源类型的第四特征，第七统计周期晚于第八统计周期。其中，在线用户数量的增长率的获取方式与评论数量的增长率的获取方式同理，具体细节不做赘述。其中，销量的增长率的获取方式与评论数量的增长率的获取方式同理，具体细节不做赘述。

[0219] 702、计算机设备根据每个资源类型的第三特征以及第四特征，生成目标图像。

[0220] 目标图像的第一坐标轴表示第三特征，第一坐标轴可以反映厂商对每个资源类型的偏好程度，即厂商角度的流行的资源类型。第一坐标轴的正方向可以为第三特征从小到大的方向，那么某个资源类型的第三特征越大，则该资源类型在第一坐标轴的坐标越大，表明厂商越偏好该资源类型，会更多地开发该资源类型的资源。

[0221] 目标图像的第二坐标轴表示第四特征。第二坐标轴可以反映用户对每个资源类型的偏好程度，即用户角度的流行的资源类型。第二坐标轴的正方向可以为第四特征从小到大的方向，例如，某个资源类型的第三特征越大，则该资源类型在第二坐标轴的坐标越大，表明用户越偏好该资源类型，会更多地使用该资源类型的资源。

[0222] 第一坐标轴和第二坐标轴可以相互垂直。例如，第一坐标轴可以是横轴，第二坐标轴可以是纵轴。第一坐标轴和第二坐标轴可以将目标图像划分为四个象限，任一资源类型可以分布在四个象限中的一个象限中。其中，目标图像的第一象限可以记为红海，红海是经济学的术语，通常是指已知的、竞争相当激烈的市场，本实施例中，目标图像中“红海”中的资源类型是厂商和用户均偏好的资源类型。第二象限可以记为蓝海，蓝海是经济学的术语，是指未知的、尚待开发的市场，本实施例中，目标图像中“蓝海”中的资源类型是用户偏好而厂商尚未跟上的资源类型。第三象限可以记为未知，目标图像中“未知”中的资源类型是厂商和用户均不偏好的资源类型。第四象限可以记为探索，目标图像中“探索”中的资源类型

是厂商偏好且用户尚未增长的资源类型。

[0223] 可选地，可以将每个资源类型的第三特征的数据格式从数值转换为角度，将每个资源类型的第四特征的数据格式也从数值转换为角度，根据角度格式的第三特征以及第四特征，生成目标图像。通过这种方式，由于数值的取值范围可以是负无穷到正无穷，跨度较大，而角度的取值范围相对较小，因此可以保证目标图像的显示效果更加直观。

[0224] 示意性地，可以通过以下公式，将第三特征以及第四特征的数据格式从数值转换为角度：

[0225] X＝Atan(RN)*180/π；

[0226] Y＝Atan(RC)*180/π；

[0227] 其中，X表示角度格式的第三特征，RN表示数值格式的第三特征，Y表示角度格式的第四特征，RN表示数值格式的第四特征。

[0228] 以资源类型为游戏玩法为例，作为示意，参见图8，对于任一个游戏玩法而言，可以根据该游戏玩法的游戏数量增长率RN，该游戏玩法的评论数量增长率RC，得出该游戏玩法的点坐标(Atan(RN)*180/π，Atan(RC)*180/π)，该点坐标为角度对的形式，可以在坐标系中，标记该游戏玩法的位置和名称，从而绘制得到图8所示的目标图像。

[0229] 图8所示的目标图像可以记为游戏玩法的红蓝海图，通过图8的横轴，可以直观的展示出厂商维度的玩法流行趋势，通过图8的纵轴，可以直观的展示出玩家维度的玩法流行趋势，其中红海分布的各个游戏玩法，例如剧情、恋爱、模拟经营等均是当前热门的玩法，蓝海分布的各个游戏玩法，例如放置、换装、换装等是有潜力的玩法，游戏开发者可以组合红海和蓝海这两个区域内的不同游戏玩法，将组合结果作为新的游戏玩法，从而来实现游戏玩法的微创新。比如将剧情和收集这两种玩法组合在一起，作为一种新创的游戏玩法。

[0230] 703、计算机设备输出目标图像。

[0231] 计算机设备可以将目标图像发布至网页，开发者通过访问网页可以浏览目标图像。或者，计算机设备可以在屏幕中显示目标图像；或者，计算机设备可以将目标图像存储至数据库，本实施例对目标图像的输出方式不做限定。

[0232] 本实施例提供了一种自动化的预测出资源类型的流行趋势的方法，分别考虑了厂商对每个资源类型的偏好程度以及用户对每个资源类型的偏好程度，以厂商维度和用户维度生成的目标图像，能够直观的展示出流行的资源类型，从而帮助开发者对资源类型进行开发和创新。

[0233] 上述实施例示出了资源类型的流行趋势的预测方法。本发明实施例中，还提供了自动整理汇总各个榜单的功能，具体参见下述图9实施例。需要说明的一点是，图9实施例的执行主体和上述各个实施例的执行主体可以为同一计算机设备或不同的计算机设备，本实施例对此不做限定。

[0234] 图9是本发明实施例提供的一种榜单整合方法的流程图，参见图9，该实施例具体可以包括以下步骤：

[0235] 901、计算机设备获取多个榜单。

[0236] 例如，可以从图3实施例中存储数据的数据库中，读取多个榜单。

[0237] 902、计算机设备为多个榜单中的每个榜单分配对应的权重。

[0238] 例如，热门榜的权重可以是1000，热销榜的权重可以是900，预约榜的权重可以是

100。

[0239] 903、计算机设备根据每个榜单的权重以及每个榜单中资源的排名，获取每个资源的权重。

[0240] 资源所属的榜单的权重越大，资源的权重越大。或者，资源在榜单的排名越靠前，资源的权重越大。示例性地，如果资源所属的榜单的权重为a，资源在榜单的排名为b，资源的权重可以＝a+(100-b)。其中，a为正数，b为正整数。

[0241] 904、计算机设备根据每个资源的权重，对多个资源进行排序。

[0242] 例如，可以对每个资源按照权重从大到小的顺序的排序。

[0243] 905、计算机设备根据排序结果，生成目标榜单。

[0244] 目标榜单为根据多个榜单整合得到的榜单，可以作为热门资源榜单，例如可以作为爆款游戏榜单。目标榜单可以包括多个资源，每个资源可以来自于各个网站发布的一个或多个榜单。其中，目标榜单中的第i名资源即为排序结果中排在第i位的资源。i为正整数。

[0245] 可选地，可以在目标榜单中，写入每个资源所属的榜单的标识，通过榜单的标识可以指明资源来自于哪个榜单。例如，可以在游戏A旁边写入“预约榜”，从而指明游戏A是预约榜中的游戏。另外，如果目标榜单中的某个资源属于多个榜单，可以写入该资源所属的多个榜单的标识，不同榜单的标识之间可以通过分隔符连接，该分隔符可以为逗号。例如，可以在游戏B旁边写入“热销榜，优惠榜”，从而指明游戏B是热销榜和优惠榜的游戏。

[0246] 906、计算机设备输出目标榜单。

[0247] 相关技术中，需要依赖人工操作，手动地浏览各个榜单，再进行整理汇总，操作十分繁琐，费时费力。

[0248] 而本实施例提供了一种自动化的整合榜单的方法，能够根据每个榜单的权重以及榜单中资源的排名，整合出一份目标榜单，该目标榜单综合了各个榜单的数据，可参考性强，能够准确全面的反映资源的热门程度。此外，免去了人工整理汇总的繁琐流程，所以提高了榜单整合的效率。

[0249] 上述所有可选技术方案，可以采用任意结合形成本公开的可选实施例，在此不再一一赘述。

[0250] 图10是本发明实施例提供的一种热门资源的预测装置的框图，参见图10，该装置包括：获取模块1001、调用模块1002、模型运算模块1003、排序模块1004以及输出模块1005。

[0251] 获取模块1001，用于获取多个资源的第一特征，该第一特征包括资源的标签数量、价格、发行时长、评测数量、访问数量中的一项或多项；

[0252] 调用模块1002，用于调用机器学习模型，该机器学习模型用于根据资源的第一特征对资源的第二特征进行预测，该第二特征能够表示资源的热门程度；

[0253] 模型运算模块1003，用于将该多个资源的第一特征输入该机器学习模型，输出每个资源的第二特征；

[0254] 排序模块1004，用于根据每个资源的第二特征，对该多个资源进行排序；

[0255] 输出模块1005，用于将该多个资源中排在前第一数目的资源输出为热门资源。

[0256] 可选地，该机器学习模型包括多个回归树，该模型运算模块1003，包括：

[0257] 回归树运算子模块，用于对于该多个资源中的每个资源，将该资源的第一特征输入该多个回归树，得到每个回归树的预测结果；

[0258] 求和子模块，用于获取该多个回归树的预测结果之和，作为该资源的第二特征。

[0259] 可选地，每个回归树包括一个或多个非叶子节点、多个分支以及多个叶子节点，该非叶子节点为第一节点、第二节点、第三节点、第四节点或第五节点；

[0260] 该第一节点用于根据标签数量进行分类，该第一节点的每个分支表示标签数量的一个取值范围；

[0261] 该第二节点用于根据价格进行分类，该第二节点的每个分支表示价格的一个取值范围；

[0262] 该第三节点用于根据发行时长进行分类，该第三节点的每个分支表示发行时长的一个取值范围；

[0263] 该第四节点用于根据评测数量进行分类，该第四节点的每个分支表示评测数量的一个取值范围；

[0264] 该第五节点用于根据访问数量进行分类，该第五节点的每个分支表示访问数量的一个取值范围；

[0265] 该回归树运算子模块，用于：对于该多个回归树中的每个回归树，根据该资源的第一特征在该回归树中经过的目标非叶子节点以及目标分支，确定该资源对应的目标叶子节点；将该目标叶子节点表示的预测结果，获取为该回归树的预测结果。

[0266] 可选地，该第二特征为该资源的销量，该多个回归树中第一个回归树的预测结果表示销量的预测值，该多个回归树中第n个回归树的预测结果表示销量的预测误差，该预测误差为第1个回归树至第n-1个回归树的误差之和，该n为大于1的正整数；或者，

[0267] 该第二特征为该资源的在线用户数量，该多个回归树中第一个回归树的预测结果表示在线用户数量的预测值，该多个回归树中第n个回归树的预测结果表示在线用户数量的预测误差。

[0268] 可选地，该获取模块1001，还用于获取多个资源类型的第三特征以及第四特征，该第三特征能够表示厂商对资源类型的偏好程度，该第四特征能够表示用户对资源类型的偏好程度；

[0269] 该装置还包括：生成模块，用于根据每个资源类型的第三特征以及第四特征，生成目标图像，该目标图像的第一坐标轴表示第三特征，该目标图像的第二坐标轴表示第四特征；

[0270] 该输出模块1005，还用于输出该目标图像。

[0271] 可选地，该获取模块1001，用于执行下述至少一项：

[0272] 根据每个资源类型的资源在第一统计周期新增的数量和在第二统计周期新增的数量，获取每个资源类型的资源数量的增长率，作为每个资源类型的第三特征，该第一统计周期晚于该第二统计周期；

[0273] 根据每个资源类型的资源在第三统计周期的评论数量和在第四统计周期的评论数量，获取每个资源类型的评论数量的增长率，作为每个资源类型的第四特征，该第三统计周期晚于该第四统计周期；

[0274] 根据每个资源类型的资源在第五统计周期的在线用户数量和在第六统计周期的在线用户数量，获取每个资源类型的在线用户数量的增长率，作为每个资源类型的第四特征，该第五统计周期晚于该第六统计周期；

[0275] 根据每个资源类型的资源在第七统计周期的销量和在第八统计周期的销量，获取每个资源类型的销量的增长率，作为每个资源类型的第四特征，该第七统计周期晚于该第八统计周期。

[0276] 可选地，该装置还包括：

[0277] 对比模块，用于对于该多个资源中的每个资源，对该资源在当前统计周期的第二特征以及上一个统计周期的第二特征进行对比，得到该资源的环比增长率；

[0278] 该排序模块1004，还用于根据每个资源的环比增长率，对该多个资源进行排序；

[0279] 该输出模块1005，还用于将该多个资源中排在前第二数目的资源输出为热门资源。

[0280] 可选地，该获取模块1001，还用于获取多个榜单；

[0281] 该装置还包括：分配模块，用于为该多个榜单中的每个榜单分配对应的权重；

[0282] 该获取模块1001，还用于根据每个榜单的权重以及每个榜单中资源的排名，获取每个资源的权重，榜单的权重越大或排名越靠前，资源的权重越大；

[0283] 排序模块1004，用于根据每个资源的权重，对该多个资源进行排序；

[0284] 生成模块，用于根据排序结果，生成目标榜单。

[0285] 可选地，该获取模块1001，包括：

[0286] 爬取子模块，用于爬取目标网站的网页内容，该目标网站包括资源发布网站、视频网站、直播网站、资讯网站中的一项或多项；

[0287] 解析子模块，用于解析该网页内容，得到多个资源的数据；

[0288] 获取子模块，用于根据该多个资源的数据，获取该多个资源的第一特征。

[0289] 需要说明的是：上述实施例提供的热门资源的预测装置在预测热门资源时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将热门资源的预测装置的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的热门资源的预测装置与热门资源的预测方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

[0290] 上述方法实施例中的计算机设备可以实现为终端或服务器，例如，图11示出了本发明一个示例性实施例提供的终端1100的结构框图。该终端1100可以是：智能手机、平板电脑、MP3播放器(Moving Picture Experts Group Audio Layer III，动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV，动态影像专家压缩标准音频层面4)播放器、笔记本电脑或台式电脑。终端1100还可能被称为用户设备、便携式终端、膝上型终端、台式终端等其他名称。

[0291] 通常，终端1100包括有：处理器1101和存储器1102。

[0292] 处理器1101可以包括一个或多个处理核心，比如4核心处理器、8核心处理器等。处理器1101可以采用DSP(Digital Signal Processing，数字信号处理)、FPGA(Field－Programmable Gate Array，现场可编程门阵列)、PLA(Programmable Logic Array，可编程逻辑阵列)中的至少一种硬件形式来实现。处理器1101也可以包括主处理器和协处理器，主处理器是用于对在唤醒状态下的数据进行处理的处理器，也称CPU(Central Processing Unit，中央处理器)；协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中，处理器1101可以在集成有GPU(Graphics Processing Unit，图像处理器)，

GPU用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中，处理器1101还可以包括AI(Artificial Intelligence，人工智能)处理器，该AI处理器用于处理有关机器学习的计算操作。

[0293] 存储器1102可以包括一个或多个计算机可读存储介质，该计算机可读存储介质可以是非暂态的。存储器1102还可包括高速随机存取存储器，以及非易失性存储器，比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中，存储器1102中的非暂态的计算机可读存储介质用于存储至少一个指令，该至少一个指令用于被处理器1101所执行以实现本发明中方法实施例提供的热门资源的预测方法。

[0294] 在一些实施例中，终端1100还可选包括有：外围设备接口1103和至少一个外围设备。处理器1101、存储器1102和外围设备接口1103之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口1103相连。具体地，外围设备包括：射频电路1104、触摸显示屏1105、摄像头组件1106、音频电路1107和电源1109中的至少一种。

[0295] 外围设备接口1103可被用于将I/O(Input/Output，输入/输出)相关的至少一个外围设备连接到处理器1101和存储器1102。在一些实施例中，处理器1101、存储器1102和外围设备接口1103被集成在同一芯片或电路板上；在一些其他实施例中，处理器1101、存储器1102和外围设备接口1103中的任意一个或两个可以在单独的芯片或电路板上实现，本实施例对此不加以限定。

[0296] 射频电路1104用于接收和发射RF(Radio Frequency，射频)信号，也称电磁信号。射频电路1104通过电磁信号与通信网络以及其他通信设备进行通信。射频电路1104将电信号转换为电磁信号进行发送，或者，将接收到的电磁信号转换为电信号。可选地，射频电路1104包括：天线系统、RF收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、编解码芯片组、用户身份模块卡等等。射频电路1104可以通过至少一种无线通信协议来与其它终端进行通信。该无线通信协议包括但不限于：万维网、城域网、内联网、各代移动通信网络(2G、3G、4G及5G)、无线局域网和/或WiFi(Wireless Fidelity，无线保真)网络。在一些实施例中，射频电路1104还可以包括NFC(Near Field Communication，近距离无线通信)有关的电路，本发明对此不加以限定。

[0297] 显示屏1105用于显示UI(User Interface，用户界面)。该UI可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏1105是触摸显示屏时，显示屏1105还具有采集在显示屏1105的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器1101进行处理。此时，显示屏1105还可以用于提供虚拟按钮和/或虚拟键盘，也称软按钮和/或软键盘。在一些实施例中，显示屏1105可以为一个，设置终端1100的前面板；在另一些实施例中，显示屏1105可以为至少两个，分别设置在终端1100的不同表面或呈折叠设计；在再一些实施例中，显示屏1105可以是柔性显示屏，设置在终端1100的弯曲表面上或折叠面上。甚至，显示屏1105还可以设置成非矩形的不规则图形，也即异形屏。显示屏1105可以采用LCD(Liquid Crystal Display，液晶显示屏)、OLED(Organic Light-Emitting Diode，有机发光二极管)等材质制备。

[0298] 摄像头组件1106用于采集图像或视频。可选地，摄像头组件1106包括前置摄像头和后置摄像头。通常，前置摄像头设置在终端的前面板，后置摄像头设置在终端的背面。在

一些实施例中，后置摄像头为至少两个，分别为主摄像头、景深摄像头、广角摄像头、长焦摄像头中的任意一种，以实现主摄像头和景深摄像头融合实现背景虚化功能、主摄像头和广角摄像头融合实现全景拍摄以及VR(Virtual Reality，虚拟现实)拍摄功能或者其它融合拍摄功能。在一些实施例中，摄像头组件1106还可以包括闪光灯。闪光灯可以是单色温闪光灯，也可以是双色温闪光灯。双色温闪光灯是指暖光闪光灯和冷光闪光灯的组合，可以用于不同色温下的光线补偿。

[0299] 音频电路1107可以包括麦克风和扬声器。麦克风用于采集用户及环境的声波，并将声波转换为电信号输入至处理器1101进行处理，或者输入至射频电路1104以实现语音通信。出于立体声采集或降噪的目的，麦克风可以为多个，分别设置在终端1100的不同部位。麦克风还可以是阵列麦克风或全向采集型麦克风。扬声器则用于将来自处理器1101或射频电路1104的电信号转换为声波。扬声器可以是传统的薄膜扬声器，也可以是压电陶瓷扬声器。当扬声器是压电陶瓷扬声器时，不仅可以将电信号转换为人类可听见的声波，也可以将电信号转换为人类听不见的声波以进行测距等用途。在一些实施例中，音频电路1107还可以包括耳机插孔。

[0300] 电源1109用于为终端1100中的各个组件进行供电。电源1109可以是交流电、直流电、一次性电池或可充电电池。当电源1109包括可充电电池时，该可充电电池可以是有线充电电池或无线充电电池。有线充电电池是通过有线线路充电的电池，无线充电电池是通过无线线圈充电的电池。该可充电电池还可以用于支持快充技术。

[0301] 在一些实施例中，终端1100还包括有一个或多个传感器1110。该一个或多个传感器1110包括但不限于：加速度传感器1111、陀螺仪传感器1112、压力传感器1113、光学传感器1115以及接近传感器1116。

[0302] 加速度传感器1111可以检测以终端1100建立的坐标系的三个坐标轴上的加速度大小。比如，加速度传感器1111可以用于检测重力加速度在三个坐标轴上的分量。处理器1101可以根据加速度传感器1111采集的重力加速度信号，控制触摸显示屏1105以横向视图或纵向视图进行用户界面的显示。加速度传感器1111还可以用于游戏或者用户的运动数据的采集。

[0303] 陀螺仪传感器1112可以检测终端1100的机体方向及转动角度，陀螺仪传感器1112可以与加速度传感器1111协同采集用户对终端1100的3D动作。处理器1101根据陀螺仪传感器1112采集的数据，可以实现如下功能：动作感应(比如根据用户的倾斜操作来改变UI)、拍摄时的图像稳定、游戏控制以及惯性导航。

[0304] 压力传感器1113可以设置在终端1100的侧边框和/或触摸显示屏1105的下层。当压力传感器1113设置在终端1100的侧边框时，可以检测用户对终端1100的握持信号，由处理器1101根据压力传感器1113采集的握持信号进行左右手识别或快捷操作。当压力传感器1113设置在触摸显示屏1105的下层时，由处理器1101根据用户对触摸显示屏1105的压力操作，实现对UI界面上的可操作性控件进行控制。可操作性控件包括按钮控件、滚动条控件、图标控件、菜单控件中的至少一种。

[0305] 光学传感器1115用于采集环境光强度。在一个实施例中，处理器1101可以根据光学传感器1115采集的环境光强度，控制触摸显示屏1105的显示亮度。具体地，当环境光强度较高时，调高触摸显示屏1105的显示亮度；当环境光强度较低时，调低触摸显示屏1105的显

示亮度。在另一个实施例中，处理器1101还可以根据光学传感器1115采集的环境光强度，动态调整摄像头组件1106的拍摄参数。

[0306] 接近传感器1116，也称距离传感器，通常设置在终端1100的前面板。接近传感器1116用于采集用户与终端1100的正面之间的距离。在一个实施例中，当接近传感器1116检测到用户与终端1100的正面之间的距离逐渐变小时，由处理器1101控制触摸显示屏1105从亮屏状态切换为息屏状态；当接近传感器1116检测到用户与终端1100的正面之间的距离逐渐变大时，由处理器1101控制触摸显示屏1105从息屏状态切换为亮屏状态。

[0307] 本领域技术人员可以理解，图11中示出的结构并不构成对终端1100的限定，可以包括比图示更多或更少的组件，或者组合某些组件，或者采用不同的组件布置。

[0308] 上述方法实施例中的服务器可以实现为终端或服务器，例如，图12是本发明实施例提供的一种服务器的结构示意图，该服务器1200可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上处理器(central processing units，CPU)1201和一个或一个以上的存储器1202，其中，该存储器1202中存储有至少一条指令，该至少一条指令由该处理器1201加载并执行以实现上述各个方法实施例提供的热门资源的预测方法。当然，该服务器还可以具有有线或无线网络接口以及输入输出接口等部件，以便进行输入输出，该服务器还可以包括其他用于实现设备功能的部件，在此不做赘述。

[0309] 在示例性实施例中，还提供了一种计算机可读存储介质，例如包括指令的存储器，上述指令可由处理器执行以完成上述实施例中的热门资源的预测方法。例如，计算机可读存储介质可以是只读存储器(Read-Only Memory，简称：ROM)、随机存取存储器(Random Access Memory，简称：RAM)、只读光盘(Compact Disc Read-Only Memory，简称：CD-ROM)、磁带、软盘和光数据存储设备等。

[0310] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，该的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0311] 以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

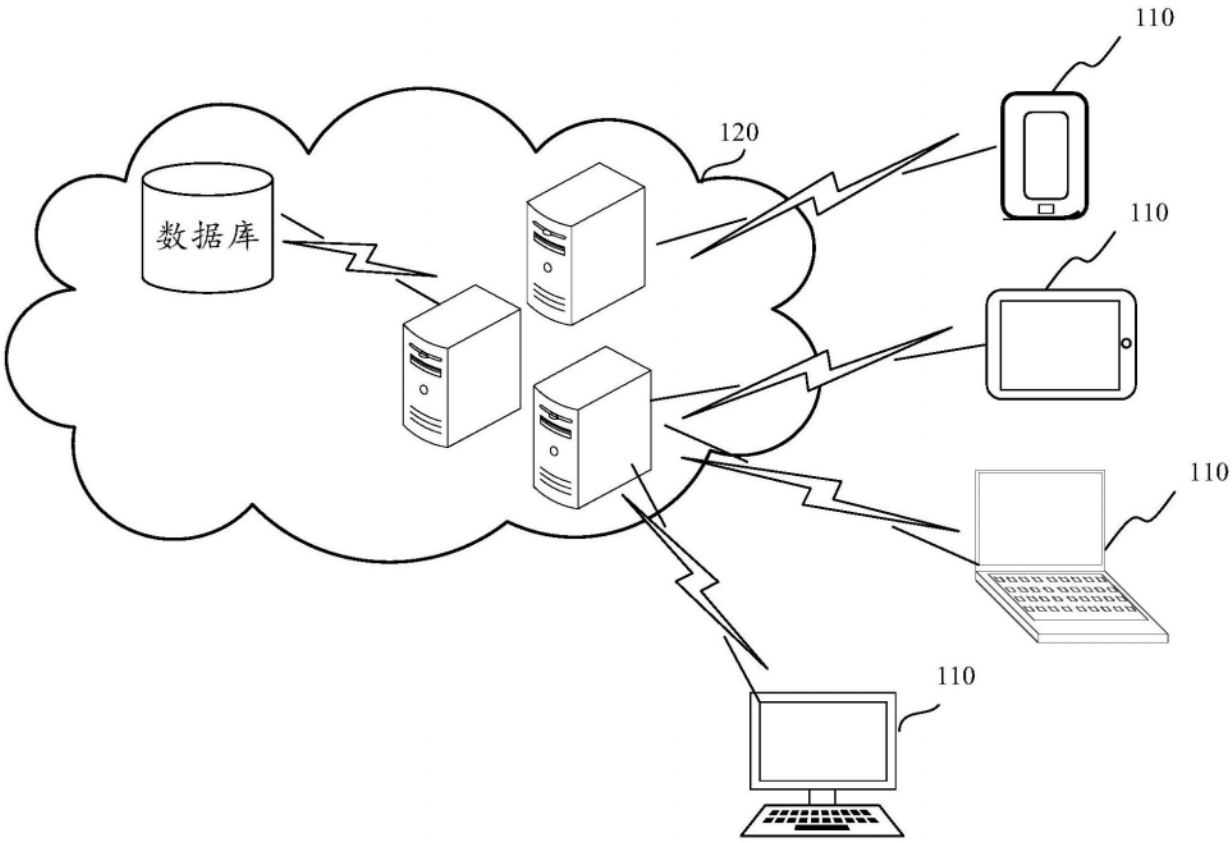


图1

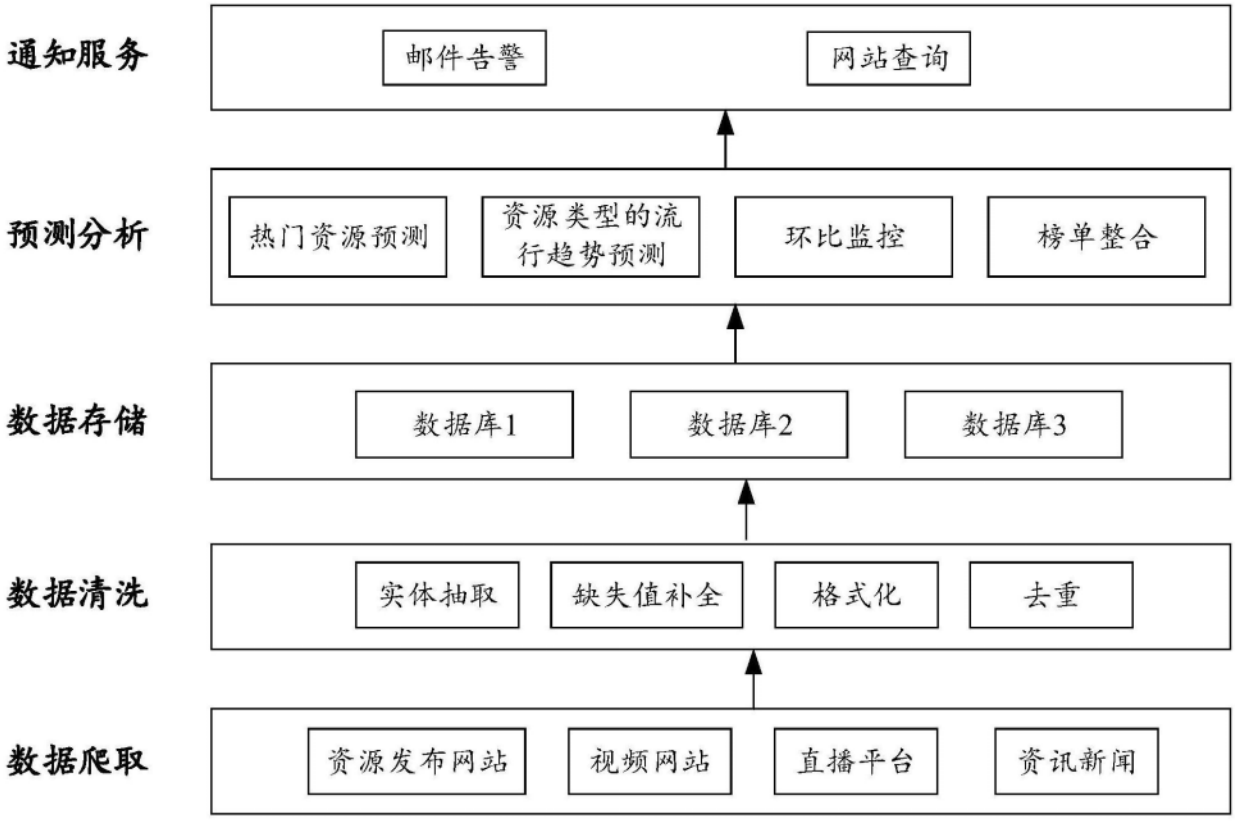

通知服务
邮件告警
网站查询
预测分析
热门资源预测
资源类型的流行趋势预测
环比监控
榜单整合
数据存储
数据库1
数据库2
数据库3
数据清洗
实体抽取
缺失值补全
格式化
去重
数据爬取
资源发布网站
视频网站
直播平台
资讯新闻

图2

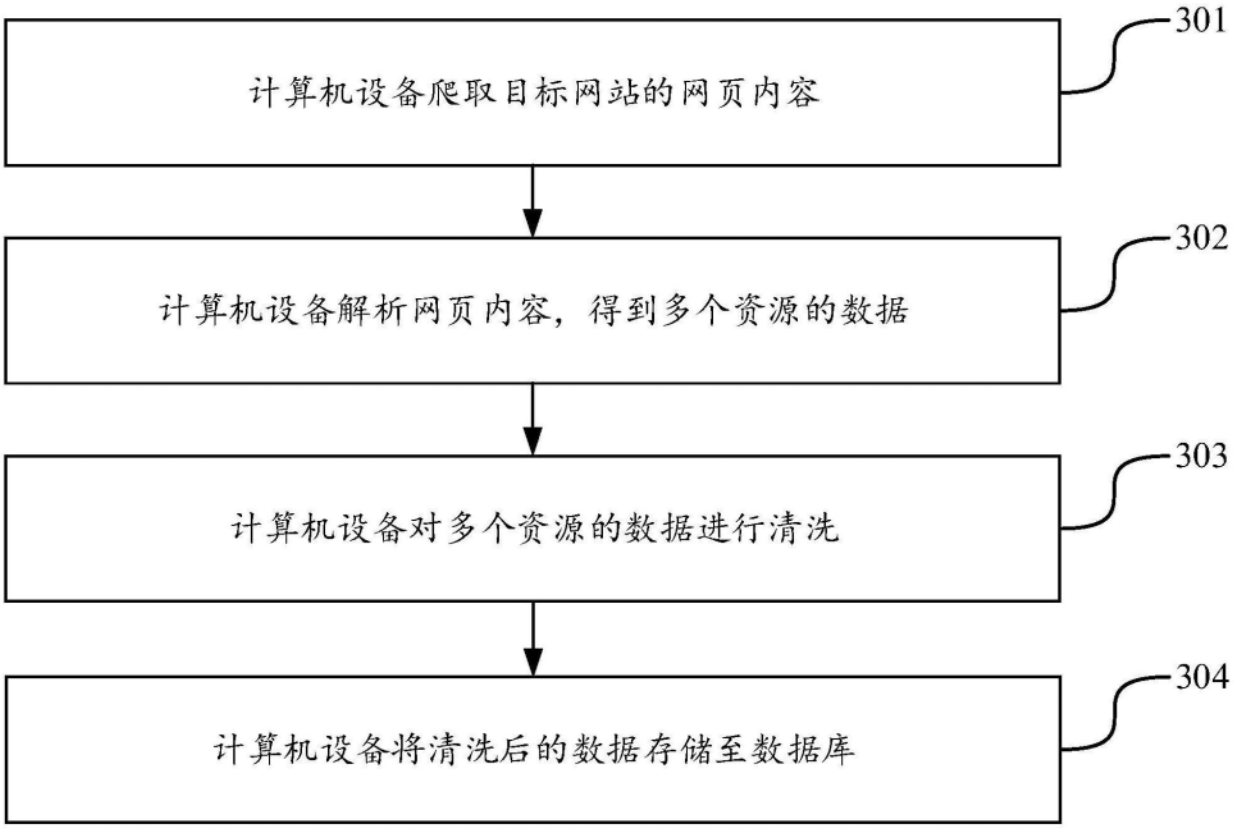

301
计算机设备爬取目标网站的网页内容
302
计算机设备解析网页内容，得到多个资源的数据
303
计算机设备对多个资源的数据进行清洗
304
计算机设备将清洗后的数据存储至数据库

图3

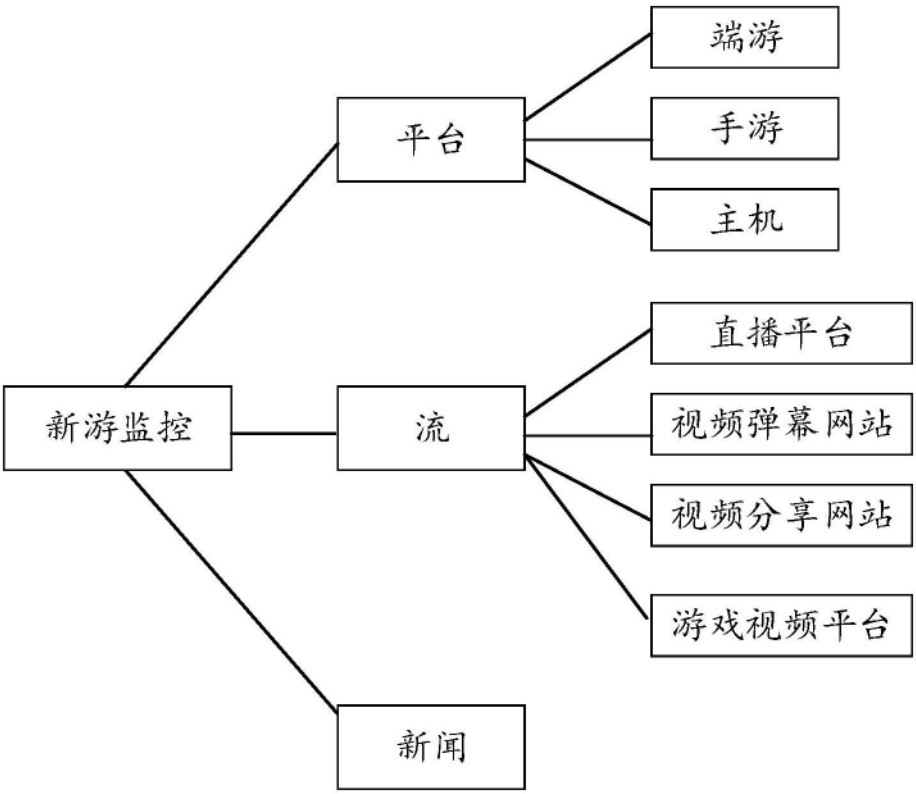


图4

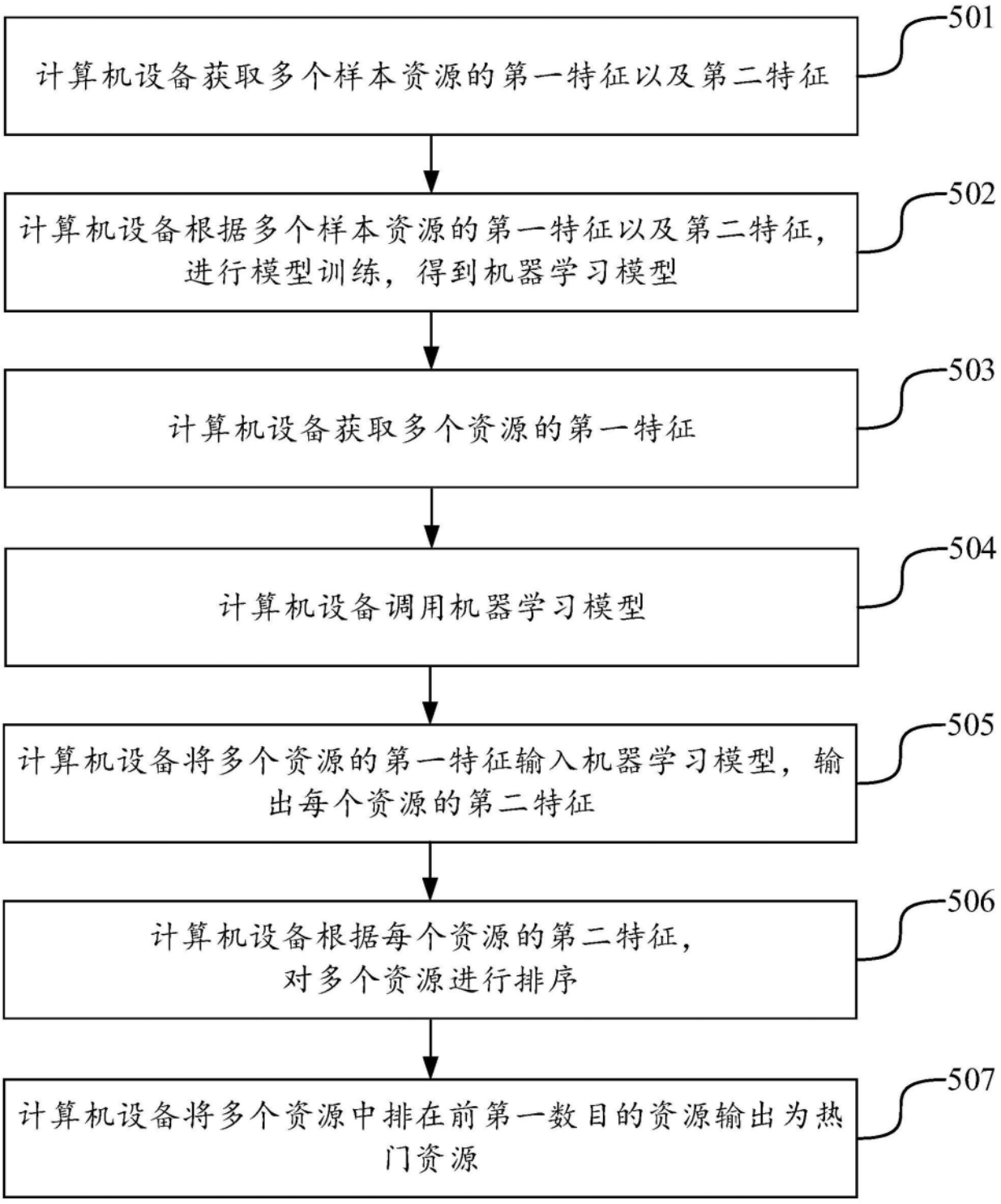

501
计算机设备获取多个样本资源的第一特征以及第二特征
502
计算机设备根据多个样本资源的第一特征以及第二特征，进行模型训练，得到机器学习模型
503
计算机设备获取多个资源的第一特征
504
计算机设备调用机器学习模型
505
计算机设备将多个资源的第一特征输入机器学习模型，输出每个资源的第二特征
506
计算机设备根据每个资源的第二特征，对多个资源进行排序
507
计算机设备将多个资源中排在前第一数目的资源输出为热门资源

图5

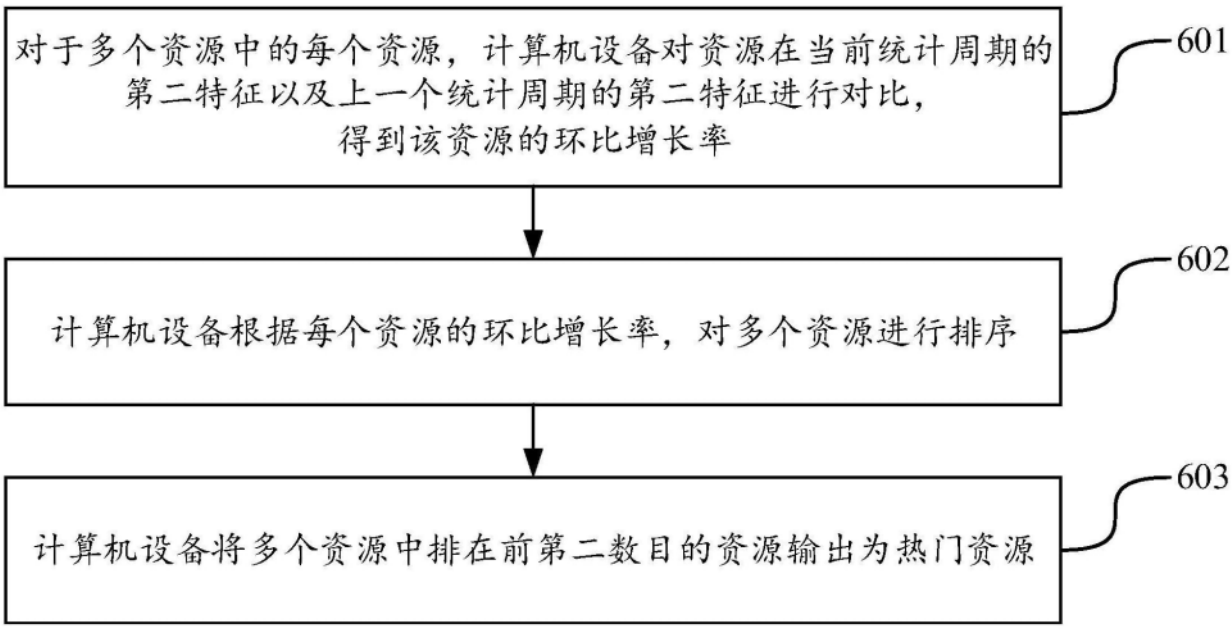

对于多个资源中的每个资源，计算机设备对资源在当前统计周期的第二特征以及上一个统计周期的第二特征进行对比，得到该资源的环比增长率
601
计算机设备根据每个资源的环比增长率，对多个资源进行排序
602
计算机设备将多个资源中排在前第二数目的资源输出为热门资源
603

图6

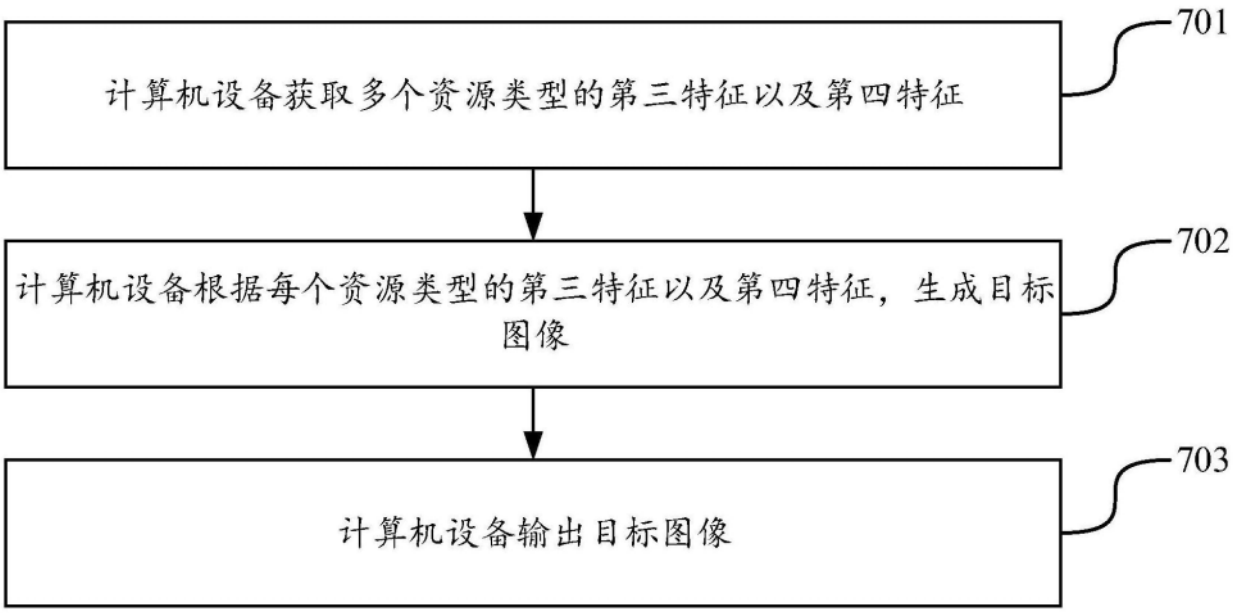

计算机设备获取多个资源类型的第三特征以及第四特征
701
计算机设备根据每个资源类型的第三特征以及第四特征，生成目标图像
702
计算机设备输出目标图像
703

图7

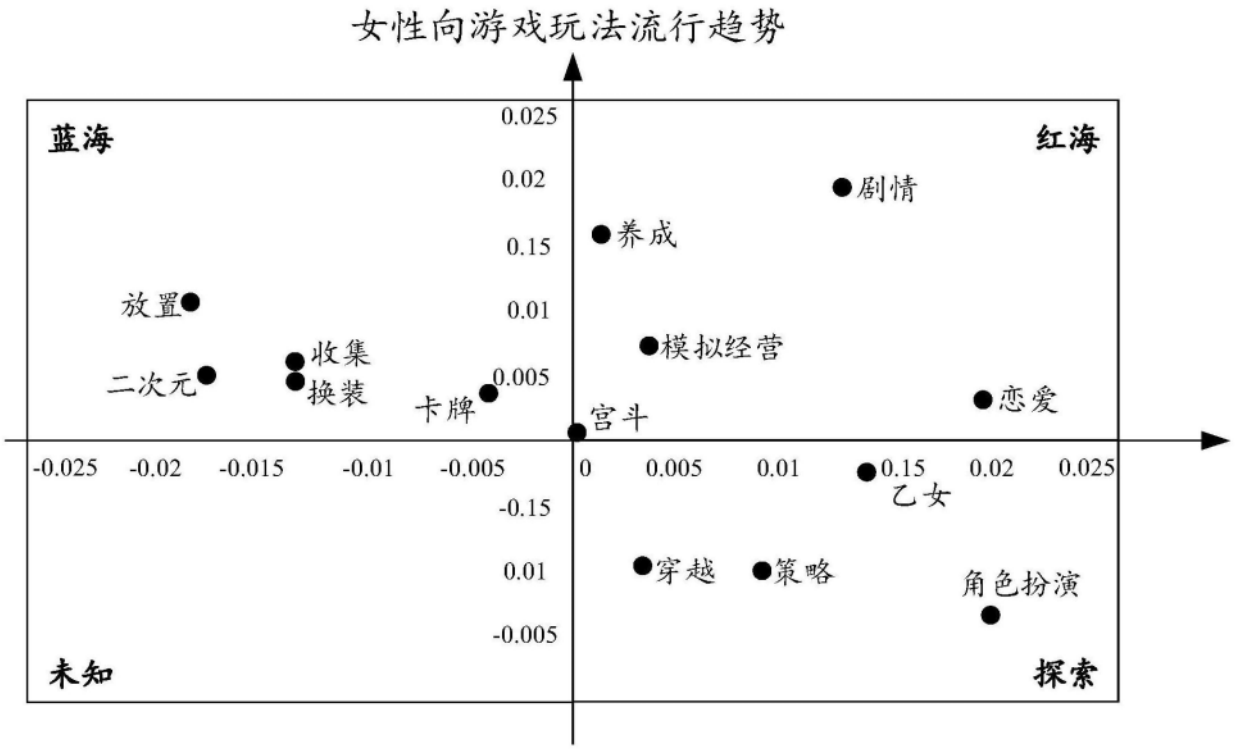

女性向游戏玩法流行趋势
蓝海
红海
未知
探索
0.025
0.02
0.15
0.01
0.005
-0.15
0.01
-0.005
-0.025
-0.02
-0.015
-0.01
-0.005
0
0.005
0.01
0.15
0.02
0.025
剧情
养成
放置
模拟经营
收集
换装
二次元
卡牌
恋爱
宫斗
乙女
穿越
策略
角色扮演

图8

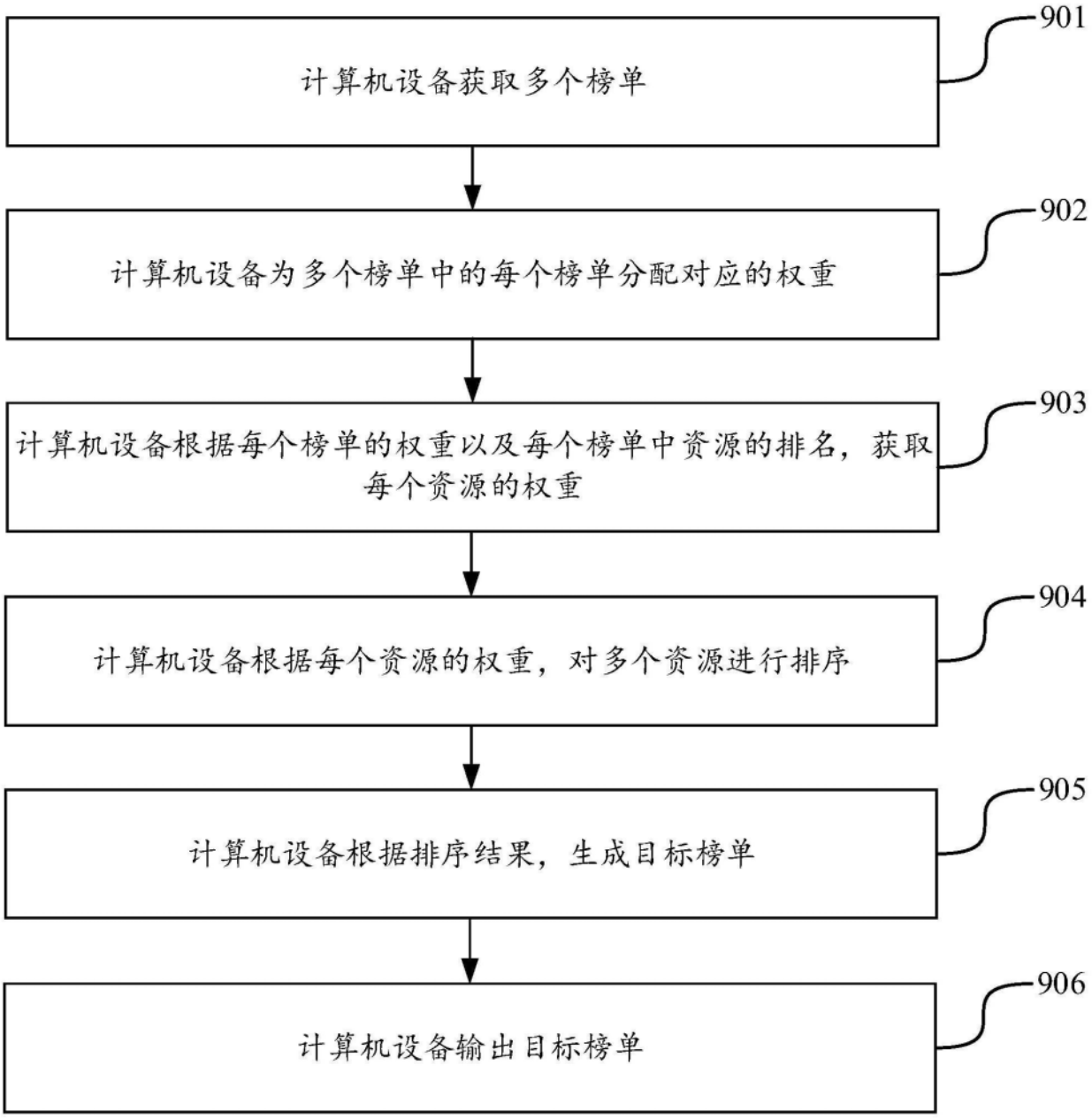

计算机设备获取多个榜单
901
计算机设备为多个榜单中的每个榜单分配对应的权重
902
计算机设备根据每个榜单的权重以及每个榜单中资源的排名，获取每个资源的权重
903
计算机设备根据每个资源的权重，对多个资源进行排序
904
计算机设备根据排序结果，生成目标榜单
905
计算机设备输出目标榜单
906

图9

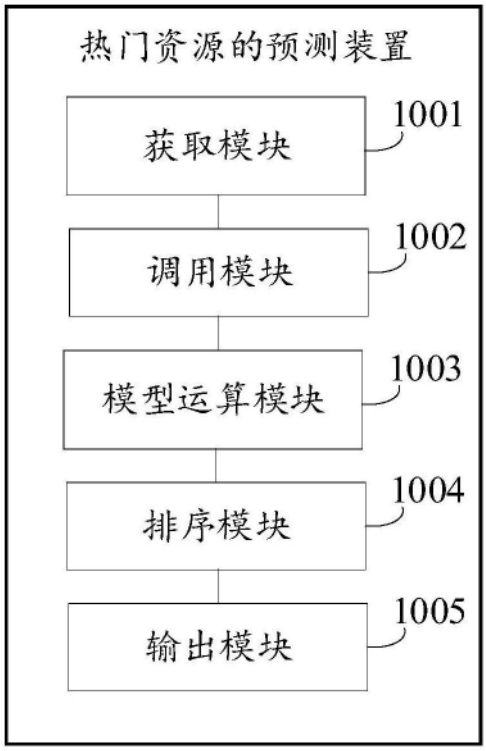

热门资源的预测装置
获取模块
1001
调用模块
1002
模型运算模块
1003
排序模块
1004
输出模块
1005

图10

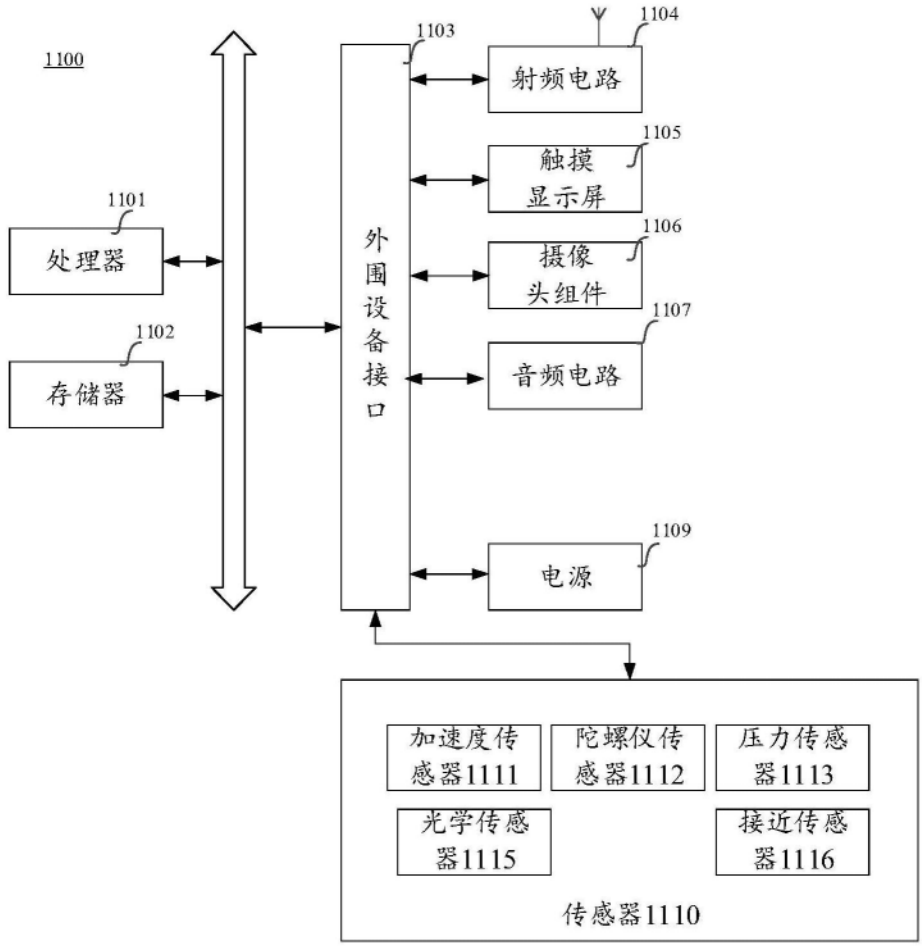

1100
1101
处理器
1102
存储器
1103
外围设备接口
1104
射频电路
1105
触摸显示屏
1106
摄像头组件
1107
音频电路
1109
电源
加速度传感器1111
陀螺仪传感器1112
压力传感器1113
光学传感器1115
接近传感器1116
传感器1110

图11

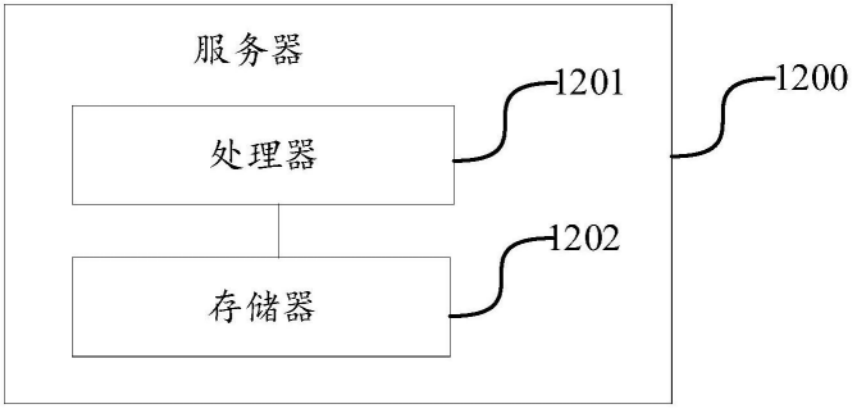

服务器
1200
1201
处理器
1202
存储器

图12