



(12)发明专利申请

(10)申请公布号 CN 108009877 A

(43)申请公布日 2018.05.08

(21)申请号 201711192555.0

(22)申请日 2017.11.24

(71)申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72)发明人 马书超

(74)专利代理机构 北京国昊天诚知识产权代理有限公司 11315

代理人 许志勇

(51) Int. Cl.

G06Q 30/06(2012.01)

G06F 17/30(2006.01)

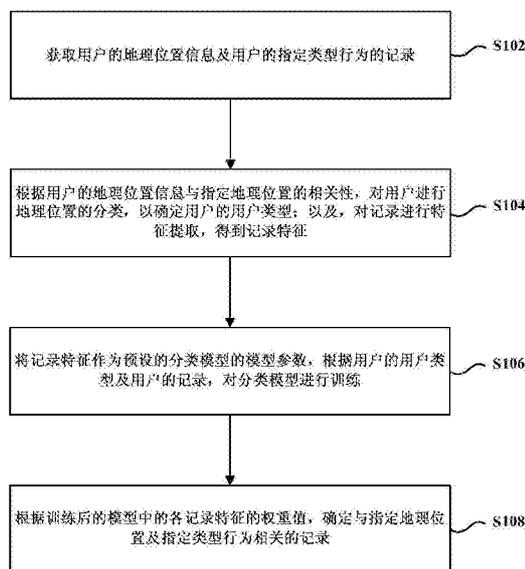
权利要求书5页 说明书19页 附图6页

(54)发明名称

信息挖掘方法及装置

(57)摘要

本说明书一个或多个实施例公开了一种信息挖掘方法及装置,用以解决现有技术中信息挖掘不够准确的问题。所述方法包括:获取用户的地理位置信息及所述用户的指定类型行为的记录;根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。



1. 一种信息挖掘方法,包括:

获取用户的地理位置信息及所述用户的指定类型行为的记录;

根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

2. 根据权利要求1所述的方法,所述对所述用户进行地理位置的分类,以确定所述用户的用户类型,包括:

若所述用户的地理位置信息与所述指定地理位置相关,则将所述用户确定为第一类用户;若所述用户的地理位置信息与所述指定地理位置不相关,则将所述用户确定为第二类用户;

相应的,所述对所述分类模型进行训练,包括:

将所述第一类用户的记录作为正样本数据,将所述第二类用户的记录作为负样本数据,对所述分类模型进行训练。

3. 根据权利要求2所述的方法,所述对所述分类模型进行训练,包括:

根据所述各记录特征与所述用户类型的相关程度,确定所述各记录特征的权重值;

其中,所述权重值与所述记录特征和所述第一类用户的相关程度正相关,与所述记录特征和所述第二类用户的相关程度负相关。

4. 根据权利要求1或3所述的方法,所述根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录,包括:

获取所述权重值最高的记录特征,或,获取所述权重值大于预设阈值的记录特征;

将获取的所述记录特征对应的记录,作为与所述指定地理位置及所述指定类型行为相关的记录。

5. 根据权利要求1所述的方法,所述记录包括主题记录;所述获取所述用户的指定类型行为的记录,包括:

获取所述用户的指定类型行为的历史行为数据;

确定所述指定类型行为对应的多个主题;

利用指定主题训练模型对所述历史行为数据及所述多个主题进行训练,得到所述用户的指定类型行为的主题记录。

6. 根据权利要求2所述的方法,所述获取所述用户的指定类型行为的记录,包括:

根据所述地理位置信息,确定所述第一类用户位于所述指定地理位置的时间;

获取所述用户在位于所述指定地理位置的时间之前的预设时间段内的指定类型行为的记录。

7. 一种信息挖掘方法,包括:

获取用户的地理位置信息及所述用户的购买记录;

根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

8. 一种信息挖掘方法,包括:

获取用户的地理位置信息及所述用户的购买记录;

根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

9. 一种信息挖掘装置,包括:

第一获取模块,获取用户的地理位置信息及所述用户的指定类型行为的记录;

第一分类及提取模块,根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

第一训练模块,将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

第一确定模块,根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

10. 根据权利要求9所述的装置,所述第一分类及提取模块包括:

第一确定单元,若所述用户的地理位置信息与所述指定地理位置相关,则将所述用户确定为第一类用户;若所述用户的地理位置信息与所述指定地理位置不相关,则将所述用户确定为第二类用户;

相应的,所述第一训练模块包括:

第一训练单元,将所述第一类用户的记录作为正样本数据,将所述第二类用户的记录作为负样本数据,对所述分类模型进行训练。

11. 根据权利要求10所述的装置,所述第一训练模块包括:

第二确定单元,根据所述各记录特征与所述用户类型的相关程度,确定所述各记录特征的权重值;

其中,所述权重值与所述记录特征和所述第一类用户的相关程度正相关,与所述记录特征和所述第二类用户的相关程度负相关。

12. 根据权利要求9或11所述的装置,所述第一确定模块包括:

第一获取单元,获取所述权重值最高的记录特征,或,获取所述权重值大于预设阈值的记录特征;

第三确定单元,将获取的所述记录特征对应的记录,作为与所述指定地理位置及所述指定类型行为相关的记录。

13. 根据权利要求9所述的装置,所述记录包括主题记录;所述第一获取模块包括:

第二获取单元,获取所述用户的指定类型行为的历史行为数据;

第四确定单元,确定所述指定类型行为对应的多个主题;

第二训练单元,利用指定主题训练模型对所述历史行为数据及所述多个主题进行训练,得到所述用户的指定类型行为的主题记录。

14. 根据权利要求10所述的装置,所述第一获取模块包括:

第五确定单元,根据所述地理位置信息,确定所述第一类用户位于所述指定地理位置的时间;

第三获取单元,获取所述用户在位于所述指定地理位置的时间之前的预设时间段内的指定类型行为的记录。

15. 一种信息挖掘装置,包括:

第二获取模块,获取用户的地理位置信息及所述用户的购买记录;

第二分类及提取模块,根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

第二训练模块,将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

第二确定模块,根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

16. 一种信息挖掘装置,包括:

第三获取模块,获取用户的地理位置信息及所述用户的购买记录;

第三分类及提取模块,根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

第三训练模块,将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

第三确定模块,根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

17. 一种信息挖掘设备,包括:

处理器;以及

被安排成存储计算机可执行指令的存储器,所述可执行指令在被执行时使所述处理器:

获取用户的地理位置信息及所述用户的指定类型行为的记录;

根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

18. 一种存储介质,用于存储计算机可执行指令,所述可执行指令在被执行时实现以下流程:

获取用户的地理位置信息及所述用户的指定类型行为的记录；

根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

19. 一种信息挖掘设备,包括:

处理器;以及

被安排成存储计算机可执行指令的存储器,所述可执行指令在被执行时使所述处理器:

获取用户的地理位置信息及所述用户的购买记录;

根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

20. 一种存储介质,用于存储计算机可执行指令,所述可执行指令在被执行时实现以下流程:

获取用户的地理位置信息及所述用户的购买记录;

根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

21. 一种信息挖掘设备,包括:

处理器;以及

被安排成存储计算机可执行指令的存储器,所述可执行指令在被执行时使所述处理器:

获取用户的地理位置信息及所述用户的购买记录;

根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

22. 一种存储介质,用于存储计算机可执行指令,所述可执行指令在被执行时实现以下流程:

获取用户的地理位置信息及所述用户的购买记录;

根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

信息挖掘方法及装置

技术领域

[0001] 本说明书涉及信息处理技术领域,尤其涉及一种信息挖掘方法及装置。

背景技术

[0002] 在信息处理技术领域,如何从大量信息中挖掘出能代表用户意图的信息始终是一个热门话题。例如,如何从用户一段时间内的出行日志中挖掘出能够代表用户出行时的购买意图的数据,进而为用户推送一些符合用户购买意图的消息。针对上述问题,现有技术中大多采用直接统计的方法,即直接统计购买物品的用户以及出行比例,这种方法往往会筛选出很多与出行毫不相关的物品,效果很差。

发明内容

[0003] 本说明书一个或多个实施例的目的是提供一种信息挖掘方法及装置,用以解决现有技术中信息挖掘不够准确的问题。

[0004] 为解决上述技术问题,本说明书一个或多个实施例是这样实现的:

[0005] 一方面,本说明书一个或多个实施例提供一种信息挖掘方法,包括:

[0006] 获取用户的地理位置信息及所述用户的指定类型行为的记录;

[0007] 根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

[0008] 将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

[0009] 根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

[0010] 可选的,所述对所述用户进行地理位置的分类,以确定所述用户的用户类型,包括:

[0011] 若所述用户的地理位置信息与所述指定地理位置相关,则将所述用户确定为第一类用户;若所述用户的地理位置信息与所述指定地理位置不相关,则将所述用户确定为第二类用户;

[0012] 相应的,所述对所述分类模型进行训练,包括:

[0013] 将所述第一类用户的记录作为正样本数据,将所述第二类用户的记录作为负样本数据,对所述分类模型进行训练。

[0014] 可选的,所述对所述分类模型进行训练,包括:

[0015] 根据所述各记录特征与所述用户类型的相关程度,确定所述各记录特征的权重值;

[0016] 其中,所述权重值与所述记录特征和所述第一类用户的相关程度正相关,与所述记录特征和所述第二类用户的相关程度负相关。

[0017] 可选的,所述根据训练后的模型中的各记录特征的权重值,确定与所述指定地理

位置及所述指定类型行为相关的记录,包括:

[0018] 获取所述权重值最高的记录特征,或,获取所述权重值大于预设阈值的记录特征;

[0019] 将获取的所述记录特征对应的记录,作为与所述指定地理位置及所述指定类型行为相关的记录。

[0020] 可选的,所述记录包括主题记录;所述获取所述用户的指定类型行为的记录,包括:

[0021] 获取所述用户的指定类型行为的历史行为数据;

[0022] 确定所述指定类型行为对应的多个主题;

[0023] 利用指定主题训练模型对所述历史行为数据及所述多个主题进行训练,得到所述用户的指定类型行为的主题记录。

[0024] 可选的,所述获取所述用户的指定类型行为的记录,包括:

[0025] 根据所述地理位置信息,确定所述第一类用户位于所述指定地理位置的时间;

[0026] 获取所述用户在位于所述指定地理位置的时间之前的预设时间段内的指定类型行为的记录。

[0027] 另一方面,本说明书一个或多个实施例提供一种信息挖掘方法,包括:

[0028] 获取用户的地理位置信息及所述用户的购买记录;

[0029] 根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0030] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0031] 根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

[0032] 再一方面,本说明书一个或多个实施例提供一种信息挖掘方法,包括:

[0033] 获取用户的地理位置信息及所述用户的购买记录;

[0034] 根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0035] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0036] 根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

[0037] 再一方面,本说明书一个或多个实施例提供一种信息挖掘装置,包括:

[0038] 第一获取模块,获取用户的地理位置信息及所述用户的指定类型行为的记录;

[0039] 第一分类及提取模块,根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

[0040] 第一训练模块,将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

[0041] 第一确定模块,根据训练后的模型中的各记录特征的权重值,确定与所述指定地

理位置及所述指定类型行为相关的记录。

[0042] 可选的,所述第一分类及提取模块包括:

[0043] 第一确定单元,若所述用户的地理位置信息与所述指定地理位置相关,则将所述用户确定为第一类用户;若所述用户的地理位置信息与所述指定地理位置不相关,则将所述用户确定为第二类用户;

[0044] 相应的,所述第一训练模块包括:

[0045] 第一训练单元,将所述第一类用户的记录作为正样本数据,将所述第二类用户的记录作为负样本数据,对所述分类模型进行训练。

[0046] 可选的,所述第一训练模块包括:

[0047] 第二确定单元,根据所述各记录特征与所述用户类型的相关程度,确定所述各记录特征的权重值;

[0048] 其中,所述权重值与所述记录特征和所述第一类用户的相关程度正相关,与所述记录特征和所述第二类用户的相关程度负相关。

[0049] 可选的,所述第一确定模块包括:

[0050] 第一获取单元,获取所述权重值最高的记录特征,或,获取所述权重值大于预设阈值的记录特征;

[0051] 第三确定单元,将获取的所述记录特征对应的记录,作为与所述指定地理位置及所述指定类型行为相关的记录。

[0052] 可选的,所述记录包括主题记录;所述第一获取模块包括:

[0053] 第二获取单元,获取所述用户的指定类型行为的历史行为数据;

[0054] 第四确定单元,确定所述指定类型行为对应的多个主题;

[0055] 第二训练单元,利用指定主题训练模型对所述历史行为数据及所述多个主题进行训练,得到所述用户的指定类型行为的主题记录。

[0056] 可选的,所述第一获取模块包括:

[0057] 第五确定单元,根据所述地理位置信息,确定所述第一类用户位于所述指定地理位置的时间;

[0058] 第三获取单元,获取所述用户在位于所述指定地理位置的时间之前的预设时间段内的指定类型行为的记录。

[0059] 再一方面,本说明书一个或多个实施例提供一种信息挖掘装置,包括:

[0060] 第二获取模块,获取用户的地理位置信息及所述用户的购买记录;

[0061] 第二分类及提取模块,根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0062] 第二训练模块,将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0063] 第二确定模块,根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

[0064] 再一方面,本说明书一个或多个实施例提供一种信息挖掘装置,包括:

[0065] 第三获取模块,获取用户的地理位置信息及所述用户的购买记录;

[0066] 第三分类及提取模块,根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0067] 第三训练模块,将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0068] 第三确定模块,根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

[0069] 再一方面,本说明书一个或多个实施例提供一种信息挖掘设备,包括:

[0070] 处理器;以及

[0071] 被安排成存储计算机可执行指令的存储器,所述可执行指令在被执行时使所述处理器:

[0072] 获取用户的地理位置信息及所述用户的指定类型行为的记录;

[0073] 根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

[0074] 将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

[0075] 根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

[0076] 再一方面,本说明书一个或多个实施例提供一种存储介质,用于存储计算机可执行指令,所述可执行指令在被执行时实现以下流程:

[0077] 获取用户的地理位置信息及所述用户的指定类型行为的记录;

[0078] 根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

[0079] 将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

[0080] 根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

[0081] 再一方面,本说明书一个或多个实施例提供一种信息挖掘设备,包括:

[0082] 处理器;以及

[0083] 被安排成存储计算机可执行指令的存储器,所述可执行指令在被执行时使所述处理器:

[0084] 获取用户的地理位置信息及所述用户的购买记录;

[0085] 根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0086] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0087] 根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

[0088] 再一方面,本说明书一个或多个实施例提供一种存储介质,用于存储计算机可执

行指令,所述可执行指令在被执行时实现以下流程:

[0089] 获取用户的地理位置信息及所述用户的购买记录;

[0090] 根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0091] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0092] 根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

[0093] 再一方面,本说明书一个或多个实施例提供一种信息挖掘设备,包括:

[0094] 处理器;以及

[0095] 被安排成存储计算机可执行指令的存储器,所述可执行指令在被执行时使所述处理器:

[0096] 获取用户的地理位置信息及所述用户的购买记录;

[0097] 根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0098] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0099] 根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

[0100] 再一方面,本说明书一个或多个实施例提供一种存储介质,用于存储计算机可执行指令,所述可执行指令在被执行时实现以下流程:

[0101] 获取用户的地理位置信息及所述用户的购买记录;

[0102] 根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0103] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0104] 根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

[0105] 采用本说明书一个或多个实施例的技术方案,通过获取用户的地理位置信息及指定类型行为的记录,并根据用户的地理位置信息确定用户的用户类型,以及对记录进行特征提取得到记录特征,进而将记录特征作为预设的分类模型的模型参数,根据用户类型及记录对分类模型进行训练,从而训练出各记录特征的权重值,并根据该权重值确定与指定地理位置及指定类型行为相关的记录。因此,该技术方案使得信息的挖掘能够依赖于用户指定类型行为的记录及用户的地理位置信息,由于指定类型行为的记录能够反映出用户的行为意图,因此最终挖掘出的记录能够很大程度上体现出用户的行为意图,提高了信息挖掘的准确度,且能够为其他后续行为(如利用挖掘出的记录进行消息推荐等)提供更加准确的信息。

附图说明

[0106] 为了更清楚地说明本说明书一个或多个实施例或现有技术中的技术方案,下面将对实施例或现有技术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说明书一个或多个实施例中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0107] 图1是根据本说明书一实施例的一种信息挖掘方法的示意性流程图;

[0108] 图2是根据本说明书一实施例的一种信息挖掘方法中主题记录的示意图;

[0109] 图3是根据本说明书另一实施例的一种信息挖掘方法的示意性流程图;

[0110] 图4是根据本说明书另一实施例的一种信息挖掘方法的示意性流程图;

[0111] 图5是根据本说明书一个实施例中一种信息挖掘装置的示意性框图;

[0112] 图6是根据本说明书另一个实施例中一种信息挖掘装置的示意性框图;

[0113] 图7是根据本说明书另一个实施例中一种信息挖掘装置的示意性框图;

[0114] 图8是根据本说明书另一个实施例中一种信息挖掘设备的示意性框图。

具体实施方式

[0115] 本说明书一个或多个实施例提供一种信息挖掘方法及装置,用以解决现有技术中信息挖掘不够准确的问题。

[0116] 为了使本技术领域的人员更好地理解本说明书一个或多个实施例中的技术方案,下面将结合本说明书一个或多个实施例中的附图,对本说明书一个或多个实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本说明书一部分实施例,而不是全部的实施例。基于本说明书一个或多个实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都应当属于本说明书一个或多个实施例保护的范畴。

[0117] 图1是根据本说明书一实施例的一种信息挖掘方法的示意性流程图,如图1所示,该方法包括:

[0118] 步骤S102,获取用户的地理位置信息及用户的指定类型行为的记录。

[0119] 步骤S104,根据用户的地理位置信息与指定地理位置的相关性,对用户进行地理位置的分类,以确定用户的用户类型;以及,对记录进行特征提取,得到记录特征。

[0120] 步骤S106,将记录特征作为预设的分类模型的模型参数,根据用户的用户类型及用户的记录,对分类模型进行训练。

[0121] 步骤S108,根据训练后的模型中的各记录特征的权重值,确定与指定地理位置及指定类型行为相关的记录。

[0122] 本实施例中,可获取一个或多个用户的地理位置信息及该一个或多个用户的指定类型行为的记录。

[0123] 例如,首次对分类模型进行训练时,可获取多个用户的地理位置信息及该多个用户的指定类型行为的记录,并通过步骤S104-S106的过程对分类模型进行训练,以使分类模型的训练能够依赖于多个样本数据(包括多个用户的地理位置信息及该多个用户的指定类型行为的记录),从而使分类模型的训练更加准确。

[0124] 若分类模型已训练完成,则可利用一个或多个样本数据对分类模型进行更新。例如,获取一个用户的地理位置信息及该用户的指定类型行为的记录,并通过步骤S104-S106的过程对分类模型进行更新,以使更新后的分类模型更加准确、更能符合用户的行为意图。

[0125] 采用本说明书实施例的技术方案,通过获取用户的地理位置信息及指定类型行为的记录,并根据用户的地理位置信息确定用户的用户类型,以及对记录进行特征提取得到记录特征,进而将记录特征作为预设的分类模型的模型参数,根据用户类型及记录对分类模型进行训练,从而训练出各记录特征的权重值,并根据该权重值确定与指定地理位置及指定类型行为相关的记录。因此,该技术方案使得信息的挖掘能够依赖于用户指定类型行为的记录及用户的地理位置信息,由于指定类型行为的记录能够反映出用户的行为意图,因此最终挖掘出的记录能够很大程度上体现出用户的行为意图,提高了信息挖掘的准确度,且能够为其他后续行为(如利用挖掘出的记录进行消息推荐等)提供更加准确的信息。

[0126] 以下针对上述步骤S102-S108进行详细说明。

[0127] 上述步骤S102-S108中所说的“记录”可包括主题记录。以下实施例均以主题记录为例进行说明。

[0128] 首先执行步骤S102,即获取用户的地理位置信息及用户的指定类型行为的主题记录。其中,可根据LBS(Location Based Service,基于位置服务)获取用户的地理位置信息。LBS是指通过电信移动运营商的无线电通讯网络或外部定位方式,获取终端用户的位置信息,在GIS(Geographic Information System,地理信息系统)平台的支持下,为用户提供相应服务的一种增值业务。

[0129] 在一个实施例中,指定类型行为如购买行为、学习行为、运动行为等。可通过以下方式获取多个用户的指定类型行为的主题记录:首先,获取用户的指定类型行为的历史行为数据。其次,确定指定类型行为对应的多个主题;最后,利用指定主题训练模型对历史行为数据及多个主题进行训练,得到用户的指定类型行为的主题记录。

[0130] 本实施例中,不同主题用于反映用户指定类型行为的不同类别,例如,指定类型行为为购买行为,考虑到用户可能购买教育类、学习类、运动类、烹饪类等各种类型的物品,因此购买行为对应的多个主题可包括教育主题、学习主题、运动主题、烹饪主题等;再例如,指定类型行为为运动行为,考虑到用户可能进行跑步、骑行、游泳等各类型的运行,因此运动行为对应的多个主题可包括跑步主题、骑行主题、游泳主题等。

[0131] 指定主题训练模型可以是LDA(Latent Dirichlet Allocation,文档主题生成)模型。在利用LDA模型训练指定类型行为的主题记录时,需将指定类型行为的历史行为数据及指定类型行为对应的多个主题作为输入数据,这里还可预先对获取到的指定类型行为的历史行为数据进行预处理及切词处理,以使经过预处理后的数据能够符合LDA模型的输入数据的格式,以及经过切词处理后的数据能够满足LDA模型的针对“词袋”进行训练的需求。

[0132] 例如,将用户的购买行为的历史行为数据进行预处理及分词后,得到处理后的历史行为数据,处理后的历史行为数据中包括多个与购买行为有关的词语(如购买物品的名称),同时给定购买行为的多个主题为:教育主题、学习主题、运动主题。将处理后的历史行为数据及购买行为的多个主题作为LDA模型的输入进行训练。LDA模型进行训练时,根据事先给定的购买行为的多个主题,迭代学习历史行为数据中各个主题的分布情况,然后按照一定的概率(可随机指定)选择其中一个主题,并确定该主题中各个词语的分布情况,重复

此步骤,最终确定每个主题中的各个词语的分布情况。根据每个主题中的各个词语的分布情况,可确定用户购买行为的主题记录,该主题记录中包括各个主题中每个词语的分布情况。

[0133] 当然,指定主题训练模型并不局限于LDA模型,其还可以是例如PLSA模型等任何一种可实现主题训练的主题模型。

[0134] 图2示出了本实施例中采用指定主题训练模型训练出的主题记录的示意图。在图2中,购买行为的主题包括旅游主题、教育主题、烹饪主题这三个主题,且在旅游主题中,词语“衣物用品”的分布占50%,词语“饰品”的分布占30%,词语“书籍”的分布占20%;在教育主题中,词语“书籍”的分布占60%,词语“电子学习机”的分布占30%,词语“电子课程”的分布占10%;在烹饪主题中,词语“书籍”的分布占60%,词语“厨具”的分布占40%。

[0135] 此外,在获取多个用户的指定类型行为的主题记录时,为使获取的主题记录能够更大限度地反映出用户与指定地理位置相关的行为意图,可首先确定曾位于指定地理位置的用户定位于指定地理位置的时间,进而获取多个用户在该时间之前的预设时间段内的指定类型行为的主题记录。其中,曾位于指定地理位置的用户指的是该用户的地理位置信息与指定地理位置相关。

[0136] 本实施例中,可针对每个曾位于指定地理位置的用户确定其定位于指定地理位置的时间,然后分别获取每个曾位于指定地理位置的用户在其各自定位于指定地理位置的时间之前的预设时间段内的指定类型行为的主题记录。也可从各个曾位于指定地理位置的用户中选择一个基准用户,并确定该基准用户定位于指定地理位置的时间,进而获取多个用户在该时间之前的预设时间段内的指定类型行为的主题记录。例如,随机选择一个曾位于指定地理位置的用户作为基准用户,或者,选择位于指定地理位置的时间最长的用户作为基准用户,或者,选择位于指定地理位置的时间最近的用户作为基准用户,等等。

[0137] 在获取到用户的地理位置信息及指定类型行为的主题记录之后,继续执行步骤S104,即对用户进行分类以及对主题记录进行特征提取。

[0138] 在一个实施例中,若用户的地理位置信息与指定地理位置相关,则将用户确定为第一类用户;若用户的地理位置信息与指定地理位置不相关,则将用户确定为第二类用户。其中,用户的地理位置信息是否与指定地理位置相关可根据地理位置信息中是否包含指定地理位置的位置信息来确定,即,若用户的地理位置信息中包含指定地理位置的位置信息,则说明该用户的地理位置信息与指定地理位置相关,确定该用户为第一类用户;若用户的地理位置信息中不包含指定地理位置的位置信息,则说明该用户的地理位置信息与指定地理位置不相关,确定该用户为第二类用户。

[0139] 例如,指定地理位置为境外,若用户的地理位置信息与境外相关,则可确定该用户为第一类用户;若用户的地理位置信息与境外无关,则可确定该用户为第二类用户。根据用户的地理位置信息是否与境外相关将用户划分为第一类用户和第二类用户,可用于后续确定第一类用户(即曾去过境外的用户)所执行的与境外相关的指定类型行为的主题记录,如确定第一类用户的出境购买物品的主题记录等。

[0140] 在一个实施例中,对主题记录进行特征提取时,可预先确定所要提取的特征类型,优选的,该特征类型与指定主题训练模型的模型参数相符。例如,若指定主题训练模型的模型参数为文字型,则可对主题记录进行文字特征的提取;若指定主题训练模型的模型参数

为字符型,则可对主题记录进行字符特征的提取。例如,对主题记录进行文字特征的提取之后,得到主题记录特征“教育”、“旅游”、“运动”等。

[0141] 对用户进行分类以及对主题记录进行特征提取之后,继续执行步骤S106,即将主题记录特征作为预设的分类模型的模型参数,并根据用户的用户类型及用户的主题记录对分类模型进行训练。该步骤中,由于用户类型可分为第一类用户和第二类用户,因此根据用户的用户类型及用户的主题记录对分类模型进行训练时,可将第一类用户的主题记录作为正样本数据,将第二类用户的主题记录作为负样本数据。预设的分类模型可以是逻辑回归模型、迭代决策树模型等任一种二分类模型,例如LR(Logistic Regression,逻辑回归)模型。

[0142] 在一个实施例中,训练得到的分类模型中包括各主题记录特征的权重值,且该权重值能够反映出各主题记录特征与用户类型的相关程度。因此,对分类模型进行训练时,可根据各主题记录特征与用户类型的相关程度,确定各主题记录特征的权重值。其中,主题记录特征权重值与该主题记录特征和第一类用户的相关程度正相关,与该主题记录特征和第二类用户的相关程度负相关。

[0143] 本实施例中,确定各主题记录特征与用户类型的相关程度时,可根据各主题记录特征在不同用户类型对应的主题记录中的出现率来确定。具体的,若主题记录特征在第一类用户对应的主题记录中的出现率越高,则该主题记录特征与第一类用户的相关程度就越高;反之,若主题记录特征在第二类用户对应的主题记录中的出现率越高,则该主题记录特征与第一类用户的相关程度就越低。

[0144] 此外,若某个主题记录特征在第一类用户对应的主题记录以及在第二类用户对应的主题记录中均出现,则可针对各用户类型设定权重,并根据各用户类型的权重以及主题记录特征分别在各用户类型对应的主题记录中的出现率,确定该主题记录特征的权重值,即确定该主题记录特征与第一类用户的相关程度。

[0145] 以LR模型为例,将主题记录特征作为LR模型的模型参数,以及将用户的用户类型及用户的主题记录作为LR模型的输入数据进行训练。在LR模型训练时,将用户的用户类型及用户的主题记录设置为因变量(即输入),并将各主题记录特征的权重值设置为自变量(即输出),针对每一个主题记录特征,给定一个初始概率(即权重值)及回归方程,然后根据因变量对该初始概率进行迭代回归,最终得出主题记录特征的权重值。其中,构建回归方程的方法为现有技术,在此不再赘述。

[0146] 对分类模型进行训练之后,继续执行步骤S108,即根据训练后的模型中的各主题记录特征的权重值,确定与指定地理位置及指定类型行为相关的主题记录。在一个实施例中,获取权重值最高的主题记录特征对应的主题记录作为与指定地理位置及指定类型行为相关的主题记录。在另一个实施例中,获取权重值大于预设阈值的主题记录特征对应的主题记录作为与指定地理位置及指定类型行为相关的主题记录。

[0147] 确定出与指定地理位置及指定类型行为相关的主题记录之后,主题记录中所包含的词语即为挖掘出的与指定地理位置及指定类型行为相关的信息。

[0148] 以下针对两个具体场景来说明本说明书提供的信息挖掘方法。

[0149] 场景一

[0150] 图3是根据本说明书一个实施例中一种信息挖掘方法的示意性流程图。在该实施

例中,指定地理位置为境外,指定类型行为为购买行为,根据用户的地理位置信息以及历史购买行为数据挖掘出用户执行的与境外相关的购买物品。如图3所示,该方法包括以下步骤:

[0151] S302,获取用户的地理位置信息及用户的购买记录。

[0152] 该步骤中,根据LBS获取用户的地理位置信息。购买记录包括购买主题记录,在获取用户的购买主题记录时,可首先获取用户的历史购买行为数据,并确定购买行为的多个主题,进而使用LDA模型等主题模型对获取到的历史购买行为数据及多个主题训练,得到用户的购买主题记录。并且,在训练购买主题记录时,需首先对历史购买行为数据进行预处理及分词,以使历史购买行为数据符合主题模型的输入数据的格式以及针对“词袋”进行训练的需求,因此,训练出的购买主题记录中包括多个与购买行为相关的词语,即购买物品的名称。

[0153] 本实施例由于意在挖掘用户执行的与境外相关的购买物品,因此在获取历史购买行为数据时,可先针对曾位于境外的用户确定其定位于境外的时间,然后获取用户在该时间之前的预设时间段内的历史购买行为数据。

[0154] 购买行为的多个主题可包括境外旅游、教育、烹饪等主题。

[0155] S304,根据用户的地理位置信息与境外的相关性,对用户进行地理位置的分类,以确定用户的用户类型;以及,对购买记录进行特征提取,得到购买记录特征。

[0156] 该步骤中,若用户的地理位置信息与境外相关,则将用户确定为第一类用户;若用户的地理位置信息与境外不相关,则将用户确定为第二类用户。用户的地理位置信息是否与境外相关可根据地理位置信息中是否包含境外位置信息来确定,即,若用户的地理位置信息中包含境外位置信息,则说明该用户的地理位置信息与境外相关,确定该用户为第一类用户;若用户的地理位置信息中不包含境外位置信息,则说明该用户的地理位置信息与境外不相关,确定该用户为第二类用户。

[0157] 对购买记录进行特征提取时,可预先确定所要提取的特征类型,优选的,若购买记录为购买主题记录,则购买主题记录的特征类型与指定主题训练模型的模型参数相符。例如,若指定主题训练模型的模型参数为文字型,则可对购买主题记录进行文字特征的提取;若指定主题训练模型的模型参数为字符型,则可对购买主题记录进行字符特征的提取。例如,对购买主题记录进行文字特征的提取之后,得到购买主题记录特征“境外旅游”、“教育”、“烹饪”等。

[0158] S306,将购买记录特征作为预设的分类模型的模型参数,根据用户的用户类型及用户的购买记录,对分类模型进行训练。

[0159] 该步骤中,购买记录包括购买主题记录,预设的分类模型可以是逻辑回归模型、迭代决策树模型等任一种二分类模型,例如LR模型。并且,在对分类模型进行训练时,将第一类用户的购买主题记录作为正样本数据,将第二类用户的购买主题记录作为负样本数据。

[0160] 训练得到的分类模型中包括各购买主题记录特征的权重值,且该权重值能够反映出各购买主题记录特征与用户类型的相关程度。因此,对分类模型进行训练时,可根据各购买主题记录特征与用户类型的相关程度,确定各购买主题记录特征的权重值。其中,购买主题记录特征权重值与该购买主题记录特征和第一类用户的相关程度正相关,与该购买主题记录特征和第二类用户的相关程度负相关。

[0161] 确定各购买主题记录特征与用户类型的相关程度时,可根据各购买主题记录特征在不同用户类型对应的购买主题记录中的出现率来确定。具体的,若购买主题记录特征在第一类用户对应的购买主题记录中的出现率越高,则该购买主题记录特征与第一类用户的相关程度就越高;反之,若购买主题记录特征在第二类用户对应的购买主题记录中的出现率越高,则该购买主题记录特征与第一类用户的相关程度就越低。

[0162] 此外,若某个购买主题记录特征在第一类用户对应的购买主题记录以及在第二类用户对应的购买主题记录中均出现,则可针对各用户类型设定权重,并根据各用户类型的权重以及购买主题记录特征分别在各用户类型对应的购买主题记录中的出现率,确定该购买主题记录特征的权重值,即确定该购买主题记录特征与第一类用户的相关程度。

[0163] S308,根据训练后的模型中的各购买记录特征的权重值,确定与境外相关的购买记录。

[0164] 该步骤中,购买记录包括购买主题记录,可获取权重值最高的购买主题记录特征对应的购买主题记录作为与境外相关的购买主题记录,也可获取权重值大于预设阈值的购买主题记录特征对应的购买主题记录作为与境外相关的购买主题记录。最终获取到的购买主题记录中包含的各词语即为用户执行的与境外相关的购买物品。

[0165] 采用本说明书实施例的技术方案,通过获取用户的地理位置信息及购买记录,并根据用户的地理位置信息确定用户的用户类型,以及对购买记录进行特征提取得到购买记录特征,进而将购买记录特征作为预设的分类模型的模型参数,根据用户类型及购买记录对分类模型进行训练,从而训练出各购买记录特征的权重值,并根据该权重值确定与境外相关的购买记录。因此,该技术方案使得用户出境购买物品的挖掘能够依赖于用户的购买记录及用户的地理位置信息,由于购买记录能够反映出用户的出境购买意图,因此最终挖掘出的购买记录能够很大程度上体现出用户的出境购买意图,提高了挖掘出境购买物品的准确度,且能够为其他后续行为(如利用挖掘出的出境购买物品进行消息推荐等)提供更加准确的信息。

[0166] 场景二

[0167] 图4是根据本说明书一个实施例中一种信息挖掘方法的示意性流程图。在该实施例中,指定地理位置为指定国家,指定类型行为为购买行为,根据用户的地理位置信息以及历史购买行为数据挖掘出用户执行的与指定国家相关的购买物品。如图4所示,该方法包括以下步骤:

[0168] S402,获取用户的地理位置信息及用户的购买记录。

[0169] 该步骤中,根据LBS获取用户的地理位置信息。购买记录包括购买主题记录,在获取用户的购买主题记录时,可首先获取用户的历史购买行为数据,并确定购买行为的多个主题,进而使用LDA模型等主题模型对获取到的历史购买行为数据及多个主题训练,得到用户的购买主题记录。并且,在训练购买主题记录时,需首先对历史购买行为数据进行预处理及分词,以使历史购买行为数据符合主题模型的输入数据的格式以及针对“词袋”进行训练的需求,因此,训练出的购买主题记录中包括多个与购买行为相关的词语,即购买物品的名称。

[0170] 本实施例由于意在挖掘用户执行的与境外相关的购买物品,因此在获取历史购买行为数据时,可先针对曾位于境外的用户确定其定位于境外的时间,然后获取用户在该时

间之前的预设时间段内的历史购买行为数据。

[0171] 购买行为的多个主题可包括境外旅游、教育、烹饪等主题。

[0172] S404,根据用户的地理位置信息与指定国家的相关性,对用户进行地理位置的分类,以确定用户的用户类型;以及,对购买记录进行特征提取,得到购买记录特征。

[0173] 该步骤中,若用户的地理位置信息与指定国家相关,则将用户确定为第一类用户;若用户的地理位置信息与指定国家不相关,则将用户确定为第二类用户。用户的地理位置信息是否与指定国家相关可根据地理位置信息中是否包含指定国家的位置信息来确定,即,若用户的地理位置信息中包含指定国家的位置信息,则说明该用户的地理位置信息与指定国家相关,确定该用户为第一类用户;若用户的地理位置信息中不包含指定国家的位置信息,则说明该用户的地理位置信息与指定国家不相关,确定该用户为第二类用户。

[0174] 对购买记录进行特征提取时,可预先确定所要提取的特征类型,优选的,若购买记录为购买主题记录,则购买主题记录的特征类型与指定主题训练模型的模型参数相符。例如,若指定主题训练模型的模型参数为文字型,则可对购买主题记录进行文字特征的提取;若指定主题训练模型的模型参数为字符型,则可对购买主题记录进行字符特征的提取。例如,对购买主题记录进行文字特征的提取之后,得到购买主题记录特征“境外旅游”、“教育”、“烹饪”等。

[0175] S406,将购买记录特征作为预设的分类模型的模型参数,根据用户的用户类型及用户的购买记录,对分类模型进行训练。

[0176] 该步骤中,购买记录包括购买主题记录,预设的分类模型可以是逻辑回归模型、迭代决策树模型等任一种二分类模型,例如LR模型。并且,在对分类模型进行训练时,将第一类用户的购买主题记录作为正样本数据,将第二类用户的购买主题记录作为负样本数据。

[0177] 训练得到的分类模型中包括各购买主题记录特征的权重值,且该权重值能够反映出各购买主题记录特征与用户类型的相关程度。因此,对分类模型进行训练时,可根据各购买主题记录特征与用户类型的相关程度,确定各购买主题记录特征的权重值。其中,购买主题记录特征权重值与该购买主题记录特征和第一类用户的相关程度正相关,与该购买主题记录特征和第二类用户的相关程度负相关。

[0178] 确定各购买主题记录特征与用户类型的相关程度时,可根据各购买主题记录特征在不同用户类型对应的购买主题记录中的出现率来确定。具体的,若购买主题记录特征在第一类用户对应的购买主题记录中的出现率越高,则该购买主题记录特征与第一类用户的相关程度就越高;反之,若购买主题记录特征在第二类用户对应的购买主题记录中的出现率越高,则该购买主题记录特征与第一类用户的相关程度就越低。

[0179] 此外,若某个购买主题记录特征在第一类用户对应的购买主题记录以及在第二类用户对应的购买主题记录中均出现,则可针对各用户类型设定权重,并根据各用户类型的权重以及购买主题记录特征分别在各用户类型对应的购买主题记录中的出现率,确定该购买主题记录特征的权重值,即确定该购买主题记录特征与第一类用户的相关程度。

[0180] S408,根据训练后的模型中的各购买记录特征的权重值,确定与指定国家相关的购买记录。

[0181] 该步骤中,购买记录包括购买主题记录,可获取权重值最高的购买主题记录特征对应的购买主题记录作为与指定国家相关的购买主题记录,也可获取权重值大于预设阈值

的购买主题记录特征对应的购买主题记录作为与指定国家相关的购买主题记录。最终获取到的购买主题记录中包含的各词语即为用户执行的与指定国家相关的购买物品。

[0182] 采用本说明书实施例的技术方案,通过获取用户的地理位置信息及购买记录,并根据用户的地理位置信息确定用户的用户类型,以及对购买记录进行特征提取得到购买记录特征,进而将购买记录特征作为预设的分类模型的模型参数,根据用户类型及购买记录对分类模型进行训练,从而训练出各购买记录特征的权重值,并根据该权重值确定与指定国家相关的购买记录。因此,该技术方案使得对用户出行指定国家时的购买物品的挖掘能够依赖于用户的购买记录及用户的地理位置信息,由于购买记录能够反映出用户出行指定国家时的购买意图,因此最终挖掘出的购买记录能够很大程度上体现出用户出行指定国家时的购买意图,提高了挖掘出行指定国家时的购买物品的准确度,且能够为其他后续行为(如利用挖掘出的出行指定国家时的购买物品进行消息推荐等)提供更加准确的信息。

[0183] 进一步地,本实施例中,通过根据用户的用户类型及用户的购买记录对分类模型进行训练,进而确定与指定国家相关的购买记录,能够极大地提高挖掘出的出行指定国家时的购买物品的覆盖率。原因在于,针对某些出行用户较少的指定国家,用户所购买的与该指定国家相关的物品也相对较少,若直接使用用户的购买行为数据来挖掘出行该指定国家时的购买物品,则会导致正样本数据(即与指定国家相关的购买行为数据)所覆盖的用户较少,进而导致训练出的与指定国家相关的购买记录准确率较低。而本方案则并非直接使用用户的购买行为数据来挖掘出行该指定国家时的购买物品,而是利用购买行为数据训练出购买记录,该购买记录能够代表用户所购买的物品种类,进而利用购买记录来挖掘出行该指定国家时的购买物品,因此大大提高了出行指定国家时的购买物品的覆盖率。

[0184] 综上,已经对本主题的特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作可以按照不同的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序,以实现期望的结果。在某些实施方式中,多任务处理和并行处理可以是有利的。

[0185] 以上为本说明书一个或多个实施例提供的信息挖掘方法,基于同样的思路,本说明书一个或多个实施例还提供一种信息挖掘装置。

[0186] 图5是根据本说明书一个实施例中一种信息挖掘装置的示意性框图。如图5所示,该装置包括:

[0187] 第一获取模块510,获取用户的地理位置信息及用户的指定类型行为的记录;

[0188] 第一分类及提取模块520,根据用户的地理位置信息与指定地理位置的相关性,对用户进行地理位置的分类,以确定用户的用户类型;以及,对记录进行特征提取,得到记录特征;

[0189] 第一训练模块530,将记录特征作为预设的分类模型的模型参数,根据用户的用户类型及用户的记录,对分类模型进行训练;

[0190] 第一确定模块540,根据训练后的模型中的各记录特征的权重值,确定与指定地理位置及指定类型行为相关的记录。

[0191] 可选的,第一分类及提取模块520包括:

[0192] 第一确定单元,若用户的地理位置信息与指定地理位置相关,则将用户确定为第一类用户;若用户的地理位置信息与指定地理位置不相关,则将用户确定为第二类用户;

- [0193] 相应的,第一训练模块530包括:
- [0194] 第一训练单元,将第一类用户的记录作为正样本数据,将第二类用户的记录作为负样本数据,对分类模型进行训练。
- [0195] 可选的,第一训练模块530包括:
- [0196] 第二确定单元,根据各记录特征与用户类型的相关程度,确定各记录特征的权重值;
- [0197] 其中,权重值与记录特征和第一类用户的相关程度正相关,与记录特征和第二类用户的相关程度负相关。
- [0198] 可选的,第一确定模块540包括:
- [0199] 第一获取单元,获取权重值最高的记录特征,或,获取权重值大于预设阈值的记录特征;
- [0200] 第三确定单元,将获取的记录特征对应的记录,作为与指定地理位置及指定类型行为相关的记录。
- [0201] 可选的,记录包括主题记录;第一获取模块510包括:
- [0202] 第二获取单元,获取用户的指定类型行为的历史行为数据;
- [0203] 第四确定单元,确定指定类型行为对应的多个主题;
- [0204] 第二训练单元,利用指定主题训练模型对历史行为数据及多个主题进行训练,得到用户的指定类型行为的主题记录。
- [0205] 可选的,第一获取模块510包括:
- [0206] 第五确定单元,根据地理位置信息,确定第一类用户位于指定地理位置的时间;
- [0207] 第三获取单元,获取用户在位于指定地理位置的时间之前的预设时间段内的指定类型行为的记录。
- [0208] 采用本说明书一个或多个实施例的装置,通过获取用户的地理位置信息及指定类型行为的记录,并根据用户的地理位置信息确定用户的用户类型,以及对记录进行特征提取得到记录特征,进而将记录特征作为预设的分类模型的模型参数,根据用户类型及记录对分类模型进行训练,从而训练出各记录特征的权重值,并根据该权重值确定与指定地理位置及指定类型行为相关的记录。因此,该技术方案使得信息的挖掘能够依赖于用户指定类型行为的记录及用户的地理位置信息,由于指定类型行为的记录能够反映出用户的行为意图,因此最终挖掘出的记录能够很大程度上体现出用户的行为意图,提高了信息挖掘的准确度,且能够为其他后续行为(如利用挖掘出的记录进行消息推荐等)提供更加准确的信息。
- [0209] 图6是根据本说明书一个实施例中一种信息挖掘装置的示意性框图。如图6所示,该装置包括:
- [0210] 第二获取模块610,获取用户的地理位置信息及用户的购买记录;
- [0211] 第二分类及提取模块620,根据用户的地理位置信息与境外的相关性,对用户进行地理位置的分类,以确定用户的用户类型;以及,对购买记录进行特征提取,得到购买记录特征;
- [0212] 第二训练模块630,将购买记录特征作为预设的分类模型的模型参数,根据用户的用户类型及用户的购买记录,对分类模型进行训练;

[0213] 第二确定模块640,根据训练后的模型中的各购买记录特征的权重值,确定与境外相关的购买记录。

[0214] 采用本说明书实施例的装置,通过获取用户的地理位置信息及购买记录,并根据用户的地理位置信息确定用户的用户类型,以及对购买记录进行特征提取得到购买记录特征,进而将购买记录特征作为预设的分类模型的模型参数,根据用户类型及购买记录对分类模型进行训练,从而训练出各购买记录特征的权重值,并根据该权重值确定与境外相关的购买记录。因此,该技术方案使得用户出境购买物品的挖掘能够依赖于用户的购买记录及用户的地理位置信息,由于购买记录能够反映出用户的出境购买意图,因此最终挖掘出的购买记录能够很大程度上体现出用户的出境购买意图,提高了挖掘出境购买物品的准确度,且能够为其他后续行为(如利用挖掘出的出境购买物品进行消息推荐等)提供更加准确的信息。

[0215] 图7是根据本说明书一个实施例中一种信息挖掘装置的示意性框图。如图7所示,该装置包括:

[0216] 第三获取模块710,获取用户的地理位置信息及用户的购买记录;

[0217] 第三分类及提取模块720,根据用户的地理位置信息与指定国家的相关性,对用户进行地理位置的分类,以确定用户的用户类型;以及,对购买记录进行特征提取,得到购买记录特征;

[0218] 第三训练模块730,将购买记录特征作为预设的分类模型的模型参数,根据用户的用户类型及用户的购买记录,对分类模型进行训练;

[0219] 第三确定模块740,根据训练后的模型中的各购买记录特征的权重值,确定与指定国家相关的购买记录。

[0220] 采用本说明书实施例的装置,通过获取用户的地理位置信息及购买记录,并根据用户的地理位置信息确定用户的用户类型,以及对购买记录进行特征提取得到购买记录特征,进而将购买记录特征作为预设的分类模型的模型参数,根据用户类型及购买记录对分类模型进行训练,从而训练出各购买记录特征的权重值,并根据该权重值确定与指定国家相关的购买记录。因此,该技术方案使得对用户出行指定国家时的购买物品的挖掘能够依赖于用户的购买记录及用户的地理位置信息,由于购买记录能够反映出用户出行指定国家时的购买意图,因此最终挖掘出的购买记录能够很大程度上体现出用户出行指定国家时的购买意图,提高了挖掘出行指定国家时的购买物品的准确度,且能够为其他后续行为(如利用挖掘出的出行指定国家时的购买物品进行消息推荐等)提供更加准确的信息。

[0221] 本领域的技术人员应可理解,图5至图7中的信息挖掘装置能够用来实现前文的信息挖掘方法,其中的细节描述应与前文方法部分描述类似,为避免繁琐,此处不另赘述。

[0222] 基于同样的思路,本说明书一个或多个实施例还提供一种信息挖掘设备,如图8所示。信息挖掘设备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的处理器801和存储器802,存储器802中可以存储有一个或一个以上存储应用程序或数据。其中,存储器802可以是短暂存储或持久存储。存储在存储器802的应用程序可以包括一个或一个以上模块(图示未示出),每个模块可以包括对信息挖掘设备中的一系列计算机可执行指令。更进一步地,处理器801可以设置为与存储器802通信,在信息挖掘设备上执行存储器802中的一系列计算机可执行指令。信息挖掘设备还可以包括一个或一个以上电源803,一

个或一个以上有线或无线网络接口804,一个或一个以上输入输出接口805,一个或一个以上键盘806。

[0223] 具体在本实施例中,信息挖掘设备包括有存储器,以及一个或一个以上的程序,其中一个或者一个以上程序存储于存储器中,且一个或者一个以上程序可以包括一个或一个以上模块,且每个模块可以包括对信息挖掘设备中的一系列计算机可执行指令,且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令:

[0224] 获取用户的地理位置信息及所述用户的指定类型行为的记录;

[0225] 根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

[0226] 将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

[0227] 根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

[0228] 可选地,计算机可执行指令在被执行时,还可以使处理器:

[0229] 若所述用户的地理位置信息与所述指定地理位置相关,则将所述用户确定为第一类用户;若所述用户的地理位置信息与所述指定地理位置不相关,则将所述用户确定为第二类用户;

[0230] 相应的,所述对所述分类模型进行训练,包括:

[0231] 将所述第一类用户的记录作为正样本数据,将所述第二类用户的记录作为负样本数据,对所述分类模型进行训练。

[0232] 可选地,计算机可执行指令在被执行时,还可以使处理器:

[0233] 根据所述各记录特征与所述用户类型的相关程度,确定所述各记录特征的权重值;

[0234] 其中,所述权重值与所述记录特征和所述第一类用户的相关程度正相关,与所述记录特征和所述第二类用户的相关程度负相关。

[0235] 可选地,计算机可执行指令在被执行时,还可以使处理器:

[0236] 获取所述权重值最高的记录特征,或,获取所述权重值大于预设阈值的记录特征;

[0237] 将获取的所述记录特征对应的记录,作为与所述指定地理位置及所述指定类型行为相关的记录。

[0238] 可选地,所述记录包括主题记录;计算机可执行指令在被执行时,还可以使处理器:

[0239] 获取所述用户的指定类型行为的历史行为数据;

[0240] 确定所述指定类型行为对应的多个主题;

[0241] 利用指定主题训练模型对所述历史行为数据及所述多个主题进行训练,得到所述用户的指定类型行为的主题记录。

[0242] 可选地,计算机可执行指令在被执行时,还可以使处理器:

[0243] 根据所述地理位置信息,确定所述第一类用户位于所述指定地理位置的时间;

[0244] 获取所述用户在位于所述指定地理位置的时间之前的预设时间段内的指定类型

行为的记录。

[0245] 具体在另一实施例中,信息挖掘设备包括有存储器,以及一个或一个以上的程序,其中一个或者一个以上程序存储于存储器中,且一个或者一个以上程序可以包括一个或一个以上模块,且每个模块可以包括对信息挖掘设备中的一系列计算机可执行指令,且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令:

[0246] 获取用户的地理位置信息及所述用户的购买记录;

[0247] 根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0248] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0249] 根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

[0250] 具体在另一实施例中,信息挖掘设备包括有存储器,以及一个或一个以上的程序,其中一个或者一个以上程序存储于存储器中,且一个或者一个以上程序可以包括一个或一个以上模块,且每个模块可以包括对信息挖掘设备中的一系列计算机可执行指令,且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令:

[0251] 获取用户的地理位置信息及所述用户的购买记录;

[0252] 根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0253] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0254] 根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

[0255] 本说明书一个或多个实施例还提出了一种计算机可读存储介质,该计算机可读存储介质存储一个或多个程序,该一个或多个程序包括指令,该指令当被包括多个应用程序的电子设备执行时,能够使该电子设备执行上述信息挖掘方法,并具体用于执行:

[0256] 获取用户的地理位置信息及所述用户的指定类型行为的记录;

[0257] 根据所述用户的地理位置信息与指定地理位置的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述记录进行特征提取,得到记录特征;

[0258] 将所述记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的记录,对所述分类模型进行训练;

[0259] 根据训练后的模型中的各记录特征的权重值,确定与所述指定地理位置及所述指定类型行为相关的记录。

[0260] 本说明书一个或多个实施例还提出了一种计算机可读存储介质,该计算机可读存储介质存储一个或多个程序,该一个或多个程序包括指令,该指令当被包括多个应用程序的电子设备执行时,能够使该电子设备执行上述信息挖掘方法,并具体用于执行:

[0261] 获取用户的地理位置信息及所述用户的购买记录；

[0262] 根据所述用户的地理位置信息与境外的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0263] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0264] 根据训练后的模型中的各购买记录特征的权重值,确定与所述境外相关的购买记录。

[0265] 本说明书一个或多个实施例还提出了一种计算机可读存储介质,该计算机可读存储介质存储一个或多个程序,该一个或多个程序包括指令,该指令当被包括多个应用程序的电子设备执行时,能够使该电子设备执行上述信息挖掘方法,并具体用于执行:

[0266] 获取用户的地理位置信息及所述用户的购买记录;

[0267] 根据所述用户的地理位置信息与指定国家的相关性,对所述用户进行地理位置的分类,以确定所述用户的用户类型;以及,对所述购买记录进行特征提取,得到购买记录特征;

[0268] 将所述购买记录特征作为预设的分类模型的模型参数,根据所述用户的用户类型及所述用户的购买记录,对所述分类模型进行训练;

[0269] 根据训练后的模型中的各购买记录特征的权重值,确定与所述指定国家相关的购买记录。

[0270] 上述实施例阐明的系统、装置、模块或单元,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的,计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[0271] 为了描述的方便,描述以上装置时以功能分为各种单元分别描述。当然,在实施本说明书一个或多个实施例时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0272] 本领域内的技术人员应明白,本说明书一个或多个实施例可提供为方法、系统、或计算机程序产品。因此,本说明书一个或多个实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本说明书一个或多个实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0273] 本说明书一个或多个实施例是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0274] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指

令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0275] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0276] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0277] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0278] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0279] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0280] 本说明书一个或多个实施例可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0281] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0282] 以上仅为本说明书一个或多个实施例而已,并不用于限制本说明书。对于本领域技术人员来说,本说明书一个或多个实施例可以有各种更改和变化。凡在本说明书一个或多个实施例的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本说明书一个或多个实施例的权利要求范围之内。

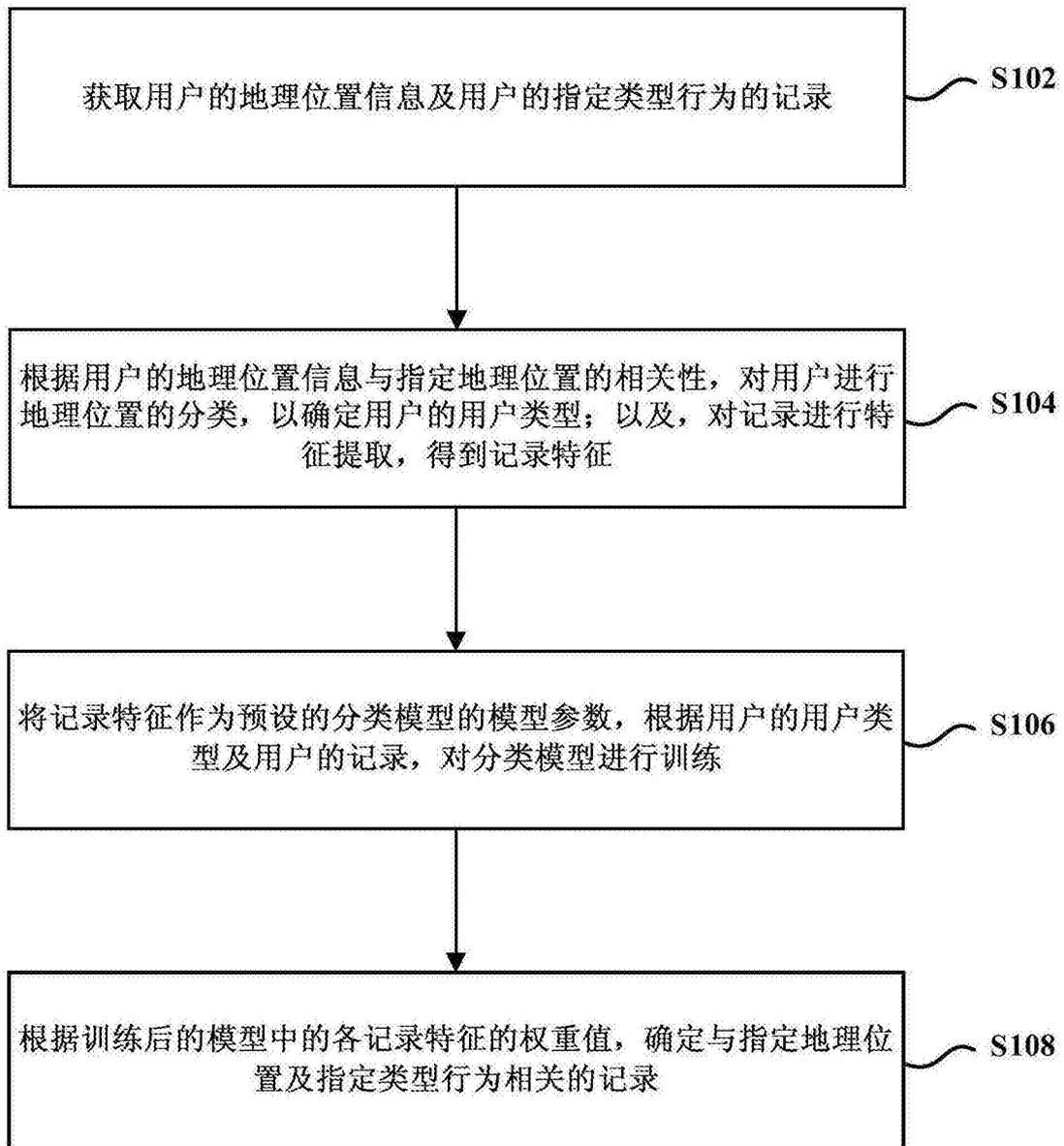


图1

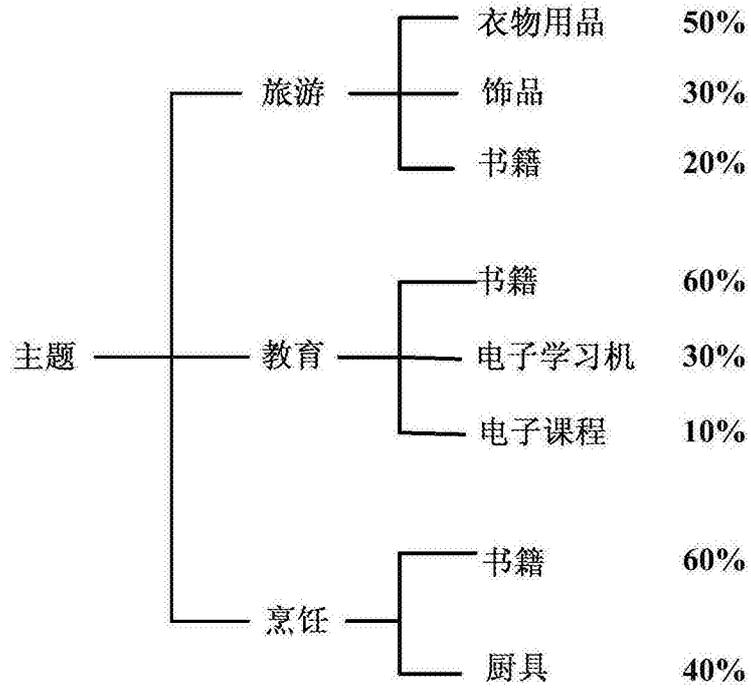


图2

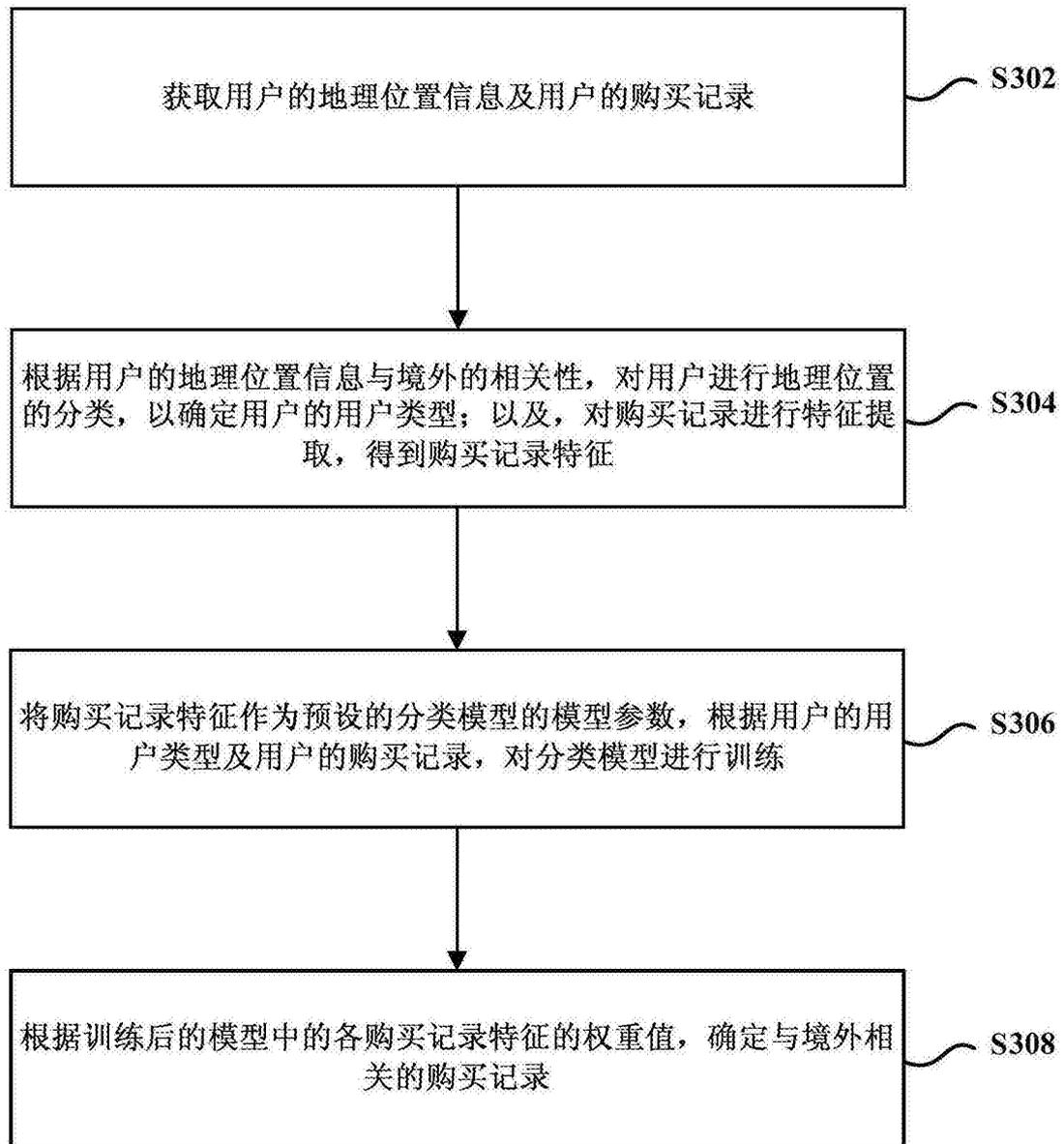


图3

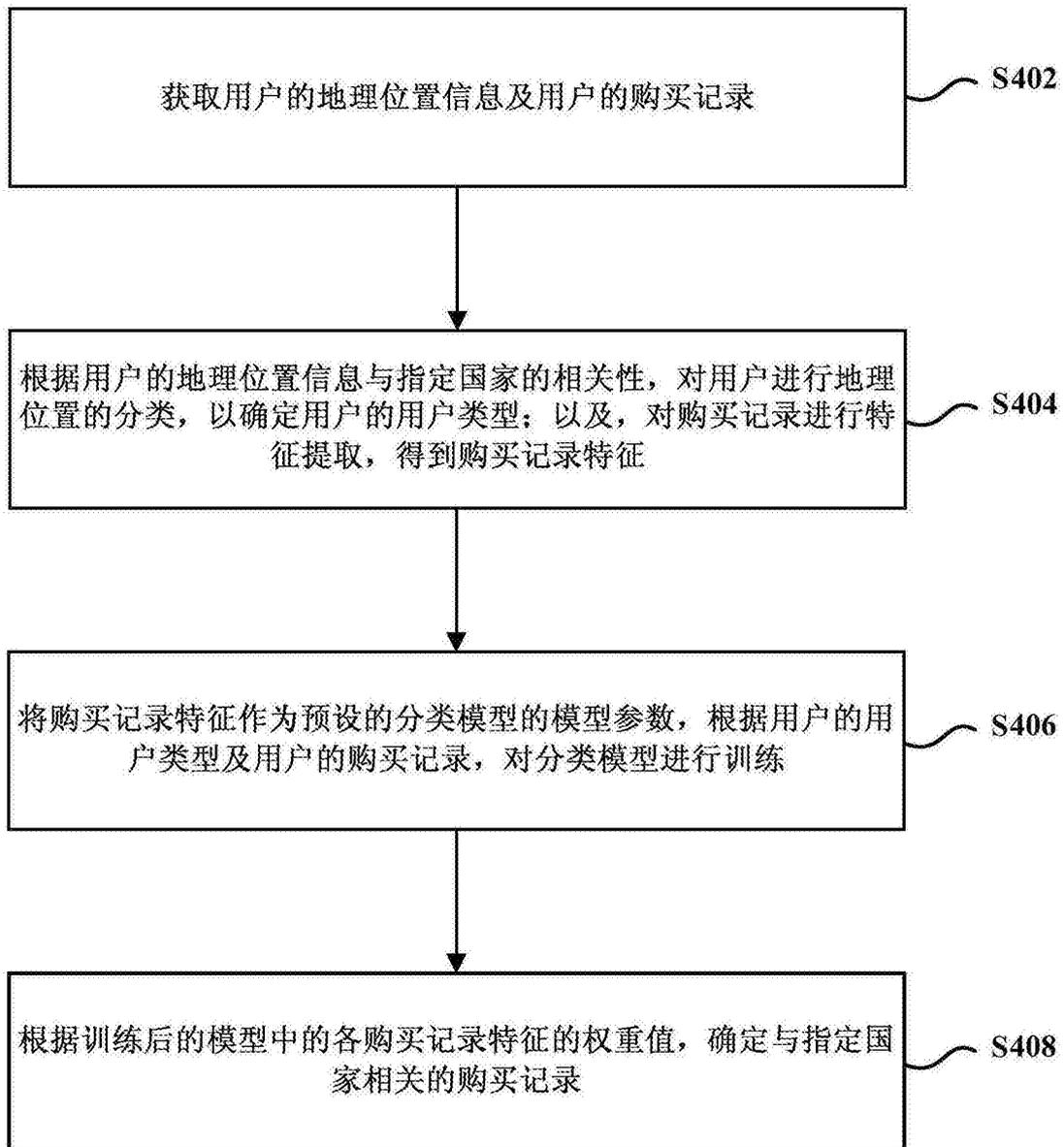


图4



图5



图6



图7

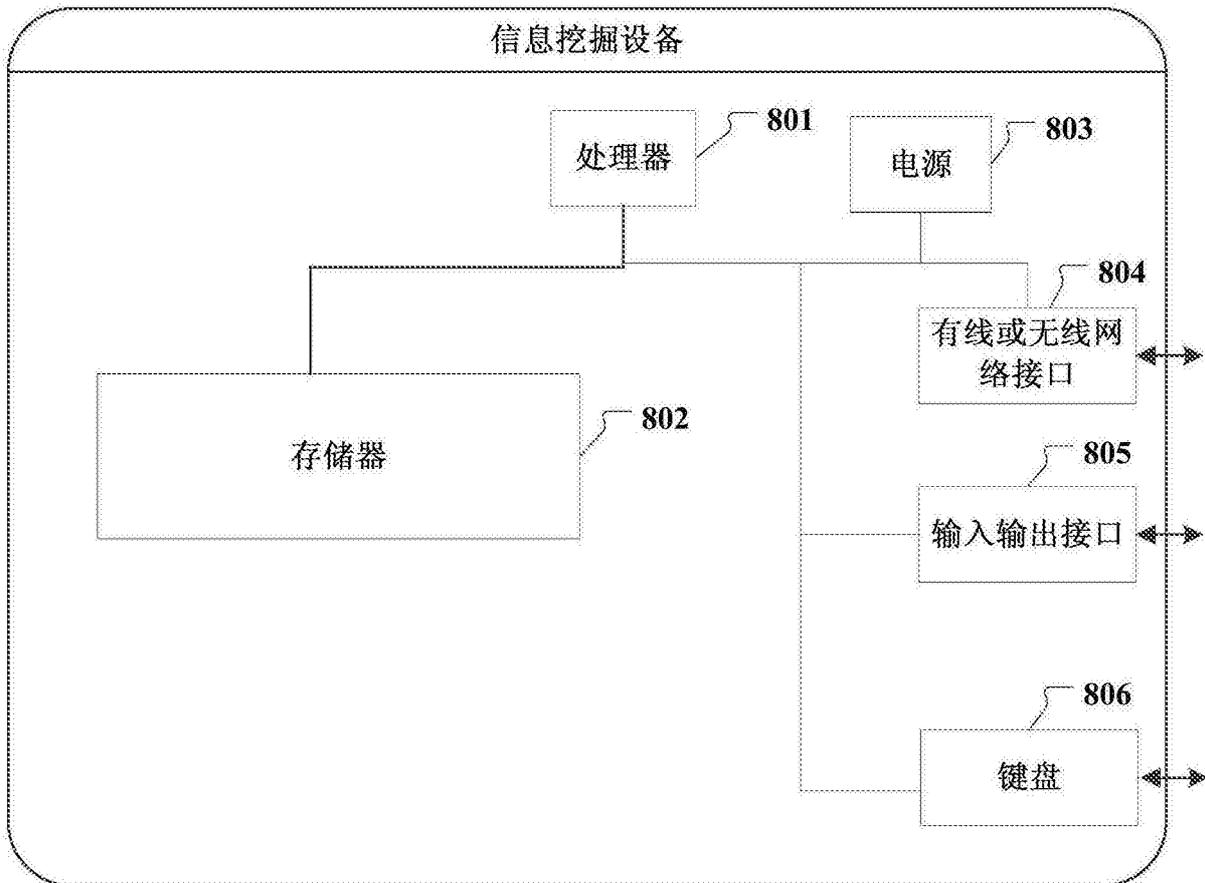


图8