



(12)发明专利申请

(10)申请公布号 CN 107590213 A

(43)申请公布日 2018.01.16

(21)申请号 201710755680.1

(22)申请日 2017.08.29

(71)申请人 重庆邮电大学

地址 400065 重庆市南岸区南山街道崇文路2号

(72)发明人 黄海辉 禹果 蔡晶 刘俊 王欣 余浩

(74)专利代理机构 重庆市恒信知识产权代理有限公司 50102

代理人 李金蓉

(51)Int.Cl.

G06F 17/30(2006.01)

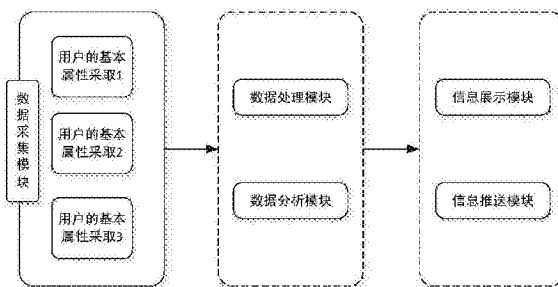
权利要求书1页 说明书4页 附图4页

(54)发明名称

基于手机大数据的混合推荐系统

(57)摘要

本发明公开了一种基于手机大数据的混合推荐系统,包括数据采集模块、数据处理模块、数据分析模块、信息展示模块与信息推送模块。将手机用户看成一个具有社会联系的圈子,以圈子中的好友兴趣度进行推荐,并使用两种推荐算法结合起来形成的混合推荐算法挖进行运算分析,以达到更好的推荐的效果。同时借助Hadoop大数据分析平台,利用其开发分布式程序,实现分布式高速处理数据的功能,对用户手机大数据进行存储和运算分析。



1. 基于手机大数据的混合推荐系统,其特征在於:包括数据采集模块、数据处理模块、数据分析模块、信息展示模块与信息推送模块;

所述数据采集模块,通过手机通信运营商大数据合作平台接口定时的采集脱敏加密后的用户手机数据,再传输到所述数据处理模块;

所述数据处理模块,存储并预处理所述数据采集模块采集到的数据;

所述数据分析模块,提取所述数据处理模块存储并预处理的数据,先根据用户消费水平将所有用户群聚类细分,再使用两种推荐算法结合起来形成的混合推荐算法进行运算分析,并将运算分析的结果输入所述信息展示模块和所述信息推送模块;

所述信息展示模块,接收所述数据分析模块的运算分析结果,提供用户群整体数据信息分析与用户个人自主查询;

所述信息推送模块,接收所述数据分析模块的运算分析结果,并根据分析结果主动向手机用户针对性的推送商品与服务,实现供求链式预测,并根据用户反馈,不断优化系统。

2. 根据权利要求1所述的基于手机大数据的混合推荐系统,其特征在於:所述混合推荐算法,由社区发现算法和关联算法这两种推荐算法构成;输入所述根据用户消费水平将所有用户群聚类细分的结果,先进行所述社区发现算法输出兴趣相似的用户群,再将不同的用户群根据所述关联算法进行挖掘,输出用户感兴趣的属性,进而得到对用户有针对性的推荐并输入所述信息展示模块和所述信息推送模块。

3. 根据权利要求2所述的基于手机大数据的混合推荐系统,其特征在於:所述社区发现算法使用的是GN算法,所述关联算法使用的是Apriori算法。

4. 根据权利要求1所述的基于手机大数据的混合推荐系统,其特征在於:所述数据处理模块和所述数据分析模块是借助Hadoop大数据分析平台,利用其开发分布式程序,实现分布式高速处理数据的功能,对用户手机大数据进行存储和运算分析。

## 基于手机大数据的混合推荐系统

### 技术领域

[0001] 本发明涉及一种基于手机大数据的混合推荐系统,主要用于分析挖掘手机大数据的用户兴趣的信息,实现对用户的个性化推荐。

### 背景技术

[0002] 近年来,由于信息技术的高速发展,科学领域与商业领域(如通信运营商)产生了规模相当巨大的数据,而且这些数据处于不断快速增加的阶段。面对海量的数据,大数据存储与处理面临巨大的挑战,如何快速有效的存储与处理是现在网络技术的一个热点。

[0003] 数据挖掘(Data Mining),是指从大量的数据中通过算法搜索隐藏于其中信息的过程,通过分类(Classification)、聚类(Clustering)、关联(Association)、预测(Predicton)和可视化(Visualization)等算法与技术找出不同的用户,分析消费者喜好与行为的方法,从而为商业和市场创造出巨大的价值。

[0004] 个性化推荐是当前数据挖掘研究的热门领域,它可以很好的提升用户体验,从而提升企业营销。单一个性化推荐算法存在一些缺点,推荐的质量和效率仍然有改善的空间,且传统的个性化推荐算法,在处理用户间关系是,忽略了用户之间的关系。

### 发明内容

[0005] 针对以上的问题,本发明提出一种基于手机大数据的混合推荐系统,将手机用户看成一个具有社会联系的圈子,以圈子中的好友兴趣度进行推荐,并运用两种推荐算法结合起来形成的混合推荐算法,使推荐效果更加准确。结合Hadoop软件平台所具有分布式并行化的结构特点,利用其开发分布式程序,实现分布式高速处理数据的功能,对用户手机大数据进行存储和运算分析。

[0006] 具体包括数据采集模块、数据处理模块、数据分析模块、信息展示模块与信息推送模块。

[0007] 所述数据采集模块,通过手机通信运营商大数据合作平台接口定时的采集脱敏加密后的用户手机数据,再传输到所述数据处理模块。

[0008] 所述数据处理模块,存储并预处理所述数据采集模块采集到的数据;

[0009] 所述数据分析模块,提取所述数据处理模块存储并预处理的数据,先根据用户消费水平将所有用户群聚类细分,再使用两种推荐算法结合起来形成的混合推荐算法进行运算分析,并将运算分析的结果输入所述信息展示模块和所述信息推送模块;

[0010] 所述信息展示模块,接收所述数据分析模块的运算分析结果,提供用户群整体数据信息分析与用户个人自主查询;

[0011] 所述信息推送模块,接收所述数据分析模块的运算分析结果,并根据分析结果主动向手机用户针对性的推送商品与服务,实现供求链式预测,并根据用户反馈,不断优化系统。

[0012] 所述混合推荐算法,由社区发现算法和关联算法这两种推荐算法构成;输入所述

根据用户消费水平将所有用户群聚类细分的结果,先进行所述社区发现算法输出兴趣相似的用户群,再将不同的用户群根据所述关联算法进行挖掘,输出用户感兴趣的属性,进而得到对用户有针对性的推荐并输入所述信息展示模块和所述信息推送模块。

[0013] 所述社区发现算法使用的是GN算法,所述关联算法使用的是Apriori算法。

[0014] 所述数据处理模块和所述数据分析模块是借助Hadoop大数据分析平台,利用其开发分布式程序,实现分布式高速处理数据的功能,对用户手机大数据进行存储和运算分析。

[0015] 本发明的优点及有益效果:

[0016] 两种推荐算法结合起来形成的混合推荐算法,首先达到了用户按兴趣分组的效果,其次在每个群组中以关联规则算法找出用户兴趣的规律,两者相结合极大提高了推荐准确度,可以达到更好的推荐的效果。

[0017] 将手机用户看成一个具有社会联系的圈子,以圈子中的好友兴趣度进行推荐,推荐效果更加准确。

[0018] 利用Hadoop平台强大的数据存储能力,实现海量级手机用户数据的存储。并利用Hadoop平台强大的数据分析计算能力,使用用户社区发现算法、关联算法实现海量级手机用户数据并行化的计算分析,极大提升了因硬件不足导致处理速度过慢的问题。最终实现向手机用户推送个性化的商品和服务,提升了手机大数据的价值。

## 附图说明

[0019] 图1是本发明实施例的总体构成图;

[0020] 图2是本发明实施例的推荐系统层次结构图;

[0021] 图3是社区发现算法的步骤过程;

[0022] 图4为GN算法边介数计算示意图;

[0023] 图5是关联算法的步骤过程;

[0024] 图6是基于社区发现与关联混合推荐算法的说明图。

## 具体实施方式

[0025] 下面将结合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、详细地描述。所描述的实施例仅仅是本发明的一部分实施例。

[0026] 如图1所示,基于手机大数据的混合推荐系统包括数据采集模块、数据处理模块、数据分析模块、信息展示模块与信息推送模块。

[0027] 数据采集模块,通过手机通信运营商大数据合作平台接口定时的采集脱敏加密后的用户手机数据。用户的基本属性采取1表示用户基本属性采集(姓名、年龄、性别、职业、手机号码、在网时长、套餐);用户的基本属性采取2表示用户的行为属性采集(上网行为信息、通话行为信息、短信数据、月流量、消费数据);用户的基本属性采取3表示用户的标签数据采集(终端偏好、APP偏好、消费特征、兴趣、位置数据)。

[0028] 数据处理模块,存储并预处理数据采集模块采集到的数据。

[0029] 数据分析模块,提取数据处理模块存储并预处理的数据,先根据用户消费水平将所有用户群聚类细分,再使用GN算法进行用户社区发现,输出兴趣相似的用户群;然后将不同的用户群分别进行Apriori算法关联规则挖掘,输出用户感兴趣的属性,进而对用户进行

针对性的推荐。

[0030] 数据处理模块和数据分析模块是借助Hadoop大数据分析平台,利用其开发分布式程序,实现分布式高速处理数据的功能,对用户手机大数据进行存储和运算分析。

[0031] 信息展示模块,接收数据分析模块的运算分析结果,提供用户群整体数据信息分析与用户个人自主查询。

[0032] 信息推送模块,接收数据分析模块的运算分析结果,并根据分析结果主动向手机用户针对性的推送商品与服务,实现供求链式预测,并根据用户反馈,不断优化系统。

[0033] 图2为推荐系统的层次结构图,包括存储层、功能层和应用层。存储层包含数据的采集与数据的存储,以及数据的预处理功能;功能层包括三个算法模块:统计模块、分析模块和挖掘模块;应用层包括两个应用模块,分别是信息展示模块与推送模块。

[0034] 其中,存储层的挖掘模块包括聚类算法和包含社区发现算法与关联算法的混合推荐算法;分析模块实现对系统各项数据的监控,并能进行异常报警功能;统计模块根据数据库数据的各种统计功能,针对不同的功能需求实现不同的统计。

[0035] 应用层的推送模块基于手机大数据以及数据挖掘算法,实现主动向手机用户群组的个性化推送,实现供求链式预测,并根据用户反馈,不断优化系统;展示模块用于对用户群整体数据信息分析与用户个人自主查询。

[0036] 存储层的数据存储,存储采取的数据,对原始数据进行数据的抽取、加载以及清洗,为进一步分析做好准备;数据采集,采集手机用户脱敏加密后的数据,其所有数据均通过手机通信运营商大数据合作平台接口定时的采集脱敏加密后的用户手机数据。存储层功能是实现数据的采集、存储以及数据的预处理功能。

[0037] 图3为社区发现算法GN的算法执行流程图,在进行社区群组划分时,进入模板度函数 $Q$ ,  $Q = \sum (e_{ij} - a_i^2)$ ,作为社区划分的衡量标准。其中,如果一个网络分为 $n$ 个社区,会形成一个 $n \times n$ 对称矩阵 $e$ ,  $e_{ij}$ 为对称矩阵中的元素,它代表着社区 $i$ 中的节点与社区 $j$ 中的节点连接的边占整个网络中所有边的比例; $e_{ii}$ 代表对角线的各元素; $a_i = \sum_j e_{ij}$ 表示的是该矩阵中的第 $i$ 行元素和。

[0038] 基于GN算法的社区发现执行步骤如下:

[0039] S11.在网络中,使用Floyd算法计算每条边相对于所有可能的源节点的边介数 $N$ ,然后删除 $N$ 值高的边,对于分裂过程中的第一次分裂,均计算一次网络的模板度 $Q$ 以及图结构。

[0040] S12.对于网络中剩余的边,重复步骤S11,直到网络中所有边均被计算,然后,将具有最大 $Q$ 值对应的网络结构作为该网络的最终分裂结果。

[0041] 图4为GN算法边介数计算示意图。搜寻网络中所有节点之间的最短路径,统计出所有边的边介数,最后统计出最大边介数然后删除,即得到最终分裂结果,也就是社区划分结果。

[0042] 如图5所示,关联算法Apriori算法的步骤过程:

[0043] S21.扫描数据库中手机大数据用户矩阵 $D$ ,计算用户所有属性的支持度(支持度是户属性 $A$ 与用户属性 $B$ 同时出现的概率,  $\text{Support}(A \rightarrow B) = P(A \cap B)$ ),并给定最小支持度阈值(阈值大小的设定的依据是将过小的支持度过滤掉),如果大于给定的最小支持度阈值,则该属性规定为频繁项,由此可以得到的所有用户属性的频繁 $1$ -项集的集合 $L_1$ (如果一个集

合中包含 $k$ 个元素,那么这个 $k$ 项集满足最小支持度阈值的集合称为频繁 $k$ 项集)。

[0044] S22.Apriori算法将在上一次生成的频繁 $(k-1)$ -项集的基础上,链接生成新的候选 $k$ -项集, $k$ 一个集合中包含 $k$ 个元素。

[0045] S23.当没有新的频繁项集产生,即 $K$ 达到最大值的时候,算法结束,得到用户的强关联规则。

[0046] 如图6所示,用户社区发现算法与关联规则推荐算法结合的混合推荐算法,输入数据,使用GN算法进行用户社区发现,输出兴趣相似的用户群;然后获取用户项目矩阵,将不同的用户群进行Apriori算法关联规则挖掘,输出用户推荐列表。

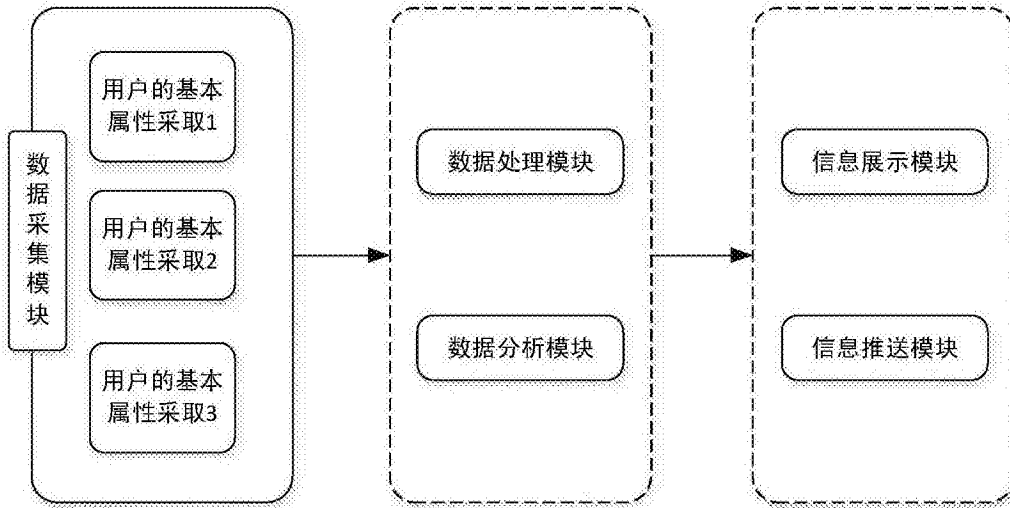


图1

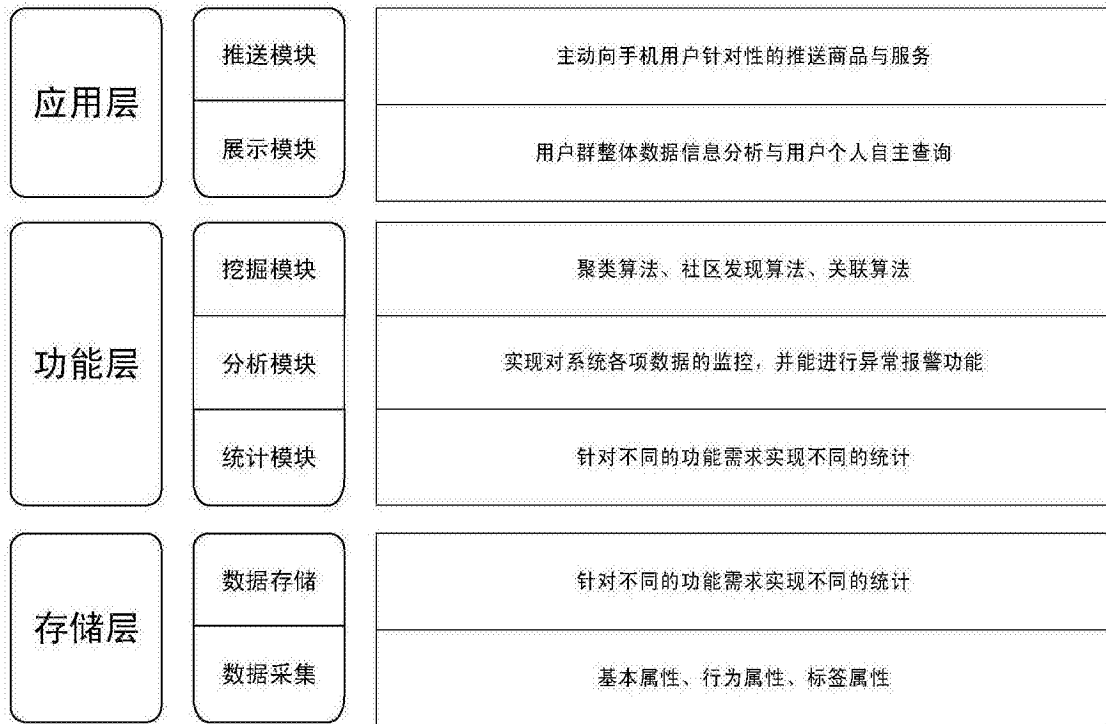


图2

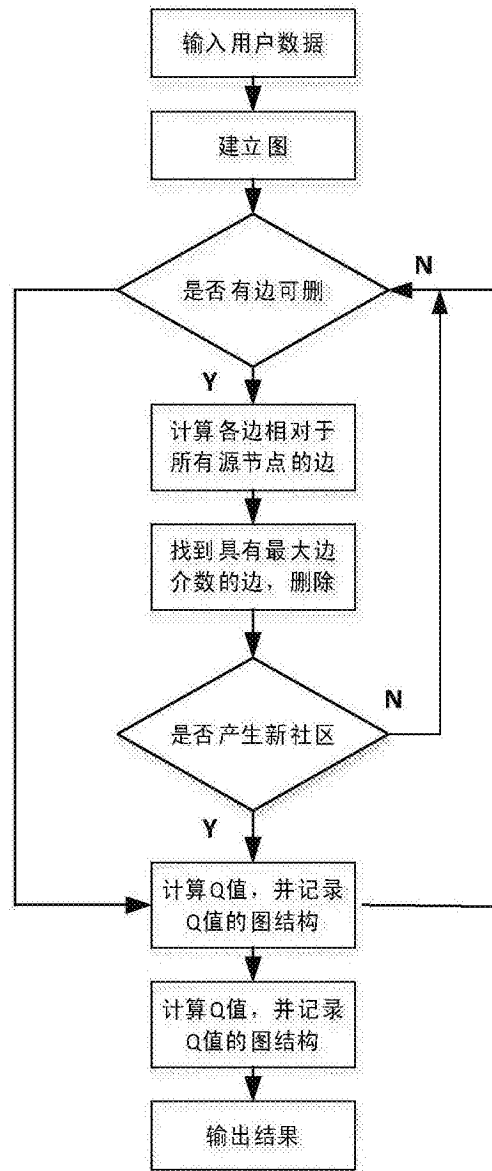


图3



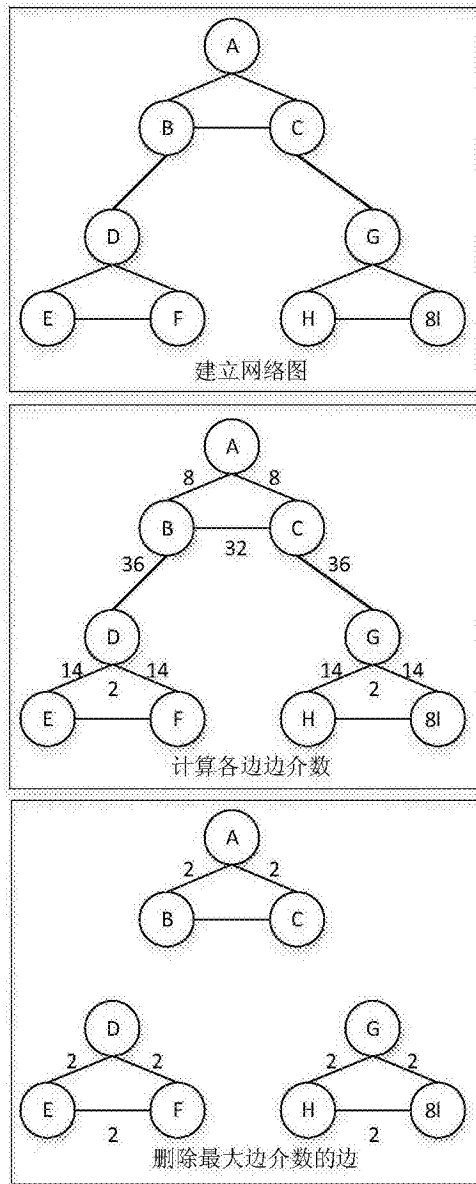


图4

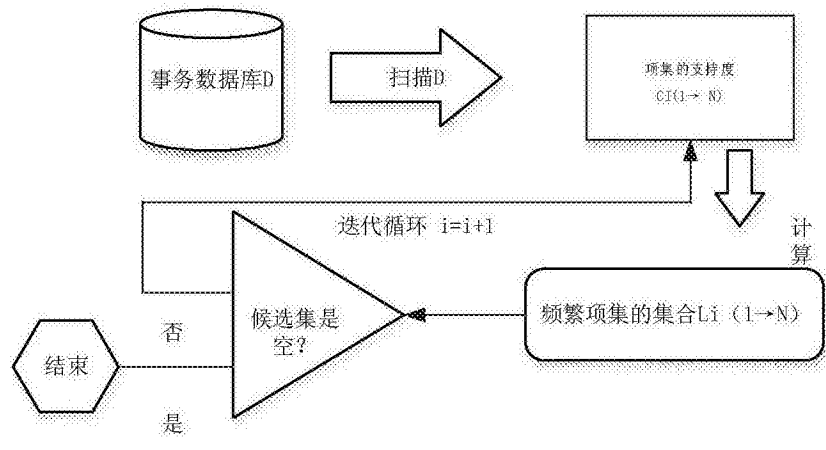


图5

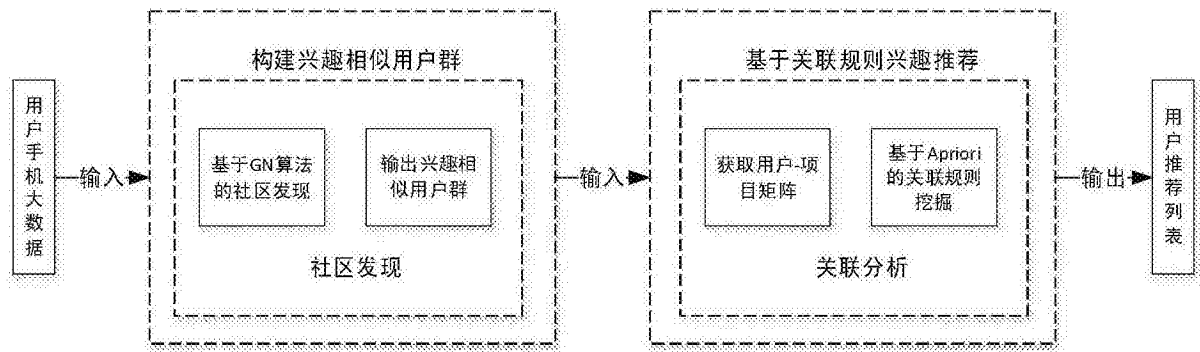


图6