(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109255586 B
(45)授权公告日 2022.03.29

(21)申请号 201810970919.1

(22)申请日 2018.08.24

(65)同一申请的已公布的文献号
申请公布号 CN 109255586 A

(43)申请公布日 2019.01.22

(73)专利权人 安徽讯飞智能科技有限公司
地址 241000 安徽省芜湖市鸠江区皖江财富广场A1座9楼

(72)发明人 水新莹 张宇光 黄亚坤

(74)专利代理机构 芜湖思诚知识产权代理有限公司 34138
代理人 项磊

(51)Int.Cl.
G06Q 10/10(2012.01)
G06Q 50/26(2012.01)
G06F 16/9535(2019.01)

(56)对比文件
CN 103345698 A,2013.10.09
CN 103744957 A,2014.04.23
US 2003140063 A1,2003.07.24
CN 105809475 A,2016.07.27
刘新跃.数字图书馆个性化信息推荐系统.《中国优秀硕士学位论文全文数据库 信息科技辑》.2012,第11-43页.
刘新跃.数字图书馆个性化信息推荐系统.《中国优秀硕士学位论文全文数据库 信息科技辑》.2012,第11-43页.

审查员 周辉

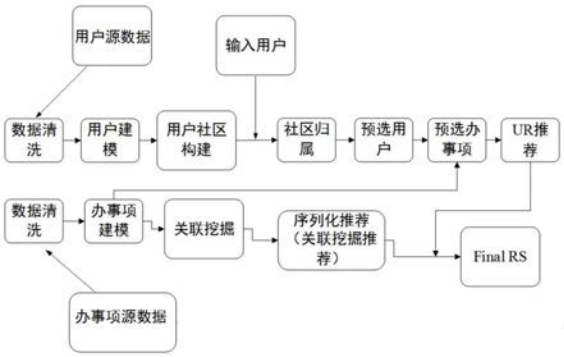
权利要求书3页 说明书7页 附图3页

(54)发明名称

一种面向电子政务办事的在线个性化推荐方法

(57)摘要

本发明公开了一种面向电子政务办事的在线个性化推荐方法,包括如下步骤:1)数据预处理与用户建模;2)建立基于用户社区的办事项推荐模型;3)建立序列模式挖掘的类Apriori算法推荐模型;4)结合用户社区和关联序列挖掘所形成的推送结果推送给目标客户。当被推荐项目具有序列或流程特性领域的推荐,本发明提供的推荐方法比传统推荐算法具有更高的推荐精度,推荐的多样性表现突出,具有一定的应用价值,在实际应用场景下能够有效提升办事服务质量与体验。



CN 109255586 B

1.一种面向电子政务办事的在线个性化推荐方法,其特征在于:包括如下步骤:

1)数据预处理与用户建模;

2)建立基于用户社区的办事项推荐模型;

3)建立序列模式挖掘的类Apriori算法推荐模型;

4)结合用户社区和关联序列挖掘所形成的推送结果推送给目标客户;

步骤3)的具体步骤如下:首先初步监测过滤难以判断办事序列的数据信息,之后对剩下的数据,办事项序列模式挖掘能够有效识别出电子政务系统中的动态系统特征,预测用户在未来一段时间内可能的办事项序列信息;D是包含一个或多个办事项序列,即与单个用户相关联的办事项有序集合;具体模型建立步骤如下:

定义序列S的支持度为S包含的所有办事项序列所占的比例,若S的支持度大于或等于阈值minsup,则S即为一个序列模式;

序列模式的挖掘枚举所有可能的序列,再进行支持度计算,对于n个办事项,依次对1个办事项,2个办事项,3个办事项直到n个办事项进行枚举;

由于先验原理对序列数据成立,因此包含特定k个办事项的任何序列必然包括该k个办事项的所有k-1个办事项的子序列,根据候选序列的支持度和上述阈值提取支持度不小于阈值的序列模式,并由此获得序列模式的关联挖掘预选推荐办事项集合RS_A;

步骤4)的具体步骤如下:在完成用户社区挖掘及相似用户预选推荐办事项集合RS_U和序列模式的关联挖掘预选推荐办事项集合RS_A之后,需要根据两部分结果的重叠程度进行优化筛选,形成最终的办事项推荐结果集合,并实时的推送给目标用户;

若假定推送给目标用户的办事项数目为K,对于RS_U和RS_A两种结果集的覆盖结果筛选,细分为三种情况:

(a).覆盖结果集能够满足$N(RS_A \cap RS_U) \geq K$,$Top(RS_A)$被选择作为最终推荐结果集合;

(b).当$N(RS_A \cap RS_U) \leq K$,覆盖部分优先作为推荐结果;对于较多未覆盖部分的办事项,分别选取$K-N(RS_A \cap RS_U)/2$个结果集作为最终结果集;

(c).完全无覆盖的推荐结果计算采用(b)的未覆盖部分的筛选结果。

2.根据权利要求1所述的一种面向电子政务办事的在线个性化推荐方法,其特征在于:步骤1)中,抽取截止当前时间的用户数据以及用户的历史办事行为数据,并对数据进行常规数据清洗和数据集成,再对用户和办事项的静态基本属性以及动态行为属性进行特征建模;

步骤2)中,构建用户办事行为关系矩阵的用户社区,用户办事行为关系网络反映了用户办事行为间关联的紧密程度,再基于用户社区的办事项推荐算法,获取到基于用户社区的办事项推荐待选集合;

步骤3)中,初步监测过滤难以判断办事序列的数据信息,建立单个用户相关联的办事项有序集合,定义序列的支持度为该序列包含的所有办事项序列所占的比例,序列模式的挖掘可以枚举所有可能的序列,再进行支持度计算,最后通过计算提取的序列模式,获得序列模式的关联挖掘预选推荐办事项集合;

步骤4)中,将步骤2)获得的基于用户社区的办事项推荐待选集合与步骤3)获得的序列模式的关联挖掘预选推荐办事项集合结合根据两部分结果的重叠程度进行优化筛选,形成最终的办事项推荐结果集合,并实时的推送给目标用户。

3.根据权利要求2所述的一种面向电子政务办事的在线个性化推荐方法，其特征在于：静态属性包括基本属性和个人隐私，基本属性包括ID、性别、年龄、婚姻和学历，个人隐私包括社保、公积金和医保信息；动态行为属性包括用户的历史行为信息和用户的反馈信息，所述历史行为信息包括已办理事和浏览行为记录，所述用户的反馈信息是用户对推送内容反馈的信息。

4.根据权利要求2所述的一种面向电子政务办事的在线个性化推荐方法，其特征在于：步骤1)的具体步骤如下：

步骤一：抽取截止当前时间的用户数据以及用户的历史办事行为数据，并对数据进行常规数据清洗，数据集成等处理，提高数据质量，同时也便于推荐中的量化计算；

步骤二：用户建模，用户模型为用户社区的划分和准确推荐提供计算基础，用户的静态和动态属性是计算用户之间相似程度的主要根据，用户社区划分依据用户自身属性特征，主要包括BI_U，SI有效属性特征向量组合；

步骤2)的具体步骤如下：

步骤三、若使用G表示用户的隐式办事关系网，G中的结点表示不同的办事用户，用户之间的链接表示存在办事记录交集，链接上的权值则反应了用户间的关系强弱，计算方式主要是基于用户的属性相似度和办事行为交集程度，对G的存储采用压缩优化的邻接矩阵V存储，电子政务用户办事数据集，用户办事关系中的权值设定不仅仅与办事行为关系有关，还与用户的基本属性有关，即用户之间的相似程度，对用户的相似度已进行归一化处理；

步骤四、基于用户社区的办事项推荐算法，根据步骤三构建的用户社区，其中，S(u,K)，是用户u最相似最高的K个用户，通过对用户对某一办事项是否有过行为进行记录，通过给定K值，即可获取用户可能性最高的待办事项集合，表示为RS_U，即获取到基于用户社区的办事项推荐待选集合；

步骤3)的具体步骤如下：

步骤五、定义序列S的支持度为S包含的所有办事项序列所占的比例，若S的支持度大于或等于阈值minsup，则S即为一个序列模式；

步骤六、序列模式的挖掘可以枚举所有可能的序列，再进行支持度计算，如对于n个办事项，依次对1个办事项，2个办事项，3个办事项直到n个办事项进行枚举；

步骤七、由于先验原理对序列数据成立，因此包含特定k个办事项的任何序列必然包括该k个办事项的所有k-1个办事项的子序列；基于Apriori算法挖掘用户办事项记录中的序列模式，最终根据支持度和步骤五提供的阈值提取序列模式，并由此获得序列模式的关联挖掘预选推荐办事项集合RS_A；

步骤4)的具体步骤如下：

步骤八、在完成用户社区挖掘及相似用户预选推荐办事项集合RS_U和序列模式的关联挖掘预选推荐办事项集合RS_A之后，需要根据两部分结果的重叠程度进行优化筛选，形成最终的办事项推荐结果集合，并实时的推送给目标用户。

5.根据权利要求4所述的一种面向电子政务办事的在线个性化推荐方法，其特征在于：所述步骤二还包括下列子步骤：

步骤2.1、用户样本可以表示为包含上述属性的n维向量$\vec{u}=(age,sex,ss,hi,pf,sa,ho,li,ca)$，所有维度的取值范围为0或1，当ss,hi,pf,ho,li,ca取值为1时，表示用户拥有社保、医保、

公积金、房产、驾照和车辆;sex取值1表示性别相同;而age取值为1表示任意两条记录的年龄属于同一年龄段,0为否;age和ss指在进行两个不同用户计算相似度时,考虑年龄或工资等级是否是同一年龄段或工资等级;如存在用户样本A与样本B,其基本属性向量$\vec{u}_A=(1,0,1,1,1,0,1,1,0)$和$\vec{u}_B=(1,0,1,1,1,1,0,1,1)$分别表示A和B性别、年龄段相同,工资等级不同以及其他属性的拥有状态;

步骤2.2、基于用户向量,若用M_{00}代表向量A和向量B都是0的维度个数;M_{01}代表向量A是0而向量B是1的维度个数;M_{10}代表向量A是1而向量B是0的维度个数;M_{11}代表向量A和向量B都是1的维度个数,n维向量的每一维都会落入上述向量中的某一类,利用$Jac(A,B)=M_{11}/M_{01}+M_{10}+M_{11}$计算出用户在基础属性$BI_U$和隐私属性SI的相似度,属性的类别区分主要是解决不同属性对相似度结果的偏好影响;接着,利用$sim(u_i,u_j)=(1-\alpha)\cdot b_sim(u_i,u_j)+\alpha\cdot s_sim(u_i,u_j)$计算融合相似度,α=0.63获得最优结果。

6.根据权利要求4所述的一种面向电子政务办事的在线个性化推荐方法,其特征在于:所述步骤三中,计算基于用户的属性相似度和办事行为交集程度的相关规则如下:

a.用户之间无任何办事行为交集,且用户基础属性相似度低,则判定为无链接行为;

b.用户间存在办事行为交集,且用户基础属性相似度低,则链接权值为办事记录的相似系数;

c.用户间存在办事行为交集,且用户基础属性相似度高,则链接权值为两种相似度之和。

7.根据权利要求4所述的一种面向电子政务办事的在线个性化推荐方法,其特征在于:所述步骤三还包括下列子步骤:

步骤3.1、利用模块度公式$Q=\frac{1}{2m}\sum_{ij}[A_{ij}-\frac{k_ik_j}{2m}]\cdot\delta(C_i,C_j)$计算社区划分的程度,其中,$A_{ij}$为连接节点i和j边的权值;m为网络中边的数量;$k_i$为节点i的度;$k_j$为节点j的度;$C_i$为i所属的社区;

步骤3.2、用户社区划分采用层次贪心算法,筛选出K个与目标用户相似度最大的用户集合,算法主要包括两个阶段,第一阶段合并社区,初始状态将每个节点视为独立社区,基于最近邻居相似度最大标准决定哪些社区应该被合并;第二阶段,将第一阶段发现的社区重新视为独立节点社区,重复构建,这两个阶段重复进行,直到网络社区划分的模块度趋于稳定。

8.根据权利要求4所述的一种面向电子政务办事的在线个性化推荐方法,其特征在于:所述步骤四还包括下列子步骤:

步骤4.1、区别于传统的音乐、电影类的评分推荐,用户与办事项之间不存在评分,仅具有办理或未办理的状态值,通过对用户对某一办事项是否有过行为进行记录,1表示用户办理或浏览办事项,状态0表示对办事项无行为记录,令$r_{i,j}=\{0,1\}$表示第i个用户对第j个项目的办事记录行为,由于$r_{i,j}$取值的特殊性,计算公式采用Jaccard相似度进行计算;

步骤4.2、通常大多数用户对于基础的热门办事项都会办理,这将造成用户的相似度差异较小,考虑在计算行为相似度时,对热门事项进行惩罚,通过给定K值,即可获取用户可能性最高的待办事项集合,表示为RS_U。

一种面向电子政务办事的在线个性化推荐方法

技术领域

[0001] 本发明涉及智慧城市和电子政务领域，具体涉及一种面向电子政务办事的在线个性化推荐方法。

背景技术

[0002] 随着以互联网为主的信息新技术在经济、社会生活各部分的扩散和应用，“互联网+政务”以电子政务服务平台为基础，以实现智慧政府为目标，对政府组织结构和办事流程进行优化重组。传统的电子政务系统缺乏面向用户个性化需求的精准服务，独立、多源、异构的政务信息增加了用户的办事难度。结合个性化推荐的电子政务系统能够根据用户画像和动态行为特征进行建模分析，推送符合该用户特征的相关项目，进一步提高了用户体验。

[0003] 传统的协同过滤推荐算法，如基于内容和项目的推荐，基于矩阵分解以及由其衍生而出的具有偏好的矩阵分解算法或结合上下文的推荐算法已在电商、音乐和电影等领域推荐取得一定成果。例如集成语义相似度与协同过滤的推荐算法提供了准确性和扩展性更高的个性化推荐服务，主要有结合模糊描述逻辑语言提出一种模糊语义的推荐服务来促进电子政务中的资源信息利用；通过集成语义相似性和传统的基于项目的协同过滤解决电子政务服务中唯一项目的推荐问题；基于增强推荐中的混合语义信息提出了一种项目语义相关性模型，并开发了智能商业定位器推荐系统原型进行验证。基于本体理论的相关算法提供了主动推送、动态、差异的个性化推荐等。

[0004] 然而，电子政务办事项的序列化特征难以直接将传统个性化推荐算法直接进行应用推广。已有推荐方法主要对推荐算法进行改进优化设计，缺乏结合电子政务办事项的业务特征向用户推荐更精准的服务。因此，综合考虑电子政务办事项的序列化特征，设计出符合电子政务特点的推荐算法是构建智慧城市的关键技术之一。

发明内容

[0005] 本发明的目的在于提供一种面向电子政务办事的在线个性化推荐方法，以解决现有技术缺乏结合电子政务办事项的业务特征向用户推荐更精准的服务能力的缺陷。

[0006] 所述的面向电子政务办事的在线个性化推荐方法，包括如下步骤：

[0007] 1)数据预处理与用户建模；

[0008] 2)建立基于用户社区的办事项推荐模型；

[0009] 3)建立序列模式挖掘的类Apriori算法推荐模型；

[0010] 4)结合用户社区和关联序列挖掘所形成的推送结果推送给目标客户。

[0011] 优选的，步骤1)中，抽取截止当前时间的用户数据以及用户的历史办事行为数据，并对数据进行常规数据清洗和数据集成，再对用户和办事项的静态基本属性以及动态行为属性进行特征建模；

[0012] 步骤2)中，构建用户办事行为关系矩阵的用户社区，用户办事行为关系网络反映了用户办事行为间关联的紧密程度，再基于用户社区的办事项推荐算法，获取到基于用户

社区的办事项推荐待选集合；

[0013] 步骤3)中，初步监测过滤难以判断的房屋账单数据，建立单个用户相关联的办事项有序集合，定义序列的支持度为该序列包含的所有办事项序列所占的比例，序列模式的挖掘可以枚举所有可能的序列，再进行支持度计算，最后通过计算提取的序列模式，获得序列模式的关联挖掘预选推荐办事项集合；

[0014] 步骤4)中，将步骤2)获得的基于用户社区的办事项推荐待选集合与步骤3)获得的序列模式的关联挖掘预选推荐办事项集合结合根据两部分结果的重叠程度进行优化筛选，形成最终的办事项推荐结果集合，并实时的推送给目标用户。

[0015] 优选的，静态属性包括基本属性和个人隐私，基本属性包括ID、性别、年龄、婚姻和学历，个人隐私包括社保、公积金和医保信息；动态行为属性包括用户的历史行为信息和用户的反馈信息，所述历史行为信息包括已办理事和浏览行为记录，所述用户的反馈信息是用户对推送内容反馈的信息。

[0016] 优选的，步骤1)的具体步骤如下：

[0017] 步骤一：抽取截止当前时间的用户数据以及用户的历史办事行为数据，并对数据进行常规数据清洗，数据集成等处理，提高数据质量，同时也便于推荐中的量化计算；

[0018] 步骤二：用户建模，用户模型为用户社区的划分和准确推荐提供计算基础，用户的静态和动态属性是计算用户之间相似程度的主要根据，用户社区划分依据用户自身属性特征，主要包括BI_U,SI有效属性特征向量组合；

[0019] 步骤2)的具体步骤如下：

[0020] 步骤三、若使用G表示用户的隐式办事关系网，G中的结点表示不同的办事用户，用户之间的链接表示存在办事记录交集，链接上的权值则反应了用户间的关系强弱，计算方式主要是基于用户的属性相似度和办事行为交集程度，对G的存储采用压缩优化的邻接矩阵V存储，电子政务用户办事数据集，用户办事关系中的权值设定不仅仅与办事行为关系有关，还与用户的基本属性有关，即用户之间的相似程度，对用户的相似度已进行归一化处理；

[0021] 步骤四、基于用户社区的办事项推荐算法，根据步骤三构建的用户社区，其中，S(u,K)，是用户u最相似最高的K个用户，通过对用户对某一办事项是否有过行为进行记录，通过给定K值，即可获取用户可能性最高的待办事项集合，表示为RS_U，即获取到基于用户社区的办事项推荐待选集合；

[0022] 步骤3)的具体步骤如下：

[0023] 步骤五、定义序列S的支持度为S包含的所有办事项序列所占的比例，若S的支持度大于或等于阈值minsup，则S即为一个序列模式；

[0024] 步骤六、序列模式的挖掘可以枚举所有可能的序列，再进行支持度计算，如对于n个办事项，依次对1个办事项，2个办事项，3个办事项直到n个办事项进行枚举；

[0025] 步骤七、由于先验原理对序列数据成立，因此包含特定k个办事项的任何序列必然包括该k个办事项的所有k-1个办事项的子序列；基于Apriori算法挖掘用户办事项记录中的序列模式，最终根据支持度和步骤五提供的阈值提取序列模式，并由此获得序列模式的关联挖掘预选推荐办事项集合RS_A；

[0026] 步骤4)的具体步骤如下：

[0027] 步骤八、在完成用户社区挖掘及相似用户预选推荐办事项集合RS_U和序列模式的关联挖掘预选推荐办事项集合RS_A之后，需要根据两部分结果的重叠程度进行优化筛选，形成最终的办事项推荐结果集合，并实时的推送给目标用户。

[0028] 优选的，所述步骤二还包括下列子步骤：

[0029] 步骤2.1、用户样本可以表示为包含上述属性的n维向量$\vec{u}=(age,sex,ss,hi,pf,sa,ho,li,ca)$，所有维度的取值范围为0或1，当ss,hi,pf,ho,li,ca取值为1时，表示用户拥有社保、医保、公积金、房产、驾照和车辆；sex取值1表示性别相同；而age取值为1表示任意两条记录的年龄属于同一年龄段，0为否；age和ss指在进行两个不同用户计算相似度时，考虑年龄或工资等级是否是同一年龄段或工资等级；如存在用户样本A与样本B，其基本属性向量$\vec{u}_A=(1,0,1,1,1,0,1,1,0)$和$\vec{u}_B=(1,0,1,1,1,1,0,1,1)$分别表示A和B性别、年龄段相同，工资等级不同以及其他属性的拥有状态；

[0030] 步骤2.2、基于上述用户向量，若用M_{00}代表向量A和向量B都是0的维度个数；M_{01}代表向量A是0而向量B是1的维度个数；M_{10}代表向量A是1而向量B是0的维度个数；M_{11}代表向量A和向量B都是1的维度个数，n维向量的每一维都会落入上述向量中的某一类，利用Jac(A,B)＝$M_{11}/M_{01}+M_{10}+M_{11}$计算出用户在基础属性$BI_U$和隐私属性SI的相似度，属性的类别区分主要是解决不同属性对相似度结果的偏好影响；接着，利用$sim(u_i,u_j)=(1-\alpha)\cdot b_sim(u_i,u_j)+\alpha\cdot s_sim(u_i,u_j)$计算融合相似度，α＝0.63获得最优结果。

[0031] 优选的，所述步骤三中，计算基于用户的属性相似度和办事行为交集程度的相关规则如下：

[0032] a.用户之间无任何办事行为交集，且用户基础属性相似度低，则判定为无链接行为；

[0033] b.用户间存在办事行为交集，且用户基础属性相似度低，则链接权值为办事记录的相似系数；

[0034] c.用户间存在办事行为交集，且用户基础属性相似度高，则链接权值为两种相似度之和。

[0035] 优选的，所述步骤三还包括下列子步骤：

[0036] 步骤3.1、利用模块度公式$Q=\frac{1}{2m}\sum_{ij}[A_{ij}-\frac{k_ik_j}{2m}]\cdot\delta(C_i,C_j)$计算社区划分的程度，其中，$A_{ij}$为连接节点i和j边的权值；m为网络中边的数量；$k_i$为节点i的度；$k_j$为节点j的度；$C_i$为i所属的社区；

[0037] 步骤3.2、用户社区划分采用层次贪心算法，筛选出K个与目标用户相似度最大的用户集合，算法主要包括两个阶段，第一阶段合并社区，初始状态将每个节点视为独立社区，基于最近邻居相似度最大标准决定哪些社区应该被合并；第二阶段，将第一阶段发现的社区重新视为独立节点社区，重复构建，这两个阶段重复进行，直到网络社区划分的模块度趋于稳定。

[0038] 优选的，所述步骤四还包括下列子步骤：

[0039] 步骤4.1、区别于传统的音乐、电影类的评分推荐，用户与办事项之间不存在评分，仅具有办理或未办理的状态值，通过对用户对某一办事项是否有过行为进行记录，1表示用

户办理或浏览办事项，状态0表示对办事项无行为记录，令$r_{i,j}$＝{0,1}表示第i个用户对第j个项目的办事记录行为，由于$r_{i,j}$取值的特殊性，计算公式采用Jaccard相似度进行计算；

[0040] 步骤4.2、通常大多数用户对于基础的热门办事项都会办理，这将造成用户的相似度差异较小，考虑在计算行为相似度时，对热门事项进行惩罚，通过给定K值，即可获取用户可能性最高的待办事项集合，表示为RS_U。

[0041] 优选的，所述步骤八中，若假定推送给目标用户的办事项数目为K，对于RS_U和RS_A两种结果集的覆盖结果筛选，可细分为图3所示的三种情况：

[0042] (a).覆盖结果集能够满足$N(RS_A \cap RS_U) \geq K$，$Top(RS_A)$被选择作为最终推荐结果集合；

[0043] (b).当$N(RS_A \cap RS_U) \leq K$，覆盖部分优先作为推荐结果；对于较多未覆盖部分的办事项，分别选取$K-N(RS_A \cap RS_U)/2$个结果集作为最终结果集；

[0044] (c).完全无覆盖的推荐结果计算采用(b)的未覆盖部分的筛选结果。

[0045] 本发明的优点在于：首先，本发明提高推荐结果的多样性，减少推荐过程中的计算量；其次，办事项的关联序列挖掘充分考虑了电子政务的业务特性，加入时间维度的办事项序列挖掘进一步提高了推荐结果的精度。此外，通过对用户脱敏后的信息基于Spark计算平台对提出的发明方法进行验证，结果表明当被推荐项目具有序列或流程特性领域的推荐，比传统推荐算法具有更高的推荐精度，推荐的多样性表现突出，具有很好的应用价值，在实际应用场景下能够有效提升办事服务质量与体验。

附图说明

[0046] 图1为本发明面向电子政务办事的在线个性化推荐方法的整体流程图；

[0047] 图2为本发明中基于用户社区的办事项推荐算法获取到基于用户社区的办事项推荐待选集合的流程图；

[0048] 图3为本发明中基于Apriori算法获取基于用户办事项记录中的序列模式的流程图。

具体实施方式

[0049] 下面对照附图，通过对实施例的描述，对本发明具体实施方式作进一步详细的说明，以帮助本领域的技术人员对本发明的发明构思、技术方案有更完整、准确和深入的理解。

[0050] 如图1-3所示，本发明提供了一种面向电子政务办事的在线个性化推荐方法。具体步骤如下：

[0051] 1)数据预处理与用户建模。

[0052] 步骤一：抽取截止当前时间的用户数据以及用户的历史办事行为数据，并对数据进行常规数据清洗，数据集成等处理，提高数据质量，同时也便于推荐中的量化计算。

[0053] 步骤二：用户建模，用户模型为用户社区的划分和准确推荐提供计算基础，用户的静态和动态属性是计算用户之间相似程度的主要根据。用户社区划分依据用户自身属性特征，主要包括BI_U,SI有效属性特征向量组合。包括以下步骤：

[0054] 步骤2.1、用户样本可以表示为包含上述属性的n维向量

$\vec{u}=(age,sex,ss,hi,pf,sa,ho,li,ca)$，所有维度的取值范围为0或1，当ss,hi,pf,ho,li,ca取值为1时，表示用户拥有社保、医保、公积金、房产、驾照和车辆；sex取值1表示性别相同；而age取值为1表示任意两条记录的年龄属于同一年龄段，0为否；age和ss指在进行两个不同用户计算相似度时，考虑年龄或工资等级是否是同一年龄段或工资等级。如存在用户样本A与样本B，其基本属性向量$\vec{u}_A=(1,0,1,1,1,0,1,1,0)$和$\vec{u}_B=(1,0,1,1,1,1,0,1,1)$分别表示A和B性别、年龄段相同，工资等级不同以及其他属性的拥有状态。

[0055] 步骤2.2、基于上述用户向量，若用M_{00}代表向量A和向量B都是0的维度个数；M_{01}代表向量A是0而向量B是1的维度个数；M_{10}代表向量A是1而向量B是0的维度个数；M_{11}代表向量A和向量B都是1的维度个数，n维向量的每一维都会落入上述向量中的某一类，利用Jac(A,B)＝$M_{11}/M_{01}+M_{10}+M_{11}$计算出用户在基础属性$BI_U$和隐私属性SI的相似度，属性的类别区分主要是解决不同属性对相似度结果的偏好影响。接着，利用$sim(u_i,u_j)=(1-\alpha)\cdot b_sim(u_i,u_j)+\alpha\cdot s_sim(u_i,u_j)$计算融合相似度，α＝0.63获得最优结果。

[0056] 2)建立基于用户社区的办事项推荐模型。

[0057] 步骤三、若使用G表示用户的隐式办事关系网，G中的结点表示不同的办事用户，用户之间的链接表示存在办事记录交集，链接上的权值则反应了用户间的关系强弱，计算方式主要是基于用户的属性相似度和办事行为交集程度。对G的存储采用压缩优化的邻接矩阵V存储。电子政务用户办事数据集，用户办事关系中的权值设定不仅仅与办事行为关系有关，还与用户的基本属性有关，即用户之间的相似程度，对用户的相似度已进行归一化处理，相关规则如下：

[0058] a.用户之间无任何办事行为交集，且用户基础属性相似度低，则判定为无链接行为；

[0059] b.用户间存在办事行为交集，且用户基础属性相似度低，则链接权值为办事记录的相似系数；

[0060] c.用户间存在办事行为交集，且用户基础属性相似度高，则链接权值为两种相似度之和。

[0061] 之后构建用户社区的步骤如下：

[0062] 步骤3.1、利用模块度公式$Q=\frac{1}{2m}\sum_{ij}[A_{ij}-\frac{k_ik_j}{2m}]\cdot\delta(C_i,C_j)$计算社区划分的程度。其中，$A_{ij}$为连接节点i和j边的权值；m为网络中边的数量；$k_i$为节点i的度；$k_j$为节点j的度；$C_i$为i所属的社区。

[0063] 步骤3.2、用户社区划分采用层次贪心算法，筛选出K个与目标用户相似度最大的用户集合。算法主要包括两个阶段，第一阶段合并社区，初始状态将每个节点视为独立社区，基于最近邻居相似度最大标准决定哪些社区应该被合并；第二阶段，将第一阶段发现的社区重新视为独立节点社区，重复构建。这两个阶段重复进行，直到网络社区划分的模块度趋于稳定。

[0064] 步骤四、基于用户社区的办事项推荐算法。根据步骤三构建的用户社区，其中，S(u,K)，是用户u最相似最高的K个用户，通过对用户对某一办事项是否有过行为进行记录。通过给定K值，即可获取用户可能性最高的待办事项集合，表示为RS_U，即获取到基于用户社

区的办事项推荐待选集合。具体步骤如下：

[0065] 步骤4.1、区别于传统的音乐、电影类的评分推荐，用户与办事项之间不存在评分，仅具有办理或未办理的状态值。通过对用户对某一办事项是否有过行为进行记录，1表示用户办理或浏览办事项，状态0表示对办事项无行为记录。令$r_{i,j}=\{0,1\}$表示第i个用户对第j个项目的办事记录行为，由于$r_{i,j}$取值的特殊性，计算公式采用Jaccard相似度进行计算。

[0066] 步骤4.2、通常，大多数用户对于基础的热门办事项都会办理，这将造成用户的相似度差异较小。考虑在计算行为相似度时，对热门事项进行惩罚，通过给定K值，即可获取用户可能性最高的待办事项集合，表示为RS_U。附图2给出了基于用户社区的办事项推荐算法。

[0067] 3)建立序列模式挖掘的类Apriori算法推荐模型。

[0068] 首先初步监测过滤难以判断办事序列的数据信息，如房屋账单数据，之后对剩下的数据，办事项序列模式挖掘能够有效识别出电子政务系统中的动态系统特征，预测用户在未来一段时间内可能的办事项序列信息。D是包含一个或多个办事项序列，即与单个用户相关联的办事项有序集合。具体模型建立步骤如下：

[0069] 步骤五、定义序列S的支持度为S包含的所有办事项序列所占的比例。若S的支持度大于或等于阈值minsup，则S即为一个序列模式。

[0070] 步骤六、序列模式的挖掘可以枚举所有可能的序列，再进行支持度计算，如对于n个办事项，依次对1个办事项，2个办事项，3个办事项直到n个办事项进行枚举。

[0071] 步骤七、由于先验原理对序列数据成立，因此包含特定k个办事项的任何序列必然包括该k个办事项的所有k-1个办事项的子序列。基于Apriori算法给出挖掘用户办事项记录中的序列模式的流程图如附图3所示。根据候选序列的支持度和步骤五提供的阈值提取支持度不小于阈值的序列模式，并由此获得序列模式的关联挖掘预选推荐办事项集合RS_A。

[0072] 4)结合用户社区和关联序列挖掘的推送结果。

[0073] 步骤八、在完成用户社区挖掘及相似用户预选推荐办事项集合RS_U和序列模式的关联挖掘预选推荐办事项集合RS_A之后，需要根据两部分结果的重叠程度进行优化筛选，形成最终的办事项推荐结果集合，并实时的推送给目标用户。

[0074] 若假定推送给目标用户的办事项数目为K，对于RS_U和RS_A两种结果集的覆盖结果筛选，可细分为图3所示的三种情况：

[0075] (a).覆盖结果集能够满足$N(RS_A \cap RS_U) \geq K$，$Top(RS_A)$被选择作为最终推荐结果集合；

[0076] (b).当$N(RS_A \cap RS_U) \leq K$，覆盖部分优先作为推荐结果；对于较多未覆盖部分的办事项，分别选取$K-N(RS_A \cap RS_U)/2$个结果集作为最终结果集；

[0077] (c).完全无覆盖的推荐结果计算采用(b)的未覆盖部分的筛选结果。

[0078] 通过Spark计算平台对提出的发明方法进行验证后表明：当被推荐项目具有序列或流程特性领域的推荐，本发明提供的推荐方法比传统推荐算法具有更高的推荐精度，推荐的多样性表现突出，具有一定的应用价值，在实际应用场景下能够有效提升办事服务质量与体验。

[0079] 在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不

必须针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0080] 上面结合附图对本发明进行了示例性描述，显然本发明具体实现并不受上述方式的限制，只要采用了本发明方法构思和技术方案进行的各种非实质性的改进，或未经改进将本发明构思和技术方案直接应用于其它场合的，均在本发明保护范围之内。

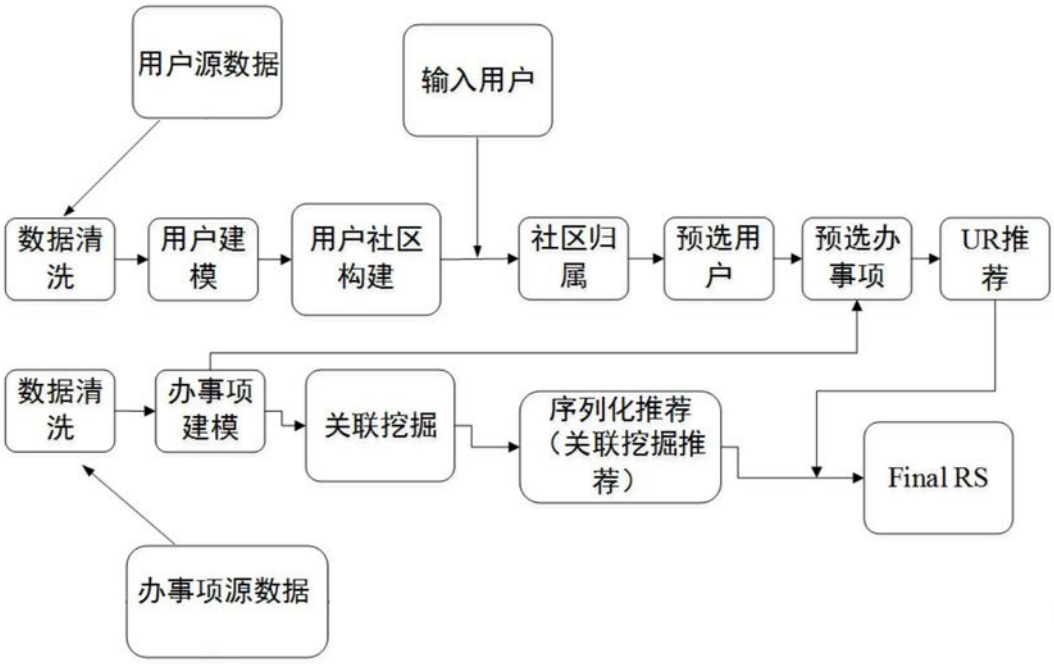

用户源数据
输入用户
数据清洗
用户建模
用户社区构建
社区归属
预选用户
预选办事项
UR推荐
数据清洗
办事项建模
关联挖掘
序列化推荐（关联挖掘推荐）
Final RS
办事项源数据

图1

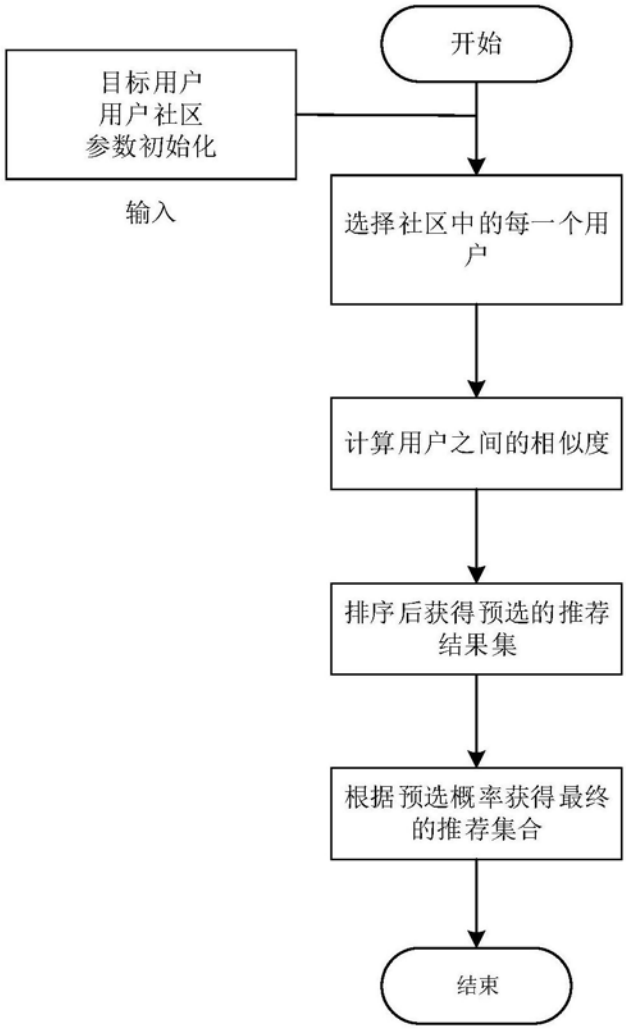

开始
目标用户
用户社区
参数初始化
输入
选择社区中的每一个用户
计算用户之间的相似度
排序后获得预选的推荐结果集
根据预选概率获得最终的推荐集合
结束

图2

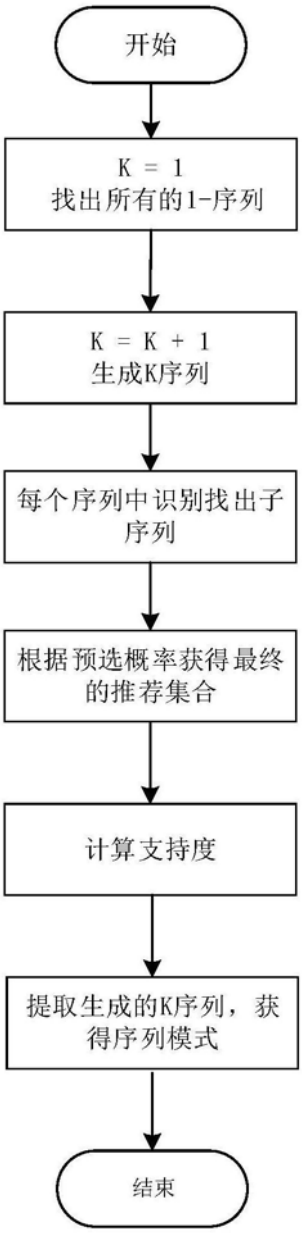

开始
K = 1
找出所有的1-序列
K = K + 1
生成K序列
每个序列中识别找出子序列
根据预选概率获得最终的推荐集合
计算支持度
提取生成的K序列，获得序列模式
结束

图3