



(12)发明专利申请

(10)申请公布号 CN 109523341 A
(43)申请公布日 2019.03.26

(21)申请号 201811189844.X

(22)申请日 2018.10.12

(71)申请人 广西师范大学

地址 541004 广西壮族自治区桂林市七星区育才路15号

(72)发明人 王利娥 李先贤 李东城 刘鹏

(74)专利代理机构 桂林市持衡专利商标事务所有限公司 45107

代理人 陈跃琳

(51) Int. Cl.

G06Q 30/06(2012.01)

G06F 16/22(2019.01)

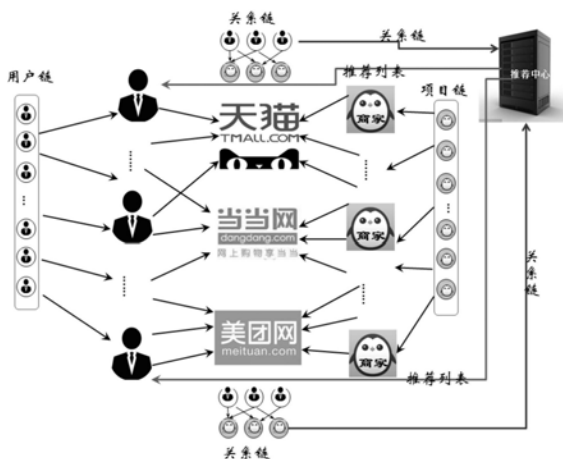
权利要求书1页 说明书5页 附图1页

(54)发明名称

基于区块链技术的匿名跨域推荐方法

(57)摘要

本发明公开一种基于区块链技术的匿名跨域推荐方法,其基于区块链技术采用异构式多链数据结构来进行数据存储,将用户、商品、交易关系数据在用户、商家和平台上分别进行存储,交易关系不存储具体的用户与商品信息,推荐中心只能获取由各平台存储的关系链,以保证攻击者无法将某个用户个体与某个商品关联起来;而针对由于数据稀疏性导致的冷启动问题,本发明基于精确的交易数据进行数据聚合和相似度计算,并采用基于用户和商品的混合推荐策略进行跨平台推荐,以实现安全,精确的双赢推荐效果。



1. 基于区块链技术的匿名跨域推荐方法,其特征是,其具体包括步骤如下:

步骤1、将用户信息存储在用户链,所有用户加入用户链,每个用户分配一个用户ID,并使用哈希函数得到哈希用户ID;所有商家的商品加入商品链,每个商品分配一个商品ID,并使用哈希函数得到哈希商品ID;

步骤2、各平台存储关系链,即哈希用户ID与商品的哈希商品ID之间的关系,并定时向推荐中心提供关系链信息;

步骤3、推荐中心先基于用户所购买商品来计算目标客户与其他用户的相似度,得到相似度最高的k个用户即相似用户;再求这k个相似用户所购买商品的并集得到候选商品集,并统计候选商品集中各购买商品出现次数;后用候选商品集与目标客户的商品集求异或,并在异或结果中选择出现次数最多的k个商品列入候选推荐列表;

步骤4、推荐中心统计关系链中商品的频繁模式,针对目标用户的所购商品选择关联度最高的k个关联商品加入步骤3所得到的候选推荐列表中;

步骤5、在步骤4所得到的候选推荐列表中,随机选择k个商品推荐给目标用户;

上述k为设定值。

2. 根据权利要求1所述的基于区块链技术的匿名跨域推荐方法,其特征是,步骤2中,所述关系链带有时间戳。

3. 根据权利要求2所述的基于区块链技术的匿名跨域推荐方法,其特征是,步骤2还进一步包括关系链信息更新过程,即:

当在一定的周期内或达到一定的数据量时,各平台主动向推荐中心推送增量关系链信息;和/或当推荐中心发送数据请求至各平台时,各平台被动向推荐中心推送增量关系链信息;

上述增量关系链信息为某个时间戳之后的关系链信息。

4. 根据权利要求1所述的基于区块链技术的匿名跨域推荐方法,其特征是,步骤3中,对于某一领域的新用户,根据该用户的其他领域信息来进行推荐。

5. 根据权利要求1所述的基于区块链技术的匿名跨域推荐方法,其特征是,步骤4中,对于某一领域的新商品,根据该商品之间的关联性来进行推荐。

基于区块链技术的匿名跨域推荐方法

技术领域

[0001] 本发明涉及区块链技术领域,具体涉及一种基于区块链技术的匿名跨域推荐方法。

背景技术

[0002] 推荐系统是应对信息呈爆炸式增长的数据筛选利器之一,它能够在海量数据中,根据用户的偏向性,主动推荐与其兴趣相一致的对象,该对象不仅可以是商品,还能够是某些商家提供的服务,如旅游服务、票务服务、订餐服务等各种休闲娱乐服务。目前,推荐系统已经流行于各个领域,在学术界也一直是研究的热点之一。但随着用户规模和商品数量的急剧增长,应用最为广泛的协同过滤推荐算法由于数据稀疏问题,导致推荐系统的冷启动问题,推荐质量大打折扣。而跨域推荐技术能够有效的解决传统推荐中的冷启动问题,成为目前学术界和商业界的研究热点。

[0003] 跨域推荐技术主要思想就是基于不同领域数据之间存在的关联性,从其它辅助领域中获取有效的用户偏好或商品特征的信息来丰富目标领域中的数据,精准地预测用户行为,以提供更合理和个性化的推荐系服务。不难发现,用户在不同的社交媒体和商务平台的数据直接或间接地反映其不同角度的兴趣偏好。但不同领域的数据之间的关系呈现出相互关联、交叉和融合的局面,呈现出高维、异构、动态、关联等特性;此外,数据中还包含大量的隐私信息,包括个人的(个人地址、电话等)、群体的(各平台营销机密等),尤其是跨域数据引入关联性之后,增强了攻击者的背景知识,其隐私问题更加复杂。因此,相比较传统的推荐技术,跨域推荐技术存在以下优势和挑战:

[0004] 1、跨域推荐能有效的解决传统推荐中的冷启动问题。随着数据量的增长,数据变得非常稀疏,使得绝大部分基于关联分析的算法(如协同过滤)效果都不好。由于新用户罕有可以利用的行为信息,所以很难给出精确的推荐;但跨域推荐技术可以很好的利用辅助领域来填补目标领域的缺省值,以有效解决冷启动问题和数据稀疏问题;

[0005] 2、跨域推荐中的数据聚合与大数据量的相似度计算问题,相比传统推荐复杂度更高。由于跨域推荐中涉及到多个不同的领域,而不同领域又拥有的百千万计的用户和商品,而且数据之间存在差异,所以如何快速高效地处理这些数据并且实现在高维数据中的相似度计算成为迫在眉睫的问题。

[0006] 3、跨域推荐中存在更为严峻的隐私安全问题。跨域推荐中涉及多维数据的交叉利用,是指整合来自不同领域的数据进行推荐。因为实际生活中每个人都会进入很多的不同的平台,并在不同的平台中留下信息。跨域推荐正是利用一个系统中的行为记录来预测另一系统中用户预期行为或者整合不同系统的信息提供推荐,以达到解决冷启动问题的目的。但数据交叉同时也带来隐私保护的难题,因为不同领域数据之间的关联为攻击者提供了更多的背景知识,增加隐私泄露的风险。因此虽然跨域数据融合有望提高推荐的效果或在更复杂的场景下进行推荐,但隐私问题却成为用户最大的顾虑。

[0007] 4、目前推荐中普遍存在的用户长期和短期的兴趣融合以及兴趣迁移变化等问题。

用户的兴趣可以分为长期兴趣和短期兴趣。长期兴趣是指在很长一段时间保持不变,比如运动等,而短期兴趣则是指特殊时期内的兴趣,比如孕期等;而且用户也会随着时间的变化或年龄的增长兴趣发生相应的迁移,比如饮食习惯、服饰偏好等,因此,如何根据商品的序列挖掘用户的兴趣变化,在保证不伤害用户长期兴趣的前提下充分满足用户的短期兴趣是推荐研究的技术难点。

发明内容

[0008] 本发明针对大量的数据存储与数据安全的问题,提供一种基于区块链技术的匿名跨域推荐方法,以实现多目标跨域推荐并保护用户隐私安全。

[0009] 为解决上述问题,本发明是通过以下技术方案实现的:

[0010] 基于区块链技术的匿名跨域推荐方法,其具体包括步骤如下:

[0011] 步骤1、将用户信息存储在用户链,所有用户加入用户链,每个用户分配一个用户ID,并使用哈希函数得到哈希用户ID;所有商家的商品加入商品链,每个商品分配一个商品ID,并使用哈希函数得到哈希商品ID;

[0012] 步骤2、各平台存储关系链,即哈希用户ID与商品的哈希商品ID之间的关系,并定时向推荐中心提供关系链信息;

[0013] 步骤3、推荐中心先基于用户所购买商品来计算目标客户与其他用户的相似度,得到相似度最高的k个用户即相似用户;再求这k个相似用户所购买商品的并集得到候选商品集,并统计候选商品集中各购买商品出现次数;后用候选商品集与目标客户的商品集求异或,并在异或结果中选择出现次数最多的k个商品列入候选推荐列表;

[0014] 步骤4、推荐中心统计关系链中商品的频繁模式,针对目标用户的所购商品选择关联度最高的k个关联商品加入步骤3所得到的候选推荐列表中;

[0015] 步骤5、在步骤4所得到的候选推荐列表中,随机选择k个商品推荐给目标用户;

[0016] 上述k为设定值。

[0017] 上述步骤2中,所述关系链带有时间戳。

[0018] 上述步骤2还进一步包括关系链信息更新过程,即:当在一定的周期内或达到一定的数据量时,各平台主动向推荐中心推送增量关系链信息;和/或当推荐中心发送数据请求至各平台时,各平台被动向推荐中心推送增量关系链信息;上述增量关系链信息为某个时间戳之后的关系链信息。

[0019] 上述步骤3中,对于某一领域的新用户,根据该用户的其他领域信息来进行推荐。

[0020] 上述步骤4中,对于某一领域的新商品,根据该商品之间的关联性来进行推荐。

[0021] 与现有技术相比,本发明具有如下特点:

[0022] 1、基于区块链技术的多链存储结构:基本思想就是将所有用户加入并存储用户链,所有商家加入并存储商品链,平台存储用户与商品之间的关系链,但不存储用户与商品的具体信息,只存储相应的哈希ID,用户和商家可以共同访问并验证自己的关系信息(比如某个用户可以验证关系链中自己买了哪些商品ID,而商家也可验证某一个商品被哪些哪些用户ID所购买),这种多链存储结构能保证用户的隐私安全并减轻数据的存储压力。

[0023] 2、多平台数据融合与相似度计算:推荐系统能获取多平台存储的关系链,并根据用户ID进行数据聚合,采用局部敏感哈希技术实现海量数据的快速相似度计算。局部敏感

哈希 (LSH, Local Sensitive Hash) 技术的基本思想是将数据压缩成紧凑的哈希码, 通过计算哈希码之间的某种距离 (比如海明距离) 即可快速估计出原始数据对间的相似度或者距离。它能保证原始特征空间中相邻的两个数据点通过相同的投影或者变换后, 仍具有较大概率相邻, 而在原始特征空间中不相邻的点仍具有较大概率不相邻。局部敏感哈希方法这种保相似度的哈希码作为对数据的一种压缩表示, 不仅实现了存储上的压缩和计算上的加速, 也可以建立哈希表实现快速的相似匿名搜索。

[0024] 3、采用混合推荐策略以实现隐私保护: 在进行推荐时不仅根据用户的相似程度进行邻居商品组合推荐, 还根据商品的关联程度进行商品关联推荐, 因此用户得到推荐列表是二者之间的随机混合, 使得获得推荐列表的用户无法推断推荐的原因是什么, 是相似用户的购买商品还是所购商品的高关联商品, 能有效的抵御伪装者攻击, 保证了邻居用户的隐私安全。

[0025] 4、增量更新与自适应推荐调整: 由于关系链带有时间戳, 很容易实现增量更新的识别, 只需要传输某一时间点之后的数据就可以; 还可以根据用户购买商品的先后顺序, 研究和挖掘用户随着时间变化的兴趣迁移, 或者区分用户的长期与短期兴趣; 此外, 推荐中心还可以根据增量更新的关系对应之前推荐研究推荐的精确程度, 并做出相应的推荐策略调整。

附图说明

[0026] 图1为基于区块链技术的匿名跨域推荐方法的原理示意图。

具体实施方式

[0027] 为使本发明的目的、技术方案和优点更加清楚明白, 以下结合具体实例, 并参照附图, 对本发明进一步详细说明。

[0028] 一种基于区块链技术的匿名跨域推荐方法, 其具体实现过程如下:

[0029] 步骤一: 构建区块链式数据存储。

[0030] 为了保证数据的安全以及减轻数据的存储压力, 本方法采用异构式存储结构, 将用户信息存储在用户链, 所有用户加入用户链, 每个用户分配一个用户ID, 并使用哈希函数得到哈希用户ID; 所有商家的商品加入商品链, 每个商品分配一个商品ID, 并使用哈希函数得到哈希商品ID。平台存储关系链, 主要是用户ID与商品ID之间的关系, 不存储用户和商品的具体信息, 只存储用户与商品的哈希ID, 用户和商家可以访问并验证, 但只能验证与自身ID相关的部分, 关系链带时间戳, 以反映时间的先后顺序。

[0031] 步骤二: 跨平台数据收集。

[0032] 由各平台定时向推荐中心提供关系链信息, 推荐中心收到信息后进行数据聚合和分析等操作, 首先根据各平台的综合数据, 采用局部敏感哈希技术分析用户的相似程度并得到相似分组, 向用户推荐其相似的邻居用户信息; 再对商品进行频繁模式挖掘, 进行商品关联程度的分析, 用于基于商品的关联推荐。对于某一领域的新用户, 可用该用户的其他领域信息来进行填补; 对于某一领域的新商品, 可以根据商品之间的关联性来进行推荐。比如, 在商品购买领域的新用户, 可根据其在旅游领域的好友所购买的商品进行推荐; 而一个新菜品的出现, 则可以根据在商品领域曾经购买过烹饪该菜品的原料的用户进行推荐, 以

有效解决冷启动问题。

[0033] 步骤三:匿名推荐。

[0034] 本发明采用混合推荐策略以加强算法的安全性,首先,基于用户的推荐策略,推荐中心基于购买商品计算目标客户与其他用户的相似度,得到相似度最高的top-k用户为候选列表,求这top-k个相似用户所够买商品的并集并统计计数;用候选集与目标客户的商品集求异或,在结果中选择计数最多的top-k个列入候选推荐列表;其次,基于商品的推荐策略,推荐中心统计关系链中商品的频繁模式,针对目标用户的所购商品选择top-k个关联商品列入候选推荐列表;最后,在基于用户相似性和商品关联性的候选推荐列表中即2k个商品中随机选择k个商品,作为最终的推荐商品列表推荐给目标用户。

[0035] 该推荐策略能有效的保证算法的安全性,推荐中心虽存储了关系链,但没有具体的用户和商品信息,无法将某个个体和具体的商品关联起来,保护了目标客户的隐私信息;用户得到推荐列表之后,无法区分该列表的数据是因为邻居购买还是由于与已购商品之间的关联性被推荐,保护了邻居用户的隐私信息。

[0036] 步骤四:增量更新与自适应推荐调整。

[0037] 为保证推荐列表的精确性,还需要根据动态的交易信息及时更新关系链,由于关系链的存储具有时间戳,在更新的时候只需要截取某个时间戳之后的增量数据进行更新就可以了。数据的收集过程具有周期性,可由各平台在一定的周期内或达到一定的数据量时主动向推荐中心推送增量信息,也可以由推荐中心发送数据请求获得。推荐中心收到增量更新数据之后还可以根据自己的推荐列表和增量交易信息之间的关系进行推荐精度和用户兴趣迁移分析,并做出相应的推荐策略调整,以实现更好的用户推荐体验。

[0038] 至此,整个推荐过程就完成了。该推荐方法采用异构式存储结构,使得用户与推荐中心不能访问具体的商品信息,从而无法推断出用户与商品之间的关联关系,很好的保护了用户的隐私信息。此外,该方法中的链式存储结构还可以根据时间的先后顺序关注用户不同时期的兴趣变化,很好的反应用户的长期兴趣与短期兴趣以及用户的兴趣迁移变化等情况,并及时的调整推荐策略,能够极大的提高了推荐精确度与用户体验;值得一提的是,该推荐方法该方法还减少了海量数据的传输和存储压力,能很好的利用多平台的交叉信息进行推荐,能很好的规避由于数据稀疏带来的冷启动等问题。

[0039] 实现上述方法的基于区块链技术的匿名跨域推荐系统,如图1所示,主要包括4个模块:

[0040] 区块链式数据存储模块。数据的存储采用异构式存储的三链结构,分为用户链,商品链、关系链,所有的用户信息构成了用户链,由用户共同保存;所有商品信息构成了商品链,由商家共同保存;用户与商品之间的关系构成了关联链,由平台保存,用户和商家可以共同访问并验证关系链信息,但用户不能访问商品链,商家不能访问用户链。平台只存储用户哈希ID和商品哈希ID,不存储具体的用户与商品信息。

[0041] 用户相似度计算与商品关联挖掘模块。不同领域数据根据用户ID一一对应拼接起来,基于局部敏感哈希技术采用指定哈希函数得到哈希编码,该哈希编码能保证原始数据相似的编码仍然相似。基于该哈希编码进行相似度计算与分组,找到多域融合的邻居用户,并进行商品频繁模式挖掘,得到不同商品之间的关联程度数据。

[0042] 匿名推荐模块。本发明主要考虑用户重叠场景,推荐中心可获得由所有平台提供

的关系链信息。本发明采用混合推荐策略,由推荐中心根据关系链中所购商品的相似程度进行基于用户的相似度推荐,而根据商品的频繁模式分析进行基于商品的关联推荐。由于在推荐过程中推荐中心并不知道具体的用户与商品信息,因此对于推荐中心来说推荐是匿名的;另一方面,用户收到推荐信息时,并不清楚该推荐项是因为邻居购买了还是因为商品关联而被推荐,所以对于用户也是匿名的。

[0043] 跨平台解决冷启动问题。当一个新用户或新商品出现时,由于没有相关的历史数据,无法进行有效推荐,这就是所谓的冷启动问题。针对新用户存在的冷启动问题,本发明可利用该用户在其他领域的相关数据进行推荐。而针对新商品的冷启动问题,本发明采用该商品在不同领域之间的关联性进行推荐。

[0044] 本发明基于区块链技术采用异构式多链数据结构来进行数据存储,将用户、商品、交易关系数据在用户、商家和平台上分别进行存储,交易关系不存储具体的用户与商品信息,推荐中心只能获取由各平台存储的关系链,以保证攻击者无法将某个用户个体与某个商品关联起来;而针对由于数据稀疏性导致的冷启动问题,本发明基于精确的交易数据进行数据聚合和相似度计算,并采用基于用户和商品的混合推荐策略进行跨平台推荐,以实现安全,精确的双赢推荐效果。

[0045] 需要说明的是,尽管以上本发明所述的实施例是说明性的,但这并非是对本发明的限制,因此本发明并不局限于上述具体实施方式中。在不脱离本发明原理的情况下,凡是本领域技术人员在本发明的启示下获得的其它实施方式,均视为在本发明的保护之内。

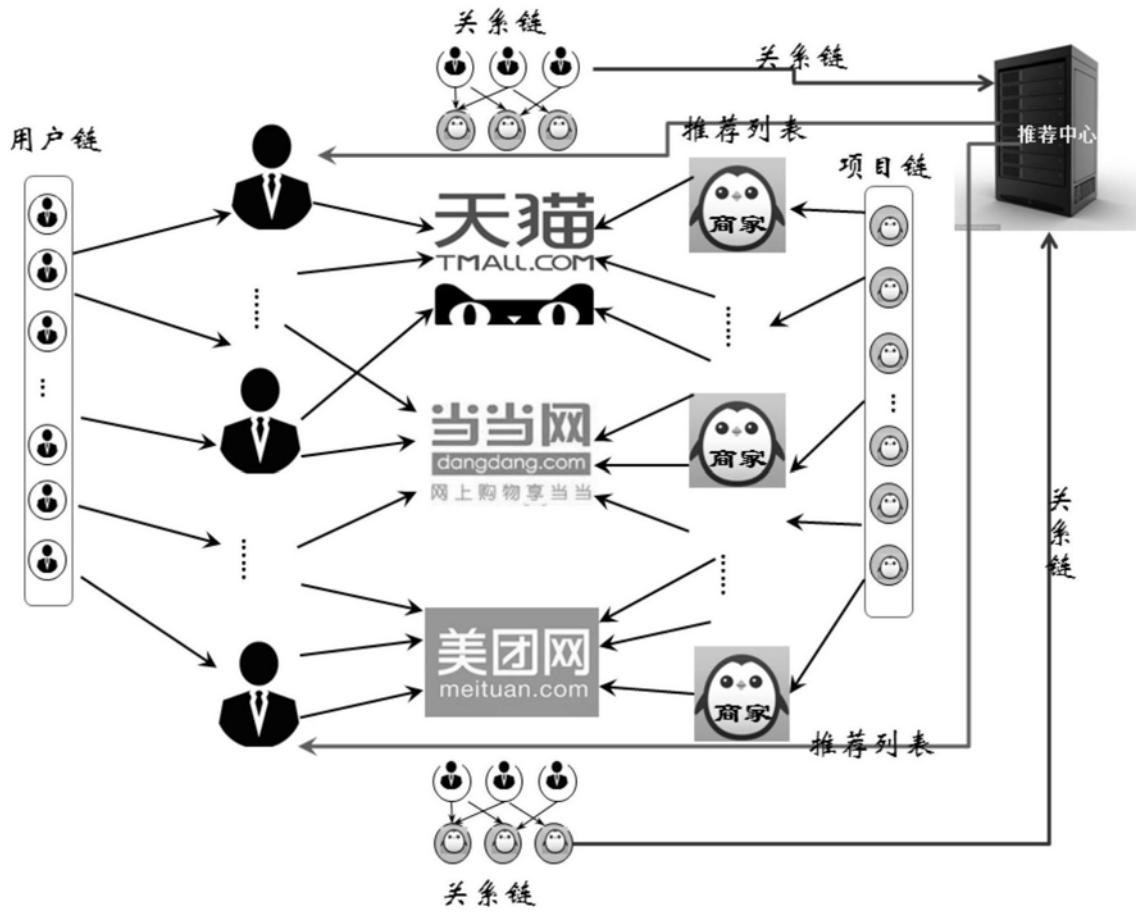


图1