



(12)发明专利申请

(10)申请公布号 CN 110910207 A

(43)申请公布日 2020.03.24

(21)申请号 201911042387.6

(22)申请日 2019.10.30

(71)申请人 苏宁云计算有限公司

地址 210042 江苏省南京市玄武区徐庄软件园苏宁大道1-1号

(72)发明人 马荣叶

(74)专利代理机构 江苏圣典律师事务所 32237

代理人 许峰

(51)Int.Cl.

G06Q 30/06(2012.01)

G06F 16/906(2019.01)

G06F 16/9536(2019.01)

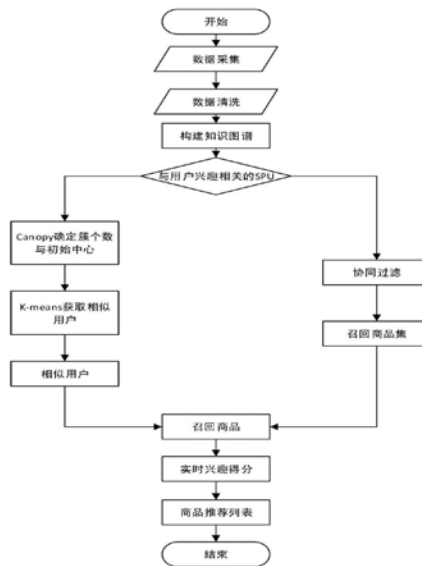
权利要求书2页 说明书7页 附图3页

(54)发明名称

一种提升商品推荐多样性的方法及系统

(57)摘要

本发明实施例公开了一种提升商品推荐多样性的方法,解决推荐商品缺乏多样性的问题,根据用户数据,利用Canopy算法与K-means算法获取相似用户;根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU;根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表。



1. 一种提升商品推荐多样性的方法,其特征在于,所述方法包括:  
根据用户数据,利用Canopy算法与K-means算法获取相似用户;  
根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU;  
根据所述相似用户和所述与用户实时兴趣关联的SPU,获取商品推荐列表。
2. 根据权利要求1所述的提升商品推荐多样性的方法,其特征在于,所述根据用户数据,利用Canopy算法与K-means算法获取相似用户,包括:  
采集用户人口统计属性数据和用户在多屏上的行为数据,对所述用户人口统计属性数据和所述行为数据进行数据清洗;  
计算用户长期兴趣偏好和用户短期兴趣偏好;  
利用Canopy算法对清洗后的用户人口统计属性数据、用户长期兴趣偏好和用户短期兴趣偏好,进行粗聚类,确定K-means算法的簇个数k和初始中心;  
利用所述K-means算法获取相似用户。
3. 根据权利要求2所述的提升商品推荐多样性的方法,其特征在于,所述确定K-means算法的簇个数k和初始中心,包括:  
设定第一距离参数T1和第二距离参数T2,且 $T1 > T2$ ,设定用户集为D,随机选取用户集D中一位用户,设该用户与用户集D中其他用户的距离d;当d小于T1时,将用户放入一个Canopy,并从用户集D中将d小于T2的用户删除;直到用户集D为空时,用户被分到了多个Canopy中;将Canopy的个数作为K-means的簇个数k,Canopy的中心作为K-means的初始中心。
4. 根据权利要求1所述的提升商品推荐多样性的方法,其特征在于,所述根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表,包括:  
将所述相似用户下的商品和所述与用户实时兴趣关联的SPU进行对比,得到第一召回SPU;  
根据用户在线行为数据,计算用户实时兴趣得分;  
根据所述用户实时兴趣得分,以及所述相似用户对第一召回SPU下商品的打分,对第一召回SPU进行CTR/CVR得分预测,再对召回商品进行去重和过滤处理;  
对经过去重和过滤处理的召回商品,进行排序,获得商品推荐列表。
5. 根据权利要求1所述的提升商品推荐多样性的方法,其特征在于,所述根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表,包括:  
若未能利用知识图谱获取与用户实时兴趣关联的SPU,则利用协同过滤组合算法,得到召回商品集;  
将所述相似用户下的商品和所述召回商品集进行对比,得到第二召回SPU;  
根据用户在线行为数据,计算用户实时兴趣得分;  
根据所述用户实时兴趣得分,以及所述相似用户对第二召回SPU下商品的打分,对第二召回SPU进行CTR/CVR得分预测,再对召回商品进行去重和过滤处理;  
对经过去重和过滤处理的召回商品,进行排序,获得商品推荐列表。
6. 一种提升商品推荐多样性的系统,其特征在于,包括:  
用户获取模块,用于根据用户数据,利用Canopy算法与K-means算法获取相似用户;  
SPU获取模块,用于根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的

SPU;

列表获取模块,用于根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表。

7.根据权利要求6所述的提升商品推荐多样性的系统,其特征在于,所述用户获取模块,包括:

数据采集子模块,用于采集用户人口统计属性数据和用户在多屏上的行为数据;

数据清洗子模块,用于对所述数据采集子模块的数据进行清洗;

偏好计算子模块,用于计算用户长期兴趣偏好和用户短期兴趣偏好;

粗聚类子模块,用于利用Canopy算法对清洗后的用户人口统计属性数据、用户长期兴趣偏好和用户短期兴趣偏好,进行粗聚类,确定K-means算法的簇个数k和初始中心;

相似用户获取子模块,用于利用所述K-means算法获取相似用户。

8.根据权利要求7所述的提升商品推荐多样性的系统,其特征在于,所述粗聚类子模块,包括:

参数计算子模块,用于确定K-means算法的簇个数k和初始中心;

设定第一距离参数T1和第二距离参数T2,且 $T1 > T2$ ,设定用户集为D,随机选取用户集D中一位用户,设该用户与用户集D中其他用户的距离d;

当d小于T1时,将用户放入一个Canopy,并从用户集D中将d小于T2的用户删除;

直到用户集D为空时,用户被分到了多个Canopy中;

将Canopy的个数作为K-means的簇个数k,Canopy的中心作为K-means的初始中心。

9.根据权利要求6所述的提升商品推荐多样性的系统,其特征在于,所述列表获取模块,包括:

第一得分计算子模块,用于根据用户在线行为数据,计算用户实时兴趣得分;

第一召回子模块,用于根据SPU获取模块,将所述相似用户下的商品和所述与用户实时兴趣关联的SPU进行对比,得到第一召回SPU;

第一得分预测子模块,用于根据所述用户实时兴趣得分,以及所述相似用户对第一召回SPU下商品的打分,对第一召回SPU进行CTR/CVR得分预测;

第一处理子模块,用于对所述召回商品进行去重和过滤处理,进行排序,获得商品推荐列表。

10.根据权利要求6所述的提升商品推荐多样性的系统,其特征在于,所述列表获取模块,包括:

第二得分计算子模块,用于根据用户在线行为数据,计算用户实时兴趣得分;

商品召回模块,用于若未能利用知识图谱获取与用户实时兴趣关联的SPU,则利用协同过滤组合算法,得到召回商品集;

第二召回子模块,用于将所述相似用户下的商品和所述召回商品集进行对比,得到第二召回SPU;

第二得分预测子模块,用于根据所述用户实时兴趣得分,以及所述相似用户对第二召回SPU下商品的打分,对第二召回SPU进行CTR/CVR得分预测;

第二处理子模块,用于对所述召回商品进行去重和过滤处理,进行排序,获得商品推荐列表。

## 一种提升商品推荐多样性的方法及系统

### 技术领域

[0001] 本发明属于商品推荐领域,尤其涉及一种提升商品推荐多样性的方法及系统。

### 背景技术

[0002] 本领域内商品推荐算法主要有以下几种:基于内容的推荐、基于数据挖掘的推荐、组合推荐和基于内存的协同过滤算法。

[0003] 基于内容的推荐技术,在生成用户特征模型之后,根据用户模型进行推荐,推荐结果直观、容易理解,能对特殊用户推荐,并且不需要领域知识,但是该方法适用范围较窄。

[0004] 基于数据挖掘的推荐技术,主要是通过生成关联规则进行推荐,不受推荐内容限制,不需要领域知识,但是该方法关联规则生成难度大,并且耗时,根据规则推荐的商品个性化程度低,商品多样性不够。

[0005] 组合推荐,通过将两种或以上的推荐方法按照特定规则进行组合,以避免或弥补组合前单一推荐技术存在的缺点,理论上有多种推荐组合方法,但是在某一具体问题种并不见得都有效,并且组合策略参数调整比较困难。

[0006] 基于内存的协同过滤算法,通过计算相似用户或相似商品进行推荐,模型更新周期短,能及时反映用户的兴趣变化,但是存在数据稀疏问题。

[0007] 基于用户协调的商品召回中用户的相似性计算,多利用K-means算法,但簇个数k和初始中心难以确定,并且计算的全量商品的相似性,数据量也是过于庞大。

[0008] 由于上述推荐算法适用内容窄、K-means算法中个数k和初始中心难以确定、组合算法间参数调整困难、处理大量数据耗时过长以及数据稀疏造成商品关系断裂,因此得到的推荐商品单一,关联度不够且缺乏多样性。

### 发明内容

[0009] 本发明提供一种提升商品推荐多样性的方法及系统,解决推荐商品缺乏多样性的问题。

[0010] 为达到上述目的,本发明的实施例采用如下技术方案:

[0011] 第一方面,本发明的实施例提供一种提升商品推荐多样性的方法,根据用户数据,利用Canopy算法与K-means算法获取相似用户;根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU;根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表。

[0012] 结合第一方面,作为第一种可实现方式,采集用户人口统计属性数据和用户在多屏上的行为数据,对用户人口统计属性数据和所述行为数据进行数据清洗,计算用户长期兴趣偏好和用户短期兴趣偏好;利用Canopy算法对清洗后的用户人口统计属性数据、用户长期兴趣偏好和用户短期兴趣偏好,进行粗聚类,确定K-means算法的簇个数k和初始中心;利用K-means算法获取相似用户;

[0013] 结合第一方面的第一种可能的实现方式,作为第二种可实现方式,设定第一距离

参数 $T1$ 和第二距离参数 $T2$ ,且 $T1>T2$ ,设定用户集为 $D$ ,随机选取用户集 $D$ 中一位用户,设该用户与用户集 $D$ 中其他用户的距离 $d$ ;当 $d$ 小于 $T1$ 时,将用户放入一个Canopy,并从用户集 $D$ 中将 $d$ 小于 $T2$ 的用户删除;直到用户集 $D$ 为空时,用户被分到了多个Canopy中;将Canopy的个数作为K-means的簇个数 $k$ ,Canopy的中心作为K-means的初始中心。

[0014] 结合第一方面,作为第三种可实现方式,将所述相似用户下的商品和所述与用户实时兴趣关联的SPU进行对比,得到第一召回SPU;根据用户在线行为数据,计算用户实时兴趣得分;根据所述用户实时兴趣得分,以及所述相似用户对第一召回SPU下商品的打分,对第一召回SPU进行CTR/CVR得分预测,再对召回商品进行去重和过滤处理;对经过去重和过滤处理的召回商品,进行排序,获得商品推荐列表。

[0015] 结合第一方面,作为第四种可实现方式,若未能利用知识图谱获取与用户实时兴趣关联的SPU,则利用协同过滤组合算法,得到召回商品集;将所述相似用户下的商品和所述召回商品集进行对比,得到第二召回SPU;根据用户在线行为数据,计算用户实时兴趣得分;根据所述用户实时兴趣得分,以及所述相似用户对第二召回SPU下的商品打分,对第二召回SPU进行CTR/CVR得分预测,再对召回商品进行去重和过滤处理;对经过去重和过滤处理的召回商品,进行排序,获得商品推荐列表。

[0016] 第二方面,本发明的实施例提供一种提升商品推荐多样性的系统,包括:

[0017] 用户获取模块,用于根据用户数据,利用Canopy算法与K-means算法获取相似用户;

[0018] SPU获取模块,用于根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU;

[0019] 列表获取模块,用于根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表。

[0020] 结合第二方面,作为第一种可能实现方式,所述用户获取模块,包括:

[0021] 数据采集子模块,用于采集用户人口统计属性数据和用户在多屏上的行为数据;

[0022] 数据清洗子模块,用于对所述数据采集子模块的数据进行清洗;

[0023] 偏好计算子模块,用于计算用户长期兴趣偏好和用户短期兴趣偏好;

[0024] 粗聚类子模块,用于利用Canopy算法对清洗后的用户人口统计属性数据、用户长期兴趣偏好和用户短期兴趣偏好,进行粗聚类,确定K-means算法的簇个数 $k$ 和初始中心;

[0025] 相似用户获取子模块,用于利用所述K-means算法获取相似用户。

[0026] 结合第二方面,作为第二种可能实现方式,所述粗聚类子模块,包括:

[0027] 参数计算子模块,用于确定K-means算法的簇个数 $k$ 和初始中心;

[0028] 设定第一距离参数 $T1$ 和第二距离参数 $T2$ ,且 $T1>T2$ ,设定用户集为 $D$ ,随机选取用户集 $D$ 中一位用户,设该用户与用户集 $D$ 中其他用户的距离 $d$ ;

[0029] 当 $d$ 小于 $T1$ 时,将用户放入一个Canopy,并从用户集 $D$ 中将 $d$ 小于 $T2$ 的用户删除;

[0030] 直到用户集 $D$ 为空时,用户被分到了多个Canopy中;

[0031] 将Canopy的个数作为K-means的簇个数 $k$ ,Canopy的中心作为K-means的初始中心。

[0032] 结合第二方面,作为第三种可能实现方式,所述列表获取模块,包括:

[0033] 第一得分计算子模块,用于根据用户在线行为数据,计算用户实时兴趣得分;

[0034] 第一召回子模块,用于根据SPU获取模块,将所述相似用户下的商品和所述与用户

实时兴趣关联的SPU进行对比,得到第一召回SPU;

[0035] 第一得分预测子模块,用于根据所述用户实时兴趣得分,以及所述相似用户对第一召回SPU下商品的打分,对第一召回SPU进行CTR/CVR得分预测;

[0036] 第一处理子模块,用于对所述召回商品进行去重和过滤处理,进行排序,获得商品推荐列表。

[0037] 结合第二方面,作为第四种可能实现方式,所述列表获取模块,包括:

[0038] 第二得分计算子模块,用于根据用户在线行为数据,计算用户实时兴趣得分;

[0039] 商品召回模块,用于若未能利用知识图谱获取与用户实时兴趣关联的SPU,则利用协同过滤组合算法,得到召回商品集;

[0040] 第二召回子模块,用于将所述相似用户下的商品和所述召回商品集进行对比,得到第二召回SPU;

[0041] 第二得分预测子模块,用于根据所述用户实时兴趣得分,以及所述相似用户对第二召回SPU下商品的打分,对第二召回SPU进行CTR/CVR得分预测;

[0042] 第二处理子模块,用于对所述召回商品进行去重和过滤处理,进行排序,获得商品推荐列表。

[0043] 本发明实施例提供一种提升商品推荐多样性的方法及系统,解决推荐商品单一,关联度不够且缺乏多样性的问题。相比于现有技术,在本发明实施例中,根据用户数据,利用Canopy算法与K-means算法获取相似用户,将Canopy 算法确定的Canopy个数和初始中心作为K-means算法的簇个数k和初始中心,解决参数调整困难的问题,计算收敛更快,能处理大量数据;根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU,通过关联计算后,数据量降低,商品的关系传递会更深,能够召回更多品类的商品;根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表,提高召回的发散度和获得商品的新鲜度,解决数据稀疏带来的商品关系的断裂,推荐效果品类更丰富。本方法计算复杂度低,并且在稀疏数据条件下的推荐性能更为优越,推荐数量更多,跨类别推荐数量也更多,与实际结果相符的有效数量、跨类别的有效数量也更多。

## 附图说明

[0044] 为了更清楚地说明本发明实施例的技术方案,下面将对本发明实施例中所需要使用的附图作简单地介绍,显而易见地,下面所描述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它的附图。

[0045] 图1为本发明实施例的流程图。

[0046] 图2为本发明实施例中Canopy算法流程图。

[0047] 图3为本发明实施例的系统框架图。

[0048] 图4为本发明实施例的系统结构框图。

## 具体实施方式

[0049] 为使本领域技术人员更好地理解本发明的技术方案,下面结合附图和具体实施方式对本发明作进一步详细描述。显然,所描述的实施例是本发明的一部分实施例,而不是全

部实施例。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的前提下所获得的实施例，都应属于本发明保护的范围。

[0050] 如图1所示，本发明实施例提供一种提升商品推荐多样性的方法，包括：

[0051] S110根据用户数据，利用Canopy算法与K-means算法获取相似用户。

[0052] 所述用户数据包括：用户人口统计属性数据和用户在多屏上的行为数据。所述用户人口统计属性数据包括但不限于：性别、年龄、常驻地址、收入和教育程度。所述用户行为数据包括但不限于：浏览、搜索、加购、购买、点评和分享。

[0053] 获得用户数据后，进行数据清洗。所述数据清洗包括：缺失值平滑，异常值剔除，重复数据去重和数据归一化。数据清洗目的在于提高数据质量，减少脏数据对模型的准确度的影响。

[0054] 利用清洗后的用户行为数据，进行用户实时兴趣识别，建立用户偏好模型，计算用户的长期兴趣偏好和短期兴趣偏好。

[0055] 长期兴趣偏好为用户比较长一段时间（例如，1个月或3个月或6个月甚至更长时间）其行为对应的兴趣空间或兴趣主题。短期兴趣偏好为用户比较短的时间（例如，7天或3天或即时）其行为对应的兴趣空间或兴趣主题。

[0056] 利用Canopy聚类算法对用户人口统计属性数据、用户长期兴趣偏好，短期兴趣偏好进行粗聚类。所述粗聚类包括：设定第一距离参数 $T1$ 和第二距离参数 $T2$ ，且 $T1 > T2$ ，设定用户集为 $D$ ，随机选取用户集 $D$ 中一位用户，设该用户与用户集 $D$ 中其他用户的距离 $d$ ；当 $d$ 小于 $T1$ 时，将用户放入一个Canopy，并从用户集 $D$ 中将 $d$ 小于 $T2$ 的用户删除；直到用户集 $D$ 为空时，用户被分到了多个Canopy中。将Canopy的个数作为K-means的簇个数，Canopy的中心作为K-means的初始中心。

[0057] 利用K-means算法获取相似用户。利用K-means算法对Canopy内的用户进行精准相似性判断，不同Canopy之间的用户不进行精准的相似性判断。采取pearson相关系数(Pearson Correlation Coefficient)来衡量用户之间的相似度。

[0058] S120根据全量订单数据，构建知识图谱，获取与用户实时兴趣关联的SPU。

[0059] 采集全量订单数据。所述全量订单数据包括但不限于：用户ID和SKU (Stock Keeping Unit, 库存量单位) 流失。优选的，对全量订单数据进行数据清洗。所述数据清洗包括：缺失值平滑，异常值剔除，重复数据去重和数据归一化。

[0060] 基于清洗后的全量订单数据，采取规则挖掘关系抽取技术，挖掘出SKU之间的销售关系，对SKU进行聚合，将关联销售的关系上升到SPU (Standard Product Unit, 标准化产品单元) 层级，将商品数据量降级，利用图数据库，存储关联销售SPU对，并结合业务知识，完成SPU知识图谱构建。

[0061] 利用知识图谱所建立的SPU与SPU的传递关系，获取用户实时兴趣对应的SPU，由此在知识图谱中获取与用户实时兴趣关联的SPU。

[0062] S130，根据所述相似用户和所述与用户实时兴趣关联的SPU，获得商品推荐列表。

[0063] 对用户在线行为数据，进行用户实时兴趣识别，不同的行为对应不同的权重，不同行为时长对应不同的权重，线性组合计算多个实时兴趣得分。

[0064] 将由相似用户协调过滤召回的SPU和知识图谱召回的SPU结合，获取第一召回SPU，第一召回SPU为仅存在于相似用户下，而不存在于与用户实时兴趣关联的SPU下的商品。

[0065] 利用用户实时兴趣得分结合相似用户对第一召回SPU下商品的打分,对第一召回SPU下的商品进行CTR/CVR(Click Through Rate,点击通过率,文中简称 CTR;ConVersion Rate,转化率,文中简称CVR)得分预测,再对召回商品进行去重和过滤处理,对处理后的召回商品按照CTR/CVR得分排序后,获得商品推荐列表。

[0066] 若未能利用知识图谱获取与用户实时兴趣关联的SPU,则利用协同过滤组合算法,得到召回商品集,将所述相似用户下的商品和所述召回商品集进行对比,得到第二召回SPU,第二召回SPU为仅存在于相似用户下,而不存在于召回商品集下的商品。

[0067] 利用用户实时兴趣得分结合相似用户对第二召回SPU下商品的打分,对第二召回SPU下的商品进行CTR/CVR得分预测,再对召回商品进行去重和过滤处理,对处理后的召回商品按照CTR/CVR得分排序后,获得商品推荐列表。

[0068] 本发明实施例提供的一种提升商品推荐多样性的方法,相比于现有技术,在本发明实施中,根据用户数据,利用Canopy算法与K-means算法获取相似用户,将Canopy算法确定的Canopy个数和初始中心作为K-means算法的簇个数k和初始中心,Canopy的个数就是K-means的簇个数K,不需要根据经验调整,距离之间的计算在Canopy内样本两两计算,不用在计算全量样本的两两距离,计算量更小,解决参数调整困难的问题,计算收敛更快,能处理大量数据;根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU,知识图谱具有传递作用,SKU量级远大于SPU,而且SPU使数据更稠密,通过关联计算后,数据量降低,商品的关系传递会更深,能够召回更多品类的商品;根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表,基于协调过滤推荐越推越窄,而知识图谱关系更发散,提高召回的发散度和获得商品的新鲜度,解决数据稀疏带来的商品关系的断裂,推荐效果品类更丰富。本方法计算复杂度低,并且在稀疏数据条件下的推荐性能更为优越,推荐数量更多,跨类别推荐数量也更多,与实际结果相符的有效数量、跨类别的有效数量也更多。

[0069] 如图4所示,本发明实施例还提供一种提升商品推荐多样性的系统,包括:

[0070] 用户获取模块,用于根据用户数据,利用Canopy算法与K-means算法获取相似用户;SPU获取模块,用于根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU;列表获取模块,用于根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表。

[0071] 根据本发明地一个实施例,所述用户获取模块,包括:

[0072] 数据采集子模块,用于采集用户人口统计属性数据和用户在多屏上的行为数据;

[0073] 数据清洗子模块,用于对所述数据采集子模块的数据进行清洗;

[0074] 偏好计算子模块,用于计算用户长期兴趣偏好和用户短期兴趣偏好;

[0075] 粗聚类子模块,用于利用Canopy算法对清洗后的用户人口统计属性数据、用户长期兴趣偏好和用户短期兴趣偏好,进行粗聚类,确定K-means算法的簇个数k和初始中心;

[0076] 相似用户获取子模块,用于利用所述K-means算法获取相似用户。

[0077] 根据本发明地一个实施例,所述粗聚类子模块,包括:

[0078] 参数计算子模块,用于确定K-means算法的簇个数k和初始中心;

[0079] 设定第一距离参数T1和第二距离参数T2,且 $T1 > T2$ ,设定用户集为D,随机选取用户集D中一位用户,设该用户与用户集D中其他用户的距离d;

[0080] 当d小于T1时,将用户放入一个Canopy,并从用户集D中将d小于T2的用户删除;



- [0081] 直到用户集D为空时,用户被分到了多个Canopy中;
- [0082] 将Canopy的个数作为K-means的簇个数k,Canopy的中心作为K-means的初始中心。
- [0083] 根据本发明地一个实施例,所述列表获取模块,包括:
- [0084] 第一得分计算子模块,用于根据用户在线行为数据,计算用户实时兴趣得分;
- [0085] 第一召回SPU子模块,用于根据SPU获取模块,将所述相似用户下的商品和所述与用户事实兴趣关联的SPU进行对比,得到第一召回SPU;
- [0086] 第一得分预测子模块,用于根据所述用户实时兴趣得分,以及所述相似用户对第一召回SPU下商品的打分,对第一召回SPU进行CTR/CVR得分预测;
- [0087] 第一处理子模块,用于对所述召回商品进行去重和过滤处理,进行排序,获得商品推荐列表。
- [0088] 根据本发明地一个实施例,所述列表获取模块,包括:
- [0089] 第二得分计算子模块,用于根据用户在线行为数据,计算用户实时兴趣得分;
- [0090] 商品召回模块,用于若未能利用知识图谱获取与用户实时兴趣关联的SPU,则利用协同过滤组合算法,得到召回商品集;
- [0091] 第二召回子模块,用于将所述相似用户下的商品和所述召回商品集进行对比,得到第二召回SPU;
- [0092] 第二得分预测子模块,用于根据所述用户实时兴趣得分,以及所述相似用户对第二召回SPU下商品的打分,对第二召回SPU进行CTR/CVR得分预测;
- [0093] 第二处理子模块,用于对所述召回商品进行去重和过滤处理,进行排序,获得商品推荐列表。
- [0094] 本发明实施例提供的一种提升商品推荐多样性的系统,解决推荐商品单一,关联度不够且缺乏多样性的问题。相比于现有技术,在本发明实施中,用户获取模块根据用户数据,利用Canopy算法与K-means算法获取相似用户,用户获取模块中的参数计算子模块将Canopy算法确定的Canopy个数和初始中心作为 K-means算法的簇个数k和初始中心,Canopy的个数就是K-means的簇个数K,不需要根据经验调整,距离之间的计算在Canopy内样本两两计算,不用在计算全量样本的两两距离,计算量更小,解决参数调整困难的问题,计算收敛更快,能处理大量数据;SPU获取模块根据全量订单数据,构建知识图谱,获取与用户实时兴趣关联的SPU,知识图谱具有传递作用,SKU量级远大于SPU,而且SPU 使数据更稠密,通过关联计算后,数据量降低,商品的关系传递会更深,能够召回更多品类的商品;列表获取模块根据所述相似用户和所述与用户实时兴趣关联的SPU,获得商品推荐列表,基于协调过滤推荐越推越窄,而知识图谱关系更发散,提高召回的发散度和获得商品的新鲜度,解决数据稀疏带来的商品关系的断裂,推荐效果品类更丰富。本系统计算复杂度低,并且在稀疏数据条件下的推荐性能更为优越,推荐数量更多,跨类别推荐数量也更多,与实际结果相符的有效数量、跨类别的有效数量也更多。
- [0095] 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,对于设备实施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。
- [0096] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以

通过计算机程序来指令相关的硬件来完成,所述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory, ROM)或随机存储记忆体(Random Access Memory, RAM)等。

[0097] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求的保护范围为准。

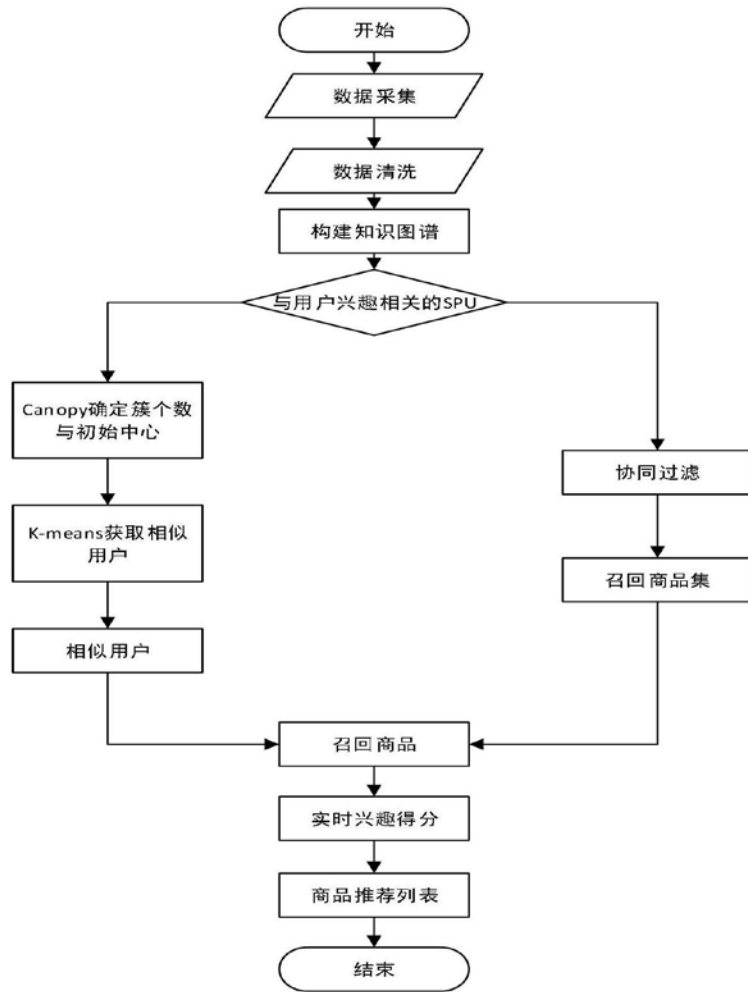


图1

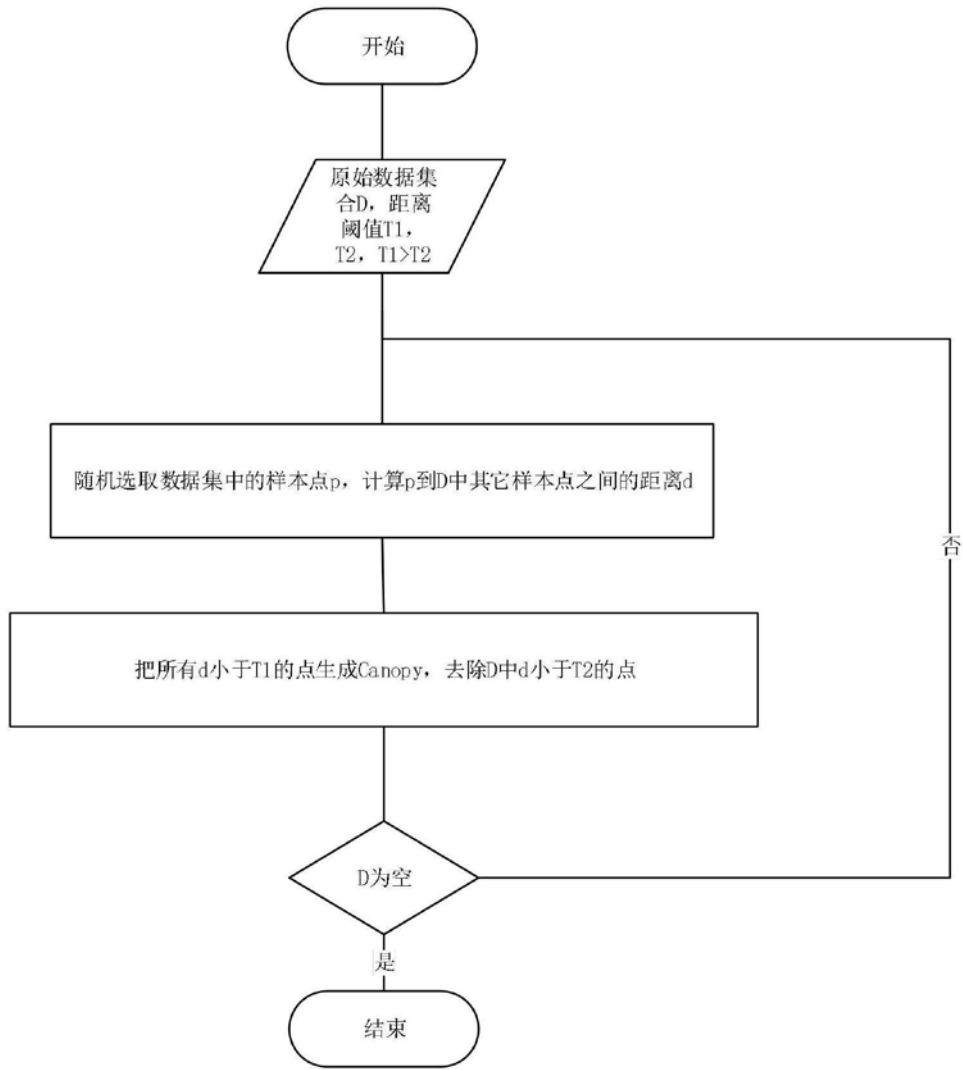


图2



图3

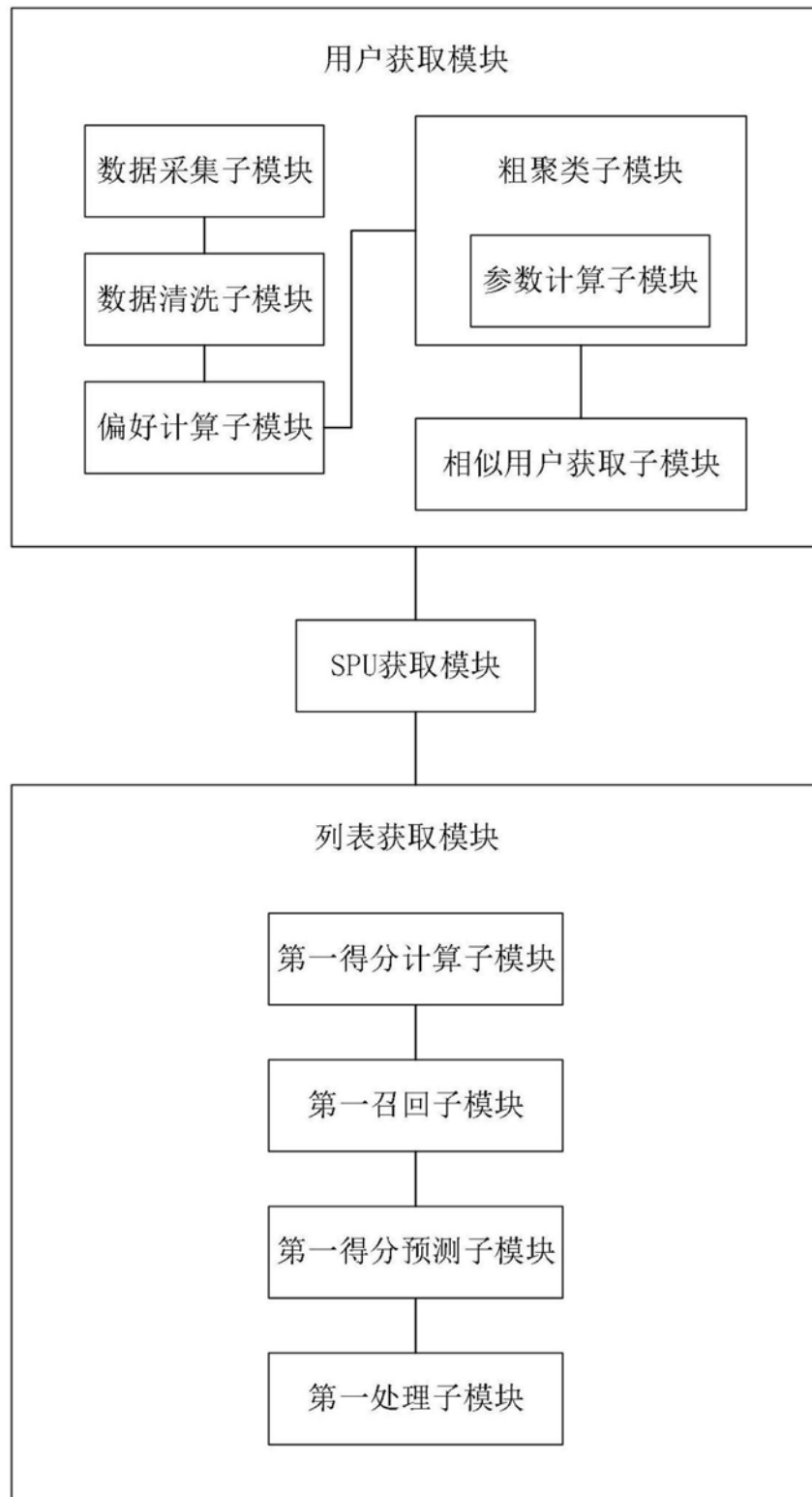


图4