

(19) 中华人民共和国国家知识产权局



## (12) 发明专利申请

(10) 申请公布号 CN 105282616 A

(43) 申请公布日 2016.01.27

(21) 申请号 201510777034.6

(22) 申请日 2015.11.13

(71) 申请人 云南大学

地址 650091 云南省昆明市翠湖北路 2 号云  
南大学科学馆 524

(72) 发明人 李浩 夏欢 康雁

(74) 专利代理机构 北京科亿知识产权代理事务  
所(普通合伙) 11350

代理人 汤东凤

(51) Int. Cl.

H04N 21/466(2011.01)

H04N 21/25(2011.01)

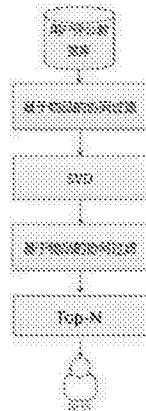
权利要求书2页 说明书4页 附图1页

(54) 发明名称

一种组合协同过滤 IPTV 节目推荐方法

(57) 摘要

本发明提供了一种组合协同过滤 IPTV 节目推荐方法，首先为节目构建用户 - 节目评分矩阵，其次利用基于物品协同过滤为原始用户 - 节目评分矩阵进行填充，然后把填充的用户 - 节目评分矩阵作为奇异值分解技术的输入，通过奇异值分解技术能够将原始用户 - 节目评分矩阵重新填充而得到一个无缺失值的用户 - 节目评分矩阵，最后将无缺失值的用户 - 节目评分矩阵再次使用基于物品的协同过滤方法预测出原始矩阵中未评分项的评分，把预测评分最高的 N 个节目，作为协同过滤的推荐结果。



1. 一种组合协同过滤 IPTV 节目推荐方法, 其特征在于, 包括以下步骤:

步骤①: 从用户评分数据库读取用户评分数据;

步骤②: 用所述步骤①中的用户评分数据定义一个  $m \times n$  的原始评分矩阵  $R$ , 表示  $m$  个用户在  $n$  个节目上的评分; 利用协同过滤方法, 计算节目之间的相似度  $\text{sim}(i, j)$ , 并求出目标用户  $u$  对节目  $j$  的预测评分值  $\text{Pre}_{u,j}$ , 最后用预测评分值  $\text{Pre}_{u,j}$  填充原始评分矩阵  $R$  得到矩阵  $R_{\text{pre}}$ ;

步骤③: 若所述步骤②中的矩阵  $R_{\text{pre}}$  没有完全填充, 首先计算出  $R_{\text{pre}}$  中的每列平均值  $\bar{R}_j$ , 之后用每列平均值  $\bar{R}_j$  填充该列的空缺项; 然后用每列的所有评分项减去目标节目所在行的行平均值  $\bar{R}_i$  得到规范化的矩阵  $R_{\text{norm}}$ ;

步骤④: 用奇异值分解方法分解所述步骤③中的矩阵  $R_{\text{norm}}$  得到矩阵  $U, S, V$ ; 它们的大小分别为  $m \times m, m \times n, n \times n$ ; 它们之间的关系是  $R_{\text{norm}} = U \times S \times V^T$ ;

步骤⑤: 通过保留所述步骤④中矩阵  $S$  的  $k$  个对角元素来获得一个  $k \times k$  的矩阵  $S_k$ , 所述步骤④中矩阵  $U$  和矩阵  $V$  也相应简化为  $U_k$  和  $V_k$ , 它们的大小分别为  $m \times k$  和  $n \times k$ ;

步骤⑥: 利用所述步骤⑤中简化的矩阵  $S_k, U_k, V_k$ , 将  $m$  个用户在  $k$  维特征空间中表示为  $U_k \times \sqrt{S_k}$ ,  $n$  个节目在  $k$  维特征空间中表示为  $\sqrt{S_k} \times V_k^T$ ; 然后运用公式  $\text{Pre}_{u,i} = \bar{R}_u + U_k \sqrt{S_k}(u) \times \sqrt{S_k} V_k^T(i)$  计算目标用户  $u$  在节目  $i$  上的预测评分值  $\text{Pre}_{u,i}$ , 其中  $\bar{R}_u$  表示目标用户  $u$  的所有已评分项目上评分的平均值,  $\sqrt{S_k} V_k^T(i)$  代表矩阵  $\sqrt{S_k} V_k^T$  的第  $i$  列,  $U_k \sqrt{S_k}(u)$  代表矩阵  $U_k \sqrt{S_k}$  的第  $u$  行; 使用本次预测评分值  $\text{Pre}_{u,i}$  填充所述步骤②中的原始评分矩阵  $R$ , 从而得到一个无缺失值的评分矩阵  $R_{\text{filled}}$ ;

步骤⑦: 基于所述步骤⑥中的评分矩阵  $R_{\text{filled}}$ , 利用协同过滤方法, 计算节目  $i$  与节目  $j$  的相似度  $\text{sim}(i, j)$ , 进而求出目标用户  $u$  对节目  $j$  的预测评分值  $\text{Pre}_{u,j}$ ; 当得到了预测评分值之后, 把预测评分值最高的  $N$  个节目, 作为此次协同过滤的推荐结果。

2. 如权利要求 1 所述的一种组合协同过滤 IPTV 节目推荐方法, 其特征在于, 所述步骤②和步骤⑦中计算相似度  $\text{sim}(i, j)$  的方法为修正的余弦相似性度量方法, 公式为以下所列公式:

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (\text{R}_{u,i} - \bar{R}_u)(\text{R}_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (\text{R}_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (\text{R}_{u,j} - \bar{R}_u)^2}}$$

其中, 其中  $\bar{R}_u$  表示目标用户  $u$  的所有已打项目评分的平均值,  $\text{R}_{u,i}$  表示用户  $u$  对节目  $i$  的评分值,  $U$  表示用户集合。

3. 如权利要求 1 所述的一种组合协同过滤 IPTV 节目推荐方法, 其特征在于, 所述步骤②和步骤⑦中计算预测评分值  $\text{Pre}_{u,j}$  的公式为以下所列公式:

$$\text{Pre}_{u,j} = \frac{\sum_{i \in \text{ratedItem}(u)} \text{sim}(i, j) \times \text{R}_{u,i}}{\sum_{i \in \text{ratedItem}(u)} \text{sim}(i, j)}$$

其中,  $\text{ratedItem}(u)$  代表用户  $u$  已评分项的集合,  $\text{R}_{u,i}$  代表原始矩阵  $R$  中用户  $u$  在节目  $i$  上的评分,  $\text{sim}(i, j)$  代表节目  $i$  与节目  $j$  的相似度。

4. 如权利要求 1 所述的一种组合协同过滤 IPTV 节目推荐方法, 其特征在于, 所述步骤⑤中参数 k 会通过实践不断优化。

5. 如权利要求 1 所述的一种组合协同过滤 IPTV 节目推荐方法, 其特征在于, 所述步骤②和步骤⑦中的协同过滤方法为基于物品的协同过滤方法。

## 一种组合协同过滤 IPTV 节目推荐方法

### 技术领域

[0001] 本发明涉及通信领域,具体涉及一种组合协同过滤 IPTV 节目推荐方法。

### 背景技术

[0002] IPTV 作为一种新兴的交互式服务技术,在电视节目推荐领域被广泛研究;IPTV 不同于传统电视,有以下特点:

[0003] 可交互:传统电视的信息传播是单向的,而 IPTV 在此基础上还允许用户与系统交互,特别是允许系统显式或隐式收集用户偏好。

[0004] 个性化:IPTV 通过用户 IP 地址来传递多媒体内容,那么 IPTV 服务商提供的服务即可以精确到个人。

[0005] 多媒体:IPTV 服务包括直播和点播电视节目、电影服务。

[0006] 基于以上 IPTV 的特点,在大量电视节目、电影(下文统称为电视节目)的背景下,为了让用户快速轻松地找到用户真正喜欢的节目。推荐系统作为一种信息过滤技术,通过收集一段时间内用户的历史数据,并从数据中分析用户的兴趣偏好,建立用户兴趣模型,从而主动给用户推荐能够满足他们兴趣和需求的信息。那么若把推荐系统集成在 IPTV 基础设施中,利用收集到的显式或隐式数据计算出推荐结果,通过 IP 宽带网把推荐结果精准地传送给用户。这将给用户挑选电视节目提供一个更快,更有趣的方式。

[0007] 目前主要有两类协同过滤推荐算法:基于用户的协同过滤推荐算法和基于物品的协同过滤算法。协同过滤的优点是能够处理多种类型的内容,包括结构化或者非结构化的,比如文本,电影,图像等;不需要用户的配置文件或者物品内容数据;可以推荐与用户过去喜欢的物品不相似的,但用户很有可能喜欢的物品。相应地,协同过滤算法也存在不少问题,冷启动问题和稀疏性问题。单纯的 SVD 算法在计算稀疏性特别大矩阵时,其推荐质量并不理想;Sarwar 等人的实验结果表明奇异值分解(SVD)方法应用于协同过滤时对于稀疏的评分矩阵效果比较好,但是如果评分矩阵过于稀疏,奇异值分解(SVD)方法的推荐结果并不如基于物品的协同过滤,并且当评分矩阵较稠密时,奇异值分解(SVD)的效果也不如基于物品的协同过滤。

### 发明内容

[0008] 为了解决现有技术中的问题,本发明提供了一种组合协同过滤 IPTV 节目推荐方法。

[0009] 本发明的一种组合协同过滤 IPTV 节目推荐方法,包括以下步骤:

[0010] 步骤①:从用户评分数据库读取用户评分数据;

[0011] 步骤②:用所述步骤①中的用户评分数据定义一个  $m \times n$  的原始评分矩阵 R,表示 m 个用户在 n 个节目上的评分;利用协同过滤方法,计算节目之间的相似度  $sim(i, j)$ ,并求出目标用户 u 对节目 j 的预测评分值  $Pre_{u,j}$ ,最后用预测评分值  $Pre_{u,j}$  填充原始评分矩阵 R 得到矩阵  $R_{pre}$ ;

[0012] 步骤③：若所述步骤②中的矩阵  $R_{pre}$  没有完全填充，首先计算出  $R_{pre}$  中的每列平均值  $\bar{R}_j$ ，之后用每列平均值  $\bar{R}_j$  填充该列的空缺项；然后用每列的所有评分项减去目标节目所在行的行平均值  $\bar{R}_i$  得到规范化的矩阵  $R_{norm}$ ；

[0013] 步骤④：用奇异值分解方法分解所述步骤③中的矩阵  $R_{norm}$  得到矩阵  $U, S, V$ ；它们的大小分别为  $m \times m, m \times n, n \times n$ ；它们之间的关系是  $R_{norm} = U \times S \times V^T$ ；

[0014] 步骤⑤：通过保留所述步骤④中矩阵  $S$  的  $k$  个对角元素来获得一个  $k \times k$  的矩阵  $S_k$ ，所述步骤④中矩阵  $U$  和矩阵  $V$  也相应简化为  $U_k$  和  $V_k$ ，它们的大小分别为  $m \times k$  和  $n \times k$ ；

[0015] 步骤⑥：利用所述步骤⑤中简化的矩阵  $S_k, U_k, V_k$ ，将  $m$  个用户在  $k$  维特征空间中表示为  $U_k \times \sqrt{S_k}$ ， $n$  个节目在  $k$  维特征空间中表示为  $\sqrt{S_k} \times V_k^T$ ；然后运用公式

$Pre_{u,i} = \bar{R}_u + U_k \sqrt{S_k}(u) \times \sqrt{S_k} V_k^T(i)$  计算目标用户  $u$  在节目  $i$  上的预测评分值  $Pre_{u,i}$ ，其中  $\bar{R}_u$  表示目标用户  $u$  的所有已评分项目上评分的平均值， $\sqrt{S_k} V_k^T(i)$  代表矩阵  $\sqrt{S_k} V_k^T$  的第  $i$  列， $U_k \sqrt{S_k}(u)$  代表矩阵  $U_k \sqrt{S_k}$  的第  $u$  行；使用本次预测评分值  $Pre_{u,i}$  填充所述步骤②中的原始评分矩阵  $R$ ，从而得到一个无缺失值的评分矩阵  $R_{filled}$ ；

[0016] 步骤⑦：基于所述步骤⑥中的评分矩阵  $R_{filled}$ ，利用协同过滤方法，计算节目  $i$  与节目  $j$  的相似度  $sim(i, j)$ ，进而求出目标用户  $u$  对节目  $j$  的预测评分值  $Pre_{u,j}$ ；当得到了预测评分值之后，把预测评分值最高的  $N$  个节目，作为此次协同过滤的推荐结果。

[0017] 所述步骤②和步骤⑦中计算相似度  $sim(i, j)$  的方法为修正的余弦相似性度量方法，公式为以下所列公式：

$$[0018] sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}$$

[0019] 其中，其中  $\bar{R}_u$  表示目标用户  $u$  的所有已打项目评分的平均值， $R_{u,i}$  表示用户  $u$  对节目  $i$  的评分值， $U$  表示用户集合。

[0020] 所述步骤②和步骤⑦中计算预测评分值  $Pre_{u,j}$  的公式为以下所列公式：

$$[0021] Pre_{u,j} = \frac{\sum_{i \in ratedItem(u)} sim(i, j) \times R_{u,i}}{\sum_{i \in ratedItem(u)} sim(i, j)}$$

[0022] 其中， $ratedItem(u)$  代表用户  $u$  已评分项的集合， $R_{u,i}$  代表原始矩阵  $R$  中用户  $u$  在节目  $i$  上的评分， $sim(i, j)$  代表节目  $i$  与节目  $j$  的相似度。

[0023] 所述步骤⑤的参数  $k$  会通过实践不断优化。

[0024] 所述步骤②和步骤⑦中的协同过滤方法为基于物品的协同过滤方法。

[0025] 本发明相对于现有技术具有以下优点：

[0026] 1、本发明巧妙结合了基于物品的协同过滤方法和奇异值分解 (SVD) 方法的优点，以特征值递增的方式预测评分，更好的解决了评分矩阵的稀疏性问题；

[0027] 2、本发明的协同过滤方法采用了基于物品的协同过滤方法，这在 IPTV 推荐领域中比基于用户的协同过滤方法精确度更高。

## 附图说明

[0028] 图 1 为本发明的流程示意图。

## 具体实施方式

[0029] 以下结合附图,对本发明的一种组合协同过滤 IPTV 节目推荐方法做进一步的详细说明。

[0030] 如图 1 所示,本发明的一种组合协同过滤 IPTV 节目推荐方法,包括以下步骤:

[0031] 步骤①:从用户评分数据库读取用户评分数据;

[0032] 步骤②:用所述步骤①中的用户评分数据定义一个  $m \times n$  的原始评分矩阵 R,表示 m 个用户在 n 个节目上的评分;利用协同过滤方法,计算节目之间的相似度  $\text{sim}(i, j)$ ,并求出目标用户 u 对节目 j 的预测评分值  $\text{Pre}_{u,j}$ ,最后用预测评分值  $\text{Pre}_{u,j}$  填充原始评分矩阵 R 得到矩阵  $R_{\text{pre}}$ ;

[0033] 步骤③:若所述步骤②中的矩阵  $R_{\text{pre}}$  没有完全填充,首先计算出  $R_{\text{pre}}$  中的每列平均值  $\bar{R}_j$ ,之后用每列平均值  $\bar{R}_j$  填充该列的空缺项;然后用每列的所有评分项减去目标节目所在行的行平均值  $\bar{R}_i$  得到规范化的矩阵  $R_{\text{norm}}$ ;

[0034] 步骤④:用奇异值分解方法分解所述步骤③中的矩阵  $R_{\text{norm}}$  得到矩阵 U, S, V;它们的大小分别为  $m \times m$ ,  $m \times n$ ,  $n \times n$ ;它们之间的关系是  $R_{\text{norm}} = U \times S \times V^T$ ;

[0035] 步骤⑤:通过保留所述步骤④中矩阵 S 的 k 个对角元素来获得一个  $k \times k$  的矩阵  $S_k$ ,所述步骤④中矩阵 U 和矩阵 V 也相应简化为  $U_k$  和  $V_k$ ,它们的大小分别为  $m \times k$  和  $n \times k$ ;

[0036] 步骤⑥:利用所述步骤⑤中简化的矩阵  $S_k$ ,  $U_k$ ,  $V_k$ ,将 m 个用户在 k 维特征空间中表示为  $U_k \times \sqrt{S_k}$ ,n 个节目在 k 维特征空间中表示为  $\sqrt{S_k} \times V_k^T$ ;然后运用公式  $\text{Pre}_{u,i} = \bar{R}_u + U_k \sqrt{S_k}(u) \times \sqrt{S_k} V_k^T(i)$  计算目标用户 u 在节目 i 上的预测评分值  $\text{Pre}_{u,i}$ ,

其中  $\bar{R}_u$  表示目标用户 u 的所有已评分项目上评分的平均值,  $\sqrt{S_k} V_k^T(i)$  代表矩阵  $\sqrt{S_k} V_k^T$  的第 i 列,  $U_k \sqrt{S_k}(u)$  代表矩阵  $U_k \sqrt{S_k}$  的第 u 行;使用本次预测评分值  $\text{Pre}_{u,i}$  填充所述步骤②中的原始评分矩阵 R,从而得到一个无缺失值的评分矩阵  $R_{\text{filled}}$ ;

[0037] 步骤⑦:基于所述步骤⑥中的评分矩阵  $R_{\text{filled}}$ ,利用协同过滤方法,计算节目 i 与节目 j 的相似度  $\text{sim}(i, j)$ ,进而求出目标用户 u 对节目 j 的预测评分值  $\text{Pre}_{u,j}$ ;当得到了预测评分值之后,把预测评分值最高的 N 个节目,作为此次协同过滤的推荐结果。

[0038] 所述步骤②和步骤⑦中计算相似度  $\text{sim}(i, j)$  的方法为修正的余弦相似性度量方法,公式为以下所列公式:

$$[0039] \text{sim}(i, j) = \frac{\sum_{u \in U} (\text{R}_{u,i} - \bar{R}_u)(\text{R}_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (\text{R}_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (\text{R}_{u,j} - \bar{R}_u)^2}}$$

[0040] 其中,其中  $\bar{R}_u$  表示目标用户 u 的所有已打项目评分的平均值,  $\text{R}_{u,i}$  表示用户 u 对节目 i 的评分值, U 表示用户集合。

[0041] 所述步骤②和步骤⑦中计算预测评分值  $\text{Pre}_{u,j}$  的公式为以下所列公式:

$$[0042] \quad Pre_{u,j} = \frac{\sum_{i \in ratedItem(u)} sim(i,j) \times R_{u,i}}{\sum_{i \in ratedItem(u)} sim(i,j)}$$

[0043] 其中, ratedItem(u) 代表用户 u 已评分项的集合,  $R_{u,i}$  代表原始矩阵 R 中用户 u 在节目 i 上的评分,  $sim(i, j)$  代表节目 i 与节目 j 的相似度。

[0044] 所述步骤⑤的参数 k 会通过实践不断优化。

[0045] 所述步骤②和步骤⑦中的协同过滤方法为基于物品的协同过滤方法。

[0046] 简而言之,首先为节目构建用户 - 节目评分矩阵,其次利用基于物品协同过滤为原始用户 - 节目评分矩阵进行填充,然后把填充的用户 - 节目评分矩阵作为奇异值分解技术的输入,通过奇异值分解技术能够将原始用户 - 节目评分矩阵重新填充而得到一个无缺失值的用户 - 节目评分矩阵,最后将无缺失值的用户 - 节目评分矩阵再次使用基于物品的协同过滤方法并预测出未原始矩阵中未评分项的评分,把预测评分最高的 N 个节目,作为组合协同过滤的推荐结果。基于物品的协同过滤被证明在用户兴趣较为固定的领域中(比如图书,电影,电子商务,电视节目等)比基于用户的方法更有优势。

[0047] 本发明结合了 SVD 与基于物品的协同过滤各自的优点,以特征值递增的方式预测评分。如图 1 所示,利用 SVD 的方法来预测未打分项的预测值,得到一个被预测值填充的矩阵,然后用被填充的评分矩阵来计算物品之间的相似度,最后再使用基于物品的协同过滤方法求出原始矩阵中未评分的项目的预测值;单纯的 SVD 技术也有自身的缺点,因为在对矩阵进行奇异值分解的时候,要求矩阵不能有空余项,通常的处理方式是把每一列的空余项填充为该列的平均值,但是这种简单的填充方式并不能准确的代表用户对某节目的具体评分值,因而首先使用协同过滤技术对原始矩阵的未评分项进行预测填充。由于协同过滤技术的限制并不是所有节目都能被预测出评分,这部分较少的空余项将以该节目所在列的平均评分值填充。

[0048] 综上,最优的方式是先使用基于物品的协同过滤的输出作为 SVD 的输入,SVD 的输出又将作为基于物品的协同过滤的输入,最终得到高质量的推荐结果。

[0049] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

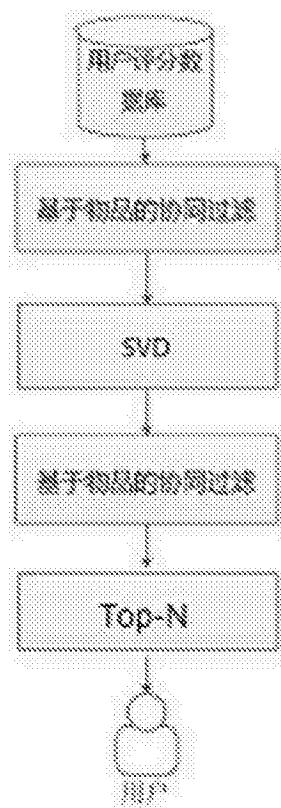


图 1