

(12) 发明专利申请

(10) 申请公布号 CN 102495864 A

(43) 申请公布日 2012. 06. 13

(21) 申请号 201110382078. 0

(22) 申请日 2011. 11. 25

(71) 申请人 清华大学

地址 100084 北京市海淀区清华园北京  
100084-82 信箱

(72) 发明人 谢峰 陈震 曹军威

(74) 专利代理机构 北京路浩知识产权代理有限  
公司 11002

代理人 王莹

(51) Int. Cl.

G06F 17/30 (2006. 01)

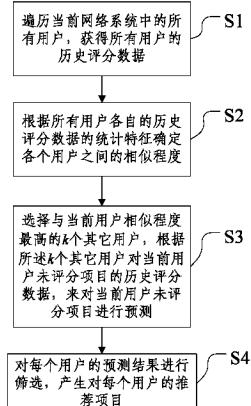
权利要求书 2 页 说明书 6 页 附图 3 页

(54) 发明名称

基于评分的协同过滤推荐方法及系统

(57) 摘要

本发明公开了一种基于评分的协同过滤推荐方法及系统，涉及协同过滤推荐技术领域，本发明利用用户历史评分数据的统计特征计算各个用户之间的相似程度，通过与当前用户相似程度较高的其他用户来计算当前用户未评价过的项目，实现了在用户和项目数量庞大的协同过滤推荐系统中用户评分数据极端稀疏情况下，解决了难以找到用户间共同评分项而导致无法计算相似性或相似性不准确的问题，能够实现准确且快速的项目推荐。



1. 一种基于评分的协同过滤推荐方法,其特征在于,包括以下步骤:

S1:遍历当前网络系统中的所有用户,获得所有用户的历史评分数据;

S2:根据所有用户各自的历史评分数据的统计特征确定各个用户之间的相似程度;

S3:选择与当前用户相似程度最高的k个其它用户,根据所述k个其它用户对当前用户未评分项目的历史评分数据,来对当前用户未评分项目进行预测;

S4:对每个用户的预测结果进行筛选,产生对每个用户的推荐项目。

2. 如权利要求1所述的方法,其特征在于,步骤S2中各个用户之间的相似程度通过下列公式进行计算,

$$\text{sim}(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| * \|\vec{v}\|} = \frac{\bar{r}_u \bar{r}_v + S_u S_v + R_u R_v}{\sqrt{\bar{r}_u^2 + S_u^2 + R_u^2} \cdot \sqrt{\bar{r}_v^2 + S_v^2 + R_v^2}}$$

其中,  $\text{sim}(u, v)$  为用户  $u$  和用户  $v$  之间的相似程度,  $\vec{u} = (\bar{r}_u, S_u, R_u)$ ,  $\vec{v} = (\bar{r}_v, S_v, R_v)$ ,  $\vec{u}$  和  $\vec{v}$  为依次对应用户  $u$  和用户  $v$  的三维向量,  $\bar{r}_u$ 、 $S_u$  和  $R_u$  为依次对应用户  $u$  的历史评分数据的均值、方差和极差,  $\bar{r}_v$ 、 $S_v$  和  $R_v$  为依次对应用户  $v$  的历史评分数据的均值、方差和极差。

3. 如权利要求2所述的方法,其特征在于,步骤S3中通过下式来对当前用户未评分项目进行预测,

$$P_{u,i} = \bar{r}_u + \frac{1}{\sum_{x \in S(u)} \text{sim}(u, x)} \sum_{x \in S(u)} \text{sim}(u, x) (r_{x,i} - \bar{r}_x)$$

其中,  $P_{u,i}$  为当前用户  $u$  的未评分项目  $i$  的预测分数,  $S(u)$  为  $k$  个其它用户中对当前用户  $u$  的未评分项目  $i$  进行过评分的用户集合,  $r_{x,i}$  为用户  $x$  对当前用户  $u$  的未评分项目  $i$  的评分, 用户  $x$  为  $S(u)$  中的某个元素。

4. 如权利要求1~3中任一项所述的方法,其特征在于,步骤S4中对每个用户的预测结果进行筛选时,通过将当前用户的未评分项目的预测分数与当前用户的历史评分数据的均值进行比较,若大于均值,则作为推荐项目。

5. 一种基于评分的协同过滤推荐系统,其特征在于,包括:

历史数据统计模块,用于遍历当前网络系统中的所有用户,获得所有用户的历史评分数据;

相似程度计算模块,用于根据所有用户各自的历史评分数据的统计特征确定各个用户之间的相似程度;

预测模块,用于选择与当前用户相似程度最高的  $k$  个其它用户,根据所述  $k$  个其它用户对当前用户未评分项目的历史评分数据,来对当前用户未评分项目进行预测;

筛选模块,用于对每个用户的预测结果进行筛选,产生对每个用户的推荐项目。

6. 如权利要求5所述的系统,其特征在于,相似程度计算模块中各个用户之间的相似程度通过下列公式进行计算,

$$\text{sim}(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| * \|\vec{v}\|} = \frac{\bar{r}_u \bar{r}_v + S_u S_v + R_u R_v}{\sqrt{\bar{r}_u^2 + S_u^2 + R_u^2} \cdot \sqrt{\bar{r}_v^2 + S_v^2 + R_v^2}}$$

其中,  $\text{sim}(u, v)$  为用户  $u$  和用户  $v$  之间的相似程度,  $\vec{u} = (\bar{r}_u, S_u, R_u)$ ,  $\vec{v} = (\bar{r}_v, S_v, R_v)$ ,  $\vec{u}$  和  $\vec{v}$  为依次对应用户  $u$  和用户  $v$  的三维向量,  $\bar{r}_u$ 、 $S_u$  和  $R_u$  为依次对应用户  $u$  的历史评分数据的均

值、方差和极差， $\bar{r}_v$ 、 $S_v$  和  $R_v$  为依次对应用户  $v$  的历史评分数据的均值、方差和极差。

7. 如权利要求 6 所述的系统，其特征在于，预测模块中通过下式对当前用户未评分项目进行预测，

$$P_{u,i} = \bar{r}_u + \frac{1}{\sum_{x \in S(u)} sim(u, x)} \sum_{x \in S(u)} sim(u, x)(r_{x,i} - \bar{r}_x)$$

其中， $P_{u,i}$  为当前用户  $u$  的未评分项目  $i$  的预测分数， $S(u)$  为  $k$  个其它用户中对当前用户  $u$  的未评分项目  $i$  进行过评分的用户集合， $r_{x,i}$  为用户  $x$  对当前用户  $u$  的未评分项目  $i$  的评分，用户  $x$  为  $S(u)$  中的某个元素。

8. 如权利要求 5～7 中任一项所述的系统，其特征在于，筛选模块中对每个用户的预测结果进行筛选时，通过将当前用户的未评分项目的预测分数与当前用户的历史评分数据的均值进行比较，若大于均值，则作为推荐项目。

## 基于评分的协同过滤推荐方法及系统

### 技术领域

[0001] 本发明涉及协同过滤推荐技术领域,特别涉及一种基于评分的协同过滤推荐方法及系统。

### 背景技术

[0002] 互联网技术的迅猛发展把我们带进了信息爆炸的时代,海量信息的同时呈现,不仅使用户很难从中发现自己感兴趣的内容,而且使得大量少为人知的信息成为网络中的“暗信息”,无法被一般用户获取。推荐系统通过建立用户与项目(例如:产品、电影、音乐、节目等)之间的二元关系,利用用户历史选择信息或相似性关系挖掘用户潜在的兴趣爱好,进而进行推荐。

[0003] 目前已存在许多经典推荐系统,协同过滤推荐系统是最早被提出并得到广泛应用的推荐系统。其核心思想就是基于评分相似的最近邻居的评分数据向目标用户产生推荐。由于最近邻居对项目的评分与目标用户非常相似,因此目标用户对未评分项目的评分可以通过最近邻居对该项目评分的加权平均值逼近。Typestry 是最早提出来的基于协同过滤的推荐系统,目标用户需要指出与自己兴趣爱好比较类似的其他用户。GroupLens 是基于用户评分的自动化协同过滤推荐系统,用于电影和新闻推荐。其他利用协同过滤方法进行推荐的系统还有 Amazon. com 的书籍推荐系统, Jester 的笑话推荐系统等等。

[0004] 与一般的推荐系统相比,协同过滤推荐系统具有两大优点:一是对推荐对象没有特殊的要求,能处理音乐、电影等难以进行文本结构化表示的对象;二是具有推荐新信息的能力,可以发现用户潜在的但自己尚未察觉的兴趣偏好。

[0005] 传统的协同过滤推荐系统利用不同用户间共同评分项目计算相似性,主流的相似性计算方法包括:余弦相似性方法和相关相似性方法;目标用户对未评分项目的评分通过相似性较大的邻居对项目的评分的加权平均值进行预测。可以看出,协同过滤推荐系统的推荐精度取决于用户间相似性计算的准确度。然而在用户和项目数量庞大的网络系统中用户评分数据极端稀疏情况下,难以找到用户间共同评分项目,从而导致用户间相似性计算结果不准确甚至无法计算相似性。协同过滤推荐系统得到了广泛应用的同时,也面临着很多问题,例如如何对新用户进行推荐或如何推荐新产品给用户即冷启动问题,评分稀疏性问题,算法可扩展性问题等。另外,传统的协同过滤推荐算法随着用户数量的增多,计算量成线性加大,实时性能越来越差,同时响应速度也越来越慢。

### 发明内容

[0006] (一) 要解决的技术问题

[0007] 本发明要解决的技术问题是:如何在用户和项目数量庞大的协同过滤推荐系统中用户评分数据极端稀疏情况下,解决难以找到用户间共同评分项而导致无法计算相似性或相似性不准确的问题。

[0008] (二) 技术方案

[0009] 为解决上述技术问题,本发明提供了一种基于评分的协同过滤推荐方法,包括以下步骤:

[0010] S1:遍历当前网络系统中的所有用户,获得所有用户的历史评分数据;

[0011] S2:根据所有用户各自的历史评分数据的统计特征确定各个用户之间的相似程度;

[0012] S3:选择与当前用户相似程度最高的k个其它用户,根据所述k个其它用户对当前用户未评分项目的历史评分数据,来对当前用户未评分项目进行预测;

[0013] S4:对每个用户的预测结果进行筛选,产生对每个用户的推荐项目。

[0014] 优选地,步骤S2中各个用户之间的相似程度通过下列公式进行计算,

$$[0015] sim(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| * \|\vec{v}\|} = \frac{\bar{r}_u \bar{r}_v + S_u S_v + R_u R_v}{\sqrt{\bar{r}_u^2 + S_u^2 + R_u^2} \cdot \sqrt{\bar{r}_v^2 + S_v^2 + R_v^2}}$$

[0016] 其中,  $sim(u, v)$  为用户  $u$  和用户  $v$  之间的相似程度,  $\vec{u} = (\bar{r}_u, S_u, R_u)$ ,  $\vec{v} = (\bar{r}_v, S_v, R_v)$ ,  $\vec{u}$  和  $\vec{v}$  为依次对应用户  $u$  和用户  $v$  的三维向量,  $\bar{r}_u$ 、 $S_u$  和  $R_u$  为依次对应用户  $u$  的历史评分数据的均值、方差和极差,  $\bar{r}_v$ 、 $S_v$  和  $R_v$  为依次对应用户  $v$  的历史评分数据的均值、方差和极差。

[0017] 优选地,步骤S3中通过下式来对当前用户未评分项目进行预测,

$$[0018] P_{u,i} = \bar{r}_u + \frac{1}{\sum_{x \in S(u)} sim(u, x)} \sum_{x \in S(u)} sim(u, x) (r_{x,i} - \bar{r}_x)$$

[0019] 其中,  $P_{u,i}$  为当前用户  $u$  的未评分项目  $i$  的预测分数,  $S(u)$  为  $k$  个其它用户中对当前用户  $u$  的未评分项目  $i$  进行过评分的用户集合,  $r_{x,i}$  为用户  $x$  对当前用户  $u$  的未评分项目  $i$  的评分, 用户  $x$  为  $S(u)$  中的某个元素。

[0020] 优选地,步骤S4中对每个用户的预测结果进行筛选时,通过将当前用户的未评分项目的预测分数与当前用户的历史评分数据的均值进行比较,若大于均值,则作为推荐项目。

[0021] 本发明还公开了一种基于评分的协同过滤推荐系统,包括:

[0022] 历史数据计算模块,用于遍历当前网络系统中的所有用户,获得所有用户的历史评分数据;

[0023] 相似程度计算模块,用于根据所有用户各自的历史评分数据的统计特征确定各个用户之间的相似程度;

[0024] 预测模块,用于选择与当前用户相似程度最高的  $k$  个其它用户,根据所述  $k$  个其它用户对当前用户未评分项目的历史评分数据,来对当前用户未评分项目进行预测;

[0025] 筛选模块,用于对每个用户的预测结果进行筛选,产生对每个用户的推荐项目。

[0026] 优选地,相似程度计算模块中各个用户之间的相似程度通过下列公式进行计算,

$$[0027] sim(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| * \|\vec{v}\|} = \frac{\bar{r}_u \bar{r}_v + S_u S_v + R_u R_v}{\sqrt{\bar{r}_u^2 + S_u^2 + R_u^2} \cdot \sqrt{\bar{r}_v^2 + S_v^2 + R_v^2}}$$

[0028] 其中,  $sim(u, v)$  为用户  $u$  和用户  $v$  之间的相似程度,  $\vec{u} = (\bar{r}_u, S_u, R_u)$ ,  $\vec{v} = (\bar{r}_v, S_v, R_v)$ ,  $\vec{u}$  和  $\vec{v}$  为依次对应用户  $u$  和用户  $v$  的三维向量,  $\bar{r}_u$ 、 $S_u$  和  $R_u$  为依次对应用户  $u$  的历史评分数据的均值、方差和极差,  $\bar{r}_v$ 、 $S_v$  和  $R_v$  为依次对应用户  $v$  的历史评分数据的均值、方差和极差。

[0029] 优选地，预测模块中通过下式对当前用户未评分项目进行预测，

$$[0030] P_{u,i} = \bar{r}_u + \frac{1}{\sum_{x \in S(u)} sim(u, x)} \sum_{x \in S(u)} sim(u, x)(r_{x,i} - \bar{r}_x)$$

[0031] 其中， $P_{u,i}$  为当前用户  $u$  的未评分项目  $i$  的预测分数， $S(u)$  为  $k$  个其它用户中对当前用户  $u$  的未评分项目  $i$  进行过评分的用户集合， $r_{x,i}$  为用户  $x$  对当前用户  $u$  的未评分项目  $i$  的评分，用户  $x$  为  $S(u)$  中的某个元素。

[0032] 优选地，筛选模块中对每个用户的预测结果进行筛选时，通过将当前用户的未评分项目的预测分数与当前用户的历史评分数据的均值进行比较，若大于均值，则作为推荐项目。

### [0033] (三) 有益效果

[0034] 本发明利用用户历史评分数据的统计特征计算各个用户之间的相似程度，通过与当前用户相似程度较高的其他用户来计算当前用户未评价过的项目，实现了在用户和项目数量庞大的协同过滤推荐系统中用户评分数据极端稀疏情况下，解决了难以找到用户间共同评分项而导致无法计算相似性或相似性不准确的问题，能够实现准确且快速的项目推荐。

## 附图说明

[0035] 图 1 是按照本发明一种实施方式的基于评分的协同过滤推荐方法的流程图；

[0036] 图 2 是本发明一种实施例中当前网络系统中仅包括 4 个用户，且这 4 个用户对 10 个项目的评分数据的示意图；

[0037] 图 3 是本发明一种实施例中经过计算后获得的统计量表征矩阵的示意图；

[0038] 图 4 是本发明一种实施例中各个用户之间相似程度的示意图；

[0039] 图 5 是本发明一种实施例中用户未评分项目的预测分数，以及各个用户得到的项目推荐结果；

[0040] 图 6 是本发明一种实施例中以“MovieLens 100K”作为数据集，随机选择 80% 作为训练集，剩下 20% 作为测试集，分别采用本发明的方法和传统协同过滤方法的平均绝对误差随所述  $k$  个其它用户的数量变化的结果；

[0041] 图 7 是以“MovieLens 100K”作为数据集，随机选择 80% 作为训练集，剩下 20% 作为测试集，分别采用本发明的方法和传统协同过滤方法的均方根绝对误差随所述  $k$  个其它用户的数量变化的结果。

## 具体实施方式

[0042] 下面结合附图和实施例，对本发明的具体实施方式作进一步详细描述。以下实施例用于说明本发明，但不用来限制本发明的范围。

[0043] 图 1 是按照本发明一种实施方式的基于评分的协同过滤推荐方法的流程图，参照图 1，本实施方式的方法包括以下步骤：

[0044] S1：遍历当前网络系统中的所有用户，获得所有用户的历史评分数据；

[0045] S2：根据所有用户各自的历史评分数据的统计特征确定各个用户之间的相似程度；本实施方式中，先构建每个用户的三维向量，再根据每个用户的三维向量通过夹角余弦

值计算各个用户之间的相似程度；将各用户特征向量间的夹角余弦值作为彼此之间的相似程度。两向量夹角越小，余弦值越大，表明用户越相似；

[0046] S3：选择与当前用户相似程度最高的 k 个其它用户，根据所述 k 个其它用户对当前用户未评分项目的历史评分数据，来对当前用户未评分项目进行预测；

[0047] S4：对每个用户的预测结果进行筛选，产生对每个用户的推荐项目。

[0048] 优选地，步骤 S2 中各个用户之间的相似程度通过下列公式进行计算，

$$[0049] \quad sim(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| * \|\vec{v}\|} = \frac{\bar{r}_u \bar{r}_v + S_u S_v + R_u R_v}{\sqrt{\bar{r}_u^2 + S_u^2 + R_u^2} \cdot \sqrt{\bar{r}_v^2 + S_v^2 + R_v^2}}$$

[0050] 其中， $sim(u, v)$  为用户 u 和用户 v 之间的相似程度， $\vec{u} = (\bar{r}_u, S_u, R_u)$ ,  $\vec{v} = (\bar{r}_v, S_v, R_v)$ ， $\vec{u}$  和  $\vec{v}$  为依次对应用户 u 和用户 v 的三维向量， $\bar{r}_u$ 、 $S_u$  和  $R_u$  为依次对应用户 u 的历史评分数据的均值、方差和极差， $\bar{r}_v$ 、 $S_v$  和  $R_v$  为依次对应用户 v 的历史评分数据的均值、方差和极差。例如， $(\bar{r}_u, S_u, R_u)$  表示用户 u 的历史评分数据的统计特征，平均值表示用户对项目的平均评分喜好，方差表示用户对项目评分的波动或偏差，极差表示用户对评分的范围喜好。在一定程度上说明这三个值将影响用户对未评分项目的未来评分。

[0051] 其中，所有用户各自的历史评分数据的平均值、方差和极差通过下列公式计算，

$$[0052] \quad \bar{r}_u = \frac{\sum_{i \in I(u)} r_{u,i}}{n}$$

$$[0053] \quad S_u = \frac{1}{size(I(u))} \sum_{i \in I(u)} (r_{u,i} - \bar{r}_u)^2$$

$$[0054] \quad R_u = \max(r_{u,i}) - \min(r_{u,i})$$

[0055] 其中， $\bar{r}_u$ ， $S_u$ ， $R_u$  分别表示用户 u 的历史评分数据的均值、方差和极差， $I(u)$  为用户 u 所有评分项目集合， $size(I(u))$  为用户 u 评分项目总数， $r_{u,i}$  为用户 u 对项目 i 的评分， $\max(r_{u,i})$  和  $\min(r_{u,i})$  分别表示用户 u 评分最大值和最小值。

[0056] 优选地，步骤 S3 中通过下式来对当前用户未评分项目进行预测，

$$[0057] \quad P_{u,i} = \bar{r}_u + \frac{1}{\sum_{x \in S(u)} sim(u, x)} \sum_{x \in S(u)} sim(u, x)(r_{x,i} - \bar{r}_x)$$

[0058] 其中， $P_{u,i}$  为当前用户 u 的未评分项目 i 的预测分数， $S(u)$  为 k 个其它用户中对当前用户 u 的未评分项目 i 进行过评分的用户集合， $r_{x,i}$  为用户 x 对当前用户 u 的未评分项目 i 的评分，用户 x 为  $S(u)$  中的某个元素。

[0059] 优选地，步骤 S4 中对每个用户的预测结果进行筛选时，通过将当前用户的未评分项目的预测分数与当前用户的历史评分数据的均值进行比较，若大于均值，则作为推荐项目。

[0060] 实施例 1

[0061] 下面用一个简单的例子说明上述算法推荐过程。参照图 2，该例子包含 4 个用户对 10 个项目的评分，其中每个用户评分过 5 个项目，我们需要对目标用户未评分的 5 个项目进行评分预测，并作出相应推荐。详细步骤说明如下：

[0062] 1、计算各用户历史评分数据的统计特征，如均值、方差和极差，得到如图 3 所示的统计量表征矩阵，即用三维向量描述用户特征。

[0063] 2、利用余弦相似性方法计算各向量间的余弦夹角值,得到用户相似性矩阵,参照图4,相似程度的值越大,说明用户之间相似程度越高,评分越相似,如例中用户  $u_1$  与用户  $u_2$  相似程度最高,而用户  $u_4$  与用户  $u_3$  相似程度最高。

[0064] 3、由于例中用户数较少,在这里我们选择3个相似程度最高的用户作为目标用户的最近邻居,并对目标用户未评分项目进行相似性加权,得到图5所示的评分预测矩阵,根据预测评分值大小对未评分项目进行排序,选择预测评分值大于目标用户平均评分的项目产生推荐,图5中分数加粗的项目即为最终推荐结果。

[0065] 图6是本发明一种实施例中以“MovieLens 100K”作为数据集,随机选择80%作为训练集,剩下20%作为测试集,分别采用本发明的方法和传统协同过滤方法的平均绝对误差随所述k个其它用户的数量变化的结果;图7是以“MovieLens 100K”作为数据集,随机选择80%作为训练集,剩下20%作为测试集,分别采用本发明的方法和传统协同过滤方法的均方根绝对误差随所述k个其它用户的数量变化的结果。

[0066] 本发明所提出的基于评分的协同过滤推荐方法,其优点如下:

[0067] (1) 利用用户评分信息的统计量计算用户间相似性,能有效克服评分稀疏情况下难以找到用户间共同评分项而导致无法计算相似性或相似性计算不准确的缺点;

[0068] (2) 针对每一个用户,都对应唯一的一组统计量(均值、方差、极差),不同用户之间的相似性计算都在三维空间中进行,避免了不同用户间共同评分项目数不同导致的维度差异;

[0069] (3) 数据统计量的计算远远快于共同评分项目的查找,该算法推荐速度比传统协同过滤推荐算法快10倍以上;同时提高了10%左右的推荐精度,使本发明在实际电子商务系统应用中获得更好的性能。

[0070] 本发明还公开了一种基于评分的协同过滤推荐系统,其特征在于,包括:

[0071] 历史数据统计模块,用于遍历当前网络系统中的所有用户,获得所有用户的历史评分数据;

[0072] 相似程度计算模块,用于根据所有用户各自的历史评分数据的统计特征确定各个用户之间的相似程度;

[0073] 预测模块,用于选择与当前用户相似程度最高的k个其它用户,根据所述k个其它用户对当前用户未评分项目的历史评分数据,来对当前用户未评分项目进行预测;

[0074] 筛选模块,用于对每个用户的预测结果进行筛选,产生对每个用户的推荐项目。

[0075] 优选地,相似程度计算模块中各个用户之间的相似程度通过下列公式进行计算,

$$[0076] sim(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| * \|\vec{v}\|} = \frac{\bar{r}_u \bar{r}_v + S_u S_v + R_u R_v}{\sqrt{\bar{r}_u^2 + S_u^2 + R_u^2} \cdot \sqrt{\bar{r}_v^2 + S_v^2 + R_v^2}}$$

[0077] 其中,  $sim(u, v)$  为用户  $u$  和用户  $v$  之间的相似程度,  $\vec{u} = (\bar{r}_u, S_u, R_u)$ ,  $\vec{v} = (\bar{r}_v, S_v, R_v)$ ,  $\vec{u}$  和  $\vec{v}$  为依次对应用户  $u$  和用户  $v$  的三维向量,  $\bar{r}_u$ 、 $S_u$  和  $R_u$  为依次对应用户  $u$  的历史评分数据的均值、方差和极差,  $\bar{r}_v$ 、 $S_v$  和  $R_v$  为依次对应用户  $v$  的历史评分数据的均值、方差和极差。

[0078] 优选地,预测模块中通过下式对当前用户未评分项目进行预测,

$$[0079] P_{u,i} = \bar{r}_u + \frac{1}{\sum_{x \in S(u)} sim(u, x)} \sum_{x \in S(u)} sim(u, x)(r_{x,i} - \bar{r}_x)$$

[0080] 其中,  $P_{u,i}$  为当前用户  $u$  的未评分项目  $i$  的预测分数,  $S(u)$  为  $k$  个其它用户中对当前用户  $u$  的未评分项目  $i$  进行过评分的用户集合,  $r_{x,i}$  为用户  $x$  对当前用户  $u$  的未评分项目  $i$  的评分, 用户  $x$  为  $S(u)$  中的某个元素。

[0081] 优选地, 筛选模块中对每个用户的预测结果进行筛选时, 通过将当前用户的未评分项目的预测分数与当前用户的历史评分数据的均值进行比较, 若大于均值, 则作为推荐项目。

[0082] 以上实施方式仅用于说明本发明, 而并非对本发明的限制, 有关技术领域的普通技术人员, 在不脱离本发明的精神和范围的情况下, 还可以做出各种变化和变型, 因此所有等同的技术方案也属于本发明的范畴, 本发明的专利保护范围应由权利要求限定。

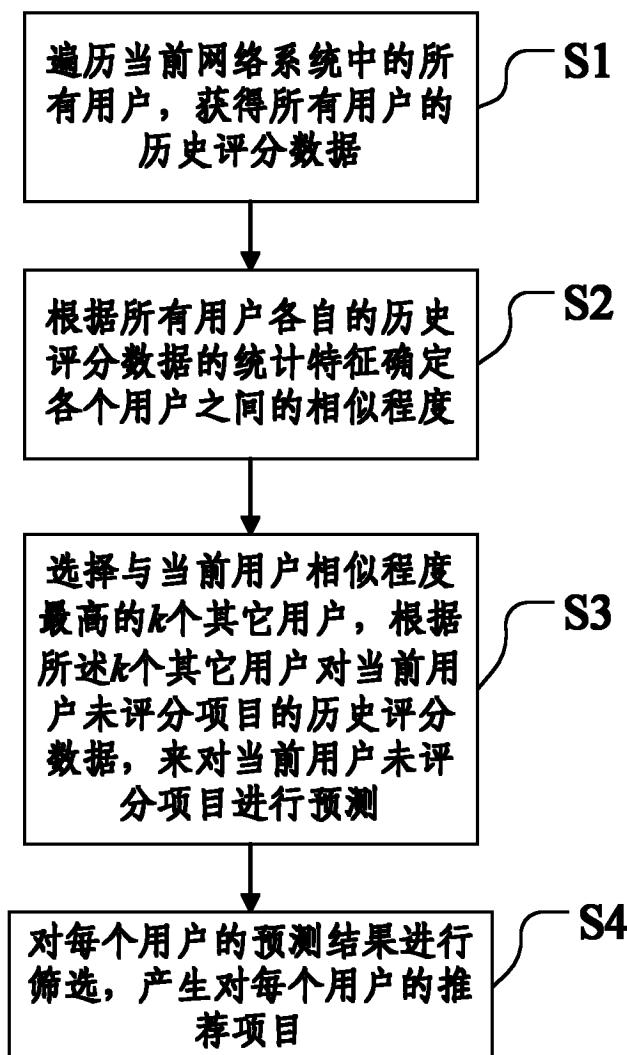


图 1

项目 用户	i <sub>1</sub>	i <sub>2</sub>	i <sub>3</sub>	i <sub>4</sub>	i <sub>5</sub>	i <sub>6</sub>	i <sub>7</sub>	i <sub>8</sub>	i <sub>9</sub>	i <sub>10</sub>
u <sub>1</sub>	4	4	0	0	0	5	0	4	0	5
u <sub>2</sub>	3	4	0	0	0	4	0	3	0	4
u <sub>3</sub>	0	0	4	5	5	0	4	0	3	0
u <sub>4</sub>	1	0	3	2	0	0	3	3	0	0

图 2

统计特征		$\mu$	$S_s$	$R$
用户				
$u_1$		4.4	0.24	1.0
$u_2$		3.6	0.24	1.0
$u_3$		4.2	0.56	2.0
$u_4$		2.8	1.08	4.0

图 3

用户		$u_1$	$u_2$	$u_3$	$u_4$
用户					
$u_1$		1.0	0.998817	0.973678	0.733729
$u_2$		0.998817	1.0	0.983581	0.765883
$u_3$		0.973678	0.983581	1.0	0.869164
$u_4$		0.733729	0.765883	0.869164	1.0

图 4

项目		$i_1$	$i_2$	$i_3$	$i_4$	$i_5$	$i_6$	$i_7$	$i_8$	$i_9$	$i_{10}$
用户											
$u_1$				3.57	3.71	5.0		4.43		3.0	
$u_2$				3.56	3.69	5.0		4.44		3.0	
$u_3$		2.73	4.0				4.5		3.36		4.5
$u_4$			4.0			5.0	4.49			3.0	4.49

图 5

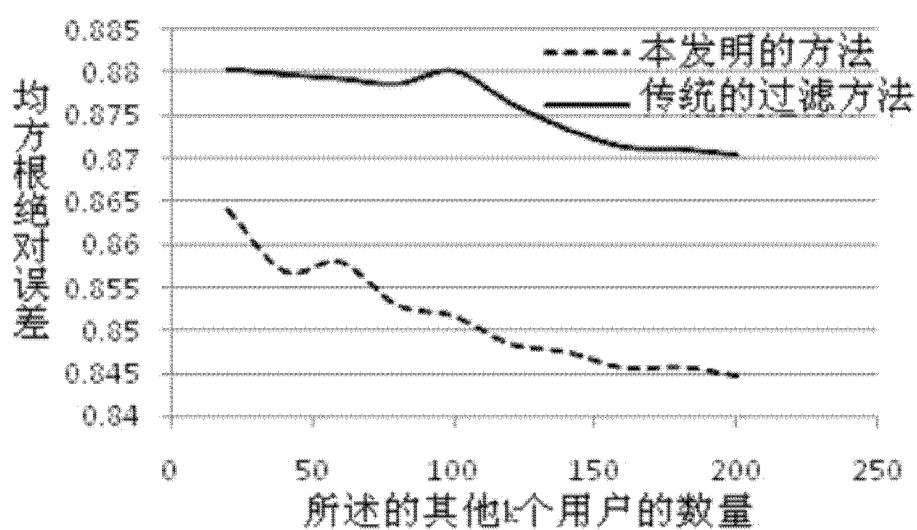


图 6

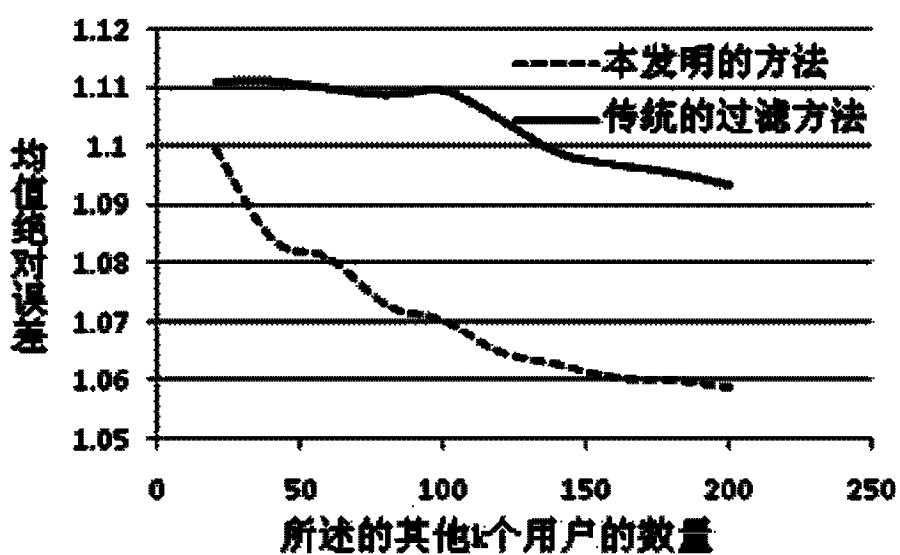


图 7