[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/30 (2006.01)
G06Q 30/00 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200810216517.9

[43] 公开日 2010 年 3 月 31 日

[11] 公开号 CN 101685458A

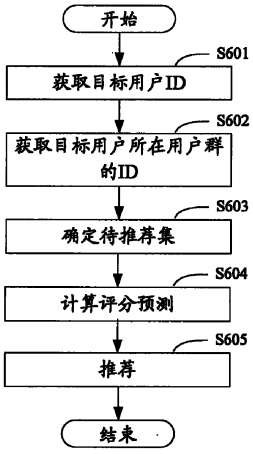
[22] 申请日 2008.9.27
[21] 申请号 200810216517.9
[71] 申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
[72] 发明人 杜家春 汪芳山 方 琦 谭卫国 钟杰萍
权利要求书 5 页 说明书 15 页 附图 6 页

[54] 发明名称

一种基于协同过滤的推荐方法和系统

[57] 摘要

一种基于协同过滤的推荐方法，包括：获取目标用户标识；查找该目标用户标识对应的用户群标识；获取根据所述用户群标识对应的用户－项目评分矩阵确定的项目间相似度；根据该项目间相似度，向目标用户推荐项目。 采用本发明提供的推荐方法能提高项目推荐的准确度，提高用户体验。 本发明同时提供一种基于协同过滤的推荐系统。



1、一种基于协同过滤的推荐方法，其特征在于，包括:

获取目标用户标识；查找所述目标用户标识对应的用户群标识；获取根据所述用户群标识对应的用户-项目评分矩阵确定的项目间相似度；根据所述项目间相似度，向目标用户推荐项目。

2、如权利要求 1 所述的方法，其特征在于，所述方法还包括:

根据用户对项目的评分，建立用户-项目评分矩阵；根据用户-项目评分矩阵进行用户间相似度计算，将用户进行分群；其中，每一用户群对应一用户群标识。

3、如权利要求 2 所述的方法，其特征在于，所述根据用户-项目评分矩阵进行用户间相似度计算采用均值聚类算法（K-means），包括：随机选取 k 个用户作为初始群组中心；对每个用户，计算所述用户与各初始群组中心的距离，将所述用户分到与其距离最近的初始群组中心所在的群组中。

4、如权利要求 3 所述的方法，其特征在于，当所有用户分群完毕后，计算每个群组的新群组中心，所述新群组中心是对应群组中所有用户评分向量的单位向量的均值。

5、如权利要求 2 所述的方法，其特征在于，所述根据用户-项目评分矩阵进行用户间相似度计算，将用户进行分群采用人工分群、机器分群或人机结合分群。

6、如权利要求 1 所述的方法，其特征在于，所述获取根据所述用户群标识对应的用户-项目评分矩阵确定的项目间相似度，包括：获取用户群标识；根据所述用户群标识获取对应用户群中所有用户对应的用户-项目评分矩阵；计算所述用户-项目评分矩阵中项目间相似度。

7、如权利要求 6 所述的方法，其特征在于，所述计算所述用户-项目评分矩阵中项目间相似度采用余弦相似度、皮尔森（Pearson）相关系数或修正的余弦相似度计算。

8、如权利要求 1 所述的方法，其特征在于，若根据所述目标用户标识未查找到对应的用户群标识，包括：获取所述目标用户标识对应目标用户的基本属性；分类器根据所述目标用户基本属性将所述目标用户分到对应用户群，并获得所述用户群对应的用户标识。

9、如权利要求 8 所述的方法，其特征在于，所述分类器的建立方法包括：在所述每个用户群中随机选出占所述用户群用户总数 a%的用户标识；获取所述 a%的用户的基本属性；根据所述 a%的用户基本属性特征构建分类器。

10、如权利要求 1 所述的方法，其特征在于，所述根据所述项目间相似度，向目标用户推荐项目，包括：

判断目标用户在所述用户群对应的用户-项目评分矩阵中是否有评分记录，若有，通过所述项目间相似度，确定与所述评分记录对应项目相似的项目作为待推荐集。

11、如权利要求 10 所述的方法，其特在于，所述根据所述项目间相似度，向目标用户推荐项目，包括：

判断目标用户在所述用户群对应的用户-项目评分矩阵中是否有评分记录，若没有，通过计算所述用户-项目评分矩阵中热点项目的评分预测，将热点项目作为待推荐集，其中，热点项目为被评分最多的前 M 个项目。

12、如权利要求 11 所述的方法，其特征在于，对所述用户-项目评分矩阵中热点项目计算基于热点项目的评分预测，包括：获取用户群标识；根据所述用户群标识获取对应用户群中所有用户对应的用户-项目评分矩阵；计算所述用户群对应用户-项目评分矩阵中热点项目热点度，热点项目热点度为所述项目所得评分的平均值，所述热点项目的热点度即为所述热点项目的评分预测。

13、如权利要求 10 所述的方法，其特征在于，所述方法进一步包括：判断所述待推荐集中项目数目是否不小于 N，若小于，则在距离目标用户

最近的其它用户群中获取待推荐集，与已确定的待推荐集取并集，直到推荐项目数大于或等于 N，或者直到所有用户群遍历完毕为止。

14、如权利要求 13 所述的方法，其特征在于，所述方法进一步包括：判断所述推荐集中项目数目是否不小于 N，若大于或等于，则计算所述推荐集中各项目的评分预测，将评分预测最高的前 N 个项目作为推荐项目向用户推荐。

15、如权利要求 14 所述的方法，其特征在于，计算所述推荐集中各项目的评分预测采用基于相似项目评分预测。

16、一种基于协同过滤的推荐系统，其特征在于，包括：

推荐控制模块，用于获取目标用户标识，调用确定待推荐集模块和生成模块向所述目标用户标识对应的目标用户推荐项目；

确定待推荐集模块，用于查找所述目标用户标识对应的用户群组标识，获取根据所述用户群标识对应的用户-项目评分矩阵确定的项目间相似度，根据所述项目间相似度确定待推荐集，或者获取根据所述用户群标识对应的用户-项目评分矩阵确定的热点项目集，将所述热点项目集作为待推荐集；

生成推荐模块，用于向用户推荐推荐集中的项目。

17、如权利要求 16 所述的系统，其特征在于，包括：数据库，所述数据库中进一步包括：用户-项目评分矩阵，用于存储每一用户对各个项目的用户-项目评分矩阵。

18、如权利要求 17 所述的系统，其特征在于，包括：用户分群模块用于根据所述数据库中所述用户-项目评分矩阵库中存储的用户-项目评分矩阵对用户进行用户分群，每个用户群对应一用户群标识和群组中心，用户分群结果存储于所述数据库中的用户群库中。

19、如权利要求 17 所述的系统，其特征在于，所述数据库中进一步包括：用户基本信息库，用于存储每一用户的基本信息。

20、如权利要求 19 所述的系统，其特征在于，包括：分类器生成模块用于根据所述用户分群结果，将每一用户群中对应用户的基本信息作为分类特征，构建一个分类器。

21、如权利要求 18 所述的系统，其特征在于，包括：热点项目热点度计算模块，用于根据所述用户分群结果和与所述用户群对应的用户-项目评分矩阵，在每一个用户群中独立找出评分最多的若干项目作为热点项目，计算所述热点项目的评分均值得到热点项目的热点度。

22、如权利要求 21 所述的系统，其特征在于，所述数据库中进一步包括：用户群项目热点度库，用于存储所述用户群组对应的热点项目的热点度。

23、如权利要求 21 所述的系统，其特征在于，包括：项目相似度计算模块，用于根据所述用户分群结果和与所述用户群组对应的用户-项目评分矩阵，在每一个用户群组中独立计算项目间相似度。

24、如权利要求 23 所述的系统，其特征在于，所述数据库中进一步包括：用户群组项目相似度库，用于存储所述用户群组对应的所述项目间相似度。

25、如权利要求 24 所述的系统，其特征在于，所述确定待推荐集模块包括：

用户所属群组确定模块，用于在用户群库中根据所述目标用户标识确定对应得用户群标识；

待推荐项目集确定模块，用于根据所述用户群标识在用户项目相似度库中获取项目间相似度，根据所述项目间相似度确定待推荐集，或者获取根据所述用户群标识对应的用户-项目评分矩阵确定的热点项目集，将所述热点项目集作为待推荐集。

26、如权利要求 16 所述的系统，其特征在于，包括：评分预测模块，用于对所述待推荐集中各项目进行基于相似项目评分的预测或基于热点

项目评分的预测，得出目标用户对于待推荐集中各项目的预测评分，将评分最高的N个项目向用户推荐。

27、如权利要求 26 所述的系统，其特征在于，所述评分预测模块包括：相似项目评分预测模块，用于对所述待推荐集中各项目进行基于相似项目评分的预测；或者，热点项目评分预测模块，用于对所述待推荐集中各项目进行基于热点项目评分的预测。

一种基于协同过滤的推荐方法和系统

技术领域

本发明涉及网络通讯技术领域，尤其涉及一种基于协同过滤的推荐方法和系统。

背景技术

推荐系统是为解决信息过载问题而提出的一种智能代理系统，能从大量信息中向用户自动推荐出符合其兴趣偏好或需求的资源。随着互联网的普及和飞速发展，推荐系统已经被广泛应用于各种领域，尤其在电子商务领域，推荐系统得到了越来越多的研究和应用。目前，几乎所有的大型电子商务网站都不同程度的使用了各种形式的推荐系统，比如Amazon、CDNOW、eBay和当当网上书店等。其中，协同过滤技术在当前推荐系统的应用中获得了较大的成功。

协同过滤算法主要有基于用户的协同过滤算法和基于项目的协同过滤算法。两种算法的输入都是用户对项目的评分矩阵，如表1所示：

表1 用户对项目的评分矩阵

项目 / 用户	项目1	项目2	项目3	项目4	项目5	项目6	项目7	项目8
用户1	5	3	4					
用户2	4	2		5				
用户3		3	5	3				
用户4	4		5	4				
用户5				5		3	5	2
用户6			3			4	5	
用户7			2	4			4	5
用户8			3	5	4	5	4	3
用户9			5	4			5	

其中，用户对项目的评分可以显式获得，例如：通过用户对项目进行评分操作；也可隐式获得，例如：通过用户对项目的搜索、浏览、购买等

行为构造评分函数计算得到。矩阵的每一行形成的向量表示该行对应用户的对各个项目的评分向量。

基于用户的协同过滤算法的基本原理是利用用户对项目评分的相似性来互相推荐用户可能感兴趣的项目。例如：对当前用户U，系统通过其评分记录及特定相似度函数，计算出与其评分行为最相近的k个用户作为用户U的最近邻居集，统计用户U的近邻用户评分过而用户U未评分的项目生成候选推荐集，然后计算用户U对候选推荐集中每个项目i的预测评分，取其中预测评分最高的N个项目作为用户U的Top-N推荐集。

基于项目的协同过滤算法则比较项目之间的相似性，根据当前用户已评分的项目集合推荐未评分的项目。由于项目之间的相似性比用户相似性稳定，因此可以离线进行计算存储并定期更新，所以基于项目的协同过滤算法相对于基于用户的协同过滤算法，推荐精度高，实时性好，对基于项目的协同过滤算法进行优化推荐准确度更高、效果更佳、更符合客户需求。

基于项目的协同推荐的基本处理流程，分为线下相似度计算和线上推荐两个部分。图1所示为基于项目的协同推荐方法中线下相似度计算流程，图2所示为基于项目的协同推荐方法中线上推荐流程。

图1中线下相似度计算流程用于计算并保存项目间的相似度。其中，步骤1：获取每一用户对每一项目的评分矩阵；步骤2：计算各个项目间相似度，可采用相似度函数为余弦相似度、皮尔森相似度（Pearson）等；步骤3、存储各个不同项目间相似度。

在预先计算存储了各个不同项目间相似度的基础上，如图2所示线上推荐流程如下：步骤11：获取待推荐的用户标识（ID），即目标用户标识（ID）；步骤12：获取目标用户ID对应的目标用户已经评分的项目集合；步骤13：根据预先存储的项目相似度数据，获取与目标用户已经评分的项目集合中各项目相似度高的项目，形成该目标用户的待推荐项目集；步骤14：根据项目间相似度，进一步计算目标用户对待推荐项目集中每个项目

的预测评分，例如：根据如下公式计算预测评分：$P_{U,i}=\frac{\sum sim(j,i)\cdot R_{U,j}}{\sum sim(j,i)}$，其中，$P_{U,i}$表示目标用户U对待项目i的预测评分，$sim(j,i)$表示项目j和项目i之间的相似度，$R_{U,j}$表示用户$U$对项目$j$的实际评分；步骤15：根据预测评分结果取评分最高的前N项作为对目标用户的推荐结果。

在基于项目的协同过滤算法流程中，项目间的相似度对最终的推荐结果有着至关重要的影响。在传统的基于项目的协同过滤推荐算法中，项目之间相似度的计算并未考虑到不同偏好用户群之间的差异。项目间相似度基于用户评分矩阵计算得到，对所有的用户而言，同样两个项目，它们之间的相似度是相同的。而现实中，对同样两个项目的看法，不同偏好的用户群观点通常不同。这势必造成推荐准确度低，质量下降。

发明内容

为了提高推荐的准确性，符合用户偏好，本发明实施例提供一种基于协同过滤的推荐方法和系统。

一种基于协同过滤的推荐方法，包括：获取目标用户标识；查找所述目标用户标识对应的用户群标识；获取根据所述用户群标识对应的用户-项目评分矩阵确定的项目间相似度；根据所述项目间相似度，向目标用户推荐项目。

一种基于协同过滤的推荐系统，包括：推荐控制模块，用于获取目标用户标识，调用确定待推荐集模块和生成模块向所述目标用户标识对应的目标用户推荐项目；确定待推荐集模块，用于查找所述目标用户标识对应的用户群组标识，获取根据所述用户群标识对应的用户-项目评分矩阵确定的项目间相似度，根据所述项目间相似度确定待推荐集，或者获取根据所述用户群标识对应的用户-项目评分矩阵确定的热点项目集，将所述热点项目集作为待推荐集；生成推荐模块，用于向用户推荐推荐集中的项目。

采用本发明实施例提供的基于协同过滤的推荐方法和系统，通过将用户

分群，使得用户群中的每个用户偏好基本相同，利用这样的用户群所包含的项目相似度信息为用户推荐，提高了推荐的准确性，体现了个性化。

附图说明

图 1 为现有技术基于项目的协同推荐方法中线下相似度计算流程;

图 2 为现有技术基于项目的协同推荐方法中线上推荐流程;

图 3 为本发明实施例一提供的一种基于协同过滤的推荐系统结构示意图;

图 4 为本发明实施例二提供的一种基于协同过滤的推荐方法流程中用户分群流程示意图;

图 5 为本发明实施例二提供的一种基于协同过滤的推荐方法流程中计算项目间相似度流程示意图;

图 6 为本发明实施例二提供的一种基于协同过滤的推荐方法流程中计算项目热点度流程示意图;

图 7 为本发明实施例二提供的一种基于协同过滤的推荐方法流程中建立分类器流程示意图;

图 8 为本发明实施例二提供的一种基于协同过滤的推荐方法流程中线上推荐流程示意图;

图9为为本发明实施例三提供的一种基于协同过滤的推荐方法流程示意图。

具体实施方式

下面通过附图和实施例，对本发明的技术方案做进一步的详细描述。

本发明实施例中提出了一种首先将用户基于用户-项目评分矩阵分群，每一个用户群仅包含该群中用户对所有项目的评分数据，然后在每一个用户群上独立计算项目间相似度，最后以目标用户所在群中计算得到的相似度作为依据对目标用户进行推荐。

如图3所示为本发明实施例一提供的一种基于协同过滤的推荐系统结构示意图。该推荐系统包括：推荐控制模块51、生成推荐模块52，确定待推荐集模块54、数据库55、评分预测模块53、以及定时器56、用户分群模块57、分类器生成模块58、项目热点度计算模块59和项目相似度计算模块60。其中、评分预测模块53中还包括相似项目评分预测模块531、热点项目评分预测模块532；确定待推荐集模块54中还包括用户所属群组确定模块541、待推荐项目集确定模块542；数据库55中还包括用户基本信息库551、用户群库552、用户群项目热点度库553、用户项目评分矩阵库555和用户群项目相似度库554。运算过程中出现并进行了五个部分数据的存储和提取，其中包括系统基础数据集和系统运算数据集。

系统基础数据集主要包括：用户-项目评分矩阵数据，具体为每一用户在业务使用过程中产生的对不同项目的评分数据；用户基本信息数据，具体为描述了用户本身的基本属性信息，包括地域、职业、性别、年龄、教育程度等。

系统运算数据集主要包括：用户群数据，包含用户基于用户-项目评分矩阵数据分群的结果，每一个用户对应一个群组，每一个群组对应一个群组中心；用户群项目热点度数据库，用于记录基于用户分群结果生成的每个用户群对应的热点项目以及热点度，其中，热点项目为被评分最多的前M(M不小于N)个项目，热点项目热点度为所述项目所得评分的平均值；用户群项目相似度数据库，用于记录基于用户分群结果生成的每个用户群对应的项目之间相似度的情况。

如下详细介绍该推荐系统中每一模块的功能及模块间的交互。该推荐系统中各个模块并非全部必要，可以根据功能或性能的强弱需要，相应增减部分模块。

推荐控制模块51为在线推荐部分的主控模块，在接收到待推荐用户ID(即目标用户ID)之后，具有调用其他各模块能力，完成整个推荐处理

流程。

确定待推荐集模块 54 用于根据待推荐用户 ID 确定对应目标用户之后，通过定位目标用户所属用户群，找到目标用户评分项目的邻居项目的集合，或者所述用户群对应的热点项目集，得到待推荐集，将此集合作为下一步评分预测模块 53 的运算基础。确定待推荐集模块可进一步细分为用户所属群组确定模块 541、待推荐项目集确定模块 542。其中，用户所属群组确定模块 541 用于确定用户所属的用户群，可以根据目标用户 ID 定位目标用户所属用户群，或者根据分类器确定目标用户所属用户群；待推荐项目集确定模块 542 用于在目标用户所属群中确定待推荐项目集合，可以通过目标用户评分项目的邻居项目的集合，或者所述用户群对应的热点项目集，得到待推荐集。如果待推荐集合中项目个数小于 N，则计算目标用户与其他群组的距离，在距离最近的群组中继续上述确定待推荐集的过程，直到推荐项目数大于或等于 N，或者直到所有用户群遍历完毕为止。

评分预测模块 53，主要用于在确定待推荐集模块 54 得到的待推荐项目集合中进行基于相似项目评分的预测或基于热点项目评分的预测，得出目标用户对于待推荐项目的预测评分。此模块可进一步细分为相似项目评分预测模块 531、热点项目评分预测模块 532。其中，相似项目评分预测模块 531 根据相似项目间的相似度计算预测评分，例如：根据如下公式计算预测评分：$P_{U,i}=\frac{\sum sim(j,i)\cdot R_{U,j}}{\sum sim(j,i)}$，其中，$P_{U,i}$表示目标用户 U 对待项目 i 的预测评分，$sim(j,i)$表示项目 j 和项目 i 之间的相似度，$R_{U,j}$表示用户 U 对项目 j 的实际评分；热点项目评分预测模块 532 用于计算基于热点项目的预测评分，例如：计算热点项目的热点度作为热点项目的预测评分。在本发明的其他实施例中也可不需要进行待推荐项目集合的进一步预测评分而直接推荐给用户。

生成推荐模块 52，主要用于根据评分预测模块 53 对待推荐项目集合

中各项目的预测评分，将评分最高的前N个项目作为对目标用户的推荐结果。

用户分群模块 57，用于根据数据库 55 中用户-项目评分矩阵库 555 中存储的全部用户的用户-项目评分矩阵进行用户分群，得到全体用户的分群结果，以及每个群组的群组中心，存储在数据库 55 的用户群库 552 中。

分类器生成模块 58，用于根据用户分群结果，以数据库 55 中用户基本信息库 551 中每一用户群中各个用户基本信息为分类特征，构建一个分类器并存储。在本发明的其他实施例中，分类训练集也可以是根据已有用户数量大小取一个合适的百分比，以此百分比为要求在每一个用户群中随机选出若干用户基本信息作为分类训练集数据。

项目热点度计算模块 59，用于根据用户分群结果和用户-项目评分矩阵，在每一个用户群中独立找出评分最多的若干项目，即热点项目，计算所得评分均值，即热点度，并存储在数据库 55 的用户群项目热点度库 553 中。

项目相似度计算模块 60，用于根据用户分群结果和用户-项目评分矩阵，在每一个用户群中独立计算项目间相似度并存储在数据库 55 的用户群项目相似度库 554 中。

在本发明其他实施例中，待推荐项目集确定模块 542 可以同时使用项目热点度计算模块 59 及项目相似度计算模块 60 中存储数据确定针对目标用户所在用户群确定待推荐项目集，也可以根据需要采用二者其中任一模块中存储的数据确定针对目标用户所在用户群确定待推荐项目集。

定时器 56，用于定时触发用户分群模块 57、分类器生成模块 58、项目热点度计算模块、项目相似度计算模块 60 对基础数据集进行处理，包括更新后的基础数据集。在本发明的其他实施例中该模块为可选模块。

根据上述对推荐系统的描述可知，推荐系统在执行具体操作时可以分

为线下和线上两部分组成。其中，线下部分由定时器 56 定时触发用户分群模块 57、分类器生成模块 58、项目热点度计算模块 59 及项目相似度计算模块 60，也可通过手动触发，主要为线上部分的运算提供数据，减轻线上计算量，提高推荐速率，以达到实时推荐目的。所需数据存储于数据库 55 中。线上部分主要完成的是对目标用户的在线推荐工作。获得目标用户所在群组、待推荐项目集合和对待推荐项目的评分预测是线上部分的重要过程，其主要任务是在推荐前为目标用户寻找与其兴趣度最类似的项目集合并预测其评分。

图 4 所示为本发明实施例二提供的一种基于协同过滤的推荐方法流程中用户分群流程详细示意图。步骤 S101，获取每一用户对各个项目的评分；步骤 S102，根据用户项目评分，建立用户-项目评分矩阵，如表 2 所示；

表2 用户-项目评分矩阵

项目 / 用户	项目1	项目2	项目3	项目4	项目5	项目6	项目7	项目8
用户1	5	3	4					
用户2	4	2		5				
用户3		3	5	3				
用户4	4		5	4				
用户5				5		3	5	2
用户6			3			4	5	
用户7			2	4			4	5
用户8			3	5	4	5	4	3
用户9			5	4			5	

步骤 S103，对用户分群，得到若干用户群和每个用户群的群组中心。本实施例中提供一种基于用户间相似度的均值聚类算法（k-means）对所有用户进行分群。在本发明的其他实施例中可采用多种分群的方法，如人工分群、机器分群、人机结合等。

其中，基于用户间相似度的 k-means 聚类算法对所有用户进行分群，

包括：（1）定义类别个数k和误差精度e，随机选取k个用户$M_1,M_2,\cdots,M_k$作为初始群组中心，分别对应类别$C_1,C_2,\cdots,C_k$；（2）对每个用户U，计算所述用户与各初始群组中心的距离$d(U,M_i)=1-sim(U,M_i),i=1,2,\cdots,k$，$sim(U,M_i)$指用户$U$与群组中心$M_i$的相似度。将所述用户分到与其距离最近的初始群组中心所在的群组中，并计算分散度$E(t)=\sum_{i=1}^{k}\sum_{U}\delta_{Ui}d(U,M_i),\delta_{Ui}=\begin{cases}1,U\in C_i\\0,U\notin C_i\end{cases}$，$t$指迭代次数；（3）计算新的聚类中心$M_i'=\sum_{U\in C_i}\frac{U}{\|U\|\cdot\|C_i\|}$，其中$\|U\|$指用户$U$的评分向量的模长，$\|C_i\|$指类别$C_i$中用户的总数；（4）重复（2）、（3）直到$|E(t+1)-E(t)|<e$终止。为每一个群赋予一个用户群标识（ID），同时记录下每一个用户群最终的群组中心。在本实施例中，以将所有用户划分为两个用户群为例进行说明。如表3所示为用户群列表。

表3 用户群列表

用户群ID \ 用户	用户ID
1	4
1	2
1	1
1	3
2	9
2	6
2	5
2	8
2	7

上述用户群1和用户群2对应的群组中心，如表4所示。

表4 用户群对应的群组中心

用户群 \ 项目	项目1	项目2	项目3	项目4	项目5	项目6	项目7	项目8
用户群1	0.46	0.30	0.46	0.44	0.00	0.00	0.00	0.00
用户群2	0.00	0.00	0.32	0.43	0.08	0.29	0.58	0.24

图 5 所示为本发明实施例二提供的一种基于协同过滤的推荐方法流程中计算项目间相似度流程示意图。步骤 S201，获取唯一标识每一用户群的用户群 ID；步骤 S202，根据该用户群 ID 获取对应用户群中所有用户对应的用户-项目评分矩阵；步骤 S203，计算该用户群对应用户-项目评分矩阵中项目间相似度并保存。在本发明的其他实施例中项目间相似度可采用：余弦相似度、Pearson 相似度、修正的余弦相似度等。在本实施例中，采用余弦相似度，得到每个用户群对应的项目间相似度，如表 5 和表 6 所示。

表5 用户群1对应的项目间相似度

项目 / 项目	项目1	项目2	项目3	项目4	项目5	项目6	项目7	项目8
项目1	1.00	0.65	0.65	0.67	0.00	0.00	0.00	0.00
项目2	0.65	1.00	0.71	0.57	0.00	0.00	0.00	0.00
项目3	0.65	0.71	1.00	0.61	0.00	0.00	0.00	0.00
项目4	0.67	0.57	0.61	1.00	0.00	0.00	0.00	0.00
项目5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
项目6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
项目7	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
项目8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

表6 用户群2对应的项目间相似度

项目 / 项目	项目1	项目2	项目3	项目4	项目5	项目6	项目7	项目8
项目1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
项目2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
项目3	0.00	0.00	1.00	0.69	0.44	0.56	0.85	0.45
项目4	0.00	0.00	0.69	1.00	0.55	0.62	0.86	0.81
项目5	0.00	0.00	0.44	0.55	1.00	0.71	0.39	0.49
项目6	0.00	0.00	0.56	0.62	0.71	1.00	0.75	0.48
项目7	0.00	0.00	0.85	0.86	0.39	0.75	1.00	0.66
项目8	0.00	0.00	0.45	0.81	0.49	0.48	0.66	1.00

步骤 S204，判断所有的用户群是否遍历完毕，若没有遍历完毕，返回

步骤 S201；若遍历完毕，结束本流程。

图 6 所示为本发明实施例二提供的一种基于协同过滤的推荐方法流程中计算项目热点度流程示意图。步骤 S301，获取唯一标识每一用户群的用户群 ID；步骤 S302，根据该用户群 ID 获取对应用户群中每一个用户对应的用户-项目评分矩阵；步骤 S302，计算该用户群对应用户-项目评分矩阵中热点项目热点度，热点项目是指被评分最多的前若干项目，项目热点度是该项目所得评分的平均值。在本实施例中，以每个用户群取 2 个热点项目为例，每个用户群对应的热点项目与项目热点度为，如表 7 和表 8 所示。

表7 用户群1对应的项目热点度

项目	热点度
项目3	3.50
项目1	3.25

表8 用户群2对应的项目热点度

项目	热点度
项目7	4.60
项目4	3.60

步骤 S304，判断所有的用户群是否遍历完毕，若没有遍历完毕，返回步骤 S301；若遍历完毕，结束本流程。

图 7 所示为本发明实施例二提供的一种基于协同过滤的推荐方法流程中建立分类器流程示意图。步骤 S401，在每个用户群中随机选出占该群用户总数 a%的用户 ID；步骤 S402，获取以上用户的基本属性；步骤 S403，分析上述用户基本属性特征构建分类器。在本发明的实施例中，可以采用决策树、神经网络等多种方法构造分类器。

上述图 4、图 5、图 6、图 7 所述流程均可在线下或离线状态完成。基于上述用户分群数据、用户群对应项目相似度数据、用户群对应项目热点度数据、分类器构建。

图 8 为本发明实施例二提供的线上推荐流程示意图。

步骤S501，确定待推荐的用户ID，通常将该用户称为目标用户，即获取目标用户ID;

步骤S502，根据该目标用户ID判断对应目标用户是否在用户群中，如果在，执行步骤S503，则获取该目标用户对应的用户群ID，否则，执行步骤S504，获取目标用户基本属性；步骤S505，利用分类器将目标用户分到对应的某个用户群，获取对应的用户群ID;

步骤S506，判断目标用户是否有项目评分记录，如果有，执行步骤S507，确定相似项目待推荐集，否则，执行步骤S508，计算目标用户对所属用户群的热点项目的评分预测，在本实施例中可以要求热点项目数不小于N;

步骤S509，利用目标用户所在用户群中项目相似度和用户项目评分为依据，选择与用户评分高的项目相似度高的且目标用户未评分的项目作为待推荐集，判断待推荐集中项目数目是否不小于N；如果否，执行步骤S511，计算目标用户与其他用户群的群组中心的距离，在离目标用户最近的其他群中选择待推荐集并与上述步骤的待推荐集作并集处理，直到待推荐集中项目数目不小于N，或者直到所有用户群遍历完毕为止；如果是，执行步骤S510，计算目标用户对待推荐集中每一个项目的评分预测。

步骤S512，将评分预测最高的N个项目作为推荐项目向目标用户推荐。

在本实施例中，步骤S504，步骤S505为了解决当新目标用户不在已有的用户群中，对新用户进行分群后进行推荐的流程，可以预见的在不考虑新目标用户的情况下。步骤S504，步骤S505为可选步骤。步骤S506给出了了当目标用户有评分记录和没有评分记录的两种推荐流程，在本发明的其他实施例中可以采用其中之一。步骤S508和步骤S507也同时给出了两种推荐的算法，可以预见在本发明的其他实施例中可以任意采用其中之一。步

骤S509，S511给出当待推荐集中项目数小于N时，在临近用户群中确定待推荐集的流程，可以预见在本发明的其他实施例中若对推荐集项目数不做限制时为可选步骤。步骤S510在本发明的其他实施例中直接将待推荐集推荐给用户时为可选步骤。步骤S510同样为提高推荐准确度的步骤在本发明的其他实施例为可选步骤。综上所述，本实施例方法流程的上述步骤可以根据推荐准确度的需要进行灵活适当的调整、取舍，均能达到提高推荐准确度的效果。

如图9所示为本发明实施例三结合一具体应用实例说明本发明方法流程。

步骤S601，获取目标用户ID，确定对应的目标用户。

在本发明的实施例中，目标用户由业务调用方提供。业务调用方给出目标用户ID，期望获取该目标用户的推荐项目列表。假设用户7为目标用户，如表9所示为用户-项目评分矩阵。

表9 用户-项目评分矩阵

项目 / 用户	项目1	项目2	项目3	项目4	项目5	项目6	项目7	项目8
用户1	5	3	4					
用户2	4	2		5				
用户3		3	5	3				
用户4	4		5	4				
用户5				5		3	5	2
用户6			3			4	5	
用户7			2	4			4	5
用户8			3	5	4	5	4	3
用户9			5	4			5	

步骤S602，获取目标用户所在用户群的ID。在本实施例中，根据表3可知用户7属于用户群2。如果目标用户是新用户，则需要利用用户基本信息将用户分类以获取该新用户所在用户群的ID。

步骤S603，确定待推荐集。首先取用户7评分高的项目，这里以用户7

评分大于等于4为评分高的标准，例如：评分大于等于4的项目为项目4、项目7、项目8，并且接下来通过查找前述实施例表6得到与项目4、项目7和项目8相似度高（这里的相似度高指所选项目与项目4、项目7和项目8的相似度的均值大于0.5）且用户7没有评分过的项目作为待推荐集，即得待推荐集为项目6和项目3。当待推荐项目集中项目数不小于N，N等于1；此时待推荐集中有两个项目，满足不小于1的条件。

如果待推荐集中项目数小于1，则需要计算目标用户与其他群组中心的距离，挑选最近的用户群，并在此用户群中挑选待推荐集，直到待推荐集中项目总数不小于1,或者直到所有用户群遍历完毕为止。

如果目标用户没有评分记录，则计算目标用户对所属群组的热点项目的评分预测。该评分预测可查阅前述实施例表7、表8的结果。

步骤S604，计算评分预测。

利用公式 $P_{U,i}=\frac{\sum sim(j,i)\cdot R_{U,j}}{\sum sim(j,i)}$ 计算，$P_{U,i}$表示目标用户U对待项目i的预测评分，$sim(j,i)$表示项目j和项目i之间的相似度，$R_{U,j}$表示用户U对项目j的实际评分。根据如上公式，用户7对待推荐项目的评分预测，如表10所示。

表10 用户7对待推荐项目的评分预测

项目	评分预测
项目3	3.79
项目6	3.73

步骤S604，将满足上述条件的项目推荐给用户。根据表10，最终将项目3推荐给用户7。

本发明实施例提供一种基于协同过滤推荐的方法和系统。该方法在线下处理的过程中，首先利用用户项目评分数据将用户分群，然后在每个用户群中独立计算项目间相似度，并且可以由分群结果建立一个分类器，使

得对新用户亦能进行较好的分类。线上推荐时，需要获取目标用户所属的群组，利用该群组相关的项目间相似度对目标用户进行基于项目的协同过滤推荐，或者利用该群组相关的热点项目的热点度为目标用户进行推荐。相比于传统的协同推荐流程，本发明先将用户分群，使得每个用户群的用户偏好基本相似，利用这样的用户群所包含的项目相似度信息为用户推荐，提高了推荐的准确性，体现了个性化。同时，分群后计算相似度也提高了线下处理的计算速度。

显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

此外，本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

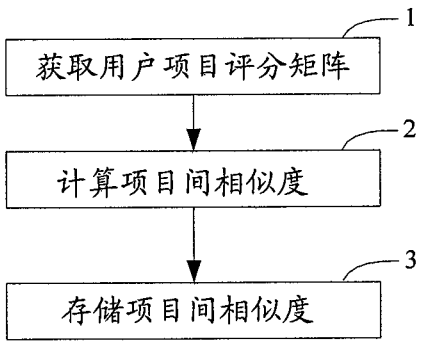

1
获取用户项目评分矩阵
2
计算项目间相似度
3
存储项目间相似度

图 1

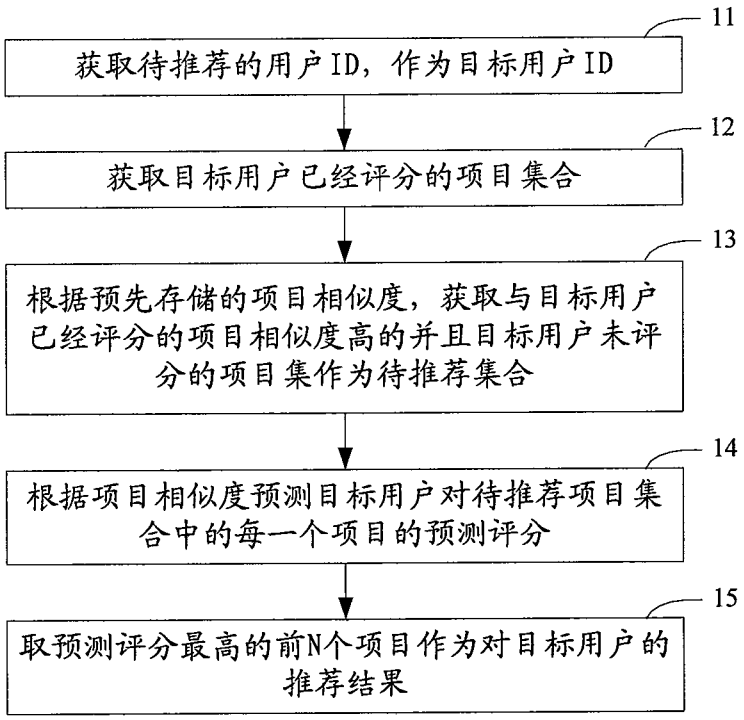

11
获取待推荐的用户ID，作为目标用户ID
12
获取目标用户已经评分的项目集合
13
根据预先存储的项目相似度，获取与目标用户已经评分的项目相似度高的并且目标用户未评分的项目集作为待推荐集合
14
根据项目相似度预测目标用户对待推荐项目集合中的每一个项目的预测评分
15
取预测评分最高的前N个项目作为对目标用户的推荐结果

图 2

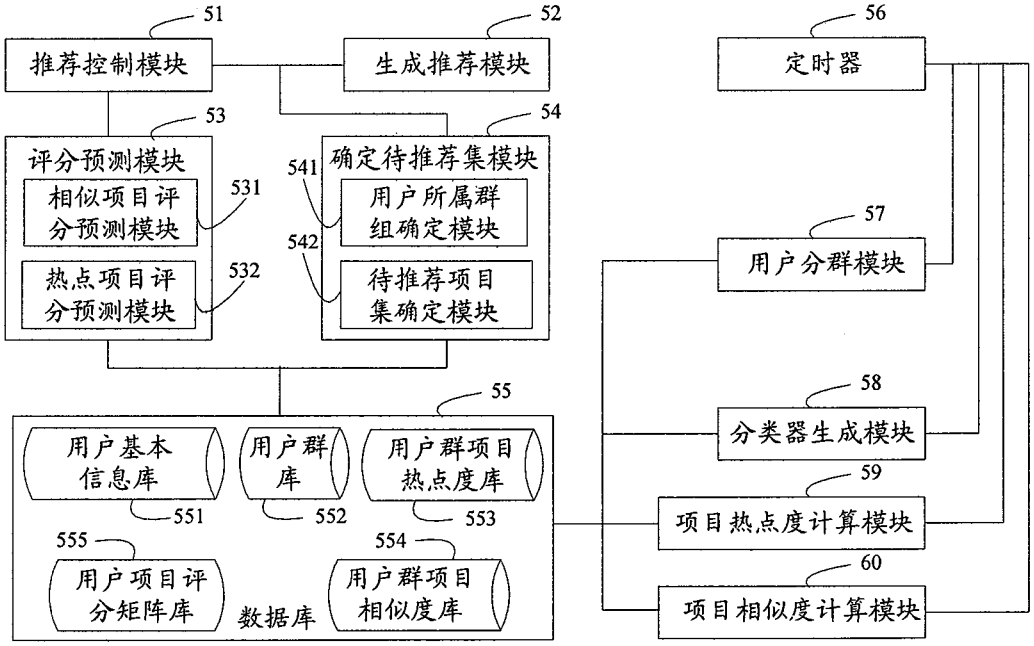

51
推荐控制模块
52
生成推荐模块
56
定时器
53
评分预测模块
相似项目评分预测模块
531
热点项目评分预测模块
532
54
确定待推荐集模块
541
用户所属群组确定模块
542
待推荐项目集确定模块
57
用户分群模块
58
分类器生成模块
59
项目热点度计算模块
60
项目相似度计算模块
55
用户基本信息库
551
用户群库
552
用户群项目热点度库
553
555
用户项目评分矩阵库
554
用户群项目相似度库
数据库

图 3

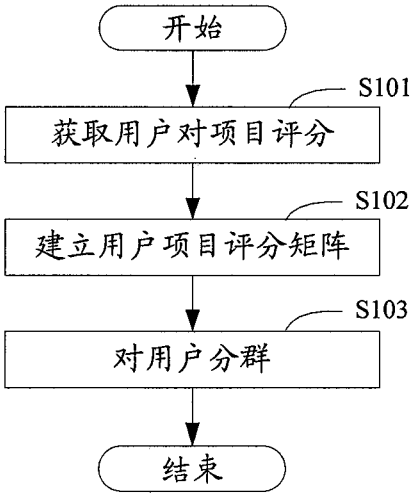

开始
S101
获取用户对项目评分
S102
建立用户项目评分矩阵
S103
对用户分群
结束

图 4

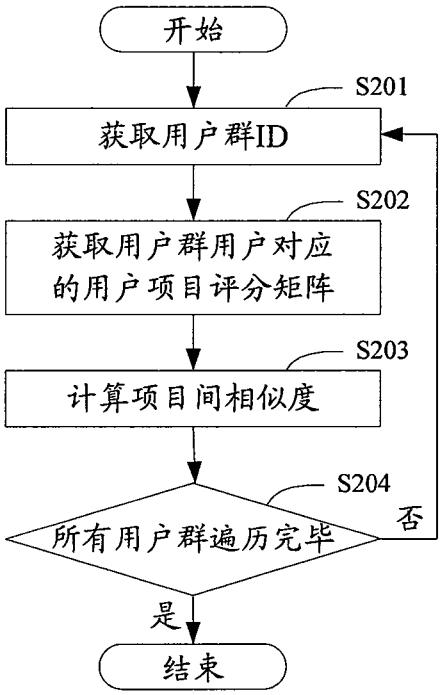

开始
S201
获取用户群ID
S202
获取用户群用户对应
的用户项目评分矩阵
S203
计算项目间相似度
S204
所有用户群遍历完毕
否
是
结束

图 5

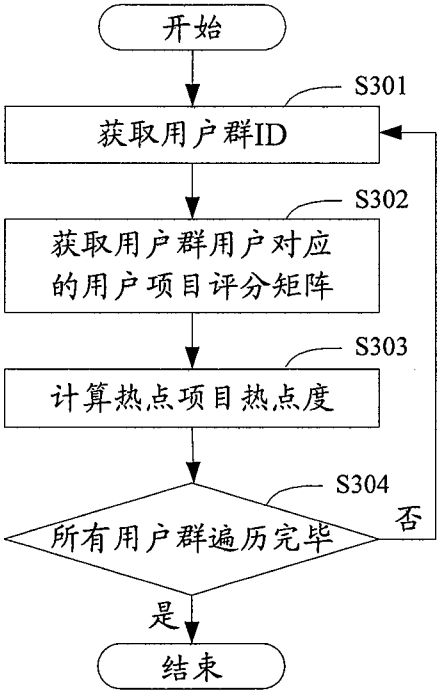

开始
S301
获取用户群ID
S302
获取用户群用户对应的用户项目评分矩阵
S303
计算热点项目热点度
S304
所有用户群遍历完毕
否
是
结束

图 6

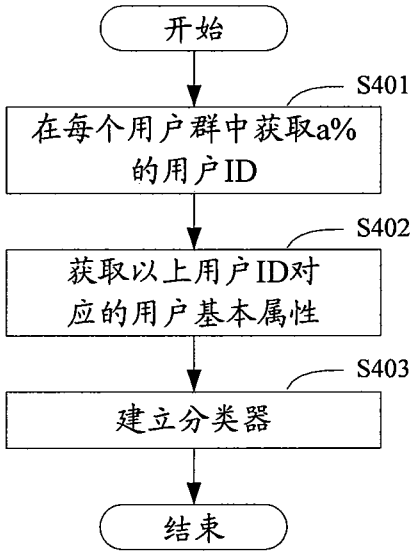

开始
S401
在每个用户群中获取a%的用户ID
S402
获取以上用户ID对应的用户基本属性
S403
建立分类器
结束

图 7

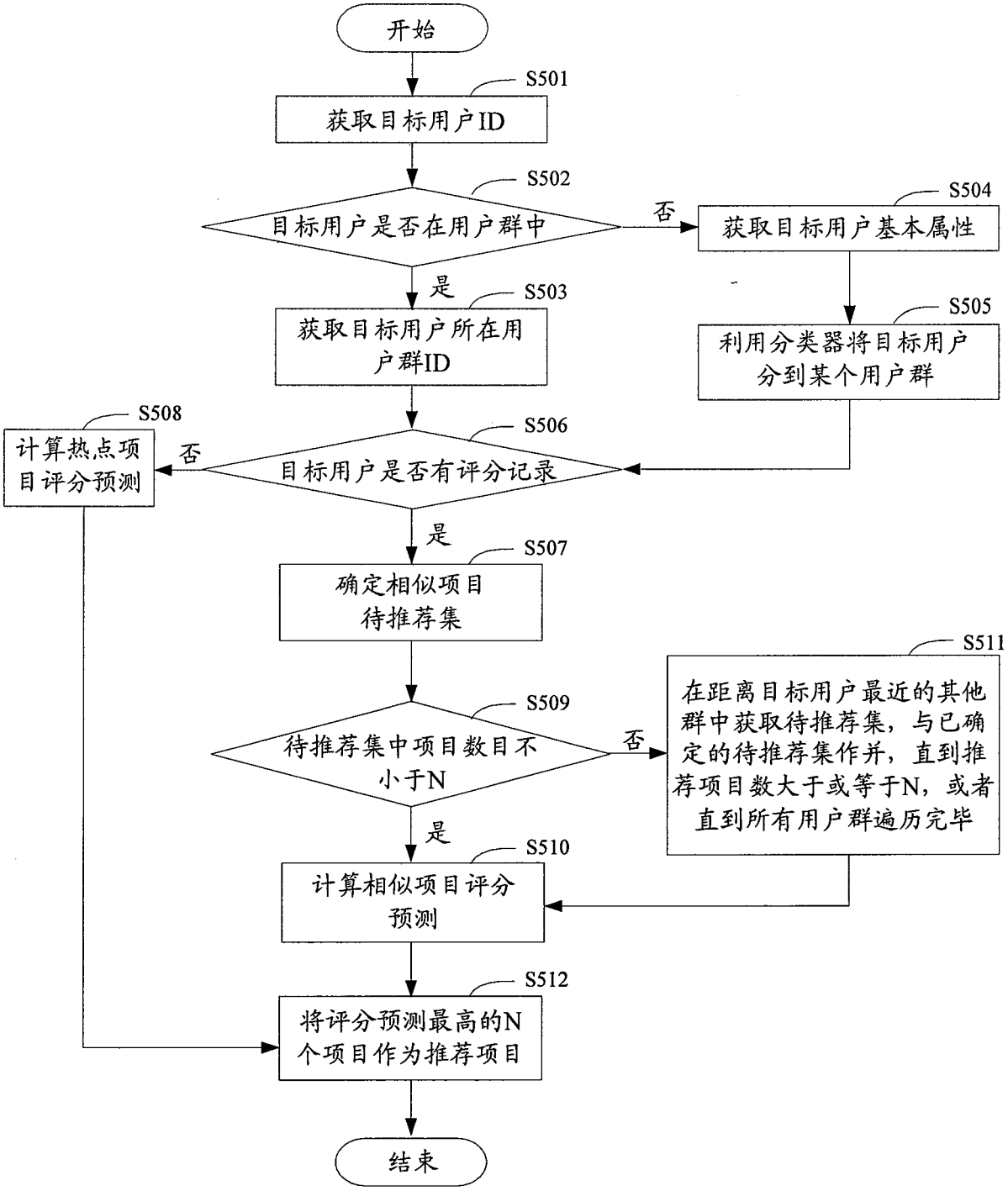

开始
S501
获取目标用户ID
S502
目标用户是否在用户群中
否
S504
获取目标用户基本属性
是
S503
获取目标用户所在用户群ID
S505
利用分类器将目标用户分到某个用户群
S508
计算热点项目评分预测
否
S506
目标用户是否有评分记录
是
S507
确定相似项目待推荐集
S511
在距离目标用户最近的其他群中获取待推荐集，与已确定的待推荐集作并，直到推荐项目数大于或等于N，或者直到所有用户群遍历完毕
S509
待推荐集中项目数目不小于N
否
是
S510
计算相似项目评分预测
S512
将评分预测最高的N个项目作为推荐项目
结束

图 8

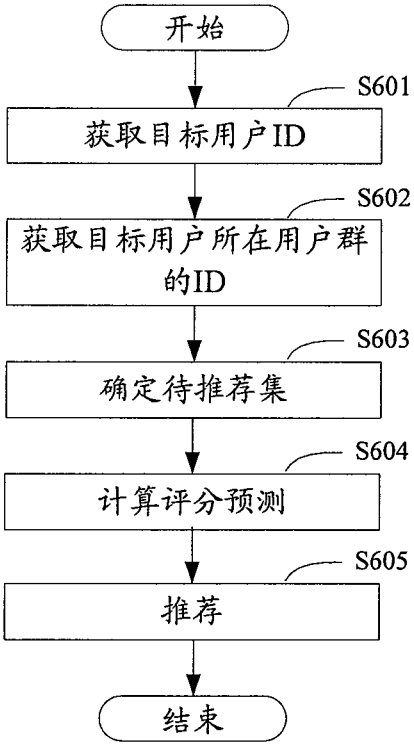

开始
S601
获取目标用户ID
S602
获取目标用户所在用户群的ID
S603
确定待推荐集
S604
计算评分预测
S605
推荐
结束

图 9