(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109063052 B
(45)授权公告日 2022.01.25

(21)申请号 201810794746.2

(22)申请日 2018.07.19

(65)同一申请的已公布的文献号
申请公布号 CN 109063052 A

(43)申请公布日 2018.12.21

(73)专利权人 北京物资学院
地址 101149 北京市通州区富河大街321号

(72)发明人 唐恒亮 薛菲 刘涛 杨玺 董晨刚

(74)专利代理机构 北京卓岚智财知识产权代理事务所(特殊普通合伙) 11624
代理人 任漱晨

(51)Int.Cl.
G06F 16/9535(2019.01)
G06F 16/9537(2019.01)
G06F 16/906(2019.01)
G06K 9/62(2006.01)

(56)对比文件
US 2010185579 A1,2010.07.22

审查员 于淼

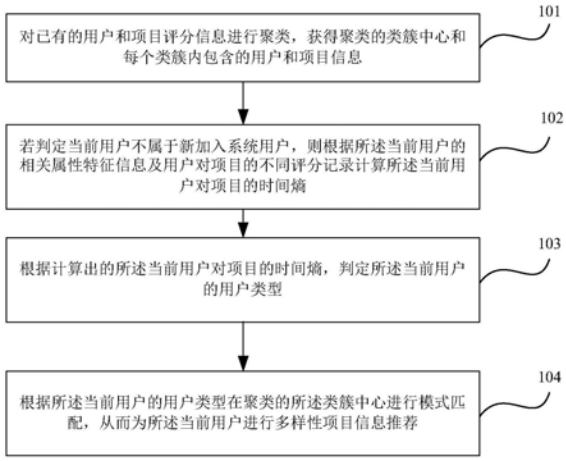
权利要求书3页 说明书10页 附图4页

(54)发明名称

一种基于时间熵的个性化推荐方法及装置

(57)摘要

本发明实施例提供一种基于时间熵的个性化推荐方法及装置,所述方法包括:对已有的用户和项目评分信息进行聚类,获得聚类的类簇中心和每个类簇内包含的用户和项目信息;若判定当前用户不属于新加入系统用户,则根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵;根据计算出的所述当前用户对项目的时间熵,判定所述当前用户的用户类型;根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配,从而为所述当前用户进行多样性项目信息推荐。本发明实施例可以提高信息推荐准确度和推荐多样性。



CN 109063052 B

1.一种基于时间熵的个性化推荐方法,其特征在于,所述方法包括:

对已有的用户和项目评分信息进行聚类,获得聚类的类簇中心和每个类簇内包含的用户和项目信息;

若判定当前用户不属于新加入系统用户,则根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵;

根据计算出的所述当前用户对项目的时间熵,判定所述当前用户的用户类型;

根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配,从而为所述当前用户进行多样性项目信息推荐;

其中,所述根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录利用如下时间熵公式计算所述当前用户对项目的时间熵:

$$H=-\sum_{i=1}^{n}\frac{m_i\cdot score_i}{M}\log\frac{m_i}{M}\ ,$$

其中,H为时间熵,n是时间间隔数,$score_i$为用户对项目i的评分,m_i代表一个时间间隔内所有用户对项目i的评分总和,M指所有m_i的总和;

根据用户兴趣偏好,判定所述当前用户的用户类型为如下四种类型之一:最近喜欢且过去也喜欢、最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢;

所述根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配,从而为所述当前用户进行多样性项目信息推荐,包括:

根据所述当前用户的用户类型在聚类的所述类簇中心的如下两种兴趣偏好模式中进行模式匹配:规律性兴趣偏好模式和非规律性兴趣偏好模式;其中,所述规律性兴趣偏好模式为最近喜欢且过去也喜欢;所述非规律性兴趣偏好模式包括:最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢;

其中,所述根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配,从而为所述当前用户进行多样性项目信息推荐,具体为:当模式匹配为规律性兴趣偏好模式时,计算用户所属聚类,簇类中运用基于项目的协同过滤算法进行推荐;否则,采用项目的多样性公式为用户进行推荐;

其中,所述多样性公式为:

$$dv(i)=\frac{fr(i)}{\alpha\cdot z(i)+1}$$

$$z(i)=\sum_{j\in Set_i}sim(i,j)$$

$$sim(i,j)=\frac{\sum_{u\in U_{ij}}(r_{ui}-\overline{r_i})(r_{uj}-\overline{r_j})}{\sqrt{\sum_{u\in U_{ij}}(r_{ui}-\overline{r_i})^2\sum_{u\in U_{ij}}(r_{uj}-\overline{r_j})^2}}$$

其中,fr(i)指标签在集合中出现的频率,z(i)是标签候选集合中所有标签的相似度之和,α为调节系数,项r_{ui}和r_{uj}分别代表用户u对项目i和项目j的评分,$\overline{r_j}$为所有评价过项目j

的用户对项目j的平均评分，$\overline{r_i}$为所有评价过项目i的用户对项目i的平均评分，U_{ij}为同时对项目i与项目j评分的用户集合，sim(i,j)的值介于[-1,1]之间。

2.如权利要求1所述基于时间熵的个性化推荐方法，其特征在于，所述方法还包括：

若判定当前用户为新加入系统用户，采用如下项目流行度计算方式将排名在前的N个项目信息推荐给所述当前用户：

$$i_{pop}=\frac{|U_i|}{\sqrt{\sum_{i\in I}|U_i|^2}},$$

其中，i_{pop}为项目流行度，U_i代表推荐系统中评价过项目i的用户集合，I为推荐系统中的所有项目数。

3.一种基于时间熵的个性化推荐装置，其特征在于，所述装置包括：

聚类单元，用于对已有的用户和项目评分信息进行聚类，获得聚类的类簇中心和每个类簇内包含的用户和项目信息；

计算单元，用于若判定当前用户不属于新加入系统用户，则根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵；

判断单元，用于根据计算出的所述当前用户对项目的时间熵，判定所述当前用户的用户类型；

匹配单元，用于根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配，从而为所述当前用户进行多样性项目信息推荐；

其中，所述计算单元，具体用于根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录利用如下时间熵公式计算所述当前用户对项目的时间熵：

$$H=-\sum_{i=1}^{n}\frac{m_i\cdot score_i}{M}\log\frac{m_i}{M},$$

其中，H为时间熵，n是时间间隔数，$score_i$为用户对项目i的评分，m_i代表一个时间间隔内所有用户对项目i的评分总和，M指所有m_i的总和；

所述判断单元，具体用于根据用户兴趣偏好，判定所述当前用户的用户类型为如下四种类型之一：最近喜欢且过去也喜欢、最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢；

所述匹配单元，具体用于根据所述当前用户的用户类型在聚类的所述类簇中心的如下两种兴趣偏好模式中进行模式匹配：规律性兴趣偏好模式和非规律性兴趣偏好模式；其中，所述规律性兴趣偏好模式为最近喜欢且过去也喜欢；所述非规律性兴趣偏好模式包括：最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢；

其中，所述匹配单元，具体用于：当模式匹配为规律性兴趣偏好模式时，计算用户所属聚类，簇类中运用基于项目的协同过滤算法进行推荐；否则，采用项目的多样性公式为用户进行推荐；

其中，所述多样性公式为：

$$dv(i)=\frac{fr(i)}{\alpha\cdot z(i)+1}$$

$$z(i)=\sum_{j\in Set_i} sim(i,j)$$

$$sim(i,j)=\frac{\sum_{u\in U_{ij}}(r_{ui}-\overline{r_i})(r_{uj}-\overline{r_j})}{\sqrt{\sum_{u\in U_{ij}}(r_{ui}-\overline{r_i})^2\sum_{u\in U_{ij}}(r_{uj}-\overline{r_j})^2}}$$

其中,fr(i)指标签在集合中出现的频率,z(i)是标签候选集合中所有标签的相似度之和,α为调节系数,项r_{ui}和r_{uj}分别代表用户u对项目i和项目j的评分,$\overline{r_j}$为所有评价过项目j的用户对项目j的平均评分,$\overline{r_i}$为所有评价过项目i的用户对项目i的平均评分,U_{ij}为同时对项目i与项目j评分的用户集合,sim(i,j)的值介于[-1,1]之间。

4.如权利要求3所述基于时间熵的个性化推荐装置,其特征在于,所述装置还包括:

推荐单元,用于若判定当前用户为新加入系统用户,采用如下项目流行度计算方式将排名在前的N个项目信息推荐给所述当前用户:

$$i_{pop}=\frac{|U_i|}{\sqrt{\sum_{i\in I}|U_i|^2}},$$

其中,i_{pop}为项目流行度,U_i代表推荐系统中评价过项目i的用户集合,I为推荐系统中的所有项目数。

一种基于时间熵的个性化推荐方法及装置

技术领域

[0001] 本发明涉及互联网智能信息推荐技术领域，尤其涉及一种基于时间熵的个性化推荐方法及装置。

背景技术

[0002] 随着互联网技术的兴起和信息技术的快速发展，互联网产生了大量的数据信息。由Excelcom公司发布的一份“互联网一分钟产生数据”的图表信息，我们可知Facebook共产生701,389账号登陆、Netflix共有69,444小时长的视频被观看、Snapchat分享了527,760张照片、App Store上51,000个app被下载、Linkedin创建了120多个新账号、Twitter发布了347,222条新推文、Instagram发布了28,194张新照片、Google产生了240万条新搜索请求，使得互联网从原来信息匮乏的时代走向了信息过载(Information overload)，这也使得用户想要从海量信息库中快速并且准确地找到其感兴趣的信息变得愈发困难。

[0003] 面对信息过载问题，普通用户往往无法适从。科学家为了更好地满足用户的信息需求，提出了推荐系统技术，该技术通过将机器学习、数据挖掘、用户行为学和人机交互等多个领域的技术进行结合，并运用大规模并行数据处理框架，进而快速并准确地为每位用户提供个性化信息服务。协同过滤是一种能够产生个性化推荐的有效技术，在各种推荐系统中都得到广泛应用，其基本任务是根据相似的偏好匹配用户，以推荐用户可能会喜欢的项目。协同过滤算法一般可以分为基于内存和基于模型。其中，基于内存的协同过滤又可分为基于用户和基于项目。前者是计算用户间相似度，得到与目标用户兴趣偏好相似的最近邻，以此为基础进行预测推荐。

[0004] 然而，传统的协同过滤推荐算法在为用户推荐项目时，通常只选用基于用户或者基于项目的推荐方法为目标用户进行推荐，这种推荐方式仅仅选用了用户对项目的评分信息，而忽略了用户的兴趣会随着时间的变化受情绪、朋友和时尚潮流等其他影响而发生变化，也就是说，用户的兴趣可能在一定时期内只关注一个或者几个项目，即兴趣迁移。因此，单一类型的推荐并不能满足其他多样性用户的实用性需求。

发明内容

[0005] 本发明实施例提供一种基于时间熵的个性化推荐方法及装置，以提高信息推荐准确度和推荐多样性。

[0006] 一方面，本发明实施例提供了一种基于时间熵的个性化推荐方法，所述方法包括：

[0007] 对已有的用户和项目评分信息进行聚类，获得聚类的类簇中心和每个类簇内包含的用户和项目信息；

[0008] 若判定当前用户不属于新加入系统用户，则根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵；

[0009] 根据计算出的所述当前用户对项目的时间熵，判定所述当前用户的用户类型；

[0010] 根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配，从而为所述

当前用户进行多样性项目信息推荐。

[0011] 另一方面，本发明实施例提供了一种基于时间熵的个性化推荐装置，所述装置包括：

[0012] 聚类单元，用于对已有的用户和项目评分信息进行聚类，获得聚类的类簇中心和每个类簇内包含的用户和项目信息；

[0013] 计算单元，用于若判定当前用户不属于新加入系统用户，则根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵；

[0014] 判断单元，用于根据计算出的所述当前用户对项目的时间熵，判定所述当前用户的用户类型；

[0015] 匹配单元，用于根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配，从而为所述当前用户进行多样性项目信息推荐。

[0016] 上述技术方案具有如下有益效果：利用基于时间熵的个性化推荐方法对用户进行推荐，一方面，对用户和项目信息进行聚类及计算项目的流行度，可以提高推荐的效率、准确度并解决用户的冷启动问题。另一方面，通过计算用户对项目的时间熵可以有效利用用户的多兴趣，从而提高推荐方法的多样性。

附图说明

[0017] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中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他的附图。

[0018] 图1为本发明实施例一种基于时间熵的个性化推荐方法流程图；

[0019] 图2为本发明实施例一种基于时间熵的个性化推荐装置结构示意图；

[0020] 图3为本发明实施例另一种基于时间熵的个性化推荐装置结构示意图；

[0021] 图4为本发明应用实例基于时间熵的个性化推荐方法的整体流程图；

[0022] 图5为本发明应用实例基于时间熵的个性化推荐方法与其它推荐算法推荐效率对比图；

[0023] 图6为本发明应用实例基于时间熵的个性化推荐方法多样性对比图。

具体实施方式

[0024] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。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前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0025] 如图1所示，为本发明实施例一种基于时间熵的个性化推荐方法流程图，所述方法包括：

[0026] 101、对已有的用户和项目评分信息进行聚类，获得聚类的类簇中心和每个类簇内包含的用户和项目信息；

[0027] 102、若判定当前用户不属于新加入系统用户，则根据所述当前用户的相关属性特

征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵；

[0028] 103、根据计算出的所述当前用户对项目的时间熵，判定所述当前用户的用户类型；

[0029] 104、根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配，从而为所述当前用户进行多样性项目信息推荐。

[0030] 优选地，所述方法还包括：

[0031] 若判定当前用户为新加入系统用户，采用如下项目流行度计算方式将排名在前的N个项目信息推荐给所述当前用户：

[0032] $$i_{pop}=\frac{|U_i|}{\sqrt{\sum_{i\in I}|U_i|^2}},$$

[0033] 其中，i_{pop}为项目流行度，U_i代表推荐系统中评价过项目i的用户集合，I为推荐系统中的所有项目数。

[0034] 优选地，根据用户兴趣偏好，判定所述当前用户的用户类型为如下四种类型之一：最近喜欢且过去也喜欢、最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢。

[0035] 优选地，所述根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配，从而为所述当前用户进行多样性项目信息推荐，包括：

[0036] 根据所述当前用户的用户类型在聚类的所述类簇中心的如下两种兴趣偏好模式中进行模式匹配：规律性兴趣偏好模式和非规律性兴趣偏好模式；其中，所述规律性兴趣偏好模式为最近喜欢且过去也喜欢；所述非规律性兴趣偏好模式包括：最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢。

[0037] 优选地，所述根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录利用如下时间熵公式计算所述当前用户对项目的时间熵：

[0038] $$H=-\sum_{i=1}^{n}\frac{m_i\cdot score_i}{M}\log\frac{m_i}{M},$$

[0039] 其中，H为时间熵，n是时间间隔数，$score_i$为用户对项目i的评分，m_i代表一个时间间隔内所有用户对项目i的评分总和，M指所有m_i的总和。

[0040] 对应于上述方法实施例，如图2所示，为本发明实施例一种基于时间熵的个性化推荐装置结构示意图，所述装置包括：

[0041] 聚类单元21，用于对已有的用户和项目评分信息进行聚类，获得聚类的类簇中心和每个类簇内包含的用户和项目信息；

[0042] 计算单元22，用于若判定当前用户不属于新加入系统用户，则根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵；

[0043] 判断单元23，用于根据计算出的所述当前用户对项目的时间熵，判定所述当前用户的用户类型；

[0044] 匹配单元24，用于根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配，从而为所述当前用户进行多样性项目信息推荐。

[0045] 优选地，如图3所示，为本发明实施例另一种基于时间熵的个性化推荐装置结构示

意图，所述装置不但包括：聚类单元21、计算单元22、判断单元23、匹配单元24，所述装置还包括：

[0046] 推荐单元25，用于若判定当前用户为新加入系统用户，采用如下项目流行度计算方式将排名在前的N个项目信息推荐给所述当前用户：

[0047] $$i_{pop}=\frac{|U_i|}{\sqrt{\sum_{i\in I}|U_i|^2}},$$

[0048] 其中，i_{pop}为项目流行度，U_i代表推荐系统中评价过项目i的用户集合，I为推荐系统中的所有项目数。

[0049] 优选地，所述判断单元23，具体用于根据用户兴趣偏好，判定所述当前用户的用户类型为如下四种类型之一：最近喜欢且过去也喜欢、最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢。

[0050] 优选地，所述匹配单元24，具体用于根据所述当前用户的用户类型在聚类的所述类簇中心的如下两种兴趣偏好模式中进行模式匹配：规律性兴趣偏好模式和非规律性兴趣偏好模式；其中，所述规律性兴趣偏好模式为最近喜欢且过去也喜欢；所述非规律性兴趣偏好模式包括：最近喜欢但过去不喜欢、最近不喜欢但过去喜欢、最近不喜欢且过去也不喜欢。

[0051] 优选地，所述计算单元22，具体用于根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录利用如下时间熵公式计算所述当前用户对项目的时间熵：

[0052] $$H=-\sum_{i=1}^{n}\frac{m_i\cdot score_i}{M}\log\frac{m_i}{M},$$

[0053] 其中，H为时间熵，n是时间间隔数，$score_i$为用户对项目i的评分，m_i代表一个时间间隔内所有用户对项目i的评分总和，M指所有m_i的总和。

[0054] 本发明实施例上述技术方案具有如下有益效果：利用基于时间熵的个性化推荐方法对用户进行推荐，一方面，对用户和项目信息进行聚类及计算项目的流行度，可以提高推荐的效率、准确度并解决用户的冷启动问题。另一方面，通过计算用户对项目的时间熵可以有效利用用户的多兴趣，从而提高推荐方法的多样性。

[0055] 本发明以上实施例提出了一种基于时间熵的个性化推荐方法(Personalized Time Collaborative Filtering，PTCF)为用户进行推荐。该方法首先对推荐系统中已有的用户和项目评分信息进行聚类，从而获得相应的聚类中心和不同的类簇信息；然后判定目标用户是否属于新加入系统用户，如果不是则根据用户的相关属性特征信息及用户对项目的不同评分记录计算用户对项目的时间熵，并判定用户属于四种用户类型中的哪一类用户，进而对推荐系统中的目标用户进行模式匹配，从而采用项目多样性计算公式为用户进行多样性推荐；否则采用项目流行度计算方式将排名在前的N个项目推荐给目标用户。

[0056] 上述这种基于时间熵的个性化推荐方法PTCF的主要推荐机制在于：在推荐算法的选择上，不仅仅选择了基于聚类的协同过滤算法进行推荐，同时考虑到用户的兴趣会随着各种因素的变化而发生变化进而对用户进行多样性推荐。这种推荐方式不仅弥补了推荐系统在原来计算全部评分信息进行推荐的推荐效率问题，同时利用用户对项目评分的多样化信息，从而有效地提高了该推荐方法的推荐准确度和推荐多样性。

[0057] 本发明实施例的主要内容是基于时间熵的个性化推荐方法PTCF的研究及应用，主要包括原始用户和项目评分信息进行聚类、怎样计算目标用户对不同项目评分的时间熵、目标用户如何进行分类和模式匹配以及针对目标用户是否为新加入系统用户进行相应的推荐。其采用的技术方案为：1)通过运用RLPSO_KM聚类算法对用户和项目评分信息进行聚类，从而获得聚类的类簇中心和相应的类簇信息；2)通过引入时间熵的定义，从而计算已加入系统的目标用户对不同项目的时间熵，进而为目标用户进行分类和模式匹配，从而为其进行有效性推荐；3)通过引入计算项目流行度方式，从而为新加入系统的目标用户进行推荐。

[0058] 本发明实施例采用的技术方案为一种基于时间熵的个性化推荐方法，该方法的实现步骤如下：

[0059] (1)对用户和项目评分信息聚类。首先，将用户对项目的评分信息进行处理；然后采用改进Kmeans聚类算法RLPSO_KM对处理后的信息进行聚类，其中包括初始聚类中心和初始聚类数目以及聚类的迭代次数等一系列参数设置；最后输出聚类的类簇中心和每个类簇内包含的用户和项目信息。

[0060] (2)基于时间熵的个性化推荐。不同用户的兴趣偏好往往不同，一些用户一直喜欢同类型的电影，而对于另一些用户则刚好相反，他们对电影的喜爱类型可能会随着他们的情绪、朋友和时尚趋势而发生改变。

[0061] 这里根据用户对项目的评分信息，我们将用户的兴趣偏好分为喜欢(Likes)和不喜欢(Dislikes)。同样地如果以时间来衡量，将其分为最近(Recent)和过去(Past)的时间，所以我们简单地将用户的兴趣偏好划分为四种类型。第一种类型为RecentLikes；PastLikes，这类用户的兴趣偏好具有规律性，也就是说此类用户通常只喜欢一种类型的电影，并且喜欢一种电影类型的持续时间较长，一般很长一段时间不会发生改变。第二种用户类型是RecentLikes;PastDislikes，第三种类型是RecentDislikes;PastLikes。对于这些用户来说，他们的兴趣偏好随着时间的改变而发生变化。最后一种是RecentDislikes；PastDislikes，这些用户评价过的项目信息呈现多样化的趋势，同时项目呈现随机性和无规律性。

[0062] 通过对用户进行分类，我们将用户的类型分为两种模式，分别为第一种模式(第一类用户)，第二种模式(第二、三、四类用户)。由于项目隐含的价值和时间信息意味着用户属于哪种模式的用户。例如，如果用户频繁访问相同类型的电影，我们认为他是遵循第一种模式的用户。这种模式很简单，同样类型的电影呈现均匀分布的趋势。相反，如果该用户属于第二种模式，则没有规律性可循。受此启发，我们提出了一种基于时间熵的新型模式挖掘方法，这种方法可以用来衡量用户历史评分记录的时间分布，定义时间熵的计算公式如下：

[0063] $$H=-\sum_{i=1}^{n}\frac{m_i\cdot score_i}{M}\log\frac{m_i}{M} \qquad (1)$$

[0064] 公式(1)中，n是时间间隔数，用户对项目i的评分为$score_i$，m_i代表一个时间间隔内所有用户对项目i的评分总和，M指所有m_i的总和。

[0065] 通常一个项目含有多个标签，可通过标签对项目进行简单描述。对于一部电影，这些标签可以代表电影的类型。对于一个用户，他可以在不同的时间评论相同的标签，即一个标签可以从属于几次评分周期。时间熵可以衡量时间戳的混乱程度，该值越高则表明用户

越喜欢这个标签。

[0066] 如表1所示，该表中包含4个用户对8个项目的评分信息。u_i(i＝1,2,3,4)为用户信息集合，i_k(k＝1,2,…,8)代表项目信息集合。用户对项目的评分值介于1到5之间，表中还罗列出用户对项目的访问次数以及对项目所属标签的评分信息。

[0067] 表1用户-项目评分信息

[0068] Table 1 User-item rating

[0069]

	$i_1(tag_1,tag_3)$	$i_2(tag_1)$	$i_3(tag_1,tag_3)$	$i_4(tag_3)$
u_1	$(5,t_1)$	$(3,t_2)$	$(2,t_1)$	$(4,t_2)$
u_2	$(2,t_1)$	$(1,t_2)$	$(3,t_1)$	$(5,t_2)$
u_3	$(4,t_1)$	$(5,t_1)$	$(3,t_1)$	$(4,t_1)$
u_4	$(5,t_1)$	$(4,t_2)$	$(5,t_2)$	$(4,t_4)$
	$i_5(tag_1, tag_2, tag_3)$	$i_6(tag_1,tag_2)$	$i_7(tag_2,tag_3)$	$i_8(tag_1,tag_3)$
u_1	$(3,t_3)$	$(2,t_3)$	$(1,t_2)$	$(4,t_3)$
u_2	$(2,t_3)$	$(3,t_3)$	$(5,t_2)$	$(1,t_3)$
u_3	$(3,t_5)$	$(2,t_5)$	$(1,t_5)$	$(1,t_5)$
u_4	$(3,t_5)$	$(3,t_4)$	$(1,t_5)$	$(3,t_2)$

[0070] 如表2所示，表中有5个时间间隔。对于用户u_1，包含3条时间分布记录。对于用户u_2，其也包含3条时间分布记录。

[0071] 表2用户对tag_1的评分信息

[0072] Table 2 Users rating for tag_1

[0073]

users	Scores with time		
u_1	$(3.5,2,t_1)$	$(3,1,t_2)$	$(3,3,t_3)$
u_2	$(2.5,2,t_1)$	$(1,1,t_2)$	$(2,3,t_3)$
u_3	$(3,3,t_1)$	$(2,3,t_5)$	
u_4	$(5,1,t_1)$	$(4,3,t_2)$	$(3,1,t_4)$

[0074]

[0075] 但是对于用户u_1和用户u_2，哪个用户更喜欢tag_1?以下是他们的时间熵：

[0076] $$H_{u_1}=-\frac{3.5*2}{6}\log\frac{2}{6}-\frac{3*1}{6}\log\frac{1}{6}-\frac{3*3}{6}\log\frac{3}{6}=3.22$$

[0077] $$H_{u_2}=-\frac{2.5*2}{6}\log\frac{2}{6}-\frac{1*1}{6}\log\frac{1}{6}-\frac{2*3}{6}\log\frac{3}{6}=1.91$$

[0078] 由于$H_{u_1}>H_{u_2}$，所以用户u_1比用户u_2更喜欢tag_1。这与用户u_1比用户u_2对tag_1评分更高的事实相一致。

[0079] 多样性推荐通过采用协调项目列表中的项目的相似点和不同之处来提高推荐性的推荐效率吸引了大量的关注。一种通用的标准衡量多样性的方法是尽可能最大化不同项目的总和。因此，对于一个给定的项目，我们设定它的多样性公式定义如下：

[0080] $$dv(i)=\frac{fr(i)}{\alpha\cdot z(i)+1} \quad (2)$$

[0081] $$z(i)=\sum_{j\in Set_i} sim(i,j) \quad (3)$$

[0082] $$sim(i,j)=\frac{\sum_{u\in U_{ij}}(r_{ui}-\overline{r_i})(r_{uj}-\overline{r_j})}{\sqrt{\sum_{u\in U_{ij}}(r_{ui}-\overline{r_i})^2\sum_{u\in U_{ij}}(r_{uj}-\overline{r_j})^2}} \quad (4)$$

[0083] 公式(2)中,fr(i)指标签tag_1在集合中出现的频率,z(i)是标签tag_1候选集合中所有标签的相似度之和,α为调节系数,这里我们选取该值为2。

[0084] 公式(4)中,项r_{ui}和r_{uj}分别代表用户u对项目i和项目j的评分,$\overline{r_j}$为所有评价过项目j的用户对项目j的平均评分,$\overline{r_i}$为所有评价过项目i的用户对项目i的平均评分,U_{ij}为同时对项目i与项目j评分的用户集合,sim(i,j)的值介于[-1,1]之间。

[0085] (3)计算项目流行度。对于系统中新加入的目标用户,我们则推荐最受用户欢迎的N个项目给目标用户。我们设定推荐系统中项目的流行度公式定义如下:

[0086] $$i_{pop}=\frac{|U_i|}{\sqrt{\sum_{i\in I}|U_i|^2}} \quad (5)$$

[0087] 公式(5)中,U_i代表推荐系统中评价过项目i的用户集合,I为推荐系统中的所有项目。

[0088] 综上,通过对协同过滤算法中的用户对项目的评分信息进行聚类,然后通过计算用户对项目的时间熵计算用户所属模式进而为用户进行多样性推荐,并对系统中为未加入系统的用户进行前N个项目的推荐。通过这种方式,一方面通过使用在类簇中计算用户的邻域提高了推荐系统的推荐效率;另一方面,通过计算用户对项目的时间熵,使得用户的兴趣得到极大的体现,弥补了传统推荐系统的单一性,使得推荐系统具有更好的多样性。

[0089] 为使本发明的目的、技术方案和特点更加清楚明白,以下结合具体实施例子,并参照附图,对本发明进行进一步的细化说明。基于时间熵的个性化推荐方法的整体流程图如图4所示。

[0090] 各个步骤说明如下:

[0091] (1)提出一种基于RLPSO_KM聚类算法对用户和评分信息进行聚类,使得推荐系统的推荐效率得到了提高。

[0092] (2)提出一种基于时间熵的个性化推荐方法,引入了时间熵计算方式,从而使得用户的多样化兴趣得到了极大体现。

[0093] (3)提出一种计算项目流行度的方式为新加入系统用户进行推荐,避免了推荐系统中项目的冷启动问题。

[0094] 实验环境如下:

[0095] 本发明应用实例通过实验验证本文提出的基于时间熵的个性化推荐方法的实际效果,实验环境为win7(64位)主机,8G内存,1T硬盘,采用的数据集为MovieLens(10M)和从Douban.com抓取的数据,并将算法10次实验的平均值作为最终的实验结果来验证推荐的准确性,同时比较基于时间熵的个性化推荐方法与其它的多样性推荐方法的多样性。

[0096] 首先,本发明应用实例对用户和项目的评分信息进行聚类。针对评分信息进行向量化或者特征化处理,使得评分信息满足聚类输入的格式;然后通过采用改进的聚类算法

RLPSO_KM对处理后的信息进行聚类，包括初始聚类中心和初始聚类数目的选择以及聚类的迭代次数等一系列参数设置；最后输出经聚类算法聚类的每个类簇中心和每个类簇内所包含的用户和项目信息。

[0097] 然后，建立基于时间熵的个性化推荐方法。通过引入时间熵的计算方式，计算用户对项目的时间熵，进而对用户进行分类和模式匹配，最后为用户进行多样性推荐。该推荐算法与其它算法的推荐效率对比图如图5和图6所示。

[0098] 最后，计算项目流行度。针对未加入系统的新用户，通过计算项目的流行度，并选取其中流行度最高的N个项目为用户进行推荐，从而解决系统中用户的冷启动问题。

[0099] 从图5中本发明提出的方法与其他算法进行对比，可以明确看到本发明提出的PTCF不论选用的实验数据集是MovieLens还是Douban数据集，相对于PMF、BPMF、SVD、McoC几种算法而言，其在精确率上都更占有优势，准确率在MovieLens和Douban数据集上相对于MCoC算法在N＝10时分别提高了1.04％和1.07％。同时，不论我们选取的Top-N值是10,20和50中的任意一种，本发明提出方法的MAP值都明显高于其他几种算法，这也说明了本发明提出方法PTCF在推荐效率上的优势。

[0100] 从图6中我们可知，通过比较在不同推荐项目列表数目的情况下，本发明提出的PTCF方法无论在实验数据集的选择上还是当推荐项目数为10,20和50时，其多样性基本保持一致，未发生明显改变。同时相对于CUTA算法而言，当N＝10时，本发明提出的个性化推荐方法的多样性在MovieLens数据集下提高了4倍，相对于CUTATime算法而言提高了1.05％。

[0101] 综合上述实验，本发明应用实例利用基于时间熵的个性化推荐方法对用户进行推荐，一方面，对用户和项目信息进行聚类及计算项目的流行度，可以提高推荐的效率、准确度并解决用户的冷启动问题。另一方面，通过计算用户对项目的时间熵可以有效利用用户的多兴趣，从而提高推荐方法的多样性。

[0102] 应该明白，公开的过程中的步骤的特定顺序或层次是示例性方法的实例。基于设计偏好，应该理解，过程中的步骤的特定顺序或层次可以在不脱离本公开的保护范围的情况下得到重新安排。所附的方法权利要求以示例性的顺序给出了各种步骤的要素，并且不是要限于所述的特定顺序或层次。

[0103] 在上述的详细描述中，各种特征一起组合在单个的实施方案中，以简化本公开。不应该将这种公开方法解释为反映了这样的意图，即，所要求保护的主题的实施方案需要比清楚地在每个权利要求中所陈述的特征更多的特征。相反，如所附的权利要求书所反映的那样，本发明处于比所公开的单个实施方案的全部特征少的状态。因此，所附的权利要求书特此清楚地被并入详细描述中，其中每项权利要求独自作为本发明单独的优选实施方案。

[0104] 为使本领域内的任何技术人员能够实现或者使用本发明，上面对所公开实施例进行了描述。对于本领域技术人员来说；这些实施例的各种修改方式都是显而易见的，并且本文定义的一般原理也可以在不脱离本公开的精神和保护范围的基础上适用于其它实施例。因此，本公开并不限于本文给出的实施例，而是与本申请公开的原理和新颖性特征的最广范围相一致。

[0105] 上文的描述包括一个或多个实施例的举例。当然，为了描述上述实施例而描述部件或方法的所有可能的结合是不可能的，但是本领域普通技术人员应该认识到，各个实施例可以做进一步的组合和排列。因此，本文中描述的实施例旨在涵盖落入所附权利要求书

的保护范围内的所有这样的改变、修改和变型。此外，就说明书或权利要求书中使用的术语“包含”，该词的涵盖方式类似于术语“包括”，就如同“包括，”在权利要求中用作衔接词所解释的那样。此外，使用在权利要求书的说明书中的任何一个术语“或者”是要表示“非排它性的或者”。

[0106] 本领域技术人员还可以了解到本发明实施例列出的各种说明性逻辑块(illustrative logical block)，单元，和步骤可以通过电子硬件、电脑软件，或两者的结合进行实现。为清楚展示硬件和软件的可替换性(interchangeability)，上述的各种说明性部件(illustrative components)，单元和步骤已经通用地描述了它们的功能。这样的功能是通过硬件还是软件来实现取决于特定的应用和整个系统的设计要求。本领域技术人员可以对于每种特定的应用，可以使用各种方法实现所述的功能，但这种实现不应被理解为超出本发明实施例保护的范围。

[0107] 本发明实施例中所描述的各种说明性的逻辑块，或单元都可以通过通用处理器，数字信号处理器，专用集成电路(ASIC)，现场可编程门阵列或其它可编程逻辑装置，离散门或晶体管逻辑，离散硬件部件，或上述任何组合的设计来实现或操作所描述的功能。通用处理器可以为微处理器，可选地，该通用处理器也可以为任何传统的处理器、控制器、微控制器或状态机。处理器也可以通过计算装置的组合来实现，例如数字信号处理器和微处理器，多个微处理器，一个或多个微处理器联合一个数字信号处理器核，或任何其它类似的配置来实现。

[0108] 本发明实施例中所描述的方法或算法的步骤可以直接嵌入硬件、处理器执行的软件模块、或者这两者的结合。软件模块可以存储于RAM存储器、闪存、ROM存储器、EPROM存储器、EEPROM存储器、寄存器、硬盘、可移动磁盘、CD-ROM或本领域中其它任意形式的存储媒介中。示例性地，存储媒介可以与处理器连接，以使得处理器可以从存储媒介中读取信息，并可以向存储媒介存写信息。可选地，存储媒介还可以集成到处理器中。处理器和存储媒介可以设置于ASIC中，ASIC可以设置于用户终端中。可选地，处理器和存储媒介也可以设置于用户终端中的不同的部件中。

[0109] 在一个或多个示例性的设计中，本发明实施例所描述的上述功能可以在硬件、软件、固件或这三者的任意组合来实现。如果在软件中实现，这些功能可以存储与电脑可读的媒介上，或以一个或多个指令或代码形式传输于电脑可读的媒介上。电脑可读媒介包括电脑存储媒介和便于使得让电脑程序从一个地方转移到其它地方的通信媒介。存储媒介可以是任何通用或特殊电脑可以接入访问的可用媒体。例如，这样的电脑可读媒体可以包括但不限于RAM、ROM、EEPROM、CD-ROM或其它光盘存储、磁盘存储或其它磁性存储装置，或其它任何可以用于承载或存储以指令或数据结构和其它可被通用或特殊电脑、或通用或特殊处理器读取形式的程序代码的媒介。此外，任何连接都可以被适当地定义为电脑可读媒介，例如，如果软件是从一个网站站点、服务器或其它远程资源通过一个同轴电缆、光纤电缆、双绞线、数字用户线(DSL)或以例如红外、无线和微波等无线方式传输的也被包含在所定义的电脑可读媒介中。所述的碟片(disk)和磁盘(disc)包括压缩磁盘、镭射盘、光盘、DVD、软盘和蓝光光盘，磁盘通常以磁性复制数据，而碟片通常以激光进行光学复制数据。上述的组合也可以包含在电脑可读媒介中。

[0110] 以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步

详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

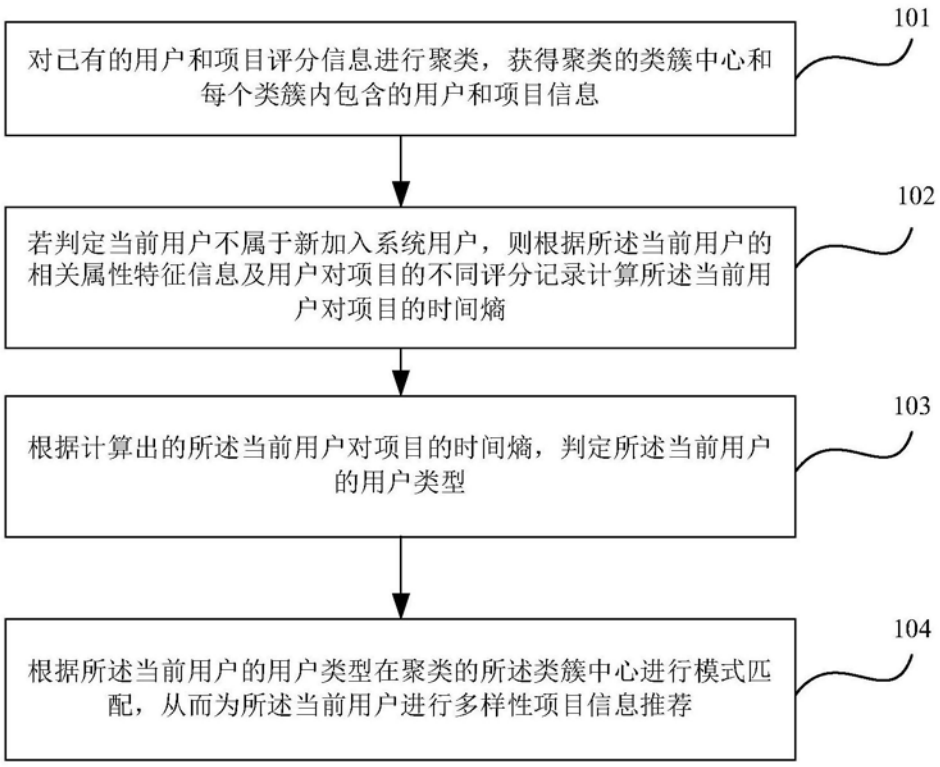

101
对已有的用户和项目评分信息进行聚类，获得聚类的类簇中心和每个类簇内包含的用户和项目信息
102
若判定当前用户不属于新加入系统用户，则根据所述当前用户的相关属性特征信息及用户对项目的不同评分记录计算所述当前用户对项目的时间熵
103
根据计算出的所述当前用户对项目的时间熵，判定所述当前用户的用户类型
104
根据所述当前用户的用户类型在聚类的所述类簇中心进行模式匹配，从而为所述当前用户进行多样性项目信息推荐

图1

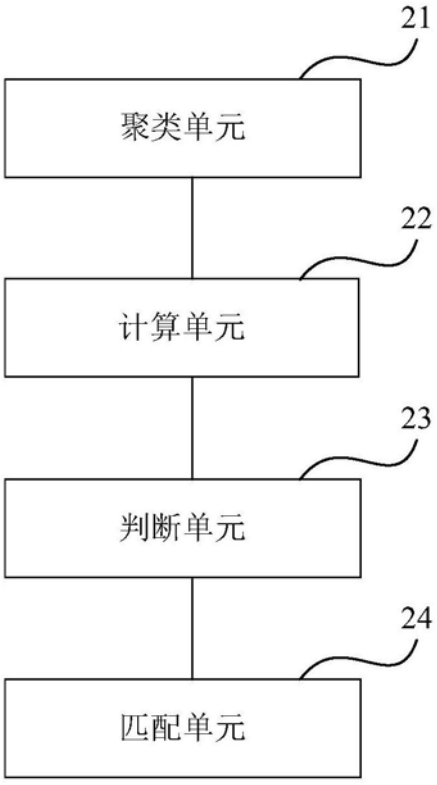

21
聚类单元
22
计算单元
23
判断单元
24
匹配单元

图2

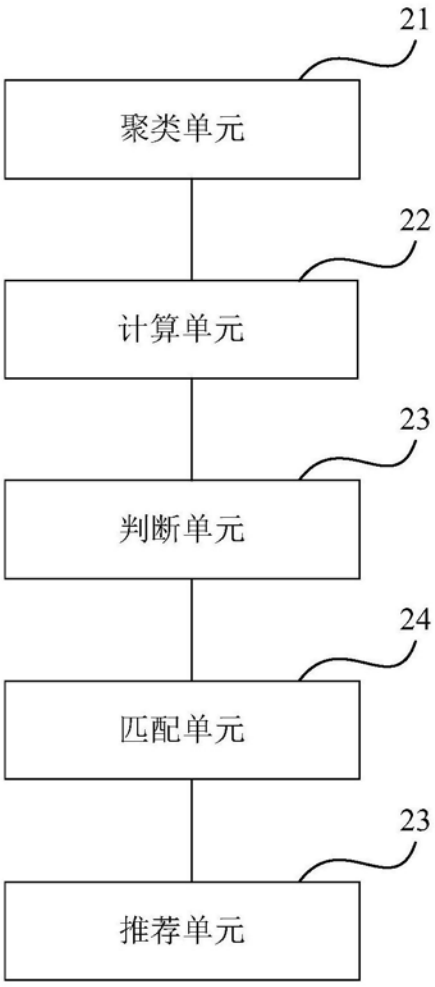


图3

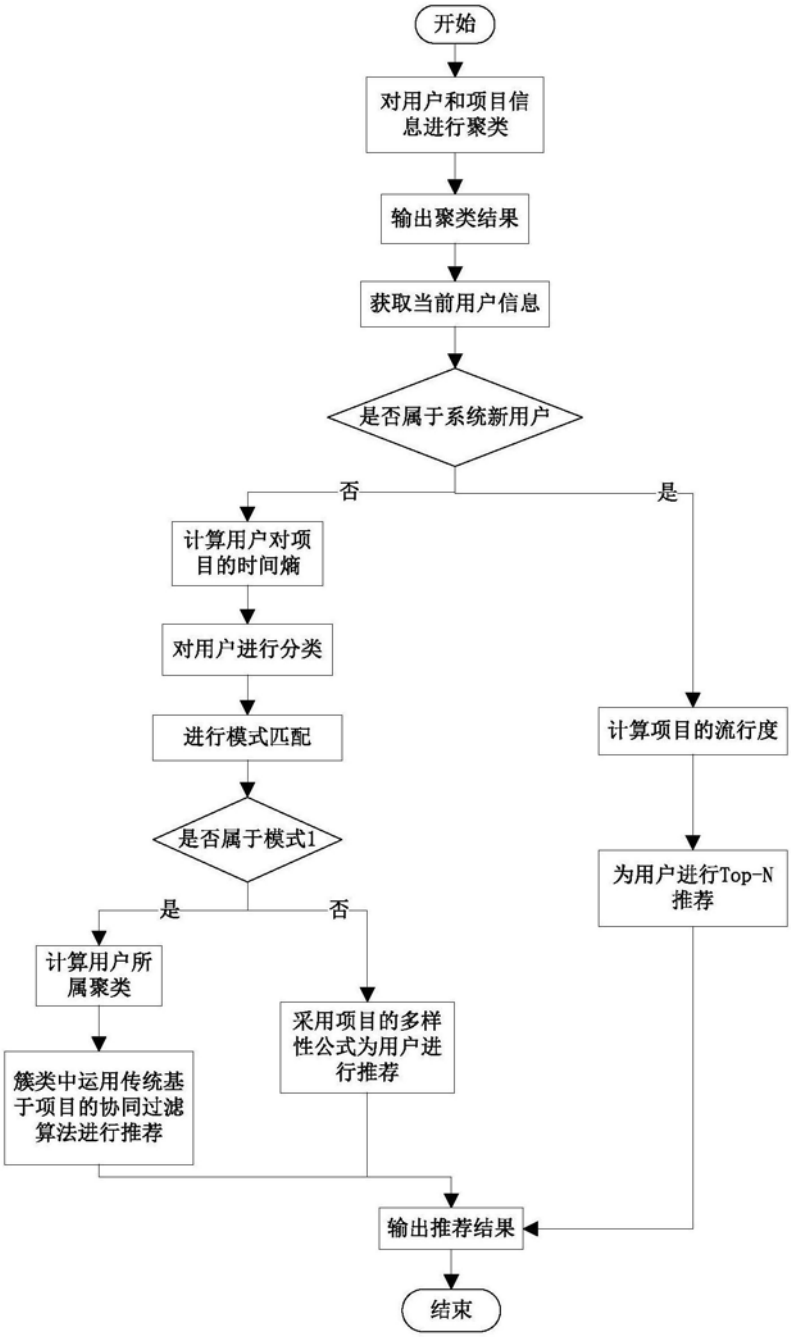

开始
对用户和项目信息进行聚类
输出聚类结果
获取当前用户信息
是否属于系统新用户
否
是
计算用户对项目的时间熵
对用户进行分类
进行模式匹配
是否属于模式1
是
否
计算用户所属聚类
簇类中运用传统基于项目的协同过滤算法进行推荐
采用项目的多样性公式为用户进行推荐
计算项目的流行度
为用户进行Top-N推荐
输出推荐结果
结束

图4

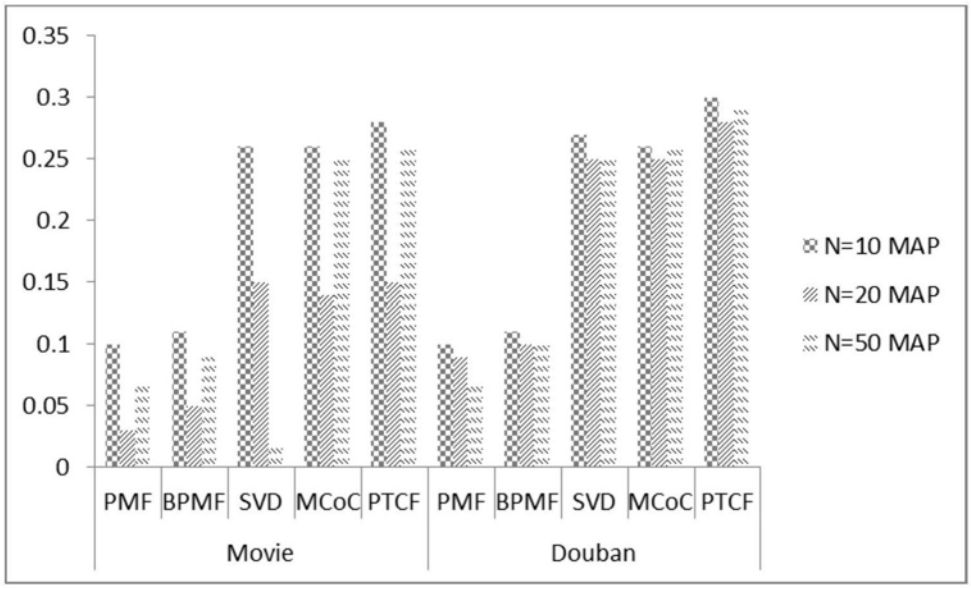

0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
PMF
BPMF
SVD
MCoC
PTCF
PMF
BPMF
SVD
MCoC
PTCF
Movie
Douban
N=10 MAP
N=20 MAP
N=50 MAP

图5

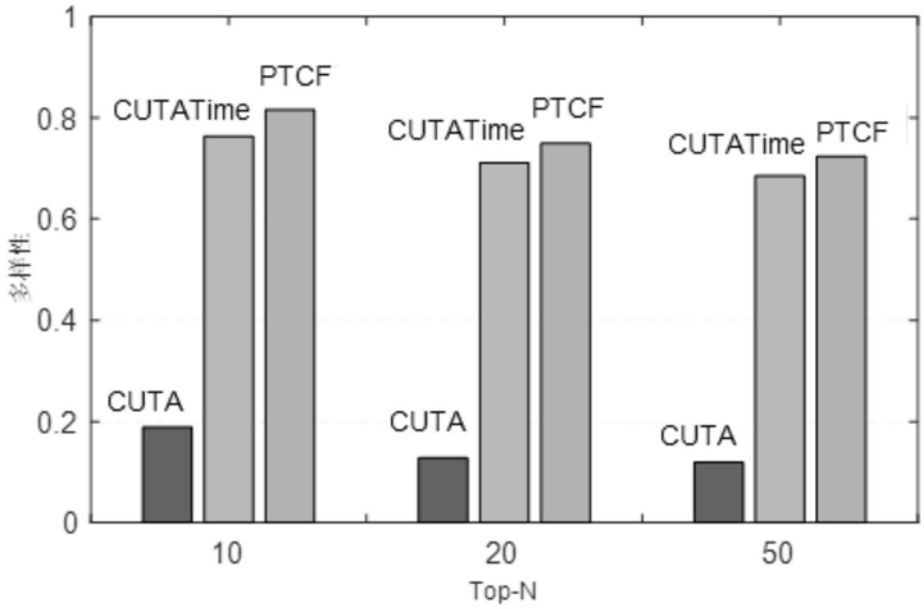

1
0.8
0.6
0.4
0.2
0
多样性
CUTA
CUTATime
PTCF
CUTA
CUTATime
PTCF
CUTA
CUTATime
PTCF
10
20
50
Top-N

图6