



开放科学  
(资源服务)  
标识码  
(OSID)

# 社交网络环境下基于用户响应的推荐方法研究

王刚 郭雪梅

天津医科大学图书馆 天津 300070

**摘要:** 在社会网络中, 通过利用用户兴趣、社交关系、历史日志等用户响应来提高推荐系统的性能。通过社交网络中用户随时间变化的响应信息实现用户的相似性度量, 并以此为基础实现对目标用户的协同过滤推荐。将本文提出的方法和传统的基于用户的协同过滤推荐方法在 Extended Epinions DataSet 进行实验比较, 实验结果发现该方法在推荐效果上优于传统方法。

**关键词:** 社交网络; 用户响应; 协同过滤推荐

**中图分类号:** G354.42

## Research on Recommendation Method Based on Users' Response in Social Network

WANG Gang GUO Xuemei

Library of Tianjin Medical University, Tianjin 300070, China

**Abstract:** In social network, the system usually improve the performance of the recommendation by using users' response, such as user interest, social relations and history log. In this study, the user similarity measurement is achieved through user response information in social network, and based on this, collaborative filtering recommendation for target users is achieved. we compare the proposed method with the traditional user collaborative filtering recommendation method in Extended Epinions DataSet, and the experimental results show that the proposed method is superior to the traditional method in recommendation effect.

**Keywords:** Social network; users' response; collaborative filtering recommendation

**基金项目:** 2018 年度天津市哲学社会科学规划项目“基于人工智能的公众医学信息服务平台构建”(TJTQ18-010); 2017 年度天津市教委科研计划项目“基于本体建模的个性化 Web 服务发现方法研究”(2017SK098)。

**作者简介:** 王刚(1980-), 硕士, 馆员, 研究方向: 信息资源管理, E-mail: wanggang80@tmu.edu.cn; 郭雪梅(1983-), 硕士, 副研究馆员, 研究方向: 学科建设, E-mail: guoxuemei@tmu.edu.cn

## 引言

目前,对于诸如微信、微博、Facebook、TWITTER 等社交网络的使用已经成为人们在线活动不可缺少的部分,同时它也成为用户信息共享和获取的重要途径<sup>[1]</sup>。然而,随着社交网络用户数量爆炸式增长,用户所产生的信息也随之激增。由于用户的信息处理能力远远落后于信息增长的速度,这必然导致信息过载的问题,使得用户获取有用信息的成本增加。

为有效解决这一问题,出现了各种推荐技术可以为用户提供优质的个性化服务。在传统的个性化推荐算法中,协同过滤算法无疑是最成功的一种<sup>[2]</sup>。协同过滤推荐算法是基于某些项目用户之间的相似性偏好,即如果用户对某些项目有相似的兴趣,则这些用户很可能对其他项目感兴趣<sup>[3]</sup>。但协同过滤算法不能基于用户动态变化的兴趣特征进行有效的推荐。

本文提出了在复杂社交网络环境下基于用户动态信息为用户提供个性化推荐的方法,通过考虑用户响应信息和时间因素的动态信息来反映用户的动态偏好特征,以此为基础设计了改进的推荐算法以及新的相似性度量方法。实验结果表明,与传统的协同过滤推荐算法相比,该算法具有更好的推荐性能。

## 1 相关研究

协同过滤推荐算法是最成功的推荐算法之一。它的核心思想可以分为三个部分:第一,计算从用户的历史信息感兴趣的用户之间的相

似性;然后,根据用户对特定项目的用户偏好的相似性选择最近邻;第三,对项目预测评分进行降序排序,将得分最高的项目推荐给目标用户<sup>[4-5]</sup>。

### 1.1 传统的相似度量方法及用户动态信息模型

协同过滤推荐的关键是度量不同用户之间的相似性,广泛采用的方法是基于用户共同历史评分数据的相似性计算。在相似性计算方法中,虽然每种方法都有其优缺点,但最常用的方法是皮尔森相关系数<sup>[6]</sup>和余弦相似性<sup>[7]</sup>两类方法,计算公式分别如公式(1)和(2)所示。

$$Sim(x, y) = \frac{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)^2} \sqrt{\sum_{i \in I_{x,y}} (r_{y,i} - \bar{r}_y)^2}} \quad (\text{公式1})$$

$$Sim(x, y) = \cos\left(\begin{matrix} \vec{x} \\ \vec{y} \end{matrix}\right) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \times \|\vec{y}\|} = \frac{\sum_{i \in I_{x,y}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2} \sqrt{\sum_{i \in I_y} r_{y,i}^2}} \quad (\text{公式2})$$

根据公式(1)或(2)相所示的相似度量方法,可以得到目标用户 $u$ 的最近邻用户集合,记为 $N_u$ 。然后利用最近邻用户 $v$ ,  $v \in N_u$ ,通过公式(3)预测目标用户 $u$ 对项目 $i$ 的评分<sup>[8]</sup>。

$$R_{u,i} = \bar{r}_u + \frac{\sum_{v \in N_u} Sim(u, v) \times (r_{v,i} - \bar{r}_v)}{\sum_{v \in N_u} (|Sim(u, v)|)} \quad (\text{公式3})$$

传统的协同过滤推荐算法没有考虑社交网络环境中用户兴趣会随时间发生变化<sup>[9]</sup>，因此传统的用户动态信息模型通过定义时间权重函数来表示用户的动态兴趣模式，然后将时间权重与推荐方法相结合。举例说明，假定用户的兴趣是随时间变化的单调递减函数，并将其与用户兴趣预测函数相结合，或者通过不同的时间段将用户的动态兴趣划分为更详细的信息，并构造相应的时间权重函数，这些方法都改进了推荐算法的推荐结果<sup>[10]</sup>。

## 1.2 社交网络环境中基于用户行为的研究

社交网络中存在海量用户，如何帮助用户得到所需的信息资源是亟待解决的问题之一，也是社交网络相关研究的重要主题。文献[10]从用户阅读的角度对学术社交网络用户内容使用行为进行探究，以科学网热门博文为分析对象，采用方差分析、相关性分析等方法，从内容基本特征、用户内容使用行为关系、内容贡献者特征等方面对用户内容使用行为特征进行研究；文献[11]提出一种基于混合图的在线社交网络个性化推荐系统，将用户社会关系网络和社会化行为融入信息推荐通过构建用户资源混合图以及构建方法并设置权重，以此为基础采用重启动随机游走进行顶点间相似度计算，得到个性化推荐列表，进行推荐；文献[12]利用三度影响力理论，从网络结构的角度进一步拓展用户关系连接，提高社交网络好友推荐的效率，通过计算用户之间的关系强度，并筛选关系强度较大的用户集合，最后通过用户共同关注的内容计算用户兴趣相似度。

## 2 推荐方法的设计

### 2.1 基于社交用户动态信息的相似度计算

在社交网络环境中，通常以图的形式表示用户之间的关系，记为  $G=(V, E)$ 。其中  $V$  为图中结点的集合，用于表示用户或者项目， $E$  为图中的边，连接图中的结点，用于表示社交网络中不同用户之间的好友关系或者用户与项目之间的行为关系。在本文中，边的集合  $E$  特指用户对于某个项目的兴趣反应。本文根据用户对于项目不同类型的响应信息和时间因素，提出了一种新的推荐算法。

在社交网络环境中，我们可以把用户的信息转发行为、收藏行为等看作是积极响应类型。因此当用户  $u$  在某个时间戳  $t$  时，其积极响应的数量可以定义为：

$$R^+(u, t) = \{i_k | t_{u, i_k} < t, (u, i_k) \in G\} \quad (\text{公式 4})$$

其中， $G$  表示社交网络中用户与项目的关系图； $U$  表示社交网络用户集合  $U = \{u_1, u_2, \dots, u_n\}$ ， $u \in U$ ； $I$  表示项目的集合  $I = \{i_1, i_2, \dots, i_n\}$ ， $i_k \in I$ ， $i_k$  为项目集合  $I$  中的第  $k$  个项目； $t$  表示用户  $u$  对于项目  $i$  给出积极响应的的时间。

同样的，如果用户对某个项目不感兴趣或者采取诸如屏蔽、取消关注等行为，我们可以将这些行为视为消极响应信息。因此，当用户在时间戳时，可以将消极响应信息的数量定义为：

$$R^-(u, t) = \{i_k | t_{u, i_k} < t, (u, i_k) \in G\} \quad (\text{公式 5})$$

基于上述响应类型定义，同时考虑用户兴趣偏好的时间效应并针对传统协同过滤算法在这方面的缺陷，本文提出了一种新的用户相似

度量方法。给定某个时间点，不同用户对于同一项目所给出的响应是不同的，这就意味着不同用户对于该项目的兴趣度也不同。通过给出用户响应的权重，能够更准确地反映用户兴趣模式的动态特征，对于用户的积极响应。用户相似性度量可以定义为

$$Sim^+(u,v) = \frac{|R^+(u,t) \cap R^+(v,t)|}{|R^+(u,t)| |R^+(v,t)|} \cdot \sum_{k=1}^n f_{time}(t) \quad (公式6)$$

其中， $|R^+(u,t)|$  和  $|R^+(v,t)|$  分别表示用户  $u$  和  $v$  各自做出积极响应的项目集合； $|R^+(u,t)| |R^+(v,t)|$  表示用户  $u$  和  $v$  均做出积极响应的项目集合， $n$  表示该集合中项目的数量； $f_{time}(t)$  用户表示用户的动态兴趣。同理，基于用户消极响应信息的相似度量公式如下所示。

$$Sim^-(u,v) = \frac{|R^-(u,t) \cap R^-(v,t)|}{|R^-(u,t)| |R^-(v,t)|} \cdot \sum_{k=1}^m f_{time}(t) \quad (公式7)$$

公式中符号含义与基于积极响应的相似度量公式符号含义类似；公式中  $m$  表示用户  $u$  和  $v$  均做出消极响应的项目集合中项目的数量。

由于考虑到社交网络用户的兴趣变化和响应信息，本文设计了随时间递减的函数来对用户社交网络中的动态兴趣特征进行建模，并将其与改进的相似性度量相结合。公式中的时间权重函数定义为

$$f_{time}(t) = \frac{R(u,t)}{|\Delta t|^\lambda} \quad (公式8)$$

其中， $|\Delta t| = |t_{u,i_k} - t_{v,i_k}|$  表示给出相同类型响应信息的两个用户的时间间隔，通过用户兴趣衰减程度参数  $\lambda$  确定用户动态兴趣的衰减率以及根据应用环境使得推荐算法性能最优。 $R(u,t)$

表示用户做出积极响应的项目数量，其值等于  $|R^+(u,t) \cap R^+(v,t)|$  所表示的集合中的项目数。

然后，将调节因子  $\alpha$  与这两种不同的相似度量方法相结合，组合相似度计算方法定义为

$$Sim'(u,v) = \alpha Sim^+(u,v) + (1-\alpha) Sim^-(u,v) \quad (公式9)$$

其中，调节参数  $\alpha$  的取值范围为  $[0,1]$ ， $\alpha$  和  $1-\alpha$  分别表示积极响应和消极响应的权重，可以采用最优权值来优化推荐结果。

将基于社交网络用户响应信息类型的相似度量与传统协同过滤推荐算法的相似度计算即皮尔逊相关系数和余弦相似度结合，计算公式分别如下所示

$$Sim(u,v) = \frac{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)^2} \sqrt{\sum_{i \in I_{x,y}} (r_{y,i} - \bar{r}_y)^2}} \cdot Sim'(u,v) \quad (公式10)$$

$$Sim(u,v) = \frac{\sum_{i \in I_{x,y}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2} \sqrt{\sum_{i \in I_y} r_{y,i}^2}} \cdot Sim'(u,v) \quad (公式11)$$

## 2.2 基于社交用户动态信息的评分预测

进行用户相似度计算之后，选择目标用户  $K$  最近邻以保证推荐系统的推荐效果。目标用户  $u$  已评分的项目集合记为  $I_u$ ，预测该用户对于项目  $i$  的评分。通过目标用户  $u$  的最近邻  $v$  对于项目  $i$  的评分实现用户  $u$  对于项目  $i$  的评分预测，计算公式如下

$$R(u,i) = \bar{r}_u + \frac{\sum_{i=1}^n [Sim(u,v)(r_{v,i} - \bar{r}_v)]}{\sum_{i=1}^n Sim(u,v)} \quad (公式12)$$

其中， $Sim(u,v)$  表示目标用户  $u$  与最近邻用户  $v$  的

相似度,  $r_{vi}$  表示用户  $v$  对于项目  $i$  的评分,  $r_u$  和  $r_v$  分别表示用户  $u$  和  $v$  所进行过项目评分的均值。

### 2.3 算法的时间复杂度分析

在推荐算法的执行过程中, 对于用户相似度计算占很大比重, 当用户历史数据数量庞大时, 推荐算法的性能至关重要。因此, 本部分

分析本文所提出的推荐算法的时间复杂度。时间复杂度针对以下两种算法进行分析, 其一是诸如皮尔逊相关系数或余弦相似度方法的传统相似度度量, 算法参见(表1); 其二是基于社会用户的动态响应信息的相似性度量, 协同过滤算法主要包括用户相似性计算和预测目标用户的评分, 算法参见(表2)。

表1 传统的用户相似度度量方法

输入: 用户集 $U$ , 项目集 $I$ , 评分矩阵 $R_{U \times I}$ 输出: 用户间相似度度量结果 $Sim(u,v)$ 步骤1: 令 $K_u=0$ 步骤2: 当 $K_u <  U $ 时, 转至步骤3, 否则, 算法结束 步骤3: 令 $K_u=K_u+1$ ; 当 $K_u <  U $ 时, 根据公式(1)或(2)计算用户之间相似度 否则, $++K_u$ , 转至步骤2
---

表2 基于社会用户响应信息的用户相似性计算和预测目标用户的评分

输入: 目标用户 $u$ , 目标项目 $i$ , 最近邻集合 $K$ , 项目集合 $I$ , 时间戳 $t$ , 用户兴趣衰减程度参数 $\lambda$ , 调节参数 $\alpha$ 输出: 目标用户 $u$ 对于项目 $i$ 的评分预测值 步骤1: 令 $k_i=0$ ; 步骤2: 当 $k_i <  K $ 时, 转至步骤3; 否则转至步骤4; 步骤3: 当 $k_i <  I $ 时, 根据公式(4)和(5)分别计算用户积极响应和消极响应的数量; 否则, $++k_i$ , 转至步骤2; 步骤4: 当 $u,v \in U$ , 根据公式(6)和(7)计算 $Sim+(u,v)$ , $Sim-(u,v)$ ; 步骤5: 根据公式(9)得到用户相似度; 步骤6: 根据公式(10)或(11)计算用户之间的组合相似度 步骤7: 根据公式(12)计算目标用户 $u$ 对于项目 $i$ 的评分预测值
--

根据算法1和2的执行过程可以知道, 当进行相似度计算的时候算法的时间复杂度主要取决于社交网络中用户集  $U$  和项目集  $I$  的大小。用户集  $U$  和项目集  $I$  的大小分别记为  $|U|$  和  $|I|$ , 根据传统的用户相似度计算方法, 计算一对用户相似度的时间复杂度为  $O(|U|)$ , 而计算任意两个不同用户相似度的时间复杂度是  $(|U| \times (|U| - 1) / 2) \times O(|I|) = O(|U|^2 |I|)$ 。考虑到传统相似度计算的时间复杂度, 根据  $K$  近邻用户集的选择和计算方法, 计算  $K$  个最近邻与目标用户之间

的相似性。令基于用户积极响应所计算用户相似度的时间复杂度为  $O(|R^+|)$ , 基于用户消极响应所计算用户相似度的时间复杂度为  $O(|R^-|)$ , 可以得到根据公式(9)所计算一对用户相似度的时间复杂度为  $O(|I|) \cdot O(\alpha |R^+| + (1-\alpha) |R^-|) = O(|I| \cdot |R|)$ , 而计算任意两个不同用户相似度的时间复杂度是  $(|K| \times (|K| - 1) / 2) \times O(|I| \cdot |R|) = O(|K|^2 \cdot |I| \cdot |R|)$ 。在用户预测评分计算中, 影响算法时间复杂度的主要因素是最近邻集的大小  $|K|$  和项目集大小  $|I|$ 。



### 3 实验比对

#### 3.1 数据准备

为了验证本文所提出的推荐方法在性能上优于传统的协同过滤推荐算法,在对比实验中使用 Extended Epinions DataSet 作为实验对象,该数据集包含约 132000 个用户,841372 条用户评论。该数据集中以 Trust (1) /Distrust (-1) 表示某个用户对于其他用户所发表内容所表示出的积极 / 消极响应。对于现有的数据集,通过随机选择将数据集平均分成 10 个子集,其中随机选取九个作为训练集,剩下的子集作为测试集。

#### 3.2 兴趣衰减程度参数 $\lambda$ 对推荐结果的影响

由于用户兴趣动态特性的递减函数参数  $\lambda$  的存在,使得参数  $\lambda$  的优化值对最终推荐结果有很大的影响。因此,在进行进一步的实验之前,首先计算参数的最佳取值,本文使用平均绝对偏差 MAE 作为评价指标来评估推荐算法的性能。MAE 值越低,推荐的结果就越准确。设定目标用户相似最近邻数量为 30,并将调节参数  $\alpha$  的取值设置为 0.5。通过改变参数  $\lambda$  取值,可以得到评价本文所提出的推荐算法的 MAE 变化情况。结果如 (图 1)。

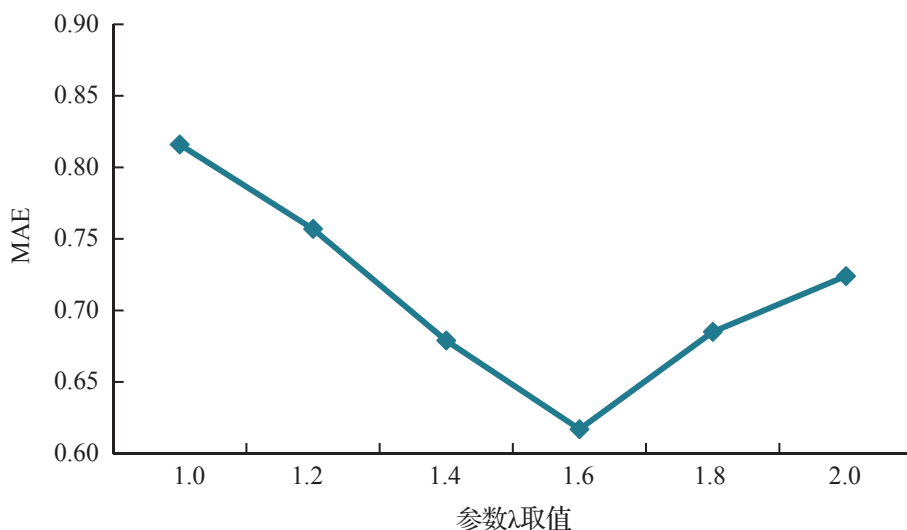


图 1 参数  $\lambda$  的取值对推荐结果的影响

从图 1 中可以看出当衰减率参数  $\lambda$  的取值为 1.6 时,本文所提出的算法性能最佳。但是,参数的最佳值可能不同于这个值,它应该根据特定的环境因素进行调整。影响推荐算法性能的主要因素之一在于数据集的特征。

#### 3.3 MAE 对比

为了验证本文提出的方法的性能以及推荐

效果,将该方法与与基于 K 最临近用户的协同过滤推荐方法<sup>[12]</sup>,记为 User-Based 和基于信任网络 and 用户评分的推荐方法<sup>[13]</sup>,记为 Trust-Based,这两种方法进行实验对比。将这三种方法在 Extended Epinions DataSet 数据集上实验结果进行比较,设定衰减率参数  $\lambda$  的取值为 1.6,并将最近邻 K 值设置为 {10,20,30,40,50,60},实验对比结果如 (图 2)。

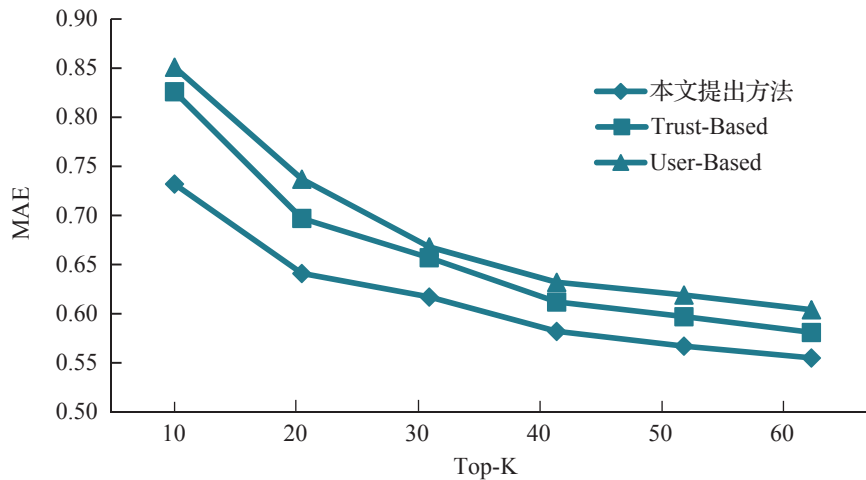


图 2 三种方法的 MAE 对比

在图 2 可以发现对于每种算法，在最近邻数不同的情况下可以发现，随着最近邻用户数的增加，三种推荐算法均得到更好的结果，即 MAE 的值呈下降趋势，但不是无限下降。它将作为一种稳定状态。此外，本文所提出的方法在预测准确度上优于其他两种方法。

#### 4 结束语

协同过滤推荐算法是一种应用比较广泛、较为成熟的推荐算法，具有很好的推荐效果。然而它在某些方面也存在不足，例如，传统的协同过滤算法在计算相似度时没有考虑用户兴趣的时间特性，因此会导致推荐精度的降低。目前，随着社会网络的兴起，社交网络用户激增，用户在社交网站上面临信息过载的问题。因此本文考虑社交网络用户的动态信息，提出改进的推荐算法实现个性化推荐服务。通过实验对比证实本文所提出的推荐算法具有更好的推荐效果。

#### 参考文献

- [1] Clough, G. Geolearners: Location-Based Informal Learning with Mobile and Social Technologies[J]. IEEE Transactions on Learning Technologies, 2010, 3(1): 33-44.
- [2] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[J]. Uncertainty in Artificial Intelligence, 1998, 98(7): 43-52.
- [3] 荣辉桂, 火生旭, 胡春华, 等. 基于用户相似度的协同过滤推荐算法[J]. 通信学报, 2014, 35(2): 16-24.
- [4] Lu Z, Dou Z, Lian J, et al. Content-Based Collaborative Filtering for News Topic Recommendation[C]. Proceedings of the twenty-Ninth AAAI Conference on Artificial Intelligence. Austin Texas, USA. 2015: 217-223.
- [5] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[J]. Uncertainty in Artificial Intelligence, 1998, 98(7): 43-52.
- [6] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]. International Conference on World Wide Web. ACM, 2001: 285-295.

- [7] 曾安, 谢杰民, 潘丹. 基于项目候选集的协同过滤算法[J]. 计算机应用研究, 2016, 33(12): 3620-3624.
- [8] 李晓静, 张晓滨. 基于LCS的用户时空行为兴趣相似性计算方法[J]. 计算机工程与应用, 2013, 49(20): 251-254.
- [9] 孙光福, 吴乐, 刘淇, 等. 基于时序行为的协同过滤推荐算法[J]. 软件学报, 2013, 24(11): 2721-2733.
- [10] 王曰芬, 贾新露, 傅柱. 学术社交网络用户内容使用行为研究——基于科学网热门博文的实证分析[J]. 现代图书情报技术, 2016(6): 63-72.
- [11] 唐晓波, 张昭. 基于混合图的在线社交网络个性化推荐系统研究[J]. 情报理论与实践, 2013, 36(2): 91-95.
- [12] 夏立新, 李重阳, 王忠义. 融合关系强度和兴趣的好友推荐方法研究[J]. 图书情报工作, 2017, 61(1): 64-71.
- [13] Dzugan N, Fannin L, Makki S K. A recommendation scheme utilizing Collaborative Filtering[C]. 2013 8th International Conference for Internet Technology and Secured Transactions, ICITST. 2013: 96-100.
- [14] Ma H, Yang H, Lyu M R, et al. SoRec: social recommendation using probabilistic matrix factorization[C]. Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM. 2008: 26-30.