

◎模式识别与人工智能◎

安全的半监督方法的协同过滤推荐算法

王玉业, 陈健美

WANG Yuye, CHEN Jianmei

江苏大学 计算机科学与通信工程学院, 江苏 镇江 212000

College of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212000, China

WANG Yuye, CHEN Jianmei. Safe semi-supervised collaborative filtering recommendation algorithm. Computer Engineering and Applications, 2018, 54(8):107-111.

Abstract: To overcome the data sparseness and ignoring items, user feature of traditional collaborative filtering algorithm which can lead to the decrease of efficiency of recommendation, this paper proposes a collaborative filtering recommendation algorithm based on safe and high confident semi-supervised method S4VM to predict the rating of unrated items effectively. Through predicting the rating of unrated items, the sparseness of data has been alleviated and the accuracy of searching nearest neighbor item has been improved simultaneously. The experimental results show that the algorithm can improve the recommended quality of the system.

Key words: collaborative filtering; recommendation system; Safe Semi-Supervised Support Vector Machine (S4VM); semi-supervised learning; confidence

摘 要:为解决传统协同过滤算法中用户评分数据稀疏性,忽视物品及用户特征,所带来的推荐质量下降的问题,提出了一种基于安全的、高置信度的半监督方法的协同过滤推荐算法,采用安全的、高置信度的半监督方法 S4VM 对没有评分的数据进行有效预测,同时考虑用户的行为信息以及物品及用户特征。通过对未评分数据进行预测,能够有效地缓解数据的稀疏性,从而提高寻找最近邻的准确度。实验结果表明,该算法能够有效地提高系统的推荐质量。

关键词:协同过滤;推荐系统;安全的半监督支持向量机(S4VM);半监督学习;置信度

文献标志码:A **中图分类号:**TP391 **doi:**10.3778/j.issn.1002-8331.1611-0140

1 引言

随着互联网和电子商务的高速发展,如何能够快速、准确地向不同用户推荐他们感兴趣的信息越来越成为很多研究者研究的焦点。个性化推荐就是解决该问题的研究方向,它通过主动挖掘出用户的兴趣爱好,为用户推送个性化的信息。

当前主流的推荐方法包括基于协同过滤、基于内容、基于知识、基于关联规则等等,其中最为成功的是基于协同过滤的算法。协同过滤的方法通过计算用户兴趣喜好的相似性,从而为用户筛选和推荐感兴趣的物品,它主要利用用户的行为信息进行推荐。

而协同过滤主要存在以下几个问题:

(1)数据稀疏性、冷启动问题:协同过滤算法主要是基于用户-项目评分矩阵,当该矩阵非常稀疏时,寻找最近邻的精确度的性能必然下降,尤其是当一个新用户进入系统时出现冷启动问题,在这种情况下,更难找到与新元素相关的数据。

(2)忽视物品、用户信息问题:协同过滤算法主要利用用户的行为信息进行推荐,忽略了物品、用户的信息,而这些信息对于提高推荐的精度有很大的帮助。

(3)可扩展性问题:随着大数据时代的到来,用户和项目的数量飞速增长,传统的协同过滤算法就会有十分严重的扩展性问题。

为了解决以上几个问题,很多研究者通过利用少量

作者简介:王玉业(1990—),女,硕士,研究方向为推荐系统、数据挖掘、机器学习等, E-mail: wyyxhy324@foxmail.com;陈健美(1962—),女,教授,研究方向为医学图像处理与分析、数据挖掘、模式识别、计算机医学应用等。

收稿日期:2016-11-08 **修回日期:**2017-01-12 **文章编号:**1002-8331(2018)08-0107-05

CNKI网络出版:2017-04-14, <http://kns.cnki.net/kcms/detail/11.2127.TP.20170414.1717.008.html>

的已标注的数据,利用机器学习的一些方法填充数据,缓解数据稀疏性,比如分类、聚类、降维等,但是这些方法都存在一个共有的问题,数据填充质量无法保证,即当标记样本很少时,只使用标记样本预测的精度不高,所以本文提出了基于安全的、高置信度的半监督方法的协同过滤推荐算法。首先半监督学习的目的就是当标记数据很少时,能够通过大量未标记的样本提高分类器的性能,而半监督学习中的直推学习(transductive learning)考虑的未标记样本是待预测的样本,这符合推荐系统中数据极度稀疏,数据填充质量无法保证的应用场景。其次,很多半监督的方法在某些情况下,性能会比只使用标记样本所得到的分类器效果更差,而所谓安全的半监督算法就是在任何情况下,都不会出现性能下降的情况,这就能够保证数据填充的质量。最后,为了更为有效地填充数据,降低噪声和可能的过拟合的影响,采用基于 k 近邻的高置信度方法对S4VM进行交叉验证,选取高质量的数据进行推荐,保证推荐的质量。基于上述分析,本文采用安全的、高置信度的S4VM,同时考虑用户的行为信息,以及物品和用户信息,对没有评分的数据进行准确的预测,从而更为有效地解决数据稀疏性、冷启动问题。

2 相关工作

传统的协同过滤算法主要分为基于用户和基于物品协同过滤算法^[1-2],或者同时基于两者的协同过滤算法^[3]。最近几年,研究者的研究焦点主要集中在解决数据稀疏性、冷启动问题、忽视物品、用户信息问题和可扩展性问题上,对于数据稀疏性、冷启动问题和忽视物品、用户信息问题,通常的处理方法主要分成两类。

(1)利用机器学习的算法对没有评分的数据进行预测,比如基于朴素贝叶斯(Naive Bayesian)^[4]、支持向量机(SVM)^[5]、神经网络(Neural Networks)^[6]、主题模型(Topic model)^[7]以及深度学习(Deep learning)^[8]协同过滤算法,这些基于机器学习中的分类方法来填充数据的算法,很大程度依赖于训练集的质量和数量,而在实际应用中,当标记样本很少时,只使用标记样本预测的精度不高,所以本文将半监督的方法引入到协同过滤中,并针对协同过滤算法存在的实际问题,对已有的半监督算法进行改进。

(2)利用用户兴趣喜好的相似性或者项目类别的相似性进行填充数据,比如对目标项目和邻居项目未共同评分的项目进行评分填充后再计算其相似性^[9],比如使用信任模型填充数据^[10],比如使用基于Dice Euclidean相似度计算相似性^[11],比如先对用户或者项目进行聚类,再计算相似性^[12]。这些方法和基于机器学习的方法存在相同的问题,需要大量的高质量的标注数据,此外这些方法没有同时考虑到用户的行为信息,以及项目和

用户信息。

此外,当前也有一些通过降维的方法来降低计算的复杂度,比如基于概率图模型的降维、隐含主题分析(PLSA)、潜在的狄里克雷分析(LDA)^[13]。比如基于矩阵分解的降维、奇异值分解(SVD)^[14]、核矩阵分解^[15],以及基于图的方法解决大规模数据^[16]。但是这些方法虽然降低了时间复杂度,但是也导致了一些信息的丢失,并且忽略了项目和用户信息。

基于上述对于最近的基于协同过滤的相关工作的优缺点分析,本文采用安全的、高置信度的半监督学习方法,同时考虑用户的行为信息,以及物品和用户信息,对没有评分的数据进行准确的预测,从而更为有效地来解决上述算法所存在的问题。

3 基于高置信度的S4VM的协同过滤算法

3.1 半监督学习

机器学习是人工智能领域重要的一个分支,在大数据时代发挥着极其重要的作用,机器学习主要包括两种学习策略:监督学习(Supervised Learning)和无监督学习(Unsupervised Learning)。监督学习是在训练过程中使用有标注的数据,然而获取有标注数据是困难且昂贵的。无监督学习则是利用无标注数据进行学习的方法,这种方法能够节省大量的人力和物力,无标注的数据也比较容易获取,但是由于缺乏正确的指导,往往准确率比较低。结合两种学习策略的半监督学习方法被提出,它是将少量有标注数据和大量无标注数据相结合,同时能够保证性能的学习框架,常用的半监督学习方法包括:自训练(self-training)、生成式模式(generative models)、协同训练(co-training)、基于图的方法(Graph-Based Methods)。半监督学习主要可以分为纯(pure)半监督学习和直推学习(transductive learning),前者假定训练数据中的未标记样本并非是待预测的数据,而后者则假定学习过程中所考虑的未标记样本恰是待预测数据,学习的目的就是在这些未标记样本上获得最优泛化性能^[17-18]。

3.2 S4VM

半监督学习的目的是,当标记样本很少时,只使用标记样本得到的分类器精度不高,希望通过探索未标记的样本来提高分类器的性能,这在很多情况下是有效的,但在某些情况下,很多半监督方法的性能会下降,甚至比只使用标记样本得到的效果更差,因此在使用无标注数据时,性能从不会下降的安全的学习方法是现在半监督学习的研究重点。S4VM就是一种安全的、直推的半监督学习方法,在使用未标记数据时,从来不会出现性能下降的情况。

S4VM是支持向量机(SVM)在半监督学习中的应用,在SVM的基础上引入了最优低密度思想,同时关注

多个可能的低密度分界线,在给定许多不同间隔(margin)较大的分界线时,优化未标记样本的类别划分,使得在最坏的情况下,相对于只使用标记样本的SVM的性能最大化,从而使得性能不会下降,因此是安全的半监督学习算法。

目标函数为:

$$h(f, \hat{y}) = \frac{\|f\|_H}{2} + c_1 \sum_{i=1}^I l(y_i, f(x_i)) + c_2 \sum_{j=1}^J l(\hat{y}_j, f(\hat{x}_j)) \quad (1)$$

S4VM算法的目标是寻找多个大间隔(margin)较大的低密度分界线 $\{f_i\}_{i=1}^T$, 以及对应的类别划分 $\{\hat{y}_i\}_{i=1}^T$, 因此使得下面函数最小化。

$$\min_{\{f, \hat{y}_i\} \in \beta_{i=1}^T} \sum_{i=1}^T h(f, \hat{y}_i) + M\Omega\left(\{\hat{y}_i\}_{i=1}^T\right) \quad (2)$$

其中 T 是低密度分界线的数量, Ω 是一个对分界线差异化进行度量的惩罚函数, M 是个大常数来保证分界线的差异性^[19]。

3.3 基于 K 近邻的置信度

在机器学习中,置信度(confidence)指的是机器学习算法在各种测试集上预测性能的稳定性,它是分类模型对未知数据预测错误的概率值,即模型预测的风险水平。关于置信度的度量方法主要包括基于贝叶斯的方法、基于数据分布的EM方法、基于 K 近邻的置信度等。在本文中,采用基于 K 近邻的置信度的度量方法,该方法必须遵循以下两个标准:(1)相似度高的无标签数据一定属于相同的类别;(2)无标签数据的类别一定与相似度高的有标签数据的类别一致。

定义1 无标注数据 x_i^U 被分类器 h 标注为某个类的置信度为 $\text{confi}(x_i^U)$, 它由 K 个无标注的近邻的相似性 A_i 和 K 个有标注的近邻的相似性 B_i 共同决定,计算公式如下:

$$\text{confi}(x_i^U) = \lambda A_i + (1 - \lambda) B_i \quad (3)$$

$$A_i = \sum_{j=1}^K \text{sim}(x_i^U, x_j^L) I(y_i^U, y_j^L) \quad (4)$$

$$B_i = \sum_{j=1}^K \text{sim}(x_i^U, x_j^U) I(y_i^U, y_j^U) \quad (5)$$

其中 $\text{sim}(x_i^U, x_j^L)$ 表示无标注数据 x_i^U 与有标注数据 x_j^L 的相似度, I 是指示函数,如果分类器 h 对于 x_i^U 的标注结果 y_i^U 与 y_j^L 相同, $I(y_i^U, y_j^L) = 1$, 否则, $I(y_i^U, y_j^L) = 0$, $\lambda \in [0, 1]$, 表示无标注的近邻的相似性 A_i 和 K 个有标注的近邻的相似性 B_i 在计算置信度中的权重。

在公式(3)中, $\text{confi}(x_i^U)$ 越大,表明 x_i^U 的标注结果与最相似的 K 个无标注数据、 K 个有标注数据越一致, x_i^U 置信度越高, $\text{confi}(x_i^U)$ 越小,表明 x_i^U 的标注结果与

最相似的 K 个无标注数据、 K 个有标注数据越不一致, x_i^U 置信度越低^[20]。

3.4 高置信度 S4VM

为了能够有效解决数据稀疏性、冷启动问题和忽视物品、用户信息问题,本文采用高置信度的 S4VM 来预测未打分数据。通过前面的分析, S4VM 是一种安全的、直推式的半监督方法,它的性能是优于只使用标记样本的有监督的方法,因此,基于 S4VM 填充数据的协同过滤算法的性能理论上应当优于之前基于朴素贝叶斯、SVM、相似性等协同过滤的算法,同时使用基于 K 近邻的置信度的方法对 S4VM 预测的结果进行交叉验证,使用高质量的填充数据进行推荐。此外,之前的填充算法通常是使用物品、用户信息作为特征,或者将用户的兴趣喜好作为特征,本文将两种特征相结合,将物品、用户信息作为 S4VM 的特征,将用户的兴趣喜好作为置信度计算的特征,提出高置信度的 S4VM 算法。具体算法如下:

算法1 高置信度 S4VM 的评分预测算法

输入:用户-项目评分矩阵 R , 项目集 $I = \{i_1, i_2, \dots, i_M\}$,

用户集 $U = \{u_1, u_2, \dots, u_N\}$, 评分标签 $I = \{c_1, c_2, \dots, c_K\}$ 。

输出:填充后的用户-项目评分矩阵 R 。

预处理数据:从训练样本中抽取出来未打分的数据,并将这些未标注数据随机分割成 L 个数据集, $D^U = \{d_1^U, d_2^U, \dots, d_L^U\}$, 将有标注的数据构建成初始化数据集 D^A 。

For(D^U 中每个未标注数据集)

(1)构建用户和项目属性特征:对于未标注数据集 d_1^U 和标注数据集 D^A , 分别从用户属性和项目属性中选择 m, n 项属性组成 S4VM 特征:

$$F^S = \{A_{u,1}, A_{u,2}, \dots, A_{u,m}, A_{i,1}, A_{i,2}, \dots, A_{i,n}\}$$

(2)评分预测:将未标注数据集 d_1^U 的特征,标注数据集 D^A 的特征和标记共同作为 S4VM 算法的训练集, S4VM 算法输出未标注数据的评分。

(3)构建用户和项目行为特征:将用户喜好向量 $v_u = \{c_1, c_2, \dots, c_M\}$ 和项目关注度向量 $v_i = \{c_1, c_2, \dots, c_N\}$ 相结合,构建用户和项目行为特征:

$$F^C = \{c_{u,1}, c_{u,2}, \dots, c_{u,M}, c_{i,1}, c_{i,2}, \dots, c_{i,N}\}$$

(4)扩充标注数据集:对于步骤(2)得到未标注数据的评分,通过步骤(3)所得的用户和项目行为特征,根据公式(3)计算这些评分的置信度,将高置信度的加上标记的数据填充到标注数据集中。

3.5 计算推荐集

在得到填充的用户-项目评分矩阵 R 后,按照以下算法计算推荐集。

算法2 基于用户的协同过滤算法

输入:用户-项目评分矩阵 R , 项目集 $I = \{i_1, i_2, \dots, i_M\}$,

用户集 $U = \{u_1, u_2, \dots, u_N\}$, 评分标签 $I = \{c_1, c_2, \dots, c_K\}$ 。

输出:推荐集 I_{new} 。

(1)用户相似性计算:利用相似性计算方法计算用户之间的相似性并将结果保存在用户相似性矩阵 $R_{\text{sim}}(N \times N)$, 计算相似性的方法主要有 Pearson 相似度、Cosine 相似度和修正的 Cosine 相似度, 他们的计算公式分别如下所示。

Pearson 相似度:

$$\text{sim}(i, j) = \frac{\sum_{c \in C_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in C_{ij}} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in C_{ij}} (R_{j,c} - \bar{R}_j)^2}} \quad (6)$$

Cosine 相似度:

$$\text{sim}(i, j) = \frac{i \cdot j}{|i| \cdot |j|} \quad (7)$$

修正的 Cosine 相似度:

$$\text{sim}(i, j) = \frac{\sum_{c \in C_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in C_i} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in C_j} (R_{j,c} - \bar{R}_j)^2}} \quad (8)$$

其中, C_{ij} 表示用户 i 和用户 j 共同评价的项目集合, $R_{i,c}$ 和 $R_{j,c}$ 分别表示用户 i, j 对项目 c 的评分, \bar{R}_i 和 \bar{R}_j 分别表示用户 i, j 的评价评分, C_i 表示用户 i 评价的项目集合, C_j 表示用户 j 评价的项目集合^[21]。在实验中, 分别使用三种计算相似性的方法, 并且比较性能的差异, 性能最好的是 Pearson 相似度。

(2)计算用户 u 未评分的项目集 $N_u = I - I_u$, I 为项目集, I_u 为用户已经评分的项目集。

(3)根据相似性矩阵 $R_{\text{sim}}(N \times N)$, 选择用户 u 的 K 个近邻。

(4)对于用户 u 未评分的项目集 N_u 中的每一项 i , 利用以下公式计算用户对项目的评分值。

$$P_{u,i} = \bar{R}_u + \frac{\sum_{u_k \in U} \text{sim}(u, u_k)(R_{u_k,i} - \bar{R}_{u_k})}{\sum_{u_k \in U} (|\text{sim}(u, u_k)|)} \quad (9)$$

其中 $\text{sim}(u, u_k)$ 表示用户 u 与其最近邻居 u_k 之间的相似性, $R_{u_k,i}$ 表示用户 u_k 对项目 i 的评分, \bar{R}_u 表示用户 u 对项目的平均评分, \bar{R}_{u_k} 表示用户 u_k 对项目的平均评分^[9]。

(5)将项目集 N_u 中的元素的预测评分按照从小到大顺序, 形成推荐集 $I_{\text{new}} = \{i_1, i_2, \dots, i_p\}$, 推荐给目标用户。

4 实验

4.1 实验数据集及度量标准

数据集采用的是 Movielens 数据集, 该数据集是由明尼苏达大学 GroupLens 研究小组通过 Movielens 收集, 包含了 943 位注册用户对 1 682 部电影的 10 000 条 1~5 分的评分数据, 每位用户至少需要对 20 部电影进行打分。为了简化问题, 将所有电影都转化为二类, 即喜欢和不喜欢, 类标签表示为 +1 和 -1, 将评分值为 4、5 的标定位 +1, 将评分值为 1~3 的标定位 0, 对于公式(5)得

到的预测评分值, 预测评分大于 0.6 的设为 +1, 预测评分小于 0.6 的设为 0, 采用的数据集是按照 80% 和 20% 的比例分割成的 base 训练集和 test 测试集, 然后进行 5-折交叉实验并取 5 组数据的平均值进行验证, 用户评分数据集的稀疏性为 0.937 0。

从用户属性中选择 Gender、Age、Occupation 作为特征属性值, 从项目属性中选择 Release Date、Category 作为特征属性, 组成特征 $A = \{\text{Gender}, \text{Age}, \text{Occupation}, \text{Release Date}, \text{Category}\}$ 。

采用平均绝对误差 (Mean Absolute Error, MAE) 评价算法的效果, 通过计算预测用户评分与实际的用户评分之间的差异来衡量预测结果的准确性, 设预测用户的评分集合表示为 p_1, p_2, \dots, p_n , 真实用户的评分作为测试集, 集合表示为 q_1, q_2, \dots, q_n , 则平均绝对误差定义为:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |p_i - q_i| \quad (i = 1, 2, \dots, n) \quad (10)$$

MAE 越小, 预测准确率越高, 推荐质量越高。

4.2 实验结果及分析

(1) MAE 性能比较

为了比较本文算法 (Proposed-CF) 的性能, 选择基于用户的协同过滤算法 (User-CF)、基于项目分类的改进协同过滤算法 (C-Item-CF)、基于朴素贝叶斯的协同过滤算法 (Bayesian-CF)、基于支持向量机的推荐算法 (SVM) 作为对比算法, 并且比较不同方法在不同邻居个数下的性能。如图 1 所示。

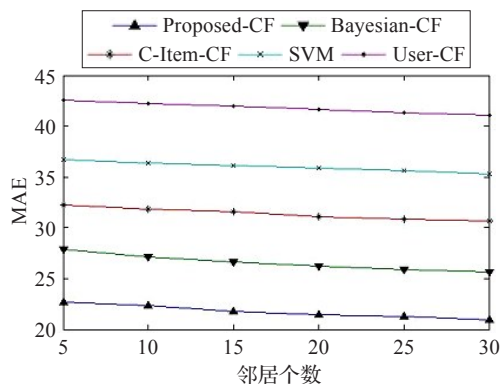


图1 推荐算法的 MAE 比较

由图 1 可知, 与对比算法相比, 提出算法的性能最好, MAE 值最小, 基于朴素贝叶斯的协同过滤方法, 是先用朴素贝叶斯的方法填充数据, 然后再使用基于用户的协同过滤进行预测, 该方法与本文方法类似, 但是仅仅使用少量的有标注数据训练分类器, 为大量的未标注数据添加标记, 分类器的性能无法保证, 最终会影响推荐的性能。基于项目分类的改进协同过滤算法, 首先利用项目分类信息为类内未评分项目预测评分值, 然后通过计算类内用户间的相似度得到目标用户的最近邻居, 最后进行推荐, 但该方法忽视了用户和项目的属性, 这是能够有效地进行推荐的重要信息。基于支持向量机

的推荐算法,是将推荐问题当成一个分类问题,但是标注数据数量有限,因此分类的性能无法保证。基于用户的协同过滤算法,是最基本的协同过滤算法,存在严重的数据稀疏性问题,因此严重影响推荐性能,而本文提出的算法综合考虑了之前算法所存在的问题。第一,综合考虑用户和项目的属性来填充数据,缓解数据稀疏性问题。第二,使用高置信度的S4VM的分类方法,该方法是安全的半监督方法,它的性能必然是优于普通的监督方法,因此,本文方法的性能有了显著的提升。

(2)高置信度S4VM中K的影响

为了研究高置信度S4VM中基于K近邻置信度中K的影响,将K分别设为{5,10,15,20,25},同时比较在不同邻居个数下MAE的差异。如图2所示。

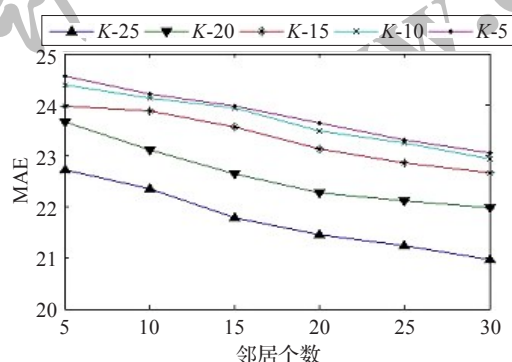


图2 不同K,不同近邻的MAE比较

由图2可知,Proposed-CF在K=25,近邻个数30下MAE值最小,从而分析可知,提出的算法(基于高置信度的协同过滤算法)在K值越大,邻居个数越多的情况下,性能越好。

5 结语

针对传统协同过滤算法存在的数据稀疏性问题,以及常见的一些方法填充数据质量无法保证的问题,本文提出一种安全、高效的半监督的方法——高置信度的S4VM,既能够有效地缓解数据稀疏性问题,又能够保证填充数据的质量,同时充分考虑用户、项目特征属性、用户行为对于推荐的影响。实验结果也表明,在用户评分数据极度稀疏的情况下,本文的方法明显优于其他方法,有效提高了推荐的质量。

参考文献:

- [1] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[C]// Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, 1998: 43-52.
- [2] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the 10th International Conference on World Wide Web, 2001: 285-295.

- [3] Wang J, De Vries A P, Reinders M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion[C]// Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006: 501-508.
- [4] 李大学, 谢名亮, 赵学斌. 基于朴素贝叶斯方法的协同过滤推荐算法[J]. 计算机应用, 2010, 30(6): 1523-1526.
- [5] 吕成成, 王维国, 丁永健. 基于KNN-SVM的混合协同过滤推荐算法[J]. 计算机应用研究, 2012, 29(5): 1707-1709.
- [6] 陈丹儿, 应玉龙. 基于项目属性和BP神经网络的协同过滤推荐[J]. 信息技术, 2015, 39(3): 70-73.
- [7] Liu N, Lu Y, Tang X J, et al. Improved collaborative filtering algorithm using topic model[C]// International Conference on Parallel & Distributed Computing, 2017.
- [8] Devooght R, Bersini H. Collaborative filtering with recurrent neural networks[J]. arXiv Preprint arXiv:1608.07400, 2016.
- [9] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9).
- [10] 王立印, 张辉, 陈勇. 一种基于Dice-Euclidean相似度计算的协同过滤算法[J]. 计算机应用研究, 2015, 32(10): 2891.
- [11] 杨兴耀, 于炯, 吐尔根. 基于信任模型填充的协同过滤推荐模型[J]. 计算机工程, 2015, 41(5): 6-13.
- [12] 熊忠阳, 刘芹, 张玉芳, 等. 基于项目分类的协同过滤改进算法[J]. 计算机应用研究, 2012, 29(2).
- [13] Hofmann T. Latent semantic models for collaborative filtering[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(1): 89-115.
- [14] 杨阳, 向阳, 熊磊. 基于矩阵分解与用户近邻模型的协同过滤推荐算法[J]. 计算机应用, 2012, 32(2): 395-398.
- [15] Rao N, Yu H F, Ravikumar P K. Collaborative filtering with graph information: Consistency and scalable methods[C]// Advances in Neural Information Processing Systems, 2015: 2107-2115.
- [16] Liu X, Aggarwal C, Li Y F, et al. Kernelized matrix factorization for collaborative filtering[C]// SIAM Conference on Data Mining, 2016: 399-416.
- [17] Chapelle O, Scholkopf B, Zien A. Semi-supervised learning (Chapelle O, et al., eds.; 2006) [Book reviews][J]. IEEE Transactions on Neural Networks, 2009, 20(3).
- [18] 周志华. 基于分歧的半监督学习[J]. 自动化学报, 2013, 39(11): 1871-1878.
- [19] Li Y F, Zhou Z H. Towards making unlabeled data never hurt[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 175-188.
- [20] Zhang Y, Wen J, Tang F, et al. Semi-supervised learning: Exploiting unlabeled data with symmetrical distribution and high confidence[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2012, 26(7).
- [21] 吴月萍, 郑建国. 改进相似性度量方法的协同过滤推荐算法[J]. 计算机应用与软件, 2011, 28(10): 7-8.