

基于码本聚类 and 因子分解机的多指标推荐算法

丁永刚^{1,2} 李石君¹ 余伟¹ 王俊¹

(武汉大学计算机学院 武汉 430072)¹ (湖北大学教育学院 武汉 430062)²

摘要 传统的协同过滤推荐算法普遍存在数据稀疏问题,且仅利用单一综合评分来计算用户相似度,无法找到在多个指标上偏好相似的用户,因而影响推荐的准确度。多指标评分推荐算法力图寻找在多个指标上偏好相似的用户,但是其评价成本高,导致数据稀疏性问题更加严重。为了找到与目标用户在多个指标上偏好相似的用户,提出基于码本聚类的思想来获取用户在各指标上的评分风格信息,然后基于评分风格信息将用户和项目在各指标上进行双向聚类,最后利用因子分解机模型(Factorization Machines, FMs)基于同一簇内的用户、项目、多指标评分信息、评分风格信息进行推荐。实验结果表明,与传统的协同过滤算法和其他多指标推荐方法相比,基于多指标评分信息的因子分解机推荐算法能够在一定程度上缓解数据稀疏问题,提高推荐的准确度。

关键词 用户偏好,多指标评分,码本聚类,因子分解机

中图分类号 TP393 文献标识码 A DOI 10.11896/j.issn.1002-137X.2017.10.034

Multi-criteria Recommendation Algorithm Based on Codebook-clustering and Factorization Machines

DING Yong-gang^{1,2} LI Shi-jun¹ YU Wei¹ WANG Jun¹

(School of Computer, Wuhan University, Wuhan 430072, China)¹

(School of Education, Hubei University, Wuhan 430062, China)²

Abstract The sparsity of user-item ratings is a common problem and the users who share similar preferences on multi-criteria cannot be found by only making use of a single overall rating to calculate the similarity of users in traditional collaborative filtering algorithm, which would affect the accuracy of recommendation. Multi-criteria recommendation algorithm tries to find users who share similar preferences on multi-criteria, but the problem of data sparsity become even worse owing to the high cost of rating. Aim at these problems, we proposed an algorithm which first obtains the information of rating style of users based on the idea of codebook-clustering, and then conducts co-clustering for users and items on each criteria. Finally, this algorithm makes recommendations by factorization machines(FMs) based on users, items, multi-criteria ratings and rating style. The experimental result shows that multi-criteria recommendation algorithm based on codebook-clustering and FMs is able to solve the problem of data sparsity to some extent, thus improving the accuracy of recommendation.

Keywords User preference, Multi-criteria ratings, Codebook-clustering, Factorization machines(FMs)

1 引言

推荐系统作为解决信息过载问题的有效手段,已在电子商务等领域^[1-3]得到广泛应用。传统的协同过滤推荐算法仅利用单一的综合评分对用户偏好进行刻画,其基本思想是通过共同评分项目的单一综合评分来计算用户间的相似度,找到与目标用户相似的其他用户,然后利用相似用户评分对目标用户评分进行预测并推荐。然而在评分数据稀疏的情况下,共同评分的项目所占比重较小,不能准确找到偏好相似的用户;同时,综合评分只能刻画用户喜欢一个商品的程度,而不能刻画用户喜欢商品的原因。但是在某些应用场景下,往

往需要考虑项目的多个指标评分才能了解用户的更详细的偏好^[4]。例如,两个用户A和B对某酒店的评分都为5,但A用户可能对酒店便捷的地理位置比较满意,而B用户则可能对酒店提供的服务比较满意。因此,为了对用户的偏好进行更细致的刻画以提高推荐结果的准确度,下一代推荐系统^[5]致力于获取用户对项目不同指标的评分信息,并将其加以利用来对目标用户做出推荐。

现有的多指标推荐系统大多通过扩展传统的协同过滤算法来利用多指标评分,然而由于多指标评价的成本较高,导致多指标评分的数据稀疏问题更加严重。这使得要找到在多指标评分上偏好相似的用户更加困难,从而影响推荐的准确度。

到稿日期:2016-08-05 返修日期:2016-12-12 本文受国家自然科学基金项目(61502350),湖北省自然科学基金项目(2014CFB289)资助。

丁永刚(1966—),女,博士,副教授,主要研究方向为数据挖掘、机器学习, E-mail: hddyg@hubu.edu.cn; 李石君(1964—),男,博士,教授,CCF会员,主要研究方向为大数据、互联网搜索与挖掘等, E-mail: shjli@whu.edu.cn; 余伟(1987—),男,博士,讲师,主要研究方向为数据质量评估、数据抽取与数据融合, E-mail: yuwei@whu.edu.cn; 王俊(1986—),男,博士,主要研究方向为数据挖掘、推荐系统, E-mail: wjwj@whu.edu.cn。

鉴于此,本文基于码本聚类的思想,从用户-项目评分矩阵中提取用户和项目在各指标上的评分/被评分风格信息,然后基于评分/被评分风格信息进行聚类,在聚类后使得同一簇内的用户对该簇内的项目具有相同或相似的指标偏好,并利用因子分解机模型(FMs)^[6-7]能够方便有效地集成附加信息的特点,将用户对项目的多指标评分信息、基于码本聚类的评分/被评分风格信息集成到 FMs 模型中,以期缓解数据稀疏问题,提高推荐的准确度。

2 相关工作

在推荐算法中利用多指评分能够更细致地刻画用户偏好,为用户提供更准确的推荐结果。现有的多指标推荐算法主要分为两类:启发式方法和基于模型的方法(包括聚合函数、概率模型和多维奇异值分解)。

启发式方法主要通过扩展传统的单一综合评分相似度计算方法来获得近邻用户以进行推荐。文献[8]提出了两种相似度计算扩展方法:1)利用 Cosine、皮尔逊相关系数等传统相似性度量方法在每一维指标上计算相似度,然后使用平均相似度或最小相似度来描述用户的整体偏好相似度;2)将多指标评分作为向量,将通过计算得到的向量间的欧几里得(Euclid)距离的逆(相似性越高,距离越小)作为用户的相似度量。文献[8]提出了3种相似度计算方法来寻找近邻用户,即分别基于每一指标上的权重(而不是评分)、每一指标上的非权重评分和每一指标上的权重评分来计算单个指标的相似度。这些方法简单直观地将多指标推荐问题转化为单指标推荐问题,但忽略了不同指标之间的评分数据存在一定关系的事实。针对该问题,文献[9-11]分析了综合评分和多指标评分之间的相关关系,提出利用多元线性回归模型^[9]、偏最小二乘回归算法^[10]和基因遗传算法^[11]来确定每一维指标评分的权重,并构建基于用户综合评分的聚合函数。但是多元线性回归方法在评分记录数少于指标数的情况下无法确定权重和在数据稀疏的情况下容易出现过拟合的问题,因此文献[12]提出使用支持向量回归方法来构建基于用户和基于商品的综合评分聚合函数,并利用随机梯度下降法训练得到的参数将这两个不同的评分值进行综合,以作为最终预测评分。与单一指标协同过滤算法相比,以上方法均不同程度地提高了推荐的准确度,但由于多指标评分数据的稀疏性问题仍未得到很好地解决,因此推荐的效果有待进一步提高。

聚类方法也是用来确定与目标用户相似的近邻用户的一种有效方法。文献[13]认为用户是否购买商品仅依据少数指标偏好,因此它根据回归算法确定这些指标偏好,并利用哈斯图(Hasse Diagram)进行聚类,然后使用同一簇内或具有相似指标偏好簇的评分信息进行推荐。文献[14-16]提出使用概率生成模型(Probabilistic Generative Model)来挖掘不同指标之间的相关关系以对用户和商品同时聚类,并通过估计得到的模型参数来预测目标用户对商品的评分。文献[17]提出利用相关-集合数据模型(Relevant-set Correlation Model)对用户与商品进行双向聚类,然后在每个簇内使用多指标协同过滤推荐算法产生推荐结果。文献[18]提出利用多维奇异值分解模型(Multi-linear Singular Value Decomposition Model,

MSDV)来挖掘用户、商品和上下文信息之间的关系,并利用训练好的模型为用户推荐餐馆。但 MSDV 计算复杂,导致推荐效率不高。本文的研究与文献[17-18]最相似,但是本文采用基于码本聚类的思想对评分风格相似的用户和被评分风格相似的项目进行双向聚类,并在聚类的基础上提出使用因子分解机模型(FMs)进行推荐的方法。实验结果表明基于 FMs 的多指标推荐方法不仅在线性的时间复杂度内产生高效的推荐,而且在数据稀疏的情况下也能得到较高的推荐准确度。

3 基于 FMs 的多指标推荐算法

3.1 基本定义

定义 1(用户集合) $U = \{u_m | m \in M\}$, U 表示用户集合,其中 M 表示用户 ID 集合。

定义 2(项目集合) $I = \{i_n | n \in N\}$, I 表示项目集合,其中 N 为项目 ID 集合。

定义 3(多指标评分矩阵) $R = \{\vec{r}_{u,i}\}$, R 表示用户对项目的多指标评分矩阵,其中 $\vec{r}_{u,i} = [r_{u,i_1}, r_{u,i_2}, \dots, r_{u,i_k}]$ 表示用户 u 对项目 i 的 k 个指标评分,评分值通常为 1~5 的整数。

定义 4(多指标推荐) 一个多指标推荐问题的效用函数 R 定义为:

$$R: U \times I \rightarrow r_0 \times r_1 \times \dots \times r_k \quad (1)$$

其中, r_0 是用户对项目的综合评分, r_1, \dots, r_k 是用户对项目的多个指标评分,且 r_0 和 r_1, \dots, r_k 之间存在以下关系:

$$r_0 = f(r_1, \dots, r_k) \quad (2)$$

3.2 基于码本聚类的用户评分风格和商品被评分风格表示

码本聚类算法主要应用于用户图像背景建模,其基本思想是使用一个码本(CodeBook, CB)结构来描述一个像素 P ,而每个码本中包含若干码元(Code Element, CE),这些码元就是该像素点 P 的一个聚类。码本聚类算法旨在构建像素的每个聚类,即码本。

传统的协同过滤算法基于项目评级来计算用户或项目之间的相似性,实际上,如果将用户-项目评分矩阵中用户和项目的信息重新构建,便可以得到用户和项目的另一种表示形式。文献[19]认为用户-项目矩阵中隐含着不同的用户评分风格(Rating Style)或项目的被评分风格(Rated Style)信息,而这种评分风格或被评分风格信息隐式地反映了用户或商品之间的相似性。例如,如果某用户总是对某些电影给出较高评分,则隐含着这些被评为高分的电影是相似的;类似地,如果某电影总是被某些用户给出较高评分,则隐含着这些用户是相似的;反之亦然。因此可以把这些具有相似性的用户或项目集聚到同一簇内。进一步地,文献[19]认为用户的评分风格或项目的被评分风格可以通过用户的评分频率和项目的被评分频率来表示,因此它首先使用规范化的评分直方图来表示评分或被评分频率,然后基于码本聚类的思想来表示用户和项目,即采用用户对评分项目的不同评分频率来表示用户,以及用项目的不同被评分频率来表示项目,并使用 C-means 聚类算法将相似用户或相似项目集聚到同一簇内。由于使用了规范化的评分直方图来表示用户或项目,因此没有共同评分但有相同或相似评分频率的用户或项目也有可能进行

相似度计算,从而可以在一定程度上缓解数据稀疏问题。

文献[20]基于用户对项目的综合评分并使用规范化的评分直方图来表示用户和项目。在多指标评分的情况下,为了更准确地刻画用户偏好,对其进行扩展,即在多个指标上对用户评分风格和项目的被评分风格进行刻画。具体说来,如果一个或多个用户对多个项目在相同指标上给出了相同或相似的不同评分频率,则这些项目极有可能是相似项目,应该将这些项目集聚到同一簇;同样地,如果一个或多个项目在相同指标上被多个用户给出了相同或相似的不同评分频率,则这些用户极有可能是相似用户,应该将这些用户集聚到同一类。据此,可以把用户 u_m 在项目指标 c 上的评分风格表示为以下向量:

$$u_{m,c} = [u_{m,c,1}, u_{m,c,2}, \dots, u_{m,c,p}] \quad (3)$$

其中, p 是用户在指标 c ($c=1, \dots, k$) 上的评分上限, $u_{m,c,r} = \sum_{i=1}^n \# r_{u_m, i, c}$ ($r=1, \dots, p$) 是用户 u_m 在所有项目的指标 c 上评分值为 r 的个数。同样地,项目 i_n 在项目指标 c 上的被评分风格表示为:

$$i_{n,c} = [i_{n,c,1}, i_{n,c,2}, \dots, i_{n,c,p}] \quad (4)$$

其中, p 是项目在指标 c ($c=1, \dots, k$) 上的被评分上限, $i_{n,c,r} = \sum_{u=1}^m \# r_{i_n, u, c}$ ($r=1, \dots, p$) 是项目 i_n 被所有用户在指标 c 上评分值为 r 的个数。

基于以上用户和项目表示,可以使用 K-means 聚类算法对用户或项目进行聚类,然后使用同一簇内的用户或项目进行预测评分。

3.3 融合多指标评分/被评分风格信息的 FMs 评分预测

3.3.1 因子分解机(FMs)

因子分解机是文献[6]提出的一种通用分解模型。与经典的分解技术中用户与项目的交互使用矩阵表示不同,因子分解机模型使用真值特征向量来表示用户与项目的交互。假设评分预测问题的数据由元组 (x, y) 的集合 S 表示,这里 $x = (x_1, \dots, x_n) \in R^n$ 是一个 n 维特征向量,则 FMs 能够使用分解交互参数对 x 的 i 个输入变量的所有嵌套交互进行 d 维度建模。当 $d=2$ 时,因子分解机模型可以表示为:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n w_{i,j} x_i x_j \quad (5)$$

其中, w_0 是全偏量, w_i 是输入变量 x_i 的一元交互参数, $w_{i,j}$ 是 v_i 和 v_j 之间的分解参数,定义为:

$$w_{i,j} = \langle v_i \cdot v_j \rangle = \sum_{f=1}^k v_{i,f} \cdot v_{j,f} \quad (6)$$

其中, k 是一个定义分解维度的超参数。

因子分解机 FMs 的一个重要特点是输入特征向量成对交互的效果可以用一个低秩矩阵来表示,因此输入向量之间的交互不使用一个独立参数 $w_{i,j}$ 来建模,而使用分解参数 $\langle v_i \cdot v_j \rangle$ 来建模。该特点使得 FMs 即使在数据稀疏的情况下也能对高维交互参数进行可靠估计,并且能够在 $O(km(x))$ ($m(x)$ 是向量 x 中非 0 元素的个数) 的线性时间内进行有效计算。

因子分解机的模型参数可以通过 3 种不同的方法来学习:随机梯度下降法(SGD)、最小二乘法(ALS)和马尔科夫链蒙特卡洛方法(MCMC)。由于 MCMC 方法具有快速学习的

特点,因此本文采用该方法来学习模型参数。

3.3.2 输入特征向量的构造

因子分解机 FMs 可以将传统的预测问题表示成效用函数: $R: U \times I \rightarrow R_0$, 而用户和项目的每一对交互 $(u, i) \in U \times I$ 都可以用一个二元特征向量 $x \in R^{U+I}$ 来表示,即用户 u 对项目 i 的评分可以用特征向量 x 表示:

$$x = (0, \dots, 0, 1, 0, \dots, 0, 0, \dots, 0, 1, 0, \dots, 0) \quad (7)$$

其中,非 0 元素对应于用户 u 和项目 i 。该特征向量也可以简化地表示为 $x(u, i) = \{(u, 1), (i, 1)\}$ 。

因子分解机 FMs 的最大优点是能够通过构建真值特征向量来集成各种附加信息,因此可以基于式(7)把用户和商品的多指标评分信息、基于码本聚类的用户评分风格信息和商品的被评分风格信息集成到 FMs 中。本文提出以下两种构造方法。

(1)分别集成单指标评分风格和被评分风格信息的输入特征向量构造方法(Sub-Criteria Rating Style, SCRS)

该方法实际上是将 k 维的多指标评分空间转换为 k 个单一指标评分空间,并将单个评分风格和被评分风格信息集成到 FMs 中进行各单一指标评分预测。因此目标函数可以表示为:

$$R: U \times I_c \times J_{u,c} \times J_{i,c} \rightarrow R_c \quad (8)$$

其中, I_c 是项目 I 在单一指标 c 上的评分空间, $J_{u,c}$ 为用户 u 在单一指标 c 上的评分风格信息, $J_{i,c}$ 为项目 i 在单一指标 c 上的被评分风格信息, R_c 为单一指标评分空间。

相应地,该效用函数的输入特征向量构造如下:

$$x(u, i) = \{(u, 1), (i, 1), (u_{c,r}, 1), (i_{c,r}, 1) : r=1, \dots, p\} \quad (9)$$

其中, $u_{c,r}$ 是用户 u 在指标 c 上的评分值为 r 的数目, $i_{c,r}$ 为项目 i 在指标 c 上的被评分值为 r 的数目。

通过利用以上方法构造特征向量预测得到的是每个指标上的评分,将 FMs 运行 k 次,即可得到 k 个指标评分,若要得到综合评分,则还需要通过线性回归训练得到每个用户在不同指标上的权重,然后将预测得到的单指标评分加权求和,即:

$$R_0' = w_1 R_1' + \dots + w_k R_k' + e \quad (10)$$

其中, w_i 是通过线性回归得到的用户在单个指标上的权重, R_c' 是通过式(6)得到的单个指标评分, e 是误差项目。

(2)集成多指标评分、评分风格信息和被评分风格信息的输入特征向量构造方法(Multi-Criteria Rating Style, MCRS)

由于综合评分与多指标评分之间存在一定关系,因此这种构造方法将多指标评分信息集成到 FMs。同时,将用户的多指标评分风格信息和项目的多指标被评分风格信息集成到 FMs 中。因此目标函数为:

$$R: U \times I \times R_c \times J_{u,c} \times J_{i,c} \rightarrow R_0 \quad (11)$$

其中, R_c 为多指标评分空间, R_0 为综合指标评分空间, $J_{u,c}$ 为用户 u 在 k 个指标上的评分风格信息, $J_{i,c}$ 为项目 i 在 k 个指标上的被评分风格信息, $c=1, \dots, k$ 。

相应地,该效用函数的输入特征向量构造如下:

$$x(u, i) = \{(u, 1), (i, 1), (u_{c,r}, 1), (i_{c,r}, 1) : c=1, \dots, k, r=1, \dots, p\} \quad (12)$$

其中, $u_{c,r}$ 是用户 u 在指标 c 上的评分值为 r 的数目, $i_{c,r}$ 为项目 i 在指标 c 上的被评分值为 r 的数目。

图 1 给出了上述两种输入特征向量构造方法的一个示例。图 1(a) 构造了一个集成单指标评分、评分风格信息和被评分

风格信息的真值特征向量,把单指标评分作为特征向量的输出;图 1(b) 构造了一个集成多指标评分、评分风格信息和被评分风格信息的真值特征向量,把综合评分作为特征向量的输出。

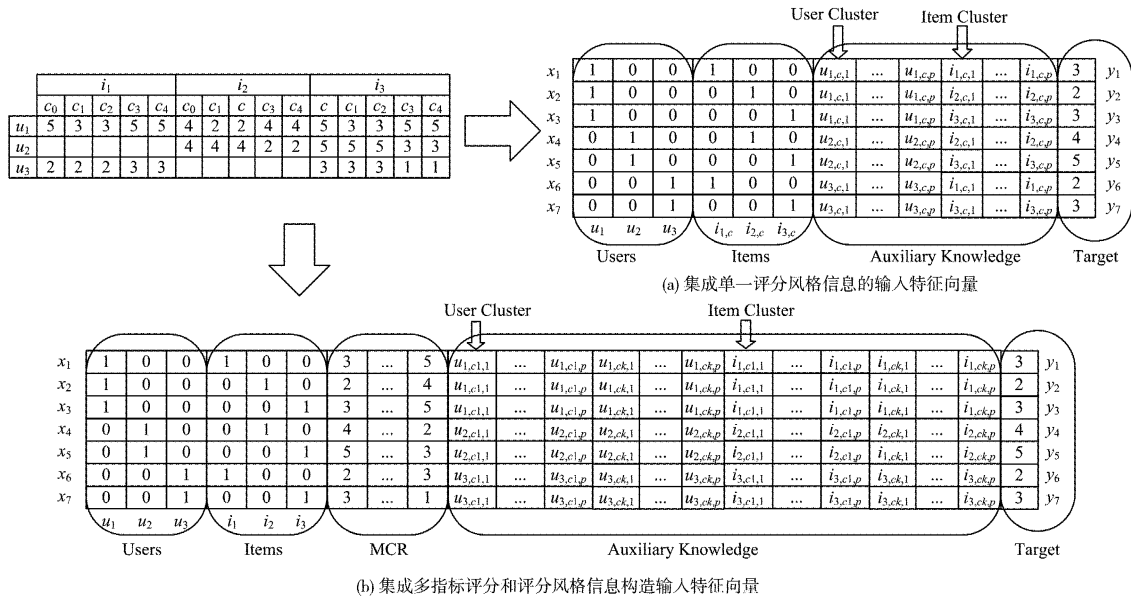


图 1 构造 FM 输入特征向量的两种方法

4 实验及结果分析

4.1 测试数据集

为了评估本文提出的两种推荐模型,选取 Tripadvisor¹⁾ 酒店评分数据作为实验数据,并与文献[20]提出的基于单一综合评分的 FM 模型、仅集成多指标评分(没有集成评分风格信息和被评分风格信息)的 FM 和文献[12-13]提出的两种目前最好的多指标推荐方法进行比较。

Tripadvisor 数据集包含对酒店价格、房间、位置、清洁、前台、服务和商务等 7 个指标的单指标评分以及 1 个综合评分,分值从 1(最不喜欢)到 5(最喜欢)。数据集共包含 357595 条评分记录,有 143038 个用户和 1850 个酒店。为了验证所提算法在数据稀疏情况下的有效性,仅过滤掉只有综合评分而没有单指标评分的数据。最后得到评分记录为 261044 条,用户数和酒店数不变。因此数据集稀疏性为 0.998。实验选择 80% 的评分数据作为训练集,剩余 20% 作为测试集。实验使用两个开源库 Weka 和 LibFM^[21]。Weka 实现多元线性回归和 K-Means 算法,LibFM 实现 FM 算法。

4.2 实验结果与分析

4.2.1 与基于单一综合评分或多指标评分的 FM 算法的比较

该实验主要比较仅集成单一综合评分(Single Rating, SR)和多指标评分(Multi-criteria Rating, MR)的 FM 算法和本文提出的两种基于码本聚类的 FM 算法在不同聚类数下的综合评分预测准确度。使用平均绝对误差(MAE)作为评价指标,实验结果如图 2 所示。从图 2 中可以看到,与 SR 和 MR 相比,在各种聚类情况下,SCRS 方法均优于 SR 方法,但

差于 MR 方法。这说明将多指标评分用于预测评分确实能在一定程度上提高推荐的准确度;而 MCRS 在各种聚类情况下均优于 SR 和 MR 方法。同时,从图 2 可以看出,SCRS 和 MCRS 方法的性能随聚类数的变化而变化,当 $K=12$ 或 $K=14$ 时,SCRS 和 MCRS 的性能分别达到最优。

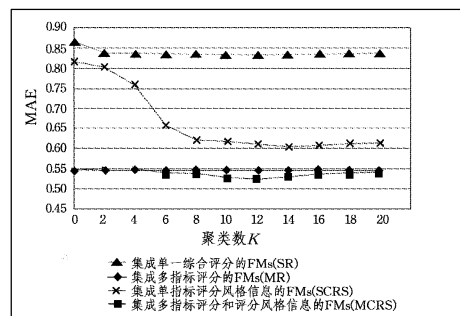


图 2 SCRS, MCRS 与 SR, MR 在不同聚类数下的 MAE 比较

4.2.2 与基于欧几里得距离聚类和权重 SVM 算法的比较

文献[12-13]分别提出的基于欧几里得距离聚类的多指标推荐方法(CECF)和权重 SVM 多指标推荐方法(W SVM)是当前最优的两种方法。本实验主要比较这两种方法与本文提出的两种基于码本聚类的 FM 算法在不同聚类数下的综合评分预测准确度,实验结果如图 3 所示。从图 3 可以看出,SCRS 和 MCRS 在各种聚类情况下均优于 CECF 和 W SVM 方法。具体说来,当聚类数 $K > 6$ 时,SCRC 优于 CECF,且在 $K=14$ 时其性能达到最优(MAE 降低了 7%),但是与 W SVM 相比,SCRC 的 MAE 仅降低了 0.54%。值得注意的是,MCRS 的性能比 CECF 和 W SVM 优得多。具体说来,与 CECF 和 W SVM 相比,在 $K=12$ 时 MCRS 的 MAE 分别降低

¹⁾ <http://www.tripadvisor.co.uk>

了 13.29% 和 6.84%。这说明多指标评分信息、评分风格信息和被评分风格信息对提高推荐的准确度均起到一定的作用。

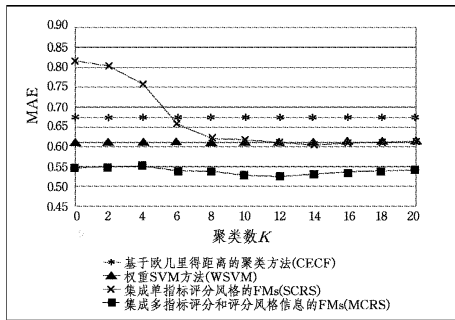


图3 CECF, WSVM 与 SCRS, MCRS 在不同聚类数下的 MAE 比较

结束语 基于码本聚类的思想和因子分解机易于扩展的特点,提出了两种集成多指标评分、评分风格信息和被评分风格信息的多指标推荐算法(SCRS 和 MCRS),并在真实的数据集上进行了实验。实验结果表明,所提出的两种多指标推荐方法均优于仅集成单一综合评分风格信息和被评分风格信息的方法(SR)与当前最好的两种多指标推荐方法(CECF 和 WSVM);同时,集成多指标评分的方法(MR)优于仅集成单指标评分风格信息和被评分风格信息的方法(SCRS),这说明通过集成多指标评分数据来预测评分确实能提高预测的准确度。此外,实验采用的数据集均比基线方法稀疏,但获得了更优的性能,说明所提出的基于码本聚类 and 因子分解机的推荐方法在数据稀疏的情况下也是有效的。但是本文提出的方法没有考虑用户在不同情况下对同一指标可能具有不同偏好这一事实,下一步将对该问题做深入研究,以进一步提高推荐的准确度。

参 考 文 献

- [1] LINDEN B G, SMITH B, YORK J. Amazon. com Recommendations Item-to-item collaborative filtering[J]. IEEE Internet Computing, 2015, 4(1): 76-80.
- [2] ZHANG W J. Research on Application of Collaborative Filtering in Electronic Commerce Recommender Systems[M]// Advances in Computer Science, Environment, Ecoinformatics, and Education. Springer Berlin Heidelberg, 2011: 539-544.
- [3] CHOI K, YOO D, KIM G, et al. A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis[J]. Electronic Commerce Research & Applications, 2012, 11(4): 309-317.
- [4] FUCHS M, ZANKER M. Multi-criteria Ratings for Recommender Systems: An Empirical Analysis in the Tourism Domain[J]. Lecture Notes in Business Information Processing, 2012, 123(2): 100-111.
- [5] ADOMAVICIUS G, TUZHILIN A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions[J]. IEEE Transactions on Knowledge & Data Engineering, 2005, 17(6): 734-749.
- [6] RENDLE S. Factorization machines [C]// Proceedings of the 10th IEEE International Conference on Data Mining, 2010: 995-1000.
- [7] RENDLE S, GANTNER Z, FREUDENTHALER C, et al. Fast context-aware recommendations with factorization machines[C]// Proceeding of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011). Beijing, China, July, 2011: 635-644.
- [8] ADOMAVICIUS G, KWON Y O. New Recommendation Techniques for Multicriteria Rating Systems[J]. Intelligent Systems IEEE, 2007, 22(3): 48-55.
- [9] MANOUSELIS N, COSTOPOULOU C. Experimental Analysis of Design Choices in Multi-Attribute Utility Collaborative Filtering[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2007, 21(2): 311-331.
- [10] TANG T Y, MCCALLA G. The Pedagogical Value of Papers: a Collaborative-Filtering based Paper Recommender[OL]. <http://journals.tdl.org/jodi/article/view/446>.
- [11] HWANG C S. Genetic Algorithms for Feature Weighting in Multi-criteria Recommender Systems [J]. Journal of Convergence Information Technology, 2010, 5(5): 126-136.
- [12] JANNACH D, KARAKAYA Z, Gedikli F. Accuracy improvements for multi-criteria recommender systems[C]// ACM Conference on Electronic Commerce. ACM, 2012: 674-689.
- [13] LIU L W, MEHANDJIEV N, XU D L. Multi-criteria service recommendation based on user criteria preferences[C]// ACM Conference on Recommender System. ACM, 2011: 77-84.
- [14] SAHOO N, KRISHNAN R, DUNCAN G, et al. Collaborative Filtering with Multicomponent Rating for Recommender Systems[C]// Proceedings of the 16th Workshop on Information Technologies and Systems (WITS'06). 2006.
- [15] ZHANG Y, ZHUANG Y, WU J, et al. Applying probabilistic latent semantic analysis to multi-criteria recommender system. [J]. Ai Communications, 2009, 22(22): 97-107.
- [16] SAHOO N, KRISHNAN R, DUNCAN G, et al. Research Note—The Halo Effect in Multicomponent Ratings and Its Implications for Recommender Systems; The Case of Yahoo! Movies[J]. Information Systems Research, 2012, 23(1): 231-246.
- [17] NNADI N. Applying relevant set correlation clustering to multi-criteria recommender systems [C]// Proceedings of the 2009 ACM Conference on Recommender Systems. 2009: 401-404.
- [18] LI Q, WANG C, GENG G. Improving personalized services in mobile commerce by a novel multicriteria rating approach[C]// International Conference on World Wide Web (WWW 2008). Beijing, China, 2008: 1235-1236.
- [19] TAN T F, NETESSINE S. Is Tom Cruise Threatened? Using Netflix Prize Data to Examine the Long Tail of Electronic Commerce[D]. Philadelphia Wharton Business School, 2009.
- [20] LONI B, SAID A, LARSON M, et al. Free lunch enhancement for collaborative filtering with factorization machines[C]// Proceedings of the 8th ACM Conference on Recommender Systems. ACM, 2014: 281-284.
- [21] RENDLE S. Factorization Machines with libFM[J]. ACM Transactions on Intelligent Systems & Technology, 2012, 3(3): 219-224.