

文章编号:1007-5321(2021)03-0106-06

DOI:10.13190/j.jbupt.2020-229

可扩展的融合多源异构数据的推荐模型

冀振燕¹, 吴梦丹^{1,2}, 杨 春¹, 李俊东¹

(1. 北京交通大学 软件学院, 北京 100044; 2. 中国科学院 软件研究所, 北京 100190)

摘要: 社交关系在生活中扮演着重要角色,用户通常会受到其好友偏好的影响,更容易选择好友购买过的物品. 为了解决推荐系统冷启动问题,对融合社交关系的推荐系统进行了研究,提出了贝叶斯个性化排序评论评分社交模型和可扩展的贝叶斯个性化排序评论评分社交模型,将评分、评论、社交关系等多源异构数据从数据源层面进行了融合,通过用户好友信任度模型将社交关系引入到推荐系统中,用基于段向量的分布式词袋模型处理评论,用全连接神经网络处理评分,用改进的贝叶斯个性化排序模型对排序结果进行优化. 实验在 Yelp 公开数据集上进行了实验,实验结果表明,所提出的 2 种模型的推荐准确度均优于其他推荐模型.

关键词: 推荐系统; 多源异构数据; 社交关系; 混合模型; 数据融合

中图分类号: TP391

文献标志码: A

Scalable Recommendation Models Fusing Multi-Source Heterogeneous Data

JI Zhen-yan¹, WU Meng-dan^{1,2}, YANG Chun¹, LI Jun-dong¹

(1. School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China;

2. Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Social relationship plays an important role in life, and users are often affected by their friends' preferences. It is easier for users to choose items that their friends have purchased. In order to solve the cold start problem of the recommended system, a recommendation system that integrates social relationships is studied, and Bayesian personalized ranking review score social model and scalable Bayesian personalized ranking review score social model are proposed. The proposed fusion recommendation models integrate multi-source heterogeneous data such as scores, reviews, and social relationships from the data source level, introduce social relationships into the recommendation system through the user friend trust model, use the paragraph vector-distributed memory model to process review, use the fully connected neural network to process rating, and use an improved Bayesian personalized ranking model to optimize the ranking results. Experiments are conducted on the Yelp public dataset. It is shown that the recommendation accuracy of the two proposed models are better than other recommendation models.

Key words: recommendation system; multi-source heterogeneous data; social relation; hybrid model; data fusion

迄今为止,推荐系统已经发展了 20 多年,其采集用户和产品信息,为用户定向推荐可能感兴趣的

收稿日期: 2020-11-12

基金项目: 国家重点研究发展计划项目(2018YFC0809300); 国家自然科学基金项目(51935002)

作者简介: 冀振燕(1972—),女,副教授,博士生导师, E-mail: zhyji@bjtu.edu.cn.

信息和商品,该系统广泛应用于电子商务、信息检索、社交媒体、音乐电台、视频网站等领域^[1-3]。网络中可收集到的信息种类多样,除了数字评分信息以外,还有文本、图像、社交关系、声音、视频等各种异构数据。

由于单一推荐模型无法处理异构数据,因此研究人员尝试混合不同模型进行推荐,但多数混合模型仅停留在算法层面,即简单地将不同算法得出的推荐结果进行融合,很少从数据源的层面对数据进行融合,无法充分利用异构数据的特征。此外,向混合模型添加新的数据源时,需要重新设计整个框架,模型的可扩展性差。由于用户的评分数据简单直观地反映了用户对物品的偏好,用户的评论数据包含了用户的使用感受、用户偏好和物品特征,蕴含着丰富的语义信息。在现实生活中,用户好友的偏好能够影响用户的选择,而用户的好友关系可以从社交网络中得到。

基于此,提出了贝叶斯个性化排序评论评分社交(BRScS, Bayesian personalized ranking review score social)模型和可扩展的贝叶斯个性化排序评论评分社交(sBRScS, scalable BRScS)模型,通过对数据源层面的评分、评论、社交网络等信息的融合以及用户好友信息的融合,实现了BRScS模型,达到了更高的推荐准确度,解决了推荐系统长期存在的冷启动问题。进一步地,通过对激活函数进行优化避免参数共享,实现了sBRScS模型,在添加新的数据源时,无需训练已有参数。

1 相关工作

推荐算法主要分为基于内容的推荐算法^[4],协同过滤推荐算法^[5]、混合推荐算法^[6]以及基于深度学习的推荐算法^[7]。

基于内容的推荐算法是对物品进行建模,准确提取出物品内容特征,计算得到与其内容相似的物品,进而为用户提供推荐。此类算法存在推荐物品种类单一、新用户冷启动等问题。

协同过滤算法本质上是通过用户和物品的历史数据进行推荐,当推荐系统中增加新的用户或物品时,由于其没有历史信息因而无法对其进行推荐,存在冷启动问题。

混合推荐算法可以分为两种^[8],一种是将不同推荐算法按照一定的策略(加权、交叉调和)混合,这种混合推荐算法的融合方式往往随着算法的改变

而改变,依赖人工对融合策略进行设计,可扩展性差,本质上只是从算法角度进行的融合;另一种混合推荐算法将各种异构数据放在统一的框架中进行训练,进而得到推荐结果,本质上是从特征角度进行的融合。由于网络中收集到的信息往往存在着数据异构、分布不均匀、多模态、数据稀疏等问题,如何从复杂的多源异构数据中有效地提取出特征,是当前混合推荐研究面对的重要挑战。

基于深度学习的推荐算法大多通过学习用户和物品特征来获得推荐结果,系统的可解释性差。当引入新的数据源时,需要重新设计整个推荐框架,系统的可扩展性差。

尽管,目前的推荐系统研究取得了一定的进展,但仍面临着推荐精度不高、冷启动风险、可扩展性差等问题^[9]。为解决上述问题,提出了可扩展的融合多源异构数据的推荐模型。

2 融合多源异构数据的推荐模型

为了充分融合异构数据的特征,提出了BRScS模型和sBRScS模型,从数据源的层面对评分、评论和社交网络等异构数据进行融合,具有较好的可扩展性和较高的推荐准确度。

2.1 推荐系统框架

所提出的BRScS模型推荐过程如图1所示。使用基于段向量的分布式词袋(PV-DBOW, distributed bag of words version of paragraph vector)模型学习得到评论信息的嵌入向量,从而获得对应的用户和物品的嵌入向量。将用户的好友引入推荐系统,使用改进的贝叶斯个性化排序(BPR, Bayesian personalized ranking)模型对排名结果进行优化,通过多层全连接神经网络学习得到评分信息的嵌入向量,进而得到对应的用户和物品的嵌入向量。在获得用户和物品的嵌入向量后,可以计算出某物品对指定用户的排序分数,进而得到针对该用户的物品推荐列表。

2.2 推荐模型

所提出的BRScS模型包括4个步骤:1)通过社交网络学习社交关系;2)构建改进的用户和物品的三元组优化BPR模型的排序结果;3)构建统一的嵌入学习框架将社交关系数据、评论数据和评分数据进行融合;4)按照排序分数由高到低的顺序,为用户提供个性化推荐列表。

2.2.1 社交关系数据学习

模型将用户的直接好友命名为用户的1度好

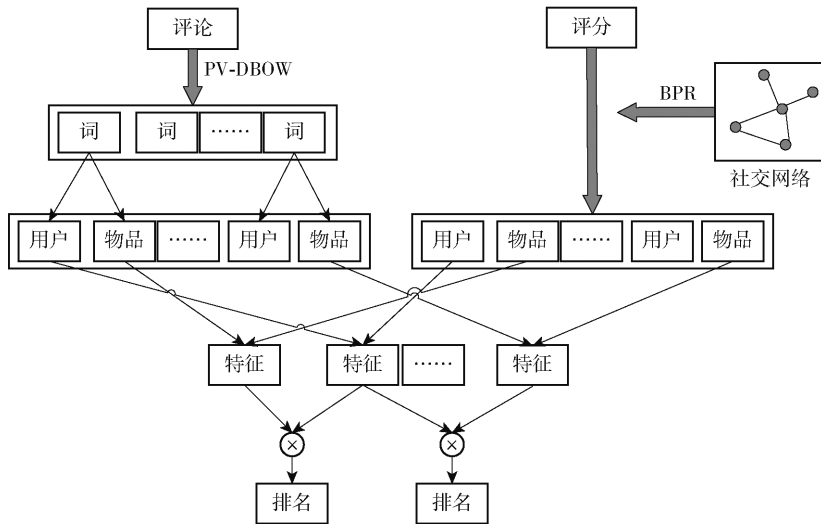


图1 推荐过程

友,将与用户距离为2的间接好友命名为用户的2度好友,将与用户距离为3的间接好友命名为用户的3度好友,以此类推。根据六度分离原理^[10],模型最多考虑距离为6的间接好友,因此,将此用户好友信任度模型命名为基于六度模型理论的用户相似度计算算法。

从社交网络数据中计算出用户间距离及用户间相似性后,可以得到不同好友偏好对用户选择的影响。将其与评论、评分等异构数据一起输入到统一的学习框架中进行训练,最后可以产生推荐结果。

2.2.2 改进的BPR模型

传统的BPR模型是通过随机采样构建的用户-物品三元组,并没有将社交关系引入到采样过程中。对传统的BPR模型进行改进,构建一个改进的用户-物品三元组 (u, m_i, m_j) 。训练数据集 T_s 定义为

$$T_s = \{(u, m_i, m_j) \mid \forall m_i, m_j \in D\} \quad (1)$$

其中: D 为物品数据集, m_i 为用户 u 及其好友在 D 中购买过的物品, m_j 为用户 u 未购买的物品。

根据贝叶斯公式,求解模型可以表示为

$$P(\theta \mid >_u) \propto P(>_u \mid \theta) P(\theta) \quad (2)$$

其中: θ 为模型的参数向量, $>_u$ 为用户 u 的偏好结构, \propto 为成正比关系。

假定所有用户的行为彼此独立,特定用户的每对商品 m_i, m_j 的排序与其他每对商品的排序无关。因此式(2)可以进一步推导得到

$$\prod P(>_u \mid \theta) = \prod P(m_i >_u m_j \mid \theta)^\alpha \times [1 - P(m_i >_u m_j \mid \theta)]^{(1-\alpha)} \quad (3)$$

$$\alpha = \begin{cases} 1, & (u, m_i, m_j) \in T_s \\ 0, & (u, m_i, m_j) \notin T_s \end{cases} \quad (4)$$

由于完整性和反对称性,式(3)可以进一步简化为

$$\prod P(>_u \mid \theta) = \prod_{(u, m_i, m_j) \in T_s} P(m_i >_u m_j \mid \theta) \quad (5)$$

其中: $P(m_i >_u m_j \mid \theta) = \gamma(x_{u, m_i, m_j}(\theta))$, $\gamma(\cdot)$ 为sigmoid函数, $x_{u, m_i, m_j}(\theta)$ 为捕获用户 u 、物品 m_i 和物品 m_j 具体关系的关于参数向量 θ 的任意实值函数。

通过以上推导可以得出BPR模型的优化函数为

$$\begin{aligned} O(\theta) &= \ln p(\theta \mid >_u) = \\ &= \ln p(>_u \mid \theta) p(\theta) = \\ &= \ln \prod_{(u, m_i, m_j) \in T_s} \gamma(x_{u, m_i, m_j}(\theta)) p(\theta) = \\ &= \sum_{(u, m_i, m_j) \in T_s} \ln \gamma(x_{u, m_i, m_j}(\theta)) + \ln p(\theta) = \\ &= \sum_{(u, m_i, m_j) \in T_s} \ln \gamma(x_{u, m_i, m_j}(\theta)) - \lambda_\theta \|\theta\|^2 \end{aligned} \quad (6)$$

其中 λ_θ 为模型的惩罚参数, $\theta \in \{1, 2\}$ 。

2.2.3 评论数据学习

在PV-DBOW模型中,每条评论首先被映射到一个语义空间,之后对其训练并进行词语预测。计算公式为

$$P(\mathbf{w} \mid \mathbf{d}_u^m) = \frac{e^{\mathbf{w}^T \mathbf{d}_u^m}}{\sum_{\mathbf{w}' \in V} e^{\mathbf{w}'^T \mathbf{d}_u^m}} \quad (7)$$

其中: \mathbf{d}_u^m 为用户 u 对物品 m 评论的嵌入向量, $m \in D$; \mathbf{w} 为 \mathbf{d}_u^m 中词的嵌入向量; V 为词库; \mathbf{w}' 为词库中的词。

为了提高计算效率,模型使用了负采样的计算策略,评论数据的目标函数定义为

$$L_1(\mathbf{w}, \mathbf{d}_u^m) = \sum_{w \in V(u, m)} \sum_{m \in R} f_{w, d_u^m} \lg \sigma(\mathbf{w}^T \mathbf{d}_u^m) + \sum_{w \in V(u, m)} \sum_{m \in R} f_{w, d_u^m} [t E_{w_N \sim P_V} \lg \sigma(-\mathbf{w}_N^T \mathbf{d}_u^m)] \quad (8)$$

其中: (u, m) 为用户-物品对, f_{w, d_u^m} 为词语 w 在评论 d_u^m 中出现的频率, P_V 为噪声分布, $E_{w_N \sim P_V} \lg \sigma(-\mathbf{w}_N^T \mathbf{d}_u^m)$ 为在 $\lg \sigma(-\mathbf{w}_N^T \mathbf{d}_u^m)$ 符合噪声分布 P_V 情况下的期望, t 为负样本的个数。

通过式(8)可以得到评论的嵌入向量 \mathbf{d}_u^m , 进而能够得到相应的用户 u 和物品 m 的嵌入向量, 记为 \mathbf{u}_1 和 \mathbf{m}_1 。

2.2.4 评分数据学习

所提模型使用两层全连接网络对评分数据进行处理, 处理后可以得到用户和物品的嵌入向量, 分别为 \mathbf{r}_u 和 \mathbf{r}_m , 评分数据的计算值为

$$\hat{r}_u^m = \phi \{ U_2 \phi [U_1(\mathbf{r}_u \cdot \mathbf{r}_m) + \mathbf{c}_1] + \mathbf{c}_2 \} \quad (9)$$

其中: $\mathbf{c}_1, \mathbf{c}_2$ 为偏差向量; U_1, U_2 为权重参数; $\phi(\cdot)$ 为 ELU 激活函数。

定义评分目标函数为评分数据计算值和实际值之间的差距:

$$L_2(\mathbf{r}_u, \mathbf{r}_m, \mathbf{c}_1, \mathbf{c}_2, U_1, U_2) = \sum_{(u, m) \in R} (\hat{r}_u^m - r_u^m)^2 \quad (10)$$

其中 r_u^m 为评分数据的实际值。通过对式(10)不断学习优化, 寻求 L_2 最小值, 可以得到嵌入向量 \mathbf{r}_u 和 \mathbf{r}_m 的最优解, 分别记为 \mathbf{u}_2 和 \mathbf{m}_2 。

2.2.5 BRScS 模型

模型使用串联^[6]的方式对评论信息和评分信息进行融合, 即用户嵌入向量 $\mathbf{u} = f(\mathbf{u}_1, \mathbf{u}_2) = [\mathbf{u}_1^T, \mathbf{u}_2^T]^T$, 物品嵌入向量 $\mathbf{m} = f(\mathbf{m}_1, \mathbf{m}_2) = [\mathbf{m}_1^T, \mathbf{m}_2^T]^T$ 。

定义排名损失函数 $\sigma(\mathbf{u}, \mathbf{m}_i, \mathbf{m}_j) = \sigma(\mathbf{u}^T \mathbf{m}_i - \mathbf{u}^T \mathbf{m}_j)$, $\sigma(\cdot)$ 为 sigmoid 函数。则 BRScS 模型的目标函数定义为

$$L = \sum_{(u, m) \in R} \sigma(\mathbf{u}^T \mathbf{m}_i - \mathbf{u}^T \mathbf{m}_j) + \lambda_1 L_1 - \lambda_2 L_2 \quad (11)$$

其中: $W = \{W_1, W_2\}$ 为模型中的参数, 参数 W_1 为评论嵌入学习模型的参数; 参数 W_2 为评分嵌入学习模型的参数, Θ 为模型中其他需要学习得到的参数, $\Theta = \{\Theta_1, \Theta_2\} = \{\{w, d_u^m\}, \{U_1, U_2, c_1, c_2, r_u, r_m\}\}$; $\lambda_1, \lambda_2 \in [0, 1]$ 分别代表不同目标函数的正则化

参数。

通过随机梯度下降法对式(11)优化求解得到相应的用户表示 \mathbf{u} 和物品表示 \mathbf{m} , 排序分数 s 计算方式为

$$s = \mathbf{u}^T \mathbf{m} \quad (12)$$

通过每件物品的排序分数, 对物品集的物品进行排序, 得到物品推荐列表。

2.2.6 sBRScS 模型

将式(11)对 Θ_k 进行优化为

$$\frac{\partial L}{\partial \Theta_k} = \lambda_k \frac{\partial L_k}{\partial \Theta_k} \quad (13)$$

其中 $k \in \{1, 2\}$ 。由式(13)可知, Θ_k 的优化只与对应的 L_k 有关。

将式(11)对 W_k 进行优化为

$$\begin{aligned} \frac{\partial L}{\partial W_k} = & \sum_{(u, m) \in R} \left(\frac{\partial \sigma}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial W_k} + \frac{\partial \sigma}{\partial \mathbf{m}_i} \frac{\partial \mathbf{m}_i}{\partial W_k} + \frac{\partial \sigma}{\partial \mathbf{m}_j} \frac{\partial \mathbf{m}_j}{\partial W_k} \right) = \\ & \sum_{(u, m) \in R} \left(\frac{\partial \sigma}{\partial \mathbf{u}} \frac{\partial f}{\partial \mathbf{u}_k} \frac{\partial \mathbf{u}_k}{\partial W_k} + \frac{\partial \sigma}{\partial \mathbf{m}_i} \frac{\partial f}{\partial (\mathbf{m}_i)_k} \frac{\partial (\mathbf{m}_i)_k}{\partial W_k} + \right. \\ & \left. \frac{\partial \sigma}{\partial \mathbf{m}_j} \frac{\partial f}{\partial (\mathbf{m}_j)_k} \frac{\partial (\mathbf{m}_j)_k}{\partial W_k} \right) \end{aligned} \quad (14)$$

对式(14)进行简化, 定义函数 $h(\cdot)$ 为

$$h(\mathbf{y}) = \frac{\partial \sigma}{\partial \mathbf{y}} \frac{\partial f}{\partial \mathbf{y}_k} \quad (15)$$

其中 $\mathbf{y} \in \{\mathbf{u}, \mathbf{m}_i, \mathbf{m}_j\}$ 。将式(15)代入式(14)中, 可以得到简化后的式子:

$$\begin{aligned} \frac{\partial L}{\partial W_k} = & \sum_{(u, m) \in R} \left[h(\mathbf{u}) \frac{\partial \mathbf{u}_k}{\partial W_k} + h(\mathbf{m}_i) \frac{\partial (\mathbf{m}_i)_k}{\partial W_k} + \right. \\ & \left. h(\mathbf{m}_j) \frac{\partial (\mathbf{m}_j)_k}{\partial W_k} \right] \end{aligned} \quad (16)$$

由 2.2.5 节可知, $\sigma(\cdot)$ 可以表示为

$$\sigma(\mathbf{u}, \mathbf{m}_i, \mathbf{m}_j) = \sum_k \sigma(\mathbf{u}_k^T (\mathbf{m}_i)_k - \mathbf{u}_k^T (\mathbf{m}_j)_k) \quad (17)$$

将式(17)带入式(15)中, 可以得到:

$$h(\mathbf{u}) = \frac{\partial \sigma}{\partial \mathbf{u}} ((\mathbf{m}_i)_k - (\mathbf{m}_j)_k)^T \quad (18)$$

$$h(\mathbf{m}_i) = \frac{\partial \sigma}{\partial \mathbf{m}_i} \mathbf{u}_k^T \quad (19)$$

$$h(\mathbf{m}_j) = \frac{\partial \sigma}{\partial \mathbf{m}_j} (-\mathbf{u}_k^T) \quad (20)$$

进一步可以得到关于 W 的优化过程:

$$\begin{aligned} \frac{\partial L}{\partial W_k} = & \sum_{(u, m) \in R} \frac{\partial \sigma}{\partial \mathbf{y}} \left[((\mathbf{m}_i)_k - (\mathbf{m}_j)_k)^T \frac{\partial \mathbf{u}_k}{\partial W_k} + \right. \\ & \left. \mathbf{u}_k^T \left(\frac{\partial (\mathbf{m}_i)_k}{\partial W_k} - \frac{\partial (\mathbf{m}_j)_k}{\partial W_k} \right) \right] \end{aligned} \quad (21)$$

由式(21)可知,在对 sBRScS 的目标函数进行优化时,处理各个数据的参数变量仅参与内部计算,并不用来共享。

因此,当向此推荐模型中引入一个新的数据源时,仅需通过相关算法对新的数据源进行训练从而得到相关算法的参数即可,无需重新训练整个框架。

3 实验设计与结果分析

3.1 环境设置

实验在 Ubuntu 16.04 操作系统下,单核 GPU GeForce GTX 1080 Ti 上进行,使用 igraph 分析用户好友关系并构建社交网络,使用 Anaconda 的 python3.6, tensorflow1.4 对模型进行训练与评估。

3.2 数据集与评价指标

实验数据来源于 Yelp 官方网站上公布的 Yelp 公开数据集。好友关系被转换为用户-好友关系对,评分和评论用于用户对商户的满意度分析。为了验证所提出模型的有效性,实验弃用了评论数小于 20 条的数据,将其余数据生成新的数据集,将其命名为 New-Yelp。

在实验中,将 New-Yelp 数据集分为两部分,其中 70% 的数据用于训练,30% 的数据用于测试。设置批处理大小为 64,迭代次数为 40,嵌入维度为 300,负样本数为 5。采用召回率、精确率、归一化折损累积增益(NDCG, normalized discounted cumulative gain)、命中率 4 个指标对推荐列表中的前 5 位(Top-5)或前 10 位(Top-10)推荐结果进行评价。

3.3 BRScS 模型对比实验

将 BRScS 模型与以下模型^[11-16]进行对比。

深度卷积神经网络评论(DR, deep conn and review)模型^[11]在深度协作神经网络模型的基础上,使用评论数据为用户生成推荐,是一种基于单点法的排序学习算法。

正则化评论社交(HRS, hdc and review and social)模型^[12]在基于词嵌入的正则化语义模型基础上,融合评论数据和社交网络信息为用户生成推荐列表。

词嵌入评论社交(SRS, sel and review and social)模型^[13]在简化的词嵌入语义模型基础上,融合评论数据和社交网络数据为用户进行推荐。

贝叶斯个性化排序评论(BR, BPR and review)模型^[14]是基于传统的 BPR 框架,使用 PV-DBOW 模型处理评论数据并为用户提供推荐列表,是一种基

于配对法的排序学习算法。

贝叶斯个性化排序评论社交(BRS, BPR and review and social)模型^[15]在改进的 BPR 框架基础上,融合评论数据和社交网络数据,引入 PV-DBOW 模型和用户信任度模型为用户进行推荐。其使用的社交数据既包含用户的直接好友,也包含用户的间接好友。

贝叶斯个性化排序评论(BRSc, BPR and review and score)模型^[16]在传统的 BPR 框架基础上,同时使用评论数据和评分数据,为用户生成推荐列表。

对使用 BRScS 模型与上述模型各自得到的 Top-5 推荐结果进行评价对比,结果如表 1 所示。由表 1 可知,BRScS 模型推荐结果的各项指标最佳。在该对比实验中,BRS,HRS 和 SRS 3 种模型均是基于评论数据和社交网络数据的推荐模型。实验结果表明,BRS 模型在推荐的精确率、召回率等方面均明显优于 HRS 模型和 SRS 模型,这表明,在推荐系统中引入 BPR 模型能够提高推荐的效果。

表 1 BRScS 模型 Top-5 对比实验结果

模型/100%	召回率	精确率	NDCG	命中率
DR	3.324	1.087	2.131	5.976
HRS	1.120	0.326	0.751	1.930
SRS	1.853	0.608	1.321	3.492
BR	3.422	1.119	2.370	6.334
BRS	4.175	1.308	2.967	7.425
BRSc	3.650	1.208	2.541	6.794
BRScS	4.363	1.352	3.084	7.710

3.4 sBRScS 模型对比实验

将 BRScS 模型与其变体进行对比,其中,可扩展的贝叶斯个性化排序评分(sBSc, scale BPR and score)模型是使用评分数据的可扩展的推荐模型;可扩展的贝叶斯个性化排序评论(sBR, scale BPR and review)模型是使用评论数据的可扩展的推荐模型,BRSc 模型是融合了评论数据和评分数据进行推荐的模型,可扩展的贝叶斯个性化排序评论评分(sBRSc, scale BPR and review and score)模型是融合了评论数据和评分数据的可扩展的推荐模型,这些模型都是基于改进的 BPR 框架。

实验对比了对推荐结果 Top-10 的评价指标,结果如表 2 所示。由表 2 可知,BRScS 在召回率、精确率、NDCG 和命中率方面均优于其他模型。尽管 sBRScS 模型的推荐性能比 BRScS 模型略有降低,但

表 2 sBRScS 模型 Top-10 对比实验结果

模型/100%	召回率	精确率	NDCG	命中率
sBSc	2. 260	0. 227	1. 627	8. 628
sBR	5. 560	1. 004	3. 103	10. 049
BRSc	5. 787	1. 057	3. 264	10. 480
sBRSc	5. 638	1. 042	3. 223	10. 268
BRScS	6. 850	1. 168	3. 915	11. 830
sBRScS	6. 726	1. 135	3. 904	10. 950

其具有更好的可扩展性,当新引入一个数据源时,无需重新设计推荐框架和重新训练模型参数,是更深层次的可扩展的融合推荐模型. 与略有降低的性能相比,sBRScS 模型不必重新训练已经训练好的参数,可以更便捷地引入新的数据源进行推荐.

4 结束语

提出 BRScS 和 sBRScS 模型,将评分数据、评论数据以及社交网络数据进行了融合. 通过在推荐系统中引入用户好友信任度模型,考虑用户好友偏好对用户选择的影响,能够在一定程度上解决冷启动和数据稀疏等问题,提升推荐的质量. 除此之外,所提模型能够很好地融合其他类型的异构数据,具有良好的可扩展性. 在未来的工作研究中,计划将包含更丰富信息的图像数据引入到现有的推荐模型中.

参考文献:

[1] Fabbri F, Bonchi F, Boratto L, et al. The effect of homophily on disparate visibility of minorities in people recommender systems[C] //International AAAI Conference on Web and Social Media. [S. l.]: AAAI, 2020: 165-175.

[2] Gao Jingyue, Wang Xiting, Wang Yasha, et al. Explainable recommendation through attentive multi-view learning [C] // AAAI Conference on Artificial Intelligence. Hawaii: AAAI, 2019: 3622-3629.

[3] Ji Zhenyan, Pi Huaiyu, Wei Wei, et al. Recommendation based on review texts and social communities: a hybrid model[J]. IEEE Access, 2019, 7: 40416-40427.

[4] Wang Chengwei, Zhou Tengfei, Chen Chen, et al. Camo: a collaborative ranking method for content based recommendation[C] // AAAI Conference on Artificial Intelligence. Hawaii: AAAI, 2019: 5224-5231.

[5] 王永, 邓江洲. 基于 KL 散度的用户相似性协同过滤算法[J]. 北京邮电大学学报, 2017, 40(2): 110-

114.

Wang Yong, Deng Jiangzhou. User similarity collaborative filtering algorithm based on KL divergence[J]. Journal of Beijing University of Posts and Telecommunications, 2017, 40(2): 110-114.

[6] Liu Hao, Li Ting, Hu Renjun, et al. Joint representation learning for multi-modal transportation recommendation [C] // AAAI Conference on Artificial Intelligence. Hawaii: AAAI, 2019: 1036-1043.

[7] Zheng Lei, Noroozi V, Yu P S. Joint deep modeling of users and items using reviews for recommendation[C] // Tenth ACM International Conference on Web Search and Data Mining. Cambridge: ACM, 2017: 425-434.

[8] Chen Wei, Niu Zhengdong, Zhao Xiangyu, et al. A hybrid recommendation algorithm adapted in e-learning environments[J]. World Wide Web, 2014, 17(2): 271-284.

[9] Zhang Shuai, Yao Lina, Sun Aixin, et al. Deep learning based recommender system: a survey and new perspectives[J]. ACM Computing Surveys (CSUR), 2019, 52(1): 1-38.

[10] Milgram, Stanley. The small world problem[J]. Psychology Today, 1967, 2(1): 60-67.

[11] Elkahky A M, Song Y, He X. A multi-view deep learning approach for cross domain user modeling in recommendation systems[C] // 24th International Conference on World Wide Web. Florence: WWW, 2015: 278-288.

[12] Krohn-Grimberghe A, Drumond L, Freudenthaler C, et al. Multi-relational matrix factorization using Bayesian personalized ranking for social network data[C] // Fifth ACM International Conference on Web Search and Data Mining. Washington: WSDM, 2012: 173-182.

[13] Le Q, Mikolov T. Distributed representations of sentences and documents[C] // International Conference on Machine Learning. Beijing: ICML, 2014: 1188-1196.

[14] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback [EB/OL]. 2012 (2012-05-09) [2020-11-12]. <https://arxiv.org/abs/1205.2618>.

[15] Al-Mashari, Majed, Zairi M. Revisiting BPR: a holistic review of practice and development [J]. Business Process Management Journal, 2000, 6(1): 10-42.

[16] He R, McAuley J. VBPR: visual Bayesian personalized ranking from implicit feedback[C] // AAAI Conference on Artificial Intelligence. Arizona: AAAI, 2016: 144-150.