

基于多上下文信息的协同过滤推荐算法



郝志峰^{1,2} 廖祥财¹ 温雯¹ 蔡瑞初¹

1 广东工业大学计算机学院 广州 510006

2 佛山科学技术学院数学与大数据学院 广东 佛山 528000

(zfhao@fosu.edu.cn)

摘要 随着电子商务和互联网的发展,数据信息呈爆炸式增长,协同过滤算法作为一种简单而高效的推荐算法,能在一定程度上有效地解决信息爆炸问题。但是传统协同过滤算法仅通过单一评分来挖掘相似用户,推荐效果并不占优势。为了提高个性化推荐的质量,如何充分利用用户(物品)的文本、图片、标签等上下文信息以使数据价值最大化是当前推荐系统亟待解决的问题。对此,提出了一种融合多种类型上下文信息的协同过滤算法。以用户商品交互信息为二部图,根据不同类型上下文的特点构建不同的相似度网络,设计目标函数在多种上下文信息网络的约束下联合矩阵分解,并学得用户商品的表示学习。在多个数据集上进行了充分实验,结果表明,融合多种类型上下文信息的协同过滤算法不仅能有效提高推荐的准确度,而且能在一定程度上解决数据稀疏性问题。

关键词: 矩阵分解; 协同过滤; 推荐系统; 多上下文信息

中图法分类号 TP181

Collaborative Filtering Recommendation Algorithm Based on Multi-context Information

HAO Zhi-feng^{1,2}, LIAO Xiang-cai¹, WEN Wen¹ and CAI Rui-chu¹

1 School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China

2 School of Mathematics and Big Data, Foshan University, Foshan, Guangdong 528000, China

Abstract With the development of e-commerce and the Internet, as well as the explosive growth of data information, collaborative filtering algorithm as a simple and efficient recommendation algorithm can effectively alleviate the problem of information explosion. However, the traditional collaborative filtering algorithm only uses a single rating to mine similar users, and the recommendation effect is not dominant. In order to improve the quality of personalized recommendations, how to make full use of the user (items) text, pictures, labels and other information to maximize the value of data is an urgent problem to be solved by the current recommendation system. Therefore, user-product interaction information is used as a bipartite graph, and different similarity networks are constructed according to the characteristics of different contexts. The design objective function is combined with matrix decomposition under the constraints of various information networks and user or item embedding can be gotten. Extensive experiments are conducted on multiple data sets, and the results show that the collaborative filtering algorithm by fusion of multiple types of information can effectively improve the accuracy of recommendations and alleviate the problem of data sparsity.

Keywords Matrix decomposition, Collaborative filtering, Recommendation system, Multi-context information

1 引言

随着信息技术的飞速发展以及 Internet 的迅速普及,以 Web2.0 为代表的互联网技术使得网络数据呈爆炸式增长,网络中充斥着越来越多的信息和服务。大数据中蕴含着巨大的价值,虽然给人类社会带来了变革性的发展,但是同时也造成了严重的“信息过载”问题。如何快速有效地从纷繁复杂的数据中获取有价值的信息成为了当前大数据发展的关键问

题。推荐算法通过用户历史行为来推断用户的兴趣爱好,并向用户推荐感兴趣的物品,帮助用户获取所需要的信息,能有效地改善信息超载问题。推荐算法应用的场景非常普遍,如电子商务、电影和视频推荐、社交网络、用户兴趣点推荐等,这将给企业带来巨大的商业价值、经济价值和服务价值。

在众多推荐算法中,协同过滤^[1]因其简单高效而受到工业界的关注。但是协同过滤算法也面临着诸多挑战:1)在推荐系统中,数据规模大,结构复杂,包含成千上万个用户和商

到稿日期:2020-07-15 返修日期:2020-09-24 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61876043,61976052);广东省科技计划(2019A141401006)

This work was supported by the National Natural Science Foundation of China(61876043,61976052) and Science and Technology Planning Project of Guangdong Province(2019A141401006).

通信作者:温雯(wwen_gdut@189.cn)

品。然而,用户只与极小部分商品产生过交互,而很大一部分商品并没有曝光给用户,这也造成了数据高维度、稀疏性等特点。2)除了用户每天浏览、评价商品等交互记录以外,还有用户商品自身的图片、文本描述、标签信息等多种类型的辅助信息,而这些辅助信息能准确反映用户和商品的属性。一般而言,用户的购买决策过程或其对某商品的评分往往受多种因素的影响,如用户喜欢某一类商品是因为它们有相同的标签属性,或者是因为用户受到其他用户的评论等文本信息的影响。因为用户兴趣受这类标签、文本等辅助信息的影响,所以这类辅助信息也被称为“上下文”。传统协同过滤模型仅利用用户与商品之间的交互关系,忽略了用户商品的辅助信息,如果充分利用这些“上下文”辅助信息,就能挖掘到用户更广泛的兴趣爱好。然而,由于这些上下文信息呈现多源性、异构性,如何融合多源异构上下文信息以缓解数据的稀疏性也给推荐算法带来了挑战。总之,推荐系统的数据高维度、稀疏性、异构性等特点,给推荐系统挖掘技术带来了前所未有的挑战。

针对上述问题,综合考虑了融合多源异构辅助信息的推荐算法并不多的情况,本文提出了一种融合多上下文联合矩阵分解的推荐算法,将用户浏览、收藏、购买商品等历史行为自然地看作二分图,根据不同的上下文信息的特点,设计不同的相似度函数以度量用户(商品)间的相似程度,并以此构建出用户(商品)间的相似度网络。比如,对于离散型的标签数据,拥有相同标签的用户之间的兴趣爱好是相似的,但它们间的相似度不能简单地用余弦距离度量。本文用 SPPMI(Shifted positive pointwise mutual information)来度量标签相似度网络中节点的相似度,对连续型属性信息(如文本、图片等)先做预处理,得到属性信息特征表示,然后根据距离函数计算节点的相似度矩阵以得到由属性构建的主题相似网络,不同的距离函数对推荐结果有影响。最后设计目标函数融合多种类型的图网络,并优化联合矩阵分解,求得解析解,得到融合了多种上下文信息的低维节点表示向量。实验结果表明优化的算法进一步提高了系统的推荐质量,能有效缓解数据稀疏性。

本文第2节介绍推荐算法的相关工作,包括基于传统的推荐方法和基于图的推荐算法的方法;第3节主要介绍问题以及对数据进行分析;第4节详细介绍本文提出的融合多种上下文节点信息的图推荐算法;第5节则在多个不同数据集上对所提算法的有效性进行验证,并展示相关的实验结果;最后对本文工作进行总结和展望。

2 相关工作

本节介绍推荐算法的相关方法,包括协同过滤算法和基于图表示学习的推荐方法。

协同过滤主要分为基于邻域的方法和基于隐语义模型的方法。基于隐语义模型的算法通过机器学习方法,使用评分集合学习出一个全局模型,通过优化一个明确的目标函数来学习用户和商品的低维表示,然后基于模型来预测缺失的评分。矩阵分解作为隐语义模型中最成功的一种实现,受到了研究者的广泛关注,常见的矩阵分解包括 SVD(Singular Value Decomposition)^[2]、PMF(Probabilistic Matrix Factorization)^[3]、NMF(Nonnegative Matrix Factorization)^[4]等方法。

为了充分利用辅助数据,改进模型预测精度,还有一类方法对多个多重相关矩阵进行联合矩阵分解^[5-7]。这些多重相关矩阵包含相同的实体,因此可以从相同实体的关联矩阵中获取额外信息。2016年Liang等^[8]受到 Word2vec^[9]以及 Levy等^[10]证明 Word2vec 等价于 PMI(Pointwise Mutual Information)矩阵隐式分解的启发,以用户历史行为记录为类比词嵌入模型,得到商品的 PMI 矩阵,联合分解评分矩阵和商品 PMI 矩阵得到用户和商品的低维特征表示。除了利用用户历史行为数据外,文献[11-12]在传统的矩阵分解模型上融合用户的近邻关系,不仅可以丰富单个用户的信息,而且能更好地对各用户进行建模。Mohsen等^[13]将用户间的社交网络信息融入矩阵分解推荐算法中,以解决用户冷启动问题。除了用户间的社会关系以外,还有基于标签的信息^[14]、基于文本的信息(用户评价)^[15]以及基于图片的信息^[16]。近年来,随着深度学习在图像识别、自然语言处理等领域取得了巨大成功,许多学者开始尝试探索如何运用深度学习来提取非结构化数据特征并与推荐系统结合。2016年,Kim等^[17]提出 ConvMF,将矩阵分解和深度学习相结合,融合了外部丰富的文本信息。

随着图表示学习方法的兴起,图表示学习算法在推荐系统中的运用也逐渐受到关注。图表示学习算法的基本思想是把用户商品交互信息看作二分图,将节点映射到一个稠密且低维的向量空间中,并在映射过程中尽量保留网络信息和节点之间的相似性。2014年,Perozzi等提出了 DeepWalk^[18],创造性地将词表示学习的方法引入到图表示学习当中。Item2vec^[19]把用户浏览的商品集合等价于 Word2vec 中单词的序列,学得了商品的表示学习。属性网络不仅包含复杂的网络结构,还包括多样化的节点属性信息。CENE^[20]将文本内容视为特殊类型的节点,并利用节点-节点链接和节点-内容链接共同学习节点表示。异构网络包括不同类型的节点或链接,为了捕获不同类型节点之间的关系,Metapath2vec^[21]设计了不同语义的元路径,在图上随机游走获取节点游走序列,并利用 Skip-gram 优化提取的序列。HIN^[22]在元路径基础上考虑到不同语义下节点表示对推荐效果有影响,设计了权重,并用聚合函数聚集不同语义下的节点表示。将知识图谱作为辅助信息引入推荐系统中也可以有效地缓解传统推荐系统存在的稀疏性和冷启动问题。CKE^[23]将推荐算法和知识图谱特征学习的目标融合,并在一个端到端的优化目标中进行训练以学得用户商品的目标函数。总之,融合多源异构的图表示学习推荐算法在近年来得到了研究人员的广泛关注。

3 问题描述

真实的推荐系统不仅包含用户浏览、收藏、购买商品等历史行为,还包含用户商品标签、文本描述等辅助信息。本节将根据这些辅助信息自身的特点设计并构建用户(商品)相似度网络,然后对在融合多类型辅助信息场景下预测用户对商品是否喜欢进行形式化描述,并对数据进行分析。

3.1 问题定义

在推荐系统中,多类型辅助信息数据通常包括结构化、半结构化和非结构化数据,其中结构化数据包括用户的隐式反

馈;半结构化数据包括用户的属性信息(如用户的年龄、所处的地点、职业等)、物品的属性信息(物品的类别、标签等)等,这些数据可通过离散化或0-1化转换成结构化数据;非结构化数据包括用户的评论信息、与商品相关的图像视频等,这些数据可以通过信息检索技术或机器学习方法来转换成结构化数据。本文详细定义了问题和与之相关的3种数据,具体内容如下所示。

定义1(标签相似网络) 用户之间的标签相似网络表示为 $G=(U, E)$, 其中 U 表示用户的集合。 $\forall u, v \in U, (u, v) \in E$ 表示用户 u 和用户 v 拥有相似的标签, 标签相似网络中 $(u, v) \in E$ 的权重 $T_{uv} \in [0, 1]$ 表示用户之间的相似程度。如果商品之间也存在这样的标签相似网络, 那么可以定义为 $G=(I, E)$ 。

定义2(用户主题相似网络) 对于给定的用户, 通常有一段文本描述, 可借助深度学习或机器学习方法对用户提取出文本特征向量 $f=(f_1, f_2, \dots, f_m)$ 。用户 u 和用户 v 之间的主题相似度定义如下:

$$S_{uv} = \exp\left(-\frac{\|f_u - f_v\|_t^2}{t}\right) \quad (1)$$

如此得到用户之间的相似程度 $S_{uv} \in [0, 1]$ 。选取一个合适的阈值 S_{thr} , 当用户间的主题相似度值大于 S_{thr} 时, 为其建立一条无向的边, 反之不建立边, 然后得到用户主题相似网络 $G=(U, E, W_u)$, 其中 W_u 表示该网络的邻接矩阵。如果商品也有文本描述, 那么依上述方法能得到商品间的主题相似网络 $G=(U, E, W_i)$ 。

定义3(用户商品二分图网络) 用户购买商品的历史行为可看做一个二分图 $G_{UI}=(U \cup I, Y_{UI})$, 其中 $Y_{ui}=1$ 表示用户 u 与商品 i 之间有连边; $Y_{ui}=0$ 表示用户 u 尚未对商品 i 进行评价, 但这并不意味着用户 u 不了解该商品。

定义4(问题的定义) 给定用户 u 、商品 i 、用户历史行为数据、用户(商品)之间的标签相似网络图 $G_u(G_i)$ 、用户(商

$$L_{co} = \underbrace{\sum_{u,i} c_{ui} (y_{ui} - \theta_u^T \beta_i)^2}_{\text{第一部分}} + \underbrace{\lambda_{u,k} \sum_{m_{uk} \neq 0} (m_{uk} - \theta_u^T \eta_k - b_u - d_k)^2 + \lambda_{i,j} \sum_{m_{ij} \neq 0} (m_{ij} - \beta_i^T \gamma_j - \omega_i - c_j)^2}_{\text{第二部分}} + \underbrace{\lambda_{u,u} \text{tr}(\theta_u^T L_u \theta_u)}_{\text{第三部分}} + \underbrace{\lambda_{i,i} \text{tr}(\beta_i^T L_i \beta_i) + \lambda_\theta \sum_u \|\theta_u\|_2^2 + \lambda_\beta \sum_i \|\beta_i\|_2^2 + \lambda_\gamma \sum_j \|\gamma_j\|_2^2 + \lambda_\eta \sum_k \|\eta_k\|_2^2}_{\text{第四部分}} \quad (2)$$

接下来分4个小节介绍目标函数的4个部分。

4.1 用户商品二分图网络

在众多传统推荐算法中, 矩阵分解因具有完善的理论基础、较高的预测精度、良好的扩展性等优点, 广受学术界与工业界的青睐。传统矩阵分解模型的基本思想是: 假设在一个购物推荐系统中存在 n 个商品和 m 个用户, 对于给定的交互矩阵 $Y \in \mathbf{R}^{m \times n}$, 寻求用户特征向量 $\theta_u \in \mathbf{R}^K (u=1, \dots, m)$ 和商品特征向量 $\beta_i \in \mathbf{R}^K (i=1, \dots, n)$, 使得两者乘积所得近似拟合 Y 。通常可基于随机梯度下降(Stochastic Gradient Descent, SGD)求出式(2)的局部最优解 θ 和 β , 具体公式如下所示:

$$L_{mf} = \sum_{u,i} c_{ui} (y_{ui} - \theta_u^T \beta_i)^2 + \lambda_\theta \sum_u \|\theta_u\|_2^2 + \lambda_\beta \sum_i \|\beta_i\|_2^2 \quad (3)$$

其中, c_{ui} 是模型的超参数, 表示模型的置信度, 也用于调节平衡样本 ($y=1$) 远远多于正样本 ($y=0$) 的情况。 $\sum_u \|\theta_u\|_2^2$,

品的文本属性, 本文的目标是通过融合上述多类型辅助信息的推荐算法来预测用户 u 是否对商品 i 感兴趣。

3.2 数据分析

本文以今日头条问答数据¹⁾为例, 为了给每个热门问题找到愿意回答的专家用户将问题推送给他们, 并保证问题有足够多的高质量回答, 头条数据提供了用户问题的标签信息以及它们的文本描述。头条问答数据中有大量的不活跃用户, 因为他们的没有回答过任何问题, 所以推荐系统仅仅根据用户回答问题的历史数据是难以捕捉到不活跃用户的兴趣爱好的。而在标签相似网络中, 拥有相同标签的用户之间的兴趣爱好相似, 如果一位不活跃用户 user1 通过标签相似网络中标签 tag 的连接寻找到相似用户 user2, 那么可以认为 user1 也可能喜欢 user2 喜欢的商品 item1, 因为他们的兴趣是相似的。经由 user1-tag-user2-item1 路径在图上多次跳转后可以寻找到不活跃用户感兴趣的问题, 同时此类跳转也可视为相近用户的兴趣传播。同样, 节点文本也能描述节点间的相似度, 因此也可以在用户主题相似网络中找到类似的多次跳转的路径。综上, 融合多类型辅助信息的推荐算法能够缓解推荐系统的数据稀疏性。

4 融合多类型辅助信息的协同过滤算法

由上述头条案例分析可知, 融合多类型辅助信息的推荐算法能够缓解推荐系统的数据稀疏性。受文献[8]启发, 本文在 Cofactor 模型的基础上提出一种融合多类型辅助信息的联合矩阵分解模型。如式(2)所示, 本文损失函数可由4个部分组成, 除第四部分以外其余部分对应3个相关矩阵。第一部分是用户商品交互网络的损失函数, 对应用户商品交互矩阵; 第二部分是标签相似网络上的损失函数, 对应标签共现矩阵; 第三部分是用户主题相似网络上的损失函数, 对应用户主题上的拉普拉斯矩阵; 第四部分是模型的正则项。

$\sum_i \|\beta_i\|_2^2$ 表示模型的正则化项, 超参数 $\lambda_\theta, \lambda_\beta$ 用于控制正则化程度。此外, 文献[24]的研究表明, 考虑用户和物品的偏置及用户的隐式反馈信息有助于提升算法的性能。

4.2 融合标签信息

在标签上下文信息中, 用户购买商品是因为用户喜欢这一类商品, 比如品牌效应。对于离散的标签数据, 需要进行离散化处理, 因此不能简单地选取余弦距离作为相似度函数。为了寻找有共同兴趣的用户, 受文献[8]启发, 以 SPPMI 定义用户间的相似程度。在 Cofactor 模型中, 商品间的相似度是通过用户历史行为数据来计算商品间共现次数而得到的。但对于冷启动商品, 由于其没有与其他任何商品共现, Cofactor 不能解决商品冷启动问题。而对于标签数据, 只要两件商品有共同的标签就发生一次共现, 即使用户没有同时购买这两

¹⁾ <https://biendata.com/competition/bytecup2016/data/>

件商品。因此基于标签信息的联合矩阵分解比 Cofactor 更适用于缓解商品的冷启动。

SPPMI 相似度函数被广泛用于自然语言处理中,在文献[10]中,PMI 矩阵中的每一行对应一个单词,每一列对应一个上下文单词,矩阵中每个元素 m_{ij} 表示单词和上下文单词之间的联系,通过降维(如奇异值分解)来学习得到词嵌入。点互信息(PMI)计算方式如式(4)所示,其中 $\#(i, j)$ 表示单词 i 和单词 j 的共现次数, $\#(i)$, $\#(j)$ 分别表示词 i 和词 j 在语料库出现的次数, D 表示语料库所有词出现的次数。式(5)中 k 是词嵌入负采样本值,因为人们很容易想到词之间的正向联系,两个词的相似度更容易被共有的正向上下文影响,而非负面的上下文则用 0 表示负值。鉴于 SPPMI 矩阵的负样本取样的优势,本文选择了相同的做法来构建用户标签共现矩阵,节点有共同标签则计共现一次,具体公式如下所示。

$$PMI(i, j) = \log \frac{\#(i, j) \cdot D}{\#(i) \#(j)} \quad (4)$$

$$SPMI(i, j) = PMI(i, j) - \log k \quad (5)$$

$$SPPMI(i, j) = \max\{PMI(i, j) - \log k, 0\} \quad (6)$$

4.3 融合文本信息

在文本上下文信息中,用户是否购买商品会受商品的评价信息的影响。具有相似节点属性的用户(商品)比具有不同节点属性的用户(商品)具有更高的相关性。为了融合文本信息,对于连续型数据,本文根据第 3.1 节中定义 2 的方法提取文本特征向量,计算节点的相似度,构建出文本主题相似性网络,并且以节点的文本构建拉普拉斯矩阵 L 作为联合矩阵分解模型的约束项,其中 $L = D - W$, D 为度矩阵, W 为邻接矩阵。如式(7)所示, θ_i , θ_j 表是两个节点的低维特征向量, W_{ij} 值越大,说明节点 i 和节点 j 的相关性越大,即 $\|\theta_i - \theta_j\|^2$ 值越小,两节点距离就越近。

$$\min \sum_{i=1}^n \sum_{j=1}^n \|\theta_i - \theta_j\|^2 W_{ij} = \text{tr}(\theta^T L \theta) \quad (7)$$

4.4 融合多类型附属信息

在现实生活中用户是否购买某物品并不是一种或两种因素所决定的,因此建模需要考虑多种信息对用户兴趣的影响。当下推荐系统亟待解决的问题是如何融合多类型数据来实现数据的价值最大化,文献[25]认为解决这一问题的关键在于数据的融合,本节提出了一种融合多源异构数据的联合矩阵分解模型。

本文定义用户-商品交互矩阵 $Y \in \mathbf{R}^{m \times n}$; 经式(6)计算得到的用户标签网络 SPPMI 矩阵 $m_{uk} \in \mathbf{R}^{m \times m}$ 及商品标签网络 SPPMI 矩阵 $m_{ij} \in \mathbf{R}^{n \times n}$; 由用户商品文本信息得到的拉普拉斯矩阵 L_u, L_i ; 经联合矩阵分解得到的用户潜在特征向量 $\theta_u \in \mathcal{R}^k$ ($u = 1, \dots, m$) 和物品潜在特征向量 $\beta_i \in \mathbf{R}^k$ ($i = 1, \dots, n$)。综合第 4.1-4.3 节的内容,本文得出联合矩阵分解的目标损失函数式(2)。式(2)分为 4 部分,后面 3 个部分可认为是对第一项的约束。在第二部分用户商品标签网络的损失函数中, η_u, γ_i 分别为用户标签相似网络及商品标签网络的上下文向量,为了防止过拟合,本文添加了用户商品偏差 b_u, w_i 和上下文偏差 d_k, c_j 。最后一部分是为了防止模型过拟合,对特征向量 $\theta_u, \beta_i, \theta_k, \gamma_i$ 进行的正则化处理。尺度参数 c_{ui} 是一个超参数,它的相对比例将会平衡模型中的矩阵分解和其他 3 部分,

本文设 $c_{ui} = l(1 + \alpha \in y_{ui})$, l 为尺度系数。当 $y_{ui} > 0$ 时, c_{ui} 关于 y_{ui} 呈线性递增,表示对于有行为的物品,行为强度越大,用户 u 对物品 i 越感兴趣;当 $y_{ui} = 0$ 时,置信度恒等于 1,表示对所有没有行为的商品,用户不感兴趣的置信度都是一样的,且比有交互行为商品的置信度低。在实证研究中,利用基于验证集的推荐结果来选择 c_{ui} 的取值,正则化参数 $\lambda_\theta, \lambda_\beta, \lambda_\gamma, \lambda_\eta$ 也是从验证数据中选出的超参数。超参数 $\lambda_{uk}, \lambda_{ij}, \lambda_{uu}, \lambda_{ii}$ 与 c_{ui} 作用类似,用于调整各个网络的平衡。目标函数中的第一到第三部分同时包含因子 θ_u, β_i , 因此可以把上述目标函数理解为对矩阵分解进行正则化处理,正则项为第二、第三部分。对于式(2)的求解,本文使用了交替最小二乘法的优化方法,取损失函数的梯度,并设置其值为 0,由此得出以下参数更新的推导,具体公式如下:

$$\theta_u \leftarrow (\sum_i c_{ui} \beta_i \beta_i^T + \sum_{k: m_{uk} \neq 0} \eta_k \eta_k^T + (L_{uu} + \lambda_\theta) I_k)^{-1} * (\sum_i c_{ui} y_{ui} \beta_i^T + \sum_{k: m_{uk} \neq 0} (m_{uk} - b_u - d_k) \eta_k) \quad (8)$$

$$\beta_i \leftarrow (\sum_u c_{ui} \theta_u \theta_u^T + \sum_{j: m_{ij} \neq 0} \gamma_j \gamma_j^T + (L_{ii} + \lambda_\beta) I_k)^{-1} * (\sum_u c_{ui} y_{ui} \theta_u^T + \sum_{j: m_{ij} \neq 0} (m_{ij} - w_i - c_j) \gamma_j) \quad (9)$$

$$\gamma_j \leftarrow (\sum_{i: m_{ij} \neq 0} \beta_i \beta_i^T + \lambda_\gamma I_k)^{-1} * (\sum_{i: m_{ij} \neq 0} (m_{ij} - w_i - c_j) \beta_i) \quad (10)$$

$$\eta_k \leftarrow (\sum_{i: m_{uk} \neq 0} \theta_u \theta_u^T + \lambda_\theta I_k)^{-1} * (\sum_{u: m_{ij} \neq 0} (m_{uk} - b_u - d_j) \beta_i) \quad (11)$$

$$w_i \leftarrow \left(\frac{1}{\{j: m_{ij} \neq 0\}} \sum_{j: m_{ij} \neq 0} m_{ij} - \theta_i^T \gamma_j - c_j \right) \quad (12)$$

$$c_j \leftarrow \left(\frac{1}{\{i: m_{ij} \neq 0\}} \sum_{i: m_{ij} \neq 0} m_{ij} - \beta_i^T \gamma_j - w_i \right) \quad (13)$$

$$b_j \leftarrow \left(\frac{1}{\{k: m_{uk} \neq 0\}} \sum_{k: m_{ij} \neq 0} m_{uk} - \theta_u^T \eta_k - d_k \right) \quad (14)$$

$$d_k \leftarrow \left(\frac{1}{\{u: m_{uk} \neq 0\}} \sum_{u: m_{ij} \neq 0} m_{uk} - \theta_u^T \eta_k - b_u \right) \quad (15)$$

为了提升计算速度,本算法在用户和物品之间进行并行化计算,显著提高了模型求解的计算效率,我们把交替最小二乘法的每一次更新看作是一次加权岭回归。

5 实验与分析

在实验设计和结果分析的部分,根据推导的求解公式实现本文所提出的融合多类型辅助信息的协同过滤算法。为了验证本文提出的算法的有效性,下面将首先说明相应的评价标准,然后介绍数据集以及对算法,最后给出所提模型与其他方法的对比实验结果,并对实验结果进行相应的分析。

5.1 度量标准

为了评估不同模型之间的性能,本文采用了归一化折扣累计增益 NDCG@M 和平均准确均值 MAP@M 两个指标作为模型性能的评价标准,计算公式如下:

$$DCG@M(u, \pi) = \sum_{i=1}^M \frac{2^{1(u(\pi(i))=1)} - 1}{\log(i+1)} \quad (16)$$

$$AP@M(u, \pi) = \sum_{i=1}^M \frac{Precision@i(u, \pi)}{(\min(i, \sum_{i'} 1\{u(\pi(i'))=1\}))} \quad (17)$$

其中, π 表示所有物品,如果用户消费了物品 $\pi(i)$,则 $u(\pi(i))$ 等于 1。DCG 的主要思想是将用户喜欢的物品排在前面比排在后面能在更大程度上提高用户体验。式(16)的分子可以理解为推荐列表中每一个物品的增益,该值随用户心理预期的排位递减(这里是指数递减),分母可以理解为排名不同的物

品的增益应该被赋予的不同的折算因子,排名越靠后则对应的折算越多,即最后的增益越小。准确率(precision)形式化可定义为 $p_n = TP / (TP + FP)$,表示的是预测为正的样本中有多少是真正的正样本。准确率通常只考虑返回结果中相关物品的个数,没有考虑物品之间的排序。考虑到结果是有序列表,需要求得结果中每个位置,因此 precision 之后对所有位置的 precision 求均值,即为公式(17)的定义。而 MAP@M 是计算所有用户平均精度(AP)的均值。

5.2 数据集

今日头条问答数据集:该数据集来源于今日头条在 bien-data 上举办的推荐算法比赛,竞赛的任务是为头条问答的问题在今日头条专家用户中匹配潜在的答案贡献者。数据集包含 27313 位用户对 8090 个问题所作的 33981 条回答记录以及用户和问题的文本描述,稀疏度为 0.015%,除此之外还包含 143 个用户标签和 20 个问题标签。每个用户平均有两个用户标签,每个问题有一个问题标签。其中至少回答过问题的用户占 21.6%,大部分用户是不活跃用户。

Amazon Software 数据集¹⁾:该数据集聚集包含 36049 位用户对 2129 个商品的 44398 条评价数据,稀疏度为 0.058%,除此之外还包含商品的本文评论以及商品之间的“also-bought”连接关系。为了吻合隐式反馈数据,实验中将评分大于 4 的样本作为正样本,其余为负样本。

5.3 实验结果与分析

我们将对比方法分为两类:

1)基于矩阵分解的模型:WMF(2008)^[26],CoF(2016),HoP-Rec(2018)^[27];

2)基于图表示学习的模型:HPE(2016)^[28],Metapath2vec(2017),MetaPath2vec++(2017)。

我们选择这些模型的理由如下:第一点,这些对比方法都是被广泛用于推荐系统的模型。第二点,为了证明融合多类型辅助信息的协同过滤算法能缓解数据稀疏性,其中的一些模型融合了辅助信息,一部分没有融合辅助信息。以下是各个模型的介绍和参数设置:

1)WMF:直接将交互矩阵低秩分解为用户特征矩阵和商品特征矩阵,没有融合辅助信息。

2)CoF:CoFactor 模型认为在同一次购买的商品之间是相似的,以用户历史行为构建商品共现矩阵,联合交互矩阵共同矩阵分解,没有融合辅助信息。

3)Hop-Rec:将 MF 与 Graph-based 模型相结合,通过用户在用户商品二分图上的随机游走,结合顶点的度,以一定概率采样不同的正样本,并为不同跳数得到的排序对赋予衰减系数,没有考虑辅助信息。

4)基于图表示的推荐算法:HPE, Metapath2vec, MetaPath2vec++都是在用户商品二分图上随机游走学得用户商品低维表示,不同的是 Metapath2vec 和 MetaPath2vec++预先定义了随机游走的路径方式,认为不同类型的游走方式会得到不同的语义,而 HPE 是随机游走,随机选取下个邻居节点。这些模型把标签信息也当做节点,考虑了不同类型的边

信息,但是没有考虑节点的辅助信息。

5)参数设置:上述方法的所有超参数和学习率都经过了网格搜索的仔细调整,表 1 和表 2 列出的结果都是在优化的超参数下得到的。在我们的实验中,文本的主题特征向量由 LDA 得到,超参数经验证集调整为最优,实验结果发现,当模型的用户商品特征向量维度设置为 100 时,效果最好。

表 1 和表 2 列出了在今日头条问答和 Amazon_Software 数据集上的 NDCG@10 值和 MAP@10 值。实验结果显示,本文提出的方法比所有基线方法的效果更好,其中在今日头条问答数据集上 NDCG 提升了 5.68%,MAP 提升了 9.23%,在 Amazon_Software 数据集上 NDCG 提升了 0.87%,MAP 提升了 3.53%,证明了融合多类型辅助信息能有效地缓解数据稀疏性问题。

表 1 在今日头条问答数据集上的实验结果

	NDCG@10	MAP@10
WMF	0.2146	0.1472
CoF	0.2042	0.1429
Hop-Rec	0.2101	0.1474
HPE	0.1877	0.1331
Metapath2vec	0.1741	0.1206
Metapath2vec++	0.1792	0.1247
Our Model	0.2268	0.1610

表 2 在 Amazon Software 数据集上的实验结果

	NDCG@10	MAP@10
WMF	0.0805	0.0650
CoF	0.0746	0.0521
Hop-Rec	0.0803	0.0631
HPE	0.0570	0.0493
Metapath2vec	0.0427	0.0373
Metapath2vec++	0.0459	0.0386
Our Model	0.0812	0.0673

结束语 本文提出了一种融合多种类型辅助信息的矩阵分解方法,并且通过实验证明了这种联合矩阵分解形式为不同的推荐指标带来了性能提升,能够有效地缓解推荐系统数据的稀疏性。在以后的工作中,我们考虑将该模型与其他模型相结合,从而提出更高性能的混合模型。

参考文献

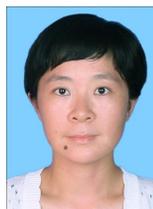
- [1] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [2] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.
- [3] MNIH A, SALAKHUTDINOV R R. Probabilistic matrix factorization[C]// Advances in Neural Information Processing Systems, 2008: 1257-1264.
- [4] LUO X, ZHOU M, XIA Y, et al. An efficient non-negative matrix-factorization-based approach to collaborative filtering for

¹⁾ <https://jmcauley.ucsd.edu/data/amazon/>

- recommender systems[J]. *IEEE Transactions on Industrial Informatics*, 2014, 10(2): 1273-1284.
- [5] SINGH A P, GORDON G J. Relational Learning via Collective Matrix Factorization[C]// *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Las Vegas, Nevada, USA, ACM, 2008.
- [6] BOUCHARD G, YIN D, GUO S. Convex collective matrix factorization[C]// *Artificial Intelligence and Statistics*. 2013: 144-152.
- [7] SINGH A P, GORDON G J. A unified view of matrix factorization models[C]// *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Berlin, Heidelberg: Springer, 2008: 358-373.
- [8] LIANG D, ALTOSAAR J, CHARLIN L, et al. Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence[C]// *Proceedings of the 10th ACM Conference on Recommender Systems*. 2016: 59-66.
- [9] MIKOLOV T. Distributed Representations of Words and Phrases and their Compositionality[J]. *Advances in Neural Information Processing Systems*, 2013, 26: 3111-3119.
- [10] LEVY O, GOLDBERG Y. Neural word embedding as implicit matrix factorization[J]. *Advances in Neural Information Processing Systems*, 2014, 3: 2177-2185.
- [11] ZHANG K H, LIANG J Y, ZHAO X W, et al. A collaborative filtering recommendation algorithm based on information of community experts[J]. *Journal of Computer Research and Development*, 2018, 55(5): 968-976.
- [12] YU Y H, GAO Y, WANG H, et al. Integrating user social status and matrix factorization for item recommendation[J]. *Journal of Computer Research and Development*, 2018, 55(1): 113-124.
- [13] MOHSEN J, MARTIN E. A matrix factorization technique with trust propagation for recommendation in social networks[C]// *Proceedings of the Fourth ACM Conference on Recommender Systems*. 2010: 135-142.
- [14] LIANG D, ZHAN M, ELLIS D P W. Content-Aware Collaborative Music Recommendation Using Pre-trained Neural Networks [C]// *ISMIR*. 2015: 295-301.
- [15] AMJAD A, KYLE K, KYUNGHYUN C, et al. Learning distributed representations from reviews for collaborative filtering [C]// *Proceedings of the 9th ACM Conference on Recommender Systems*. 2015: 147-154.
- [16] HE R N, MCAULEY J L J. VBPR: visual bayesian personalized ranking from implicit feedback[J]. *arXiv: 1510. 01784*, 2015.
- [17] KIM D H, PARK C Y, OH J, et al. Convolutional matrix factorization for document context-aware recommendation [C]// *Proceedings of the 10th ACM Conference on Recommender Systems*. 2016: 233-240.
- [18] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: Online learning of social representations [C]// *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2014: 701-710.
- [19] BARKAN O, KOENIGSTEIN N. Item2vec: neural item embedding for collaborative filtering[C]// *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2016: 1-6.
- [20] SUN X, GUO J, DING X, et al. A general framework for content-enhanced network representation learning[J]. *arXiv: 1610. 02906*, 2016.
- [21] DONG Y, CHAWLA N V, SWAMI A. Scalable representation learning for heterogeneous networks[C]// *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017: 135-144.
- [22] SHI C, HU B, ZHAO W X, et al. Heterogeneous information network embedding for recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 31(2): 357-370.
- [23] ZHANG F, YUAN N J, LIAN D, et al. Collaborative Knowledge Base Embedding for Recommender Systems[J]. *KDD*, 2016: 353-362.
- [24] KOREN Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model[C]// *Proc of the 14th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining (KDD'08)*. New York: ACM, 2008: 426-434.
- [25] MENG X F, DU Z J. Research on the big data fusion: Issues and challenges[J]. *Journal of Computer Research and Development*, 2016, 53(2): 231-246.
- [26] HU Y, KOREN Y, VOLINSKY C. Collaborative filtering for implicit feedback datasets[C]// *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008: 263-272.
- [27] YANG J H, CHEN C M, WANG C J, et al. HOP-rec: high-order proximity for implicit recommendation[C]// *Proceedings of the 12th ACM Conference on Recommender Systems*. 2018: 140-144.
- [28] CHEN C M, TSAI M F, LIN Y C, et al. Query-based music recommendations via preference embedding[C]// *Proceedings of the 10th ACM Conference on Recommender Systems*. 2016: 79-82.



HAO Zhi-feng, born in 1968, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include various aspects of algebra, machine learning, data mining and evolutionary algorithms.



WEN Wen, born in 1981, Ph.D, professor, is a member of China Computer Federation. Her main research interests include machine learning, graph embedding and sequential data analysis.