



计算机科学

COMPUTER SCIENCE

基于概率元学习的矩阵补全预测融合算法

齐秀秀, 王佳昊, 李文雄, 周帆

引用本文

齐秀秀, 王佳昊, 李文雄, 周帆. [基于概率元学习的矩阵补全预测融合算法](#) [J]. 计算机科学, 2022, 49(7): 18-24.

QI Xiu-xiu, WANG Jia-hao, LI Wen-xiong, ZHOU Fan. [Fusion Algorithm for Matrix Completion Prediction Based on Probabilistic Meta-learning](#) [J]. Computer Science, 2022, 49(7): 18-24.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于神经架构搜索的点击率预测模型](#)

Click-Through Rate Prediction Model Based on Neural Architecture Search

计算机科学, 2022, 49(7): 10-17. <https://doi.org/10.11896/jsjcx.210600009>

[基于数据增广和模型集成策略的图神经网络在抑郁症识别上的应用](#)

Application of Graph Neural Network Based on Data Augmentation and Model Ensemble in Depression Recognition

计算机科学, 2022, 49(7): 57-63. <https://doi.org/10.11896/jsjcx.210800070>

[一种改进的融合相似度和信任度的协同过滤算法](#)

Improved Collaborative Filtering Algorithm Combining Similarity and Trust

计算机科学, 2022, 49(6A): 238-241. <https://doi.org/10.11896/jsjcx.210400088>

[基于点割集图分割的矩阵变换与分解的推荐算法](#)

Matrix Transformation and Factorization Based on Graph Partitioning by Vertex Separator for Recommendation

计算机科学, 2022, 49(6A): 272-279. <https://doi.org/10.11896/jsjcx.210600159>

[基于遗憾探索的竞争网络强化学习智能推荐方法研究](#)

Study on Intelligent Recommendation Method of Dueling Network Reinforcement Learning Based on Regret Exploration

计算机科学, 2022, 49(6): 149-157. <https://doi.org/10.11896/jsjcx.210600226>

基于概率元学习的矩阵补全预测融合算法

齐秀秀 王佳昊 李文雄 周帆

电子科技大学信息与软件工程学院 成都 610054

(xiuxqihm@gmail.com)

摘要 随着互联网社交媒体规模的飞速发展,利用推荐算法对海量信息进行有效建模筛选和过滤,成为了研究用户行为偏好、热点倾向和网络安全态势等问题的关键。随着深度学习的发展,图神经网络模型在解决推荐系统应用中的密集型图结构数据时取得了较好效果。协同过滤算法作为得到最广泛应用的推荐算法,其利用用户-项目的群体交互数据来预测用户未来的偏好与项目评级。但现有的推荐算法仍面临着数据稀疏和冷启动问题,且缺少对不确定性的良好量化。文中提出了一种基于概率元学习的归纳矩阵补全预测融合算法(MetaIMC),该算法从贝叶斯推断的角度重新对元学习进行表征,构建了稳健的图深度神经网络元学习模型,充分利用数据先验知识提出从稀疏数据中学习新任务的解决方案。首先,MetaIMC可以有效地利用变分贝叶斯推理获得先验分布,缓解元模型任务训练中的不确定性和模糊性问题,进一步提升了模型的泛化能力;其次,在不借助任何用户边信息的情况下,实现新用户推荐的冷启动;最后,在传统矩阵补全及用户冷启动两个场景下,利用 Flixster, Douban 和 Yahoo_music 3 个公开数据集对模型的性能进行了评估,验证了 MetaIMC 在面对传统矩阵补全任务时的有效性,并在冷启动问题上达到了最优的效果。

关键词: 推荐系统; 矩阵补全; 图神经网络; 元学习; 变分贝叶斯推断

中图法分类号 TP181

Fusion Algorithm for Matrix Completion Prediction Based on Probabilistic Meta-learning

QI Xiu-xiu, WANG Jia-hao, LI Wen-xiong and ZHOU Fan

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

Abstract With the rapid development of Internet social media, using recommendation algorithms to effectively model and filter massive amounts of information has become the key to predict user behavior preferences, hot spot tendency, network security situation and other issues. At the same time, with the development of deep learning, graph neural network model has achieved good results in solving the dense graph structure data in recommendation system. Collaborative filtering algorithm, as the most widely used recommendation algorithm, uses user-item group interaction data to predict users' future preferences and item ratings. However, existing recommendation algorithms still face the problems of data sparseness and cold start, and lack of a good quantification of uncertainty. This paper proposes an inductive matrix completion prediction fusion algorithm based on probabilistic meta-learning (MetaIMC), which re-characterizes meta-learning from the perspective of Bayesian inference, builds a robust GNN-meta-learning model, and makes full use of data priors to build solutions for learning new tasks from sparse data. Firstly, MetaIMC can effectively use variational Bayesian inference to obtain the prior distribution, alleviate the uncertainty and ambiguity in the meta-model task training, and further improve the generalization ability of the model. Secondly, MetaIMC can implement new user recommendations and solve the cold start problem without any user side information. Finally, in the two scenarios of traditional matrix completion and user cold start, the performance of the model is evaluated by using three public datasets of Flixster, Douban and Yahoo_music, which verifies the effectiveness of MetaIMC on traditional matrix completion task, and achieves the best performance on the cold start problem.

Keywords Recommendation system, Matrix completion, Graph neural network, Meta-learning, Variational Bayesian inference

到稿日期:2021-06-16 返修日期:2021-10-21

基金项目:电子科技大学-智小金-智能家居联合研究中心项目(H04W210180);内江市科技孵化和成果转化专项资金(2021KJFH004)

This work was supported by the UESTC-ZHIXIAOJING Joint Research Center of Smart Home(H04W210180) and Neijiang Technology Incubation and Transformation Funds(2021KJFH004).

通信作者:王佳昊(wangjh@uestc.edu.cn)

1 引言

随着互联网的飞速发展,网络用户的社交行为日益丰富,数据交互信息量呈爆炸式增长,“信息过载”的问题日益突出。推荐算法作为解决资讯过滤的技术手段之一,获得了研究者们越来越多的关注。利用推荐算法对海量信息进行有效的建模筛选和过滤,成为了研究用户行为偏好、热点倾向和网络安全态势等问题的关键。现有的推荐系统利用用户的历史行为数据和社交关系来帮助用户进行信息和数据的有效过滤,并建立用户兴趣偏好模型。其中,基于内容的推荐系统^[1]以用户和商品的内容信息为输入信息,以预测用户下一次将购买的商品或对商品的评级。协同过滤算法^[2]作为应用最广泛的推荐算法,其通过集体交互数据来预测用户未来将购买的物品或对项目的偏好评级,该评级预测任务也被当作矩阵补全(Matrix Completion, MC)任务,即基于矩阵中已有的数据来填补矩阵中的未知矩阵值。

随着深度学习的发展,基于图神经网络(Graph Neural Network, GNN)^[3]的模型在网络表示学习领域取得了成功,并被广泛应用于推荐系统领域。Kalofolias 等^[4]提出的几何矩阵补全(Geometric Matrix Completion, GMC)模型引入了矩阵补全模型的正则化,构建用户图和项目图并添加边信息(Side-information)。Rao 等^[5]将一种更有效的交替最小二乘优化方法引入图正则化矩阵补全问题。Monti 等^[6]开发了一种多图 CNN(Multi-graph CNN, MGCNN)模型,在矩阵补全中加入基于图的边信息,从用户和项目各自的近邻网络中提取潜在特征,并结合递归神经网络对动态评级生成过程进行建模。Berg 等^[7]提出了一种基于图的矩阵补全自动编码器框架——图卷积矩阵补全(Graph Convolutional Matrix Completion, GC-MC)。但 MGCNN 需要不能泛化到新任务的图拉普拉斯数,而 GC-MC 使用节点 ID 的一次性热编码作为初始特征,并将其输入到 GNN,因此两种模型都不能泛化到新的用户或项目。基于迁移学习和对矩阵补全的归纳研究,Zhang 等提出了针对预测目标提取局部子图并进行归纳预测的矩阵补全模型(Inductive Matrix Completion, IMC)^[8]。

已有的推荐系统通过收集大量用户的偏好信息来评估用户偏好,并基于与目标用户评级相似的其他用户的现有评级进行推荐。然而,真实的社交媒体中以下 3 个方面存在局限性:1)不同用户偏好的条目数量极不平衡,大多数用户仅利用极少评价行为来预测项目评级,这使得用户很难了解用户偏好,无法准确推荐项目;2)在实际推荐系统中,高质量的内容并非总是可用的,新用户可能只有稀疏的用户-项目交互,甚至没有任何边信息,面对数据稀疏和冷启动问题,一般的归纳矩阵模型较难向新用户推荐新项目;3)此外,为了在矩阵补全中获得最好的可能结果,除了良好的预测精度外,模型还需具有对不确定性的良好度量,即当预测评级接近正确评级时,表示高度自信,当预测结果的可靠性低时,表示自信不足。

为了解决上述实践问题,本文结合元学习算法,将与模型无关的元学习^[9](Model-Agnostic Meta-Learning algorithm, MAML)应用到图结构数据,提出了一种基于概率元学习的归纳矩阵补全预测融合算法(MetaIMC)。该算法使用变分贝

叶斯推理来减小模型的不确定性,充分利用先验知识提出了从稀疏数据中学习新任务的解决方案,同时强化了模型对任务的确定能力,减少了在信息量不足的情况下的模糊边界,进一步增强了图模型应用的泛化能力。本文的主要创新点和贡献如下:

(1)提出了一种基于元学习和贝叶斯推断的融合算法,并结合矩阵补全任务提出了归纳矩阵补全预测模型 MetaIMC。MetaIMC 可以有效地利用变分贝叶斯推理获得先验分布,并以先验知识和经验为基础迅速适应新用户的回归预测,进一步提升模型的泛化能力。

(2)MetaIMC 在不借助任何用户边信息的情况下,通过训练基于评价矩阵的图神经网络来生成目标对(用户-项目)的局部子图,实现基于概率元学习的归纳矩阵补全预测任务。

(3)在 Flixster, Douban 和 Yahoo_music^[10-12] 3 个公开数据集上对模型的性能进行了评估,并选取了多种基线算法,在传统矩阵补全及用户冷启动两个场景下与 MetaIMC 模型进行对比,实验结果证明了 MetaIMC 在两个任务上的有效性和先进性。

2 相关工作

2.1 基于图神经网络的矩阵补全模型

矩阵补全^[13]是一种常见的推荐系统公式,其中矩阵的行和列分别代表用户和项目,预测用户对项目的兴趣评级相当于填补评级矩阵中缺失的条目。在观测数据非常稀疏的条件下本文将考虑矩阵补全,预测形状为 $N_u \times N_o$ 的评级矩阵 \mathbf{R} 的值 $r_{i,j}$, $i \in \{1, \dots, N_u\}$, $j \in \{1, \dots, N_o\}$, 其中 N_u 为用户数量, N_o 为项目数量。

在以往的研究中, Berg 等提出的图卷积矩阵补全^[7](Graph Convolutional Matrix Completion, GC-MC)是一种基于图的矩阵补全自动编码器框架,其将矩阵补全视为一个图上的链接预测问题。协同过滤中的交互数据可以通过用户和项目节点之间的二部图来表示,自动编码器通过在二部交互图上传递消息的形式,产生用户和项目节点的潜在特征,观察到的评级用链接表示,内容信息可以自然地以节点特性的形式包含在这个框架中,将预测评级问题转化为预测二部图中用户-项目的标记连接。

2.2 基于二部图模式的链路预测问题

学习监督启发式图模式已被应用于简单图中的链路预测。其中, Zhang 等^[14]提出的 Weisfeiler-Lehman Neural Machine(WLNM)使用子图邻接矩阵上的全连接神经网络学习图结构特征,然后用 GNN 代替全连接的神经网络,获得了最优的链路预测结果。与该方法不同,本文研究不使用任何预定义的启发式方法,而是使用 GNN 学习一般的图结构特征。与本文研究相似, Li 等^[15]利用图核来学习图结构特征。但图核计算和存储核矩阵需要更高维的时间复杂度和空间复杂度,因此不适合现代推荐系统。

转化为链路预测问题后,评级矩阵 \mathbf{R} 中的值 $r_{i,j}$ 可以对应用户-项目交互的评级(例如,评级在 1~5 之间,未评级则默认为 0), \mathbf{R} 可转化为具有二部结构的用户-项目交互图,其中边对应交互事件,边的权值表示用户对特定物品的评级。

2.3 概率元学习

元学习 (Meta-Learning, Learn to learn)^[16] 又称学会学习。元学习方法的目标是学习一个可以从少量数据中学习新任务的模型。在众多元学习方法中,与模型无关的元学习被广泛应用于分类和回归任务,其通过提取来自同一任务族的不同任务之间的共享结构,来学习新任务的先验知识,以增强预测模型的学习能力。在已有任务中学习先验知识,元学习器能够在有限数据的情况下快速适应新任务。针对每个任务,在最大似然设置下的 MAML 目标为:

$$\mathcal{L}_{\eta_i}(\theta) = [-\log p(Q_i | \theta - \alpha \nabla_{\theta} - \log p(\mathcal{S}_i | \theta))] \quad (1)$$

变分贝叶斯 (Variational Bayesian, VB)^[17] 用于近似贝叶斯推理和计算复杂 (Intractable) 积分。概率元学习的层次变

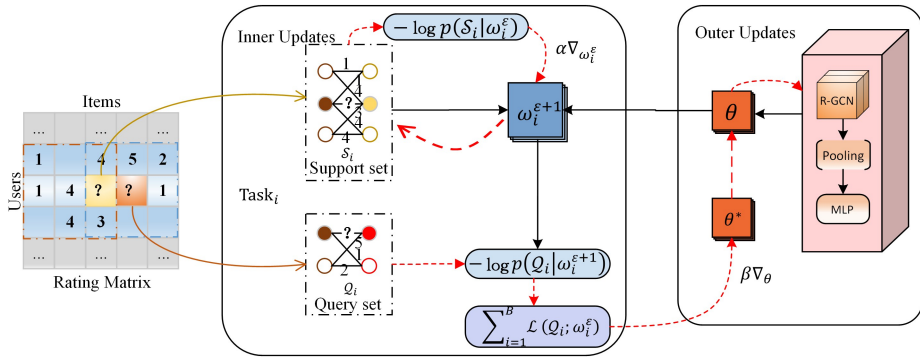


图 1 MetaIMC 的模型构架

Fig. 1 Model architecture of MetaIMC

3.1 基础网络组成

3.1.1 局部封闭子图的特征提取

在稀疏评级矩阵中,为每个未知评级矩阵值 $r_{i,j}$ 提取一个 h -跳封闭子图^[8] (h -hop Enclosing Subgraph, h -SG),局部封闭子图可形式化定义为:

$$\begin{bmatrix} r_{i-h,j-h} & \cdots & r_{i-h,j+h} \\ \vdots & r_{i,j} & \vdots \\ r_{i+h,j-h} & \cdots & r_{i+h,j+h} \end{bmatrix}$$

为此,给定 $G = (U, E, \mathcal{R}, \mathcal{N}_r(u))$,用于表示由给定的评级矩阵 \mathbf{R} 构造的无向加权二部图。1) U 代表节点集,节点代表 \mathcal{R} 中的用户 u_i 或项 v_j ; 2) 每条边 $e_{i,j}(u_i, v_j) \in E$ 有一个评级权重 $r_{i,j}$,代表用户 u_i 对项目 v_j 的评级; 3) \mathcal{R} 表示所有可能评级的集合, $\mathcal{N}_r(u)$ 表示关系 $r_{i,j}$ 下连接到 u_i 的邻居集合。对于每个观察评级 $r_{i,j}$,提取 h 跳封闭子图 $G_{i,j}^h$,本文研究取 $h=1$ 。

3.1.2 图神经网络链路预测模型

MetaIMC 使用一个图级神经网络^[23] 作为元学习器 f_{θ} ,构建围绕 (u_i, v_j) 的局部封闭子图,并将其直接映射到评级。元学习器主要由信息传递层和池化层两部分组成。

具体而言,MetaIMC 采用关系图卷积算子 (Graph Convolutional Operator, R-GCN)^[24] 作为消息传递层,为子图中的每个节点提取一个特征向量:

$$\mathbf{h}_i^{l+1} = \sigma \left(\sum_{r \in \mathcal{R}} \sum_{k \in \mathcal{N}_r(i)} \frac{1}{|\mathcal{N}_r(i)|} \mathbf{W}_k^l \mathbf{h}_k^l + \mathbf{W}_0^l \mathbf{h}_i^l \right) \quad (2)$$

其中, \mathbf{h}_i^l 表示节点在第 l 层的特征向量 (隐藏状态), σ 表示激活函数。 \mathbf{h}_0^l 和 \mathbf{h}_k^l ($r \in \mathcal{R}$) 为神经网络的可学习参数 (权重矩阵), r 代表评级矩阵 \mathbf{R} 中节点 i 的评级。由于邻居 k 与节点 i 的

分贝叶斯公式^[18-22] 的提出,为处理从少量数据中学习时产生的不确定性和模糊性提供了新的解决思路。本文研究从贝叶斯推断的角度重新对元学习算法进行表征,通过贝叶斯层次模型来形式化元学习,它不直接优化参数,而是使用贝叶斯推断进行学习,最终为元知识 (参数) 提供不确定性度量。MetaIMC 使用平摊变分推论^[18] 来估计网络参数的后验分布,增强基于元学习的回归归纳矩阵补全。

3 基于概率元学习的矩阵补全预测的融合算法

MetaIMC 的模型框架如图 1 所示,该算法利用局部图模式,提出了一种新的不使用任何边信息的归纳矩阵补全方法,该方法通过构造二部图的链路预测模型来预测未知的评级。

联系拥有不同的评级值/边 r ,并由不同的参数矩阵 \mathbf{W}_r^l 处理,因此元学习器能够从局部子图丰富的图形模式中进行学习。 $|\mathcal{N}_r(i)|$ 是一个特定于问题的规范化常量, \tanh 表示激活函数,用于在两层之间叠加传递 L 层的消息。通过将不同层的特征向量连接在一起,得到节点 i 的最终表示 $x_i = ([\mathbf{h}_i^1, \mathbf{h}_i^2, \dots, \mathbf{h}_i^L])$ 。

基础图神经网络框架的第二个组件是池化层,即对来自节点表示的子图特征向量进行总结。目标用户-项目的最终表示可形式化定义为:

$$\mathbf{p} = ([\mathbf{x}_u, \mathbf{x}_v]) \quad (3)$$

值得注意的是,尽管该池化层操作简单,但文献[22]已经验证它在矩阵补全任务中相比求和或其他高级池层的性能更好。在得到最终的图形表示之后,模型使用 MLP 来输出预测评级,即:

$$\hat{r}_{ij} = \mathbf{W}_{a2} \sigma(\mathbf{W}_{a1} \mathbf{p}) \quad (4)$$

其中, \mathbf{W}_{a1} 和 \mathbf{W}_{a2} 是 MLP 的参数,将图表示 \mathbf{p} 映射到标量等级 \hat{r}_{ij} , $\sigma(\cdot) = \text{RELU}(\cdot)$ 是激活函数。

3.2 融合算法推导

本文基于贝叶斯推断和梯度的元学习,提出了一种融合算法,该算法利用变分贝叶斯推理来推断元学习任务中网络权重的先验分布 (以及相应的后验分布),以进一步进行不确定性估计,增强模型的泛化能力,提高稀疏矩阵在用户冷启动问题下的适应能力。

3.2.1 任务抽样算法

本文遵循传统少样本学习的设置,以任务 (Task) 为采样训练单位,对原始数据集进行重新采样,形成一组任务簇 $\mathcal{T} =$

$\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_M\}$, 单任务 $\mathcal{T} = \mathcal{S} \cup \mathcal{Q}$ 又被划分成支持集 \mathcal{S} 以及查询集 \mathcal{Q} , 分别用于训练模型和检验训练是否有效。具体地, 将每个用户节点 u_i (对应于 \mathbf{R} 中的一行) 视为一个任务, 则在矩阵补全元训练过程中的任务抽样算法可分为以下步骤:

(1) 随机采样任务子集, $Y_{i,j} \sim y, |Y_{i,j}| = N$, 对于每个用户节点 u_i , 随机抽样 K 个目标评级 \mathcal{R}_i^S , 并对每个用户-项目对 (u_i, v_j) 提取 h -hop 封闭子图, 构成支持集 $\mathcal{G}_i^S = \{(G_{ij}^s, r_{ij}^s)\}_{j=1}^K$;

(2) 从用户 u_i 中剩余的用户-项目对 (u_i, v_j) 中提取 P 个 h -hop 封闭子图 $\mathcal{R}_i^Q (\notin \mathcal{R}_i^S)$, 构成查询集 $\mathcal{Q}_i = \{(G_{ik}^q, r_{ik}^q)\}_{k=1}^P$;

(3) 重复步骤 1-步骤 2 | \mathcal{T} | = M 次。

上述采样过程满足了 K -shot 的元学习回归训练设置。参数更新方式使用与模型无关的元学习算法作为基本元学习算法 f_θ , 则元目标更新为:

$$\min_{\theta} \sum_{\mathcal{T}_i \in \mathcal{P}(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\omega_i}) = \sum_{\mathcal{T}_i \in \mathcal{P}(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta - \alpha \nabla_{\mathcal{T}_i}(f_\theta)}) \quad (5)$$

其中, ω_i 为任务特定参数 (Task-specific Parameters), 表示与任务 \mathcal{T}_i 相关的步长为 α 的负对数似然在初始化全局隐变量 θ 上进行单批梯度下降后更新的参数。由于 ω_i 是从 θ 开始的梯度下降迭代过程, 因此每个 ω_i 都与 θ 具有相同的维数, 本文也将计算 ω_i 的内梯度下降法称为快速适应 (Fast-adaptation)。对于每个任务 \mathcal{T}_i , 算法通过梯度的 $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ 对初始参数 θ 进行局部更新。 $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ 代表带有参数 θ 的任务 i 上的训练损失。局部更新 (Globally-updates) 完成后, 对于所有采样任务, 算法基于 $\mathcal{L}_{\mathcal{T}_i}(f_\theta)$ 对参数 θ 进行全局更新, $\mathcal{L}_{\mathcal{T}_i}$ 是在参数 θ_i 下对任务 i 的测试损失, 其使全局更新后的参数 θ 适用于不同的任务。

3.2.2 元学习与贝叶斯的融合推断

本文提出了一种基于梯度的元学习和贝叶斯推断的融合算法, 其从贝叶斯的角度重新对元学习进行表征, 将 MAML 理解为层次概率模型中的经验贝叶斯, 通过变分推理来推断先验分布, 同时引入了一个最大后验推理概率图模型, 用于匹配和缓解元学习中的任务模糊问题。

受概率元学习最新进展的启发, 本文使用平推变分推理来估计网络参数的后验分布。具体地, 给出 \mathcal{S}_i 来确定变分分布 $q(\omega_i)$, 并通过评估相应的 \mathcal{Q}_i 的变分分布来度量模型的性能。即对于每个任务 \mathcal{T}_i , 只能使用数据 \mathcal{S}_i 来计算变分分布, 变分目标为:

$$\log \left(\prod_{i=1}^B p(\mathcal{S}_i, \mathcal{Q}_i) \right) \geq \mathbb{E}_{\theta \sim q_\theta(\theta)} \left[\sum_{i=1}^B (\mathbb{E}_{\omega_i \sim q_{\omega_i}} \log(p(\mathcal{Q}_i | \omega_i) p(\mathcal{S}_i | \omega_i)) - \text{KL}(q_{\omega_i}(\omega_i) \| p(\omega_i | \theta))) \right] - \text{KL}(q_\theta(\theta) \| p(\theta)) \quad (6)$$

其中, KL 表示 Kullback-Leibler 散度。

通过在支持数据 \mathcal{S}_i 上进行固定的迭代次数 L 来得到近似后验 $p(\omega_i | \theta)$ 。这是通过样本在 \mathcal{S}_i 中的梯度下降来实现的, 其对应于高斯先验 $p(\omega_i | \theta)$ 下的最大后验 (Maximum A Posteriori, MAP) 推理, 然而, 难以处理和得到 $p(\omega_i | \theta)$ 的精确形式, 因此考虑以 $p(\omega_i | \mathcal{S}_i, \theta)$ 的形式推断 ω_i 的后验。则上述目标重新表述为:

$$\arg \min \mathbb{E}_{\theta \sim q_\theta} \left[- \sum_{i=1}^B (\mathbb{E}_{\omega_i \sim q_{\omega_i}} (\omega_i | \mathcal{S}_i) \log(p(\mathcal{Q}_i | \omega_i) + \text{KL}(q_{\omega_i}(\omega_i | \mathcal{S}_i) \| p(\omega_i | \mathcal{S}_i, \theta))) \right] + \text{KL}(q_\theta(\theta) \| p(\theta)) \quad (7)$$

其中, ψ 是 θ 上近似后验的变分参数。对于式 (7) 用 $q_\theta(\omega_i | \mathcal{S}_i)$ 方案近似 $p(\omega_i | \mathcal{S}_i, \theta)$ 的要求, 可以通过用支持集更新参数来获得, 且 ω_i 对应的最大后验估计响应后验 $p(\omega_i | \mathcal{S}_i, \theta)$ 的全局模式。图 2 给出了基于梯度的元学习和变分贝叶斯推断的融合算法参数更新过程, 该过程通过随机初始化变分参数的小批量梯度下降来进行, 包括内部更新和外部更新两部分, 虚线表示变分近似。

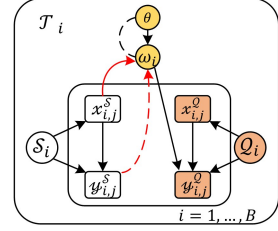


图 2 融合算法的参数更新过程

Fig. 2 Parameter update process of fusion algorithm

基于支持集 \mathcal{S}_i 的内部更新过程如下: 在支持集数据 \mathcal{S}_i 的内部更新 (Inner Updates with \mathcal{S}_i) 中, 推理表现对损失 $\mathcal{L}(\mathcal{S}_i; \omega_i^s)$ 进行固定次数 (ϵ) 的迭代:

$$\mathcal{L}(\mathcal{S}_i; \omega_i^s) = -\log p(\mathcal{Q}_i | \omega_i^s) \quad (8)$$

在每个任务中, 元参数 ϵ 次的迭代更新过程如式 (9) 所示:

$$\text{for } \epsilon = 1, \dots, \epsilon; \omega_i^{\epsilon+1} = \omega_i^\epsilon - \alpha \nabla_{\omega_i^\epsilon} \mathcal{L}(\mathcal{S}_i; \omega_i^\epsilon) \quad (9)$$

其中, $\omega_i^0 = \theta$, α 为任务的学习速率 (Task-Based Learning Rate)。

本文在完成对支持集任务的所有内环更新的基础上, 通过最小化基础网络计算的目标集损失来工作, 以参数在查询集的外部更新过程 (Outer Updates with \mathcal{Q}_i) 来实现优化初始参数值 θ 的目标。在每个任务中, 模型参数 θ 在整体元目标 $\mathcal{L}(\mathcal{Q}_i; \omega_i^s)$ 中使用梯度下降进行更新优化, 初始化参数为:

$$\theta^* = \theta - \beta \nabla_{\theta} \sum_{i=1}^B \mathcal{L}(\mathcal{Q}_i; \omega_i^s) \quad (10)$$

其中, β 为元学习速率。

3.3 变分推理缩放在元参数中的应用与实现

对于融合算法在矩阵补全任务上的实现, 我们设定模型 f_θ 的初始化参数遵循均值为 μ_θ 、方差为 σ_θ^2 的高斯分布, 如式 (11) 所示:

$$p(\theta) = \mathcal{N}(\mu_\theta, \sigma_\theta^2) \quad (11)$$

当逼近后验 $p(\omega_i | \mathcal{S}_i, \theta)$ 时, 在初始参数 $\omega_i^0 = \theta = \{\mu_\theta, \sigma_\theta^2\}$ 的内环更新过程中, 在支持集 \mathcal{S}_i 下经过 ϵ 次迭代来计算 $q_\theta(\omega_i | \mathcal{S}_i)$, 可记为:

$$q_\theta(\omega_i^\epsilon) = q_\theta(\mu_{\omega_i^\epsilon}, \tau_{\omega_i^\epsilon}) \quad (12)$$

其中, 对于每次迭代 $\epsilon = 1, \dots, \epsilon$, τ 为精度, 如式 (13) 所示:

$$\tau_{\omega_i^\epsilon} = \frac{1}{\sigma_{\omega_i^\epsilon}^2(\epsilon)} \quad (13)$$

假设一个平均场变分 (Mean Field Variational) 近似于真后验 $p(\omega_i | \mathcal{S}_i, \theta)$ 。最优因子 $q_\theta(\mu_{\omega_i^\epsilon})$ 和 $q_\theta(\tau_{\omega_i^\epsilon})$ 根据一般变分推理 (General Variational Inference) [25] 有解析解, 并结合外部更新, 可用于下一轮迭代初始化 $p(\theta)$ 。根据高斯-伽马共轭先验分布 (Gaussian-Gamma Conjugate Prior Distribution) 得到:

$$p(\theta) = \mathcal{N}(\mu_\theta; 0, \mathbf{I}) \Gamma(\tau_\theta; \delta, \delta) \quad (14)$$

其中, δ 和 δ 是 Gamma 分布的 alpha 和 beta 参数。MetaIMC 元训练参数的更新过程如算法 1 所示。

算法 1 MetaIMC 训练算法

输入: 训练数据 $(\mathcal{S}, \mathcal{Q})$, 任务学习率 α , 元学习率 β , 内循环更新次数 L
输出: 优化后的模型参数 θ^*

1. 随机初始化模型参数 $\theta = \{\mu_\theta, \sigma_\theta^2\}$;
2. $p(\theta) = \mathcal{N}(\mu_\theta; 0, \mathbf{I}) \Gamma(\tau_\theta; a, b)$;
3. while not done do
4. 取样一个小批次任务, $\mathcal{T}_i \sim \mathcal{T}, i = [1, \dots, B]$;
5. for each \mathcal{T}_i do
6. $\mu_{\omega_i}^0 = \mu_\theta; \sigma_{\omega_i}^{2(0)} = \sigma_\theta^2$;
7. /* 在支持集 \mathcal{S}_i 上的训练 */
8. for $\epsilon = 0, \dots, \epsilon$ do
9. $\omega_i^\epsilon = \{\mu_{\omega_i}^\epsilon, \sigma_{\omega_i}^{2(\epsilon)}\}$;
10. 计算均值和方差
11. $\mu_{\omega_i}^{\epsilon+1} = \mu_{\omega_i}^\epsilon - \alpha \nabla_{\mu_{\omega_i}} \mathcal{L}(\mathcal{S}_i; \omega_i^\epsilon)$;
12. $\sigma_{\omega_i}^{2(\epsilon+1)} = \sigma_{\omega_i}^{2(\epsilon)} - \alpha \nabla_{\sigma_{\omega_i}^2} \mathcal{L}(\mathcal{S}_i; \omega_i^\epsilon)$;
13. 在查询集 \mathcal{Q}_i 上评估加权查询损失 $\mathcal{L}(\mathcal{Q}_i; \omega_i^\epsilon)$;
14. end
15. end
16. 计算总损失: $\sum_{i=1}^B \mathcal{L}(\mathcal{Q}_i; \omega_i^\epsilon)$;
17. 更新参数:
- $\theta^* = \theta - \beta \nabla_{\theta} \sum_{i=1}^B \mathcal{L}(\mathcal{Q}_i; \omega_i^\epsilon)$.
18. end

4 实验结果及分析

4.1 数据集

为了评估 MetaIMC 的有效性, 本文在 3 个公共的矩阵补全数据集 (Flixster, Douban 和 Yahoo_music) 上进行了实验。数据集包含用户的社交关系及其对相关项目 (电影和音乐等) 的评级数据。

在真实的社交媒体上, 不同用户对已有项目的偏好评级数量极不平衡, 大量用户仅存在对极少的项目的评级。缺少用户历史偏好数据, 推荐系统更难准确地为用户推荐项目, 而这往往出现在社交媒体新用户的推荐场景中, 造成了新用户冷启动问题。

表 1 列出了 3 个数据集的统计数据。对于这 3 个数据集, 使用 $3\,000 \times 3\,000$ 的子矩阵 (Submatrix) 进行比较, 对于缺少的用户或项目, 采用补 0 的方法填充到所需子矩阵。

表 1 数据集信息

Table 1 Statistics of datasets

Datasets	Yahoo_music	Flixster	Douban
Users-Items	1357-1363	2341-2956	2999-3000
Ratings	4802/533	23556/2617	123202/13689
Density	0.0006	0.0029	0.0152
Rating types	1~100	0.5~5.0	1~5

4.2 实验评价指标

本文将预测评级和真实评级之间的均方根误差 (Root Mean Square Error, RMSE) 作为实验评价指标来评估后续

两个实验中模型的效果, 其值越小, 代表推荐算法的准确性越高。均方根误差如式 (15) 所示:

$$RMSE = \sqrt{\frac{\sum_{(u_i, v_j)} \|r_{ij} - \tilde{r}_{ij}\|^2}{N}} \quad (15)$$

其中, r_{ij} 和 \tilde{r}_{ij} 分别表示用户 u_i 对项目 v_j 的真实评级和预测评级, N 代表评级的数量。

4.3 实验设置与分析

实验所有基线算法都保持它们的初始设置^[8]。为了突出 MetaIMC 的有效性和先进性, 在算法的基础网络部分采用了相同的参数设置 (封闭子图跳数 $h=1$), 仅对算法中的元学习设置进行优化。对于 MetaIMC, 在 Flixster, Douban 和 Yahoo_music 上的每个元训练任务, 参数 K/P 的值分别设置为 4/2, 5/15 和 35/5, MetaIMC 中的批大小 B 设置为 16。学习速率 α 和 β 分别设置为 5×10^{-5} 和 5×10^{-4} 。

4.3.1 实验 1: 传统的矩阵补全预测

为了验证 MetaIMC 的性能, 选择以下 6 种传统矩阵补全方法作为基准模型进行比较。

(1)GRALS^[26]: 其是一种基于 Hessian-vector 乘法的具有结构信息的矩阵补全图可扩展算法。GRALS 使用了用户和项目的边信息, 通过计算用户间的欧几里得距离来确定距离目标用户最近的 10 个用户的偏好, 并将其作为推荐参考。

(2)sRGCNN^[27]: 一种基于特殊设计的多图卷积神经网络体系结构的矩阵补全的换向节点级 GNN 方法。sRGCNN 使用了用户和项目的边信息。在实验设置中, 批大小为 10, 学习率为 1×10^{-3} 。

(3)GC-MC^[7]: 一种用于矩阵补全任务的基于换能图的自编码框架, 该框架从图上的链路预测的角度考虑矩阵补全。GC-MC 使用了用户和项目的边信息。在实验设置中, 批大小为 10, 学习率为 1×10^{-2} 。

(4)IGC-MC^[7]: 一种基于节点级 GNN 的电感式矩阵补全方法模型。它用内容特征代替节点 IDs 的一热编码, 使之具有归纳性; 使用可交换矩阵层而不是边信息来完成归纳矩阵补全。IGC-MC 未使用用户和项目的边信息。在实验设置中, 批大小为 50, 学习率为 1×10^{-3} 。

(5)F-EAE^[28]: 使用可交换矩阵层而不是边信息来执行归纳矩阵补全。F-EAE 未使用用户和项目的边信息。在实验设置中, 批大小为 16, 学习率为 5×10^{-4} 。

(6)IGMC^[8]: 一个基于归纳图的矩阵补全模型。它仅基于评级矩阵生成的用户和项目来对周围的局部子图训练 GNN, 并将子图映射到相应的评级。IGMC 未使用用户和项目的边信息。在实验设置中, 批大小设置为 50, 学习速率为 1×10^{-3} 。

实验 1 是在传统矩阵补全场景中, 即将数据集根据评级随机划分为训练集与测试集, 使得所有已知用户都可得到训练。表 2 中的“s-info”表示方法是否使用边信息, 这些信息在实验中以用户图和项目图的形式呈现。MetaIMC 通过学习交叉任务的通用性来提升性能, 即使每个任务在训练中只有 K -shot 样本。实验结果如表 2 所列。在不使用任何

边信息的情况下,MetaIMC 可以达到与基线方法相近甚至更好的性能,验证了 MetaIMC 在传统测试方法中的有效性。

表2 MC 实验的 RMSE 结果对比

Table 2 Comparison of RMSE results of MC experiment

Method	s-info	Yahoo_mMusic	Flixster	Douban
GRALS	✓	38.0	1.245	0.833
sRGCNN	✓	22.4	0.926	0.801
GC-MC	✓	20.5	0.917	0.734
IGC-MC	×	20.3	0.937	0.908
F-EAE	×	20.0	0.908	0.738
IGMC	×	19.1	0.872	0.721
MetaIMC	×	19.7	0.886	0.752

4.3.2 实验2:用户冷启动场景分析

为了进一步证明元学习设置下 MetaIMC 对于新用户冷启动问题的优越性,根据表2在传统矩阵补全场景中的结果对比,选取了表现最优的传统 MC 方法作为基线,并添加了两个基准模型。

(1)MAML^[9]:一个基于元学习的模型,随机初始化神经网络参数。MAML 未使用用户和项目的边信息。在实验设置中,批大小设置为 16,学习速率 α 和 β 分别设置为 1×10^{-2} 和 5×10^{-4} 。

ABML^[29]:该模型考虑了分层变分推理的元学习,以及任务特定参数的后验。与 ABML 相比,MetaIMC 还考虑了任务的不确定性,并通过随机推理稳定了内部参数的更新。在实验设置中,批大小设置为 10,学习速率 α 和 β 分别设置为 1×10^{-1} 和 1×10^{-3} 。

实验2是元训练期间新用户对项目的评级预测情况,相比实验1,此时 MetaIMC 需要解决一个典型的元学习问题。在这种情况下,当预测来自新区域的用户时,模型仍然会使用 D^{test} 中的支持数据来更新参数(内部更新),以达到通过极少数据训练快速适应新任务(用户)的目的。

对新用户进行泛化实验,即为新用户推荐新项目。随机选取 U_{tr} 用户(评级矩阵中的行)和相应的项目评级 R_{tr} 进行元训练,其他 U_{te} 用户进行元测试。相关实验设置如表3所列,值得注意的是,元测试中的用户在元训练期间不可见,即 $U_{tr} \cap U_{te} = \emptyset$ 。

表3 元学习设置下的数据分配情况

Table 3 Data allocation under meta-learning settings

Datasets	U_{tr}/U_{te}^a	R_{tr}/R_{te}^a
Yahoo_music	183/20	2921/315
Flixster	352/50	12904/1756
Douban	1836/180	102402/10240

注:^a◎/⊙,被用于区分元训练/测试的用户或评级

表4列出了实验结果,其中的“s-info”表示基线方法是否使用边信息。在没有使用任何边信息的情况下,添加了元学习的归纳补全模型 MetaIMC 在用户冷启动场景中表现得更好。特别是在数据密度最小的 Yahoo_music 数据集上,MetaIMC 表现得最好,这表明 MetaIMC 在只利用稀疏的用户-项目交互数据时,能够比其他模型更快地收敛到最佳结果。

表4 冷启动实验的 RMSE 结果对比

Table 4 Comparison of RMSE results of cold start experiment

Method	s-info	Yahoo_music	Flixster	Douban
sRGCNN	✓	22.736	0.909	0.808
F-EAE	×	23.314	1.102	0.831
IGMC	×	22.937	1.007	0.810
MAML	×	18.131	0.947	0.853
ABML	×	17.761	0.813	0.802
MetaIMC	×	15.423	0.740	0.787

4.3.3 超参数的敏感性分析

在 Yahoo_music 数据集上进行超参数敏感性分析,通过分析模型中两个重要参数(单任务下支持集中的样本个数和内部循环迭代更新次数)对 MetaIMC 性能的影响。图3给出了每个任务中的支持集的样本数量 K 和内部更新迭代次数 L 的影响。

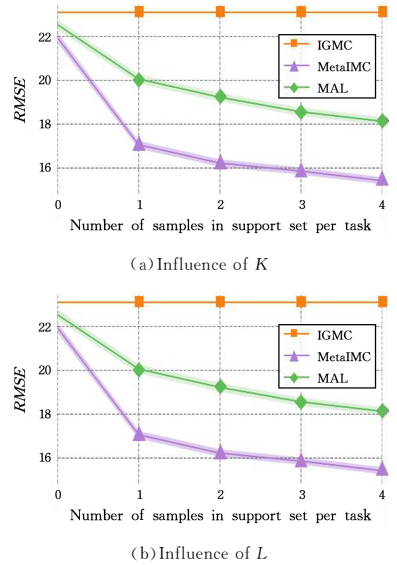


图3 超参数对 MetaIMC 性能的影响

Fig. 3 Influence of hyperparameters on MetaIMC

对于参数 K ,由图3(a)可知, $K=4$ 时 MetaIMC 的性能达到最佳。此外,在不过拟合的情况下,增加支持样本数量能够使 MetaIMC 回归模型的性能继续提升。

对于参数 L ,研究了训练期间迭代次数的变化对模型性能的影响。由图3(b)可知,使用元学习、先验学习的模型, RMSE 在一次迭代后会达到最优的性能,表明 MetaIMC 可以快速适应新用户并处理用户冷启动场景。

结束语 结合元学习的算法于图模型虽然在一定程度上缓解了上述少样本稀疏的问题且节约了计算资源,但往往缺乏对不确定性的良好量化,而这对于缺乏数据的实际图模型应用却至关重要。本文提出了一种基于概率元学习的归纳矩阵补全预测的融合算法,该方法构建了与模型无关的图深度神经网络学习模型,充分利用数据先验知识提出了从稀疏数据中学习新任务的解决方案,通过贝叶斯解释可以有效地推理出少样本学习任务的不确定性;通过元学习目标的训练,进一步增强图模型在推荐任务中的泛化能力。为了验证模型在传统矩阵补全及用户冷启动两个场景下的效果,将其与多个基准模型方法在3个数据集上进行了对比实验。实验结果表明,本文提出的 MetaIMC 模型在数据稀疏和冷启动问题上

效果更优。未来的工作将考虑研究模型行为的可解释性,包括研究基于梯度的元学习和变分贝叶斯推断的融合算法在其他图学习任务上的应用,同时也将考虑对贝叶斯网络的敏感性进行更加深入的研究;此外,也会考虑针对算法中的基础网络进行进一步优化,以及尝试利用更大的数据集对模型性能进行更全面的评估。

参 考 文 献

- [1] PAZZANI M J, BILLSUS D. Content-based recommendation systems[M] // The adaptive Web. Berlin: Springer, 2007: 325-341.
- [2] GOLDBERG D, NICHOLS D, OKI B M, et al. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61-70.
- [3] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C] // Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 1025-1035.
- [4] KALOFOLIAS V, BRESSON X, BRONSTEIN M, et al. Matrix completion on graphs[J]. arXiv:1408.1717, 2014.
- [5] RAO N. Collaborative Filtering with Graph Information; Consistency and Scalable Methods[C] // Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015: 2107-2115.
- [6] MONTI F, BRONSTEIN M M, BRESSON X. Geometric matrix completion with recurrent multi-graph neural networks[C] // Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 3700-3710.
- [7] BERG R, KIPF T N, WELLING M. Graph convolutional matrix completion [EB/OL]. (2017-10-25) [2018-07-28]. <http://arxiv.org/abs/1706.02263>.
- [8] ZHANG M, CHEN Y. Inductive matrix completion based on graph neural networks[C] // International Conference on Learning Representations, 2020.
- [9] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C] // International Conference on Machine Learning. PMLR, 2017: 1126-1135.
- [10] JAMALI M, ESTER M. 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.
- [11] MA H, ZHOU D, LIU C, et al. Recommender systems with social regularization[C] // Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011: 287-296.
- [12] DROR G, KOENIGSTEIN N, KOREN Y, et al. The Yahoo! Music Dataset and KDD-Cup'11[C] // Proceedings of KDD Cup 2011. PMLR, San Diego, CA, USA, 2012: 3-18.
- [13] CANDÈS E J, RECHT B. Exact matrix completion via convex optimization[J]. Foundations of Computational Mathematics, 2009, 9(6): 717-772.
- [14] ZHANG M, CHEN Y. Weisfeiler-lehman neural machine for link prediction[C] // Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017: 575-583.
- [15] LI X, CHEN H. Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach[J]. Decision Support Systems, 2013, 54(2): 880-890.
- [16] HUISMAN M, VAN RIJN J N, PLAAT A. A survey of deep meta-learning[J]. Artificial Intelligence Review, 2021, 54(6): 4483-4541.
- [17] BEAL M J. Variational algorithms for approximate Bayesian inference [D]. London: UCL (University College London), 2003.
- [18] RAVI S, BEATSON A. Amortized Bayesian Meta-Learning [C] // International Conference on Learning Representations (ICLR), 2019.
- [19] AMIT R, MEIR R. Meta-learning by adjusting priors based on extended PAC-Bayes theory[C] // International Conference on Machine Learning. PMLR, 2018: 205-214.
- [20] FINN C, XU K, LEVINE S. Probabilistic model-agnostic meta-learning[C] // International Conference on Neural Information Processing Systems (NIPS), 2018: 9537-9548.
- [21] YOON J, KIM T, DIA O, et al. Bayesian model-agnostic meta-learning[C] // Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018: 7343-7353.
- [22] GRANT E, FINN C, LEVINE S, et al. Recasting gradient-based meta-learning as hierarchical bayes[C] // International Conference on Learning Representations (ICLR), 2018.
- [23] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model [J]. IEEE transactions on neural networks, 2008, 20(1): 61-80.
- [24] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C] // European semantic web conference. Cham: Springer, 2018: 593-607.
- [25] BISHOP C M, NASRABADI N M. Pattern recognition and machine learning[M]. New York: Springer, 2006.
- [26] RAO N, YU H F, RAVIKUMAR P, et al. Collaborative Filtering with Graph Information; Consistency and Scalable Methods [C] // NIPS, 2015.
- [27] MONTI F, BRONSTEIN M M, BRESSON X. Geometric matrix completion with recurrent multi-graph neural networks[C] // International Conference on Neural Information Processing Systems, 2017: 3700-3710.
- [28] HARTFORD J, GRAHAM D, LEYTON-BROWN K, et al. Deep models of interactions across sets[C] // International Conference on Machine Learning. PMLR, Stockholm, 2018: 1909-1918.
- [29] RAVI S, BEATSON A. Amortized bayesian meta-learning[C] // International Conference on Learning Representations, 2019.



QI Xiu-xiu, born in 1992, postgraduate. Her main research interests include meta learning, few-shot learning, social network knowledge discovery and data mining.



WANG Jia-hao, born in 1978, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include IoT, information security and data mining.