计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191000172

基于卷积神经网络与约束概率矩阵分解的推荐算法

马海江
华侨大学计算机科学与技术学院　厦门 361021

摘　要　用户评分数据的稀疏性和上下文的信息缺失,往往导致基于矩阵分解(Matrix Factorization,MF)的推荐算法在准确性方面有所欠缺。针对此问题,文中提出了一种基于卷积神经网络(Convolutional Neural Networks,CNN)与约束概率矩阵分解(Constrained Probabilistic Matrix Factorization,CPMF)的推荐算法。首先,构建卷积神经网络模型,对用户上下文辅助信息进行识别,获得文本潜在向量,并叠加高斯噪声,初始化项目特征矩阵;然后,根据用户评分信息,利用约束矩阵来约束用户特征,并叠加补偿矩阵,初始化用户特征矩阵;接着,利用初始化的用户特征矩阵和项目特征矩阵拟合评分矩阵,对评分矩阵进行矩阵分解,并利用坐标下降算法更新参数;最后,预测用户对项目的评分,实现项目推荐。在 Movielens 和 Amazon 数据集上的实验结果表明,该推荐算法显著优于传统的推荐模型,有效地提高了推荐结果的准确率。

关键词:卷积神经网络;矩阵分解;推荐算法;上下文信息
中图法分类号　TP391

Recommendation Algorithm Based on Convolutional Neural Network and Constrained Probability Matrix Factorization

MA Hai-jiang
School of Computer Science and Technology Huaqiao University,Xiamen 361021,China

Abstract　Due to the sparsity of user rating data and the lack of context information,the recommendation algorithm based on matrix factorization is often lacking in accuracy. To solve this problem,a recommendation algorithm based on convolutional neural network and constrained probability matrix factorization is proposed. Firstly,a convolutional neural network model is constructed to identify the contextual auxiliary information of users,obtain the text potential vector,superimpose gaussian noise,and initialize the project characteristic matrix. Then,according to the user rating information,the user characteristics are constrained by the constraint matrix,and the user characteristic matrix is initialized by superimposing the compensation matrix. Then,the initialized user characteristic matrix and project characteristic matrix are used to fit the rating matrix,the rating matrix is decomposed by matrix,and the coordinate descent algorithm is used to update the parameters. Finally,predict the user's score on the project and implement the project recommendation. Experimental results on Movielens and Amazon data sets show that this recommendation algorithm is significantly superior to the traditional recommendation model and effectively improves the accuracy of recommendation results.

Keywords　Convolutional neural networks,Matrix factorization,Recommendation algorithm,Contextual information

1　引言

在网络信息迅速发展的年代,推荐系统在网络服务中发挥着重要作用。在不同的推荐技术中,基于协同过滤(Collaborative Filtering,CF)的方法[1]利用用户历史行为或偏好,取得了显著的效果。然而,用户和商品数量不断增加导致用户对商品的评级数据的稀疏性,使 CF 的性能受到了限制。此外,CF 算法只关注用户和商品、商品和商品以及用户和用户之间的关联信息,而不关注商品的内容和用户的个人信息及用户所处的上下文信息,从而影响用户的决策和推荐结果。

学者们综合考虑了评级信息和上下文辅助信息,提出了一系列的解决方法来提升推荐的准确性。Zhang 等[2]综合考虑文本评论信息与评分等级,以提升推荐模型对潜在评级预测的准确性,提出了将层叠降噪自动编码器(Stacked Denoising Auto Encoder,SDAE)与隐含因子模型(Latent Factor Model,LFM)结合的推荐算法。Li 等[3]将项目属性之间的耦合关系作为隐含信息,提出了一种基于项目属性耦合性的矩阵分解模型。Wang 等[4]将 SDAE 集成到概率矩阵分解(Probabilistic Matrix Factorization,PMF)中,提出了协同深度学习(Collaborative Deep Learning,CDL),从而在评分预测方面生成更为准确的潜在模型。Gao 等[5]提出了一种基于用户认知行为的上下文感知偏好获取方法,分别提取单维与多

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:国家社会科学基金资助项目(19BXW110);福建省社会科学规划项目(FJ2017B073);华侨大学科研启动项目(600005-Z16Y0005)
This work was supported by the Project National Social Science Foundation (19BXW110),Social Science Planning Project of Fujian Province (FJ2017B073) and Huaqiao University Research Startup Project (600005-Z16Y0005).
通信作者:马海江(751219377@qq.com)

维上下文环境下的用户偏好。

近年来,深度学习方法在自然语言处理等领域已经得到很好的应用。一方面,深度学习通过深层次非线性网络结构,可以从用户和项目相关的海量数据中,获取用户和项目的深层次特征表示。另一方面,与矩阵分解的特征向量内积方式相比,神经网络具有较强的非线性映射能力,将不同数据映射到一个相同的隐空间,能够获得数据的统一表征[6]。Wer等[7]提出了融合协同过滤与深度学习的推荐模型,该模型首先通过降噪自动编码器对项目信息进行编码,其次根据编码计算其他项目与该项目的 Person 相似度和基于相似度权重的评分值,最后将其与 timeSVD++(时间感知的隐因子模型)的评分线性相加,作为最终评分值。

现有的研究表明,卷积神经网络能够有效利用多源异构数据,获取用户和项目的潜在特征,有效缓解数据稀疏和冷启动问题。因此,本文提出了一种基于卷积神经网络与约束概率矩阵分解的推荐算法。该模型首先通过卷积神经网络捕获项目描述文本的上下文信息,获取项目的潜在特征,再将项目潜在特征融入到约束概率矩阵分解中,即使在评分数据非常稀疏的情况下,也能较准确地预测用户对项目的未知评分。

2 相关工作

2.1 矩阵分解推荐模型

矩阵分解推荐模型是一种基于模型的协同过滤算法,也是目前应用比较广泛的推荐算法之一。在基于模型的协同过滤算法中,利用用户和项目的历史数据训练得到模型,并利用该模型实现实时推荐。矩阵分解方法的基本思想是假设每个用户和每个项目都有各自的特征,用户的兴趣受少数因素影响,在 MF 中,可以从用户-项目交互矩阵中分解出用户特征矩阵和项目特征矩阵[8]。假设有 N 个用户,M 个项目和一个用户-项目评分矩阵 $\boldsymbol{R}\in R^{N\times M}$,用户 i 和项目 j 的潜在模型有 D 个隐含变量,其中 $u_i\in R^D$,$v_j\in R^D$,则用户 i 对项目 j 的评级 R_{ij} 等于用户 i 和项目 j 的对应潜在模型的内积($R_{ij}=u_i^{\mathrm{T}}v_j$)。为了避免出现过拟合现象,对某些参数作一些限制,如对损失函数加入权重衰减因子(即正则化表示)[9],该损失函数定义为:

$$\ell=\sum_{i}^{N}\sum_{j}^{M}I_{ij}(R_{ij}-u_i^{\mathrm{T}}v_j)^2+\lambda_u\sum_{i}^{N}\|u_i\|^2+\lambda_v\sum_{j}^{M}\|v_j\|^2 \quad (1)$$

其中,I_{ij} 表示用户 i 对项目 j 是否有评分,如果有评分则 I_{ij} 为 1,否则 I_{ij} 为 0;R_{ij} 表示用户 i 对项目 j 的实际评分;λ_u 和 λ_v 分别控制用户和项目特征矩阵正则化在整个模型中的比例。

2.2 卷积神经网络模型

卷积神经网络是一种带有卷积结构的深度神经网络,其本质是一种输入到输出的映射,能够学习大量的输入与输出之间的映射关系。其包括两个关键的操作:1)用于提取局部特征的卷积层;2)特征映射的池化层。

CNN 通过挖掘数据在空间上的相关性,来减少网络中的可训练参数的数量,缓解了模型的过拟合问题。Vanden等[10]提出了一种基于 CNN 获得的项目潜在模型来预测评分的模型,利用 CNN 分析了歌曲的相关信息。Cheng 等[11]提出了一种宽深度学习模型,利用联合训练的宽线性模型和深度神经网络对推荐系统进行改进,使模型具有较好的泛化能力。Zheng 等[12]使用卷积神经网络挖掘评论信息,从而获取用户和项目的隐特征,通过因子分解机(Factorization Machine,FM)[13]方法对评分进行预测。Wang 等[14]提出了一种深度知识感知网络(DKN),将知识图谱表示方法融入到新闻推荐中,把实体作为多个通道并在卷积过程中显式地保持它们的对齐关系。

综上所述,研究者在不断地优化传统的推荐算法,同时将深度学习融入到推荐算法中,提高了推荐系统的准确性。

3 卷积约束概率矩阵分解

约束概率矩阵分解是基于模型的协同过滤推荐策略,是一种潜在的特征因子分解算法。该算法通过用户对项目的评分记录来学习每个用户和项目的潜在特征向量,利用低维度的潜在特征向量矩阵计算出用户对项目的预测评分,进而产生推荐。Mnih 等[15]考虑到评分项目集合相似的用户可能具有相似的兴趣,对用户的潜在特征向量进行改进,提出利用约束矩阵来约束用户的特征,即给每个项目添加一个除特征向量外的约束矩阵,由用户参与过评分的所有项目的约束矩阵的平均值对项目的特征向量产生影响。从而可以得到,评分行为相似的用户会有相似的特征向量。实验结果表明,CPMF 在数据稀疏性和推荐准确性上,相较于只利用用户-项目评分矩阵的推荐算法取得更好的效果。

CNN 在文本处理中发挥着越来越重要的作用,并能取得较好的效果。CNN 可以有效地捕获文本的局部特征,有助于更深入的理解文本信息,也能够更深入地从项目描述信息中提取特征,从而提高系统的评分预测精度。

根据上述思路,本文提出了一种基于卷积神经网络与约束概率矩阵分解的推荐算法(记为卷积约束概率矩阵分解,ConvCPMF),其基本框架如图 1 所示。首先,以文本评论描述信息 X 作为输入,通过卷积神经网络模型 P 获得文本潜在向量,并叠加高斯噪声,初始化项目特征矩阵 $\boldsymbol{V}$;然后,利用约束矩阵 $\boldsymbol{W}$,根据用户评分信息 $\boldsymbol{I}$,来约束用户特征,并叠加补偿矩阵 $\boldsymbol{Y}$,初始化用户特征矩阵 $\boldsymbol{U}$;其次,利用初始化的用户特征矩阵 $\boldsymbol{U}$ 和项目特征矩阵 $\boldsymbol{V}$ 生成评分矩阵 $\boldsymbol{R}$,并利用坐标下降算法更新参数 $\boldsymbol{Y}$,$\boldsymbol{W}$ 和 $\boldsymbol{P}$;最后,预测用户对项目的评分,实现项目推荐。

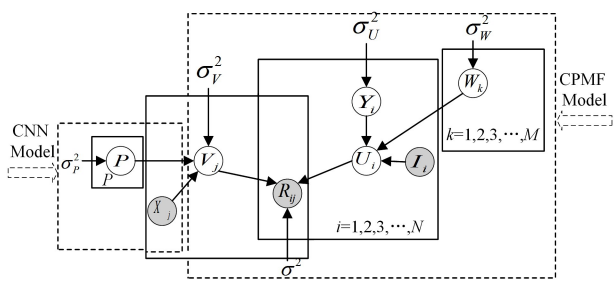

图 1 ConvCPMF 框架图

Fig. 1 ConvCPMF frame digram

3.1 约束概率矩阵分解(CPMF)

假设有 N 个用户,M 个项目,得到一个 $N\times M$ 维的评分矩阵 $\boldsymbol{R}$,矩阵 $\boldsymbol{R}$ 中的元素 $\boldsymbol{R}_{ij}$ 表示用户 i 对项目 j 的评分。假设存在潜在特征维数为 D,那么 $D\times N$ 维的矩阵 $\boldsymbol{U}$ 表示用户的潜在特征矩阵,$\boldsymbol{U}_i$ 表示用户 i 的潜在特征向量;$D\times M$ 维的矩阵 $\boldsymbol{V}$ 表示项目的潜在特征矩阵,$\boldsymbol{V}_j$ 表示项目 j 的潜在特征向量;$D\times M$ 维的矩阵 $\boldsymbol{W}$ 表示约束矩阵,用来约束用户特征,即对每个项目除特征向量之外的一个约束向量,让每个用户评分过的所有项目约束向量的均值来影响其特征向量。因此

评分行为相近的用户会有相似的特征向量。$D\times N$ 维的矩阵 $\boldsymbol{Y}$ 表示用户潜在特征的一个补偿矩阵，它可以被看作是添加到先验分布平均值的偏移量，用来初始化用户 i 的特征向量 $\boldsymbol{U}_i$ 即：

$$\boldsymbol{U}_i = Y_i + \frac{\sum_{k=1}^{M} I_{ik} W_k}{\sum_{k=1}^{M} I_{ik}} \tag{2}$$

其中，I_{ik} 表示指标函数，如果用户 i 已经对项目 k 进行了评分，则为 1，否则为 0；当用户 i 没有评分时 $U_i = Y_i$；$\boldsymbol{W}\in R^{D\times M}$ 表示潜在相似约束矩阵；$\boldsymbol{Y}\in R^{D\times N}$ 表示用户潜在特征的一个补偿矩阵，Y_i 表示 $\boldsymbol{Y}$ 的第 i 列，其服从于均值为 0，协方差矩阵为 $\sigma_Y I$ 的高斯分布：

$$p(\boldsymbol{Y}|\sigma_Y^2) = \prod_{i=1}^{N} N(Y_i|0,\sigma_Y^2 I) \tag{3}$$

因此，整个评分矩阵的条件分布定义为：

$$p(\boldsymbol{R}|\boldsymbol{Y},\boldsymbol{V},\boldsymbol{W},\sigma^2) = \prod_{i=1}^{N}\prod_{j=1}^{M}[N(R_{ij}|([Y_i + \frac{\sum_{k=1}^{M} I_{ik} W_k}{\sum_{k=1}^{M} I_{ik}}]^{\mathrm{T}} V_j),\sigma^2)]^{I_{ij}} \tag{4}$$

其中，N 表示高斯正态分布的概率密度函数；潜在约束矩阵 $\boldsymbol{W}$ 中的每个权重 W_k，均值为 0，协方差矩阵为 $\sigma_W I$ 的高斯分布，即：

$$p(W|\sigma_W^2) = \prod_{k=1}^{M} N(W_k|0,\sigma_W^2 I) \tag{5}$$

假设一个项目潜在模型由 3 个变量生成：1)CNN 中的内部权重 P；2)X_j 表示第 j 项文本；3)ε 变量为高斯噪声，这能够进一步优化项目的潜在评级模型。因此，项目潜在模型为：

$$V_j = cnn(P,X_j) + \varepsilon_j, \varepsilon_j \sim N(0,\sigma_V^2 I) \tag{6}$$

其中，$cnn(P,X_j)$ 表示项目评价信息经过卷积神经网络识别的值，ε_j 服从均值为 0，协方差矩阵为 $\sigma_V^2 I$ 的正态分布；cnn 内部权重 P 中的每个权重 p_k，均值为 0，方差为 σ_P^2 的高斯先验分布为：

$$p(P|\sigma_P^2) = \prod_{k} N(p_k|0,\sigma_P^2) \tag{7}$$

项目潜在模型的条件分布为：

$$p(V|P,X,\sigma_V^2) = \prod_{j}^{M} N(V_j|cnn(P,X_j),\sigma_V^2 I) \tag{8}$$

其中，X 表示项目描述文本集。

从 CNN 模型中获得文本潜在向量，叠加高斯噪声，初始化项目潜在模型，将高斯分布的均值作为 CNN 和 CPMF 融合的桥梁，有利于充分分析项目描述文本信息和用户评分信息。

3.2 卷积神经网络(CNN)架构

CNN 架构是用来处理项目描述文本的上下文信息，用于生成项目的潜在向量，由输入层、卷积层、池化层和输出层组成。本文采用 Kim 等[16]的 CNN 层次结构图，如图 2 所示。

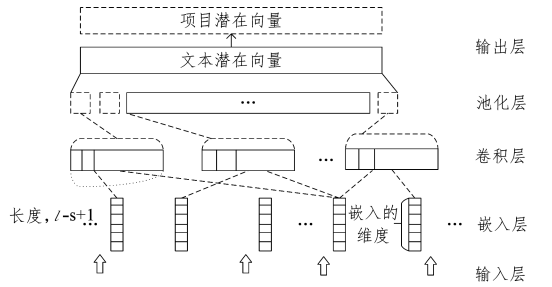

图 2 CNN 层次结构图

Fig. 2 CNN hierarchy chart

(1)输入层/嵌入层

输入层/嵌入层将用户对项目的原始描述文本信息转化成文本矩阵，作为下一个卷积层输入。具体来说，本文将一个文本作为 T 个单词的序列，然后将单词向量连接起来组成文本矩阵，该矩阵可以任意初始化或用已训练的单词初始化，嵌入到模型中。文本矩阵 $\boldsymbol{D}$ 为：

$$\boldsymbol{D} = [\cdots, \boldsymbol{d}_{i-1}, \boldsymbol{d}_i, \boldsymbol{d}_{i+1}, \cdots] \in R^{k\times T} \tag{9}$$

其中，T 表示文本的长度；$\boldsymbol{d}_i$ 是文本中第 i 个单词的词向量；k 表示每个单词的嵌入维数。

(2)卷积层

卷积层用来提取文本的上下文特征，通过对输入的文本矩阵进行卷积运算操作，可以使得原始文本的某些特征增强，降低噪声，获得新的特征。上下文特征 $c_i^j \in R$ 由第 j 个共享权重 $P_c^j \in R^{k\times s}$ 提取，其中窗口大小 s 确定周围单词的数量：

$$c_i^j = f(P_c^j * D_{(:,i:(i+s-1))} + b_c^j) \tag{10}$$

其中，b_c^j 表示偏置项；$f()$ 表示非线性激活函数；$*$ 表示卷积操作；$D_{(:,i:(i+s-1))}$ 表示滑动窗口取 D 中所有行的第 i 列到 $i+s-1$ 列作为输入。非线性激活函数主要有 Tanh，Sigmoid 和 Relu，本文采用修正线性单元激励函数 Relu 来防止过拟合及梯度消失的问题，得到上下文特征向量 $\boldsymbol{c}^j \in R^{T-s+1}$：

$$\boldsymbol{c}^j = [c_1^j, c_2^j, \cdots, c_{T-s+1}^j] \tag{11}$$

其中，$\boldsymbol{c}^j$ 为第 j 个共享权重 P_c^j 提取的长度为 T−s+1 的上下文特征向量。

由 CNN 的特性可知，一个共享权重只能捕获到一种类型的上下文特征。因此，需要使用多个共享权重来捕获多种类型的上下文特征，最终能够生成与 P_c 的数目 n_c 一样多的上下文特征向量。

(3)池化层

池化层通常被用在连续的卷积层之间，其主要作用是减少特征和参数数量，以减少网络的计算量，从而控制过拟合，提取具有代表性的特征。从卷积层提取上下文特征向量，文本被表示为具有 n_c 个上下文特征的向量，其中每个上下文特征向量具有可变长度（特征向量的长度为 $T-s+1$）。因此，利用最大池化(max-pooling)提取上下文特征向量中的最大值作为文本特征向量，将文本的表示形式减少为 n_c 固定长度向量：

$$\boldsymbol{z}_f = [\max(\boldsymbol{c}^1), \max(\boldsymbol{c}^2), \max(\boldsymbol{c}^3), \cdots, \max(\boldsymbol{c}^j), \cdots, \max(\boldsymbol{c}^{n_c})] \tag{12}$$

其中，$\boldsymbol{z}_f$ 表示池化后的文本特征向量；$\boldsymbol{c}^j$ 表示上下文特征向量。

(4)输出层

将池化层输出的结果作为输出层的输入，通过传统的非线性投影产生文本潜在特征向量：

$$\boldsymbol{e} = \tanh(\boldsymbol{W}_{f2}\{\tanh(\boldsymbol{W}_{f1}\boldsymbol{z}_f + \boldsymbol{b}_{f1})\} + \boldsymbol{b}_{f2}) \tag{13}$$

其中，$\boldsymbol{W}_{f1}\in R^{f\times n_c}$，$\boldsymbol{W}_{f2}\in R^{f\times D}$ 表示投影矩阵；$\boldsymbol{b}_{f1}\in R^f$，$\boldsymbol{b}_{f2}\in R^D$ 表示 $\boldsymbol{W}_{f1}$，$\boldsymbol{W}_{f2}$ 的偏置向量；D 表示用户、项目潜在特征的列维度；f 表示隐层节点神经元数量。

CNN 模型以原始文本作为输入，经过卷积池化处理后在输出层返回每个文本的潜在向量作为输出：

$$\boldsymbol{o}_j = cnn(\boldsymbol{P},X_j) \tag{14}$$

其中，X_j 表示第 j 项的原始文本；$\boldsymbol{o}_j$ 表示第 j 项的文本潜在向量。

综上所知,卷积神经网络其实是一个函数。它将原始给定的文本信息的特征值作为输出值,而该输出值用于表示用户对项目评级信息的评分。

3.3 模型损失函数

本文使用最大后验估计(Maximum A Posteriori Probability Estimate,MAP)来优化用户潜在模型、项目潜在模型、CNN 权重和偏执项等变量:

$$\max_{\boldsymbol{Y},\boldsymbol{V},\boldsymbol{W},\boldsymbol{P}} p(\boldsymbol{Y},\boldsymbol{V},\boldsymbol{P},\boldsymbol{W}|\boldsymbol{R},X,\sigma^2,\sigma_Y^2,\sigma_V^2,\sigma_P^2,\sigma_W^2)$$
$$= \max_{Y,V,W,P} [p(\boldsymbol{R}|\boldsymbol{Y},\boldsymbol{V},\sigma^2)p(Y|\sigma_Y^2) p(\boldsymbol{V}|\boldsymbol{P},\sigma_V^2)p(\boldsymbol{P}|\sigma_P^2)p(\boldsymbol{W}|\sigma_W^2)] \quad (15)$$

对式(15)取负对数得到模型损失函数:

$$E(Y,V,P,W)=\sum_i^N\sum_j^M\frac{I_{ij}}{2}(R_{ij}-(Y_i+\frac{\sum_{k=1}^M I_{ik}W_k}{\sum_{k=1}^M I_{ik}})^{\mathrm{T}}V_j)^2+\frac{\lambda_Y}{2}\sum_i^N\|Y_i\|^2+\frac{\lambda_V}{2}\sum_j^M\|V_j-cnn(P,X_j)\|^2+\frac{\lambda_P}{2}\sum_k^{|p_k|}\|p_k\|^2+\frac{\lambda_W}{2}\sum_k^M\|W_k\|^2 \quad (16)$$

其中,$E(Y,V,P,W)$表示损失函数;$\lambda_Y=\frac{\sigma^2}{\sigma_Y^2}$,$\lambda_V=\frac{\sigma^2}{\sigma_V^2}$,$\lambda_P=\frac{\sigma^2}{\sigma_P^2}$和$\lambda_W=\frac{\sigma^2}{\sigma_W^2}$分别表示正则化系数,一般取较小值;$\|Y_i\|^2$,$\|p_k\|^2$和$\|W_k\|^2$分别表示对应矩阵向量中的欧几里得范数。

本文采用坐标下降法,优化潜在特征向量。坐标下降法属于一种非梯度优化的方法,它在每步迭代中沿一个坐标的方向进行线性搜索,通过循环使用不同的坐标方法来达到目标函数的局部极小值。相比梯度下降法而言,坐标下降法不需要计算目标函数的梯度,它在每一次迭代过程,会按照一定的顺序对每一个参数进行更新,直到收敛。

坐标下降法就是每次只对一个维度进行变换,将其他维度进行固定,如此循环迭代,最后得到最优解。应用于 ConvCPMF 模型就是假设 V 和 W(或 Y)是一个常量,对式(16)变为关于 Y(或 W,V)的一个二次函数。然后将优化函数 E 关于 Y(或 W,V)进行微分,计算 Y(或 W,V)的最优解:

$$Y_i \longleftarrow (V\boldsymbol{I}_iV^T+\lambda_Y\boldsymbol{I}_D)^{-1}V\boldsymbol{R}_i \quad (17)$$

$$V_j \longleftarrow (U\boldsymbol{I}_jU^T+\lambda_Y\boldsymbol{I}_D)^{-1}(U\boldsymbol{R}_j+\lambda_V cnn(P,X_j)) \quad (18)$$

$$W_k \longleftarrow (V\boldsymbol{I}_kV^T+\lambda_W I_D)^{-1}(V\boldsymbol{R}_k-V\boldsymbol{I}_kY^T) \quad (19)$$

其中,$\boldsymbol{I}_i$ 是一个对角矩阵,其对角元素为 $\boldsymbol{I}_{ij}$,$j=1,\cdots,M$;$\boldsymbol{I}_D$ 表示 D 维的对角矩阵;$\boldsymbol{R}_i$ 表示用户 i 的$(R_{ij})_{j=1}^M$的向量;I_j,I_k,R_j 和 R_k 分别类似地定义为 I_i 和 R_i;$U_i=Y_i+\frac{\sum_{k=1}^M I_{ik}W_k}{\sum_{k=1}^M I_{ik}}$,$U=[U_1,U_2,\cdots,U_i,\cdots,U_N]$。式(18)表示 CNN 模型生成项目潜在向量对 V_j 的影响。

权重矩阵 $\boldsymbol{P}$ 和 CNN 模型的特征息息相关,当 $\boldsymbol{Y}$,$\boldsymbol{W}$ 和 $\boldsymbol{V}$ 恒定时,可以将损失函数 E 看做算是具有 L_2 正则项的平方误差函数:

$$L(P)=\frac{\lambda_V}{2}\sum_j^M\|V_j-cnn(P,X_j)\|^2+\frac{\lambda_P}{2}\sum_k^{|p_k|}\|p_k\|^2+cons\tan t \quad (20)$$

其中,$cons\tan t$ 表示 $\boldsymbol{Y}$,$\boldsymbol{W}$ 和 $\boldsymbol{V}$ 恒定量。

本文采用反向传播算法,优化权重矩阵 P。反向传播算法是适合于多层神经元网络的一种学习算法,它建立在梯度下降法的基础上。通过优化 $\boldsymbol{Y}$,$\boldsymbol{W}$ 和 $\boldsymbol{P}$,最终可以预测用户对项目的未知评分:

$$R_{ij}\approx E\left[R_{ij}|\left(Y_i+\frac{\sum_{k=1}^M I_{ik}W_k}{\sum_{k=1}^M I_{ik}}\right)^{\mathrm{T}}V_j,\sigma^2\right]$$
$$=(Y_i+\frac{\sum_{k=1}^M I_{ik}W_k}{\sum_{k=1}^M I_{ik}})^{\mathrm{T}}(cnn(P,X_j)+\varepsilon_j) \quad (21)$$

3.4 算法描述

ConvCPMF 算法的整体流程如步骤算法 1 所示。

算法 1 ConvCPMF 算法

输入:每个项目描述文本 N 和数据集 S

输出:用户对项目的未知评分

1. 将文本 N 通过 word2vec 处理成词向量
2. 嵌入层将上一步的词向量嵌入成 k 维,得到词序列 $D\in R^{k\times T}$
3. 卷积层使用 3 个不同的窗口大小的特征抽取器,对上一层的输出做卷积操作
4. 池化层将得到的每个特征向量做最大池化,提取特征
5. 将池化结果压平操作,即把多维的输入一维化
6. 将上述得到的结果输入到全连接层,进行防过拟合操作,再进行降维处理,经输出层输出结果
7. CNN 输出文本的潜在向量,叠加高斯噪声,初始化项目特征矩阵
8. 初始化用户特征矩阵,并利用用户特征矩阵和项目特征矩阵来生成评分矩阵 $\boldsymbol{R}$
9. 由数据计算得到 CPMF 结果,利用坐标下降算法更新参数

4 实验及结果分析

本节将在真实数据集上对 ConvCMPF 模型的性能指标进行评估与验证,对实验结果进行分析,和目前较好的推荐模型进行对比,并评估其项目预测的准确性。

4.1 实验数据和设置

本文采用的基础实验数据集有 3 个,如表 1 所列。MovieLens 是一个关于电影评分的数据集,包含多个用户对多部电影的评级数据,也包括电影数据信息和用户属性信息。MovieLens 的 ML-1m 数据集包含 6 040 个用户对 3 952 部电影的 993 482 条评分数据;ML-10m 数据集包含 69 787 个用户对 10 073 部电影的 9 945 875 条评分数据[17]。加州大学圣迭戈分校提供 Amazon 的 AIV 电影评分数据集包含 29 757 个用户对 15 149 部电影的 135 188 条评论数据[18]。

表 1 实验数据集统计表

Table 1 Experimental data sets statistics

数据集	ML-1m	ML-10m	AIV
用户数	6 040	69 787	29 757
项目数	3 544	10 073	15 149
评论数	993 482	9 945 875	135 188
密度/%	4.641	1.141	0.030

为了提高实验结果的准确性,文中对数据集的描述文本进行数据预处理:

(1)清除常用的停用词;

(2)将项目的原始评价文本的长度设置为 300 个单词；

(3)使用 TF-IDF 计算文本中每个词的得分；

(4)删除原始文本数据中的非词汇部分；

(5)为了提高模型的总体性能，本文将表 2 所列的每个数据集随机划分为训练集、验证集和测试集。我们使用数据集的 80%作为训练集，10%作为验证集，10%作为测试集。

除了数据预处理外，本文的实验环境设置如下：

(1)Windows10 64 位操作系统；

(2)GPU 用 NVIDIA CUDA 9.0；

(3)搭建好 python 中 pytorch 库环境，使用 Pycharm & python3.6 代码管理平台进行训练。本文的推荐算法 ConvCPMF 的训练参数设置如表 2 所列。

表 2 参数设置表

Table 2 Table porameter setting

参数名称	参数设置	说明
k	200	词向量维数
s	[3,4,5]	3 个不同的窗口大小长度
D	50	用户、项目潜在特征的列维度
max_df	0.5	文本中单词出现的频率阈值
$num\_kernel\_per\_s$	100	通道数大小
$droupout$	0.2	丢弃率
$learning\_rate$	0.001	学习率
$batch\_size$	128	批处理尺寸
λ_Y	0.01	正则化系数
λ_V	0.01	正则化系数
λ_P	0.05	正则化系数
λ_W	0.05	正则化系数

4.2 评价指标

推荐系统的评价指标通常采用平均绝对误差(Mean Absolute Error，MSE)、均方根误差(Root Mean Square，RMSE)和召回率(Recall Ratio，RCALL)来评价推荐系统在预测评分值时的准确性。本文采用均方根误差(RMSE)对实验结果进行评估。RMSE 是所有项目真实评分和预测评分误差的平方和的平均值的根值：

$$RMSE=\sqrt{\frac{1}{Q}\sum_{i}^{N}\sum_{j}^{M}(r_{ij}-\hat{r}_{ij})^2} \tag{22}$$

其中，Q 表示用户评价数量；r_{ij} 表示原始评分矩阵的值；$\hat{r}_{ij}$ 表示模型的预测打分值。RMSE 值越小，说明模型预测得分越接近真实值，模型性能就越好。

4.3 实验结果分析

为了验证本文推荐模型的有效性，本文选取目前在推荐上相对较好的算法与 ConvCPMF 进行对比分析。

(1)CPMF(约束概率矩阵分解)[15]是一种利用项目评级的协同过滤模型，通过用户对项目的评分记录来学习每个用户和项目的潜在特征向量，最后利用低维度的潜在特征向量矩阵计算出用户对项目的预测评分，进而产生推荐。

(2)PMF(概率矩阵分解)[15]是一种标准型的评分预测模型，只使用了用户的评分。

(3)ConvMF(卷积概率矩阵分解)[16]是一种协同深度学习模型，将卷积神经网络集成到 PMF 中，用来提高评分预测准确性。

(4)NMF(非负矩阵分解)[19]是在矩阵中所有元素均为非负数约束条件之下的矩阵分解方法。

图 3 展示了不同方法在 3 个数据集上的平均 *RMSE* 结果。

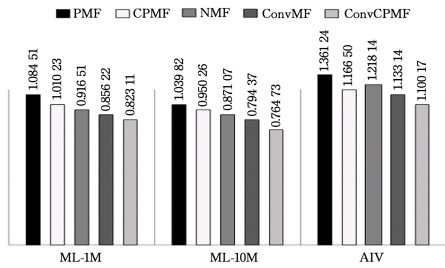

图 3 测试集上 3 个数据集的平均 RMSE

Fig. 3 Average RMSE of 3 datasets on test set

由图 3 知，在 3 个数据集上，ConvCPMF 的均方根误差分别为 0.823 11，0.764 73 和 1.100 17，比其他文献的相关模型 PMF，CPMF 和 NMF 在 3 个数据集上的最优性能分别提升了 9.34%，10.63%和 1.90%，说明卷积神经网络与约束概率矩阵分解相结合显著。

相对于较优的 ConvMF，本模型在 ML-1m 数据集上 RMSE 提高了 3.3%，在 ML-10m 数据集上 RMSE 提高了 2.86%。在一个相对比较稀疏和倾斜的亚马逊数据集上，本文的混合模型相对于 ConvMF 性能提高了 3.27%。这表明本文提出的推荐模型通过有效地分析文本信息能构建准确的项目潜在模型，同时也表明卷积神经网络已经很好地整合到约束概率矩阵分解中，即有效地利用了用户关于项目的评价信息。此外，本文用于实验的 3 个数据集稀疏度差异较为明显，从侧面也说明了本文提出的模型在数据稀疏问题上具有很好的泛化能力。

图 4－图 6 显示了在 3 个数据集上比较 PMF，CPMF，NMF，ConvMF 和本文模型 ConvCPMF 的 RMSE 结果趋势折线图。可以看出，ConvCPMF 表现较优，PMF 是较差的模型。此外，CPMF 的性能在 MovieLens 数据集上的表现比 NMF，ConvMF 和 ConvCPMF 差；而在稀疏的亚马逊数据集上，CPMF 的性能比 NMF 好。

综上所述，通过卷积和用户-项目评价矩阵的矩阵因式分解相结合，可以处理稀疏的用户-项目评级矩阵和额外的信息，为每个用户和项目学习更有效的潜在因素，从而提供了更准确的推荐。

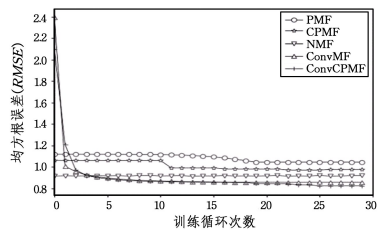

图 4 ML-1M 数据集上的 RMSE 折线图

Fig. 4 RMSE on ML-1M dataset

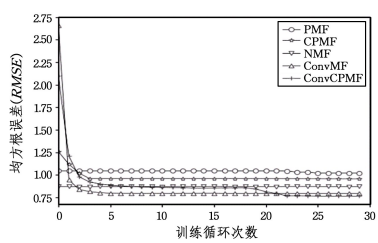

图 5 ML-10M 数据集上的 RMSE 折线图

Fig. 5 RMSE on ML-10M dataset

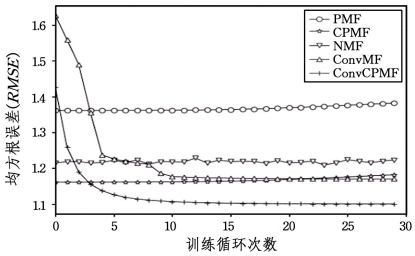

图 6 AIV 数据集上的 RMSE 折线图

Fig. 6 RMSE on AIV dataset

结束语 在丰富的社交网络资源和服务的今天,用户和商家对推荐系统精准性的需求变得愈发强烈。本文将协同过滤推荐算法与深度学习结合,主要是在评分稀疏性和上下文信息获取上进行模型的研究,提出了一种融合 CPMF 和 CNN 的推荐算法,将 CNN 集成到 CPMF 中,旨在评分预测的描述文本上获取上下文信息。实验结果表明,ConvCPMF 在数据信息稀疏性上相较于其他推荐模型,取得了显著的效果。我们下一步的研究主要围绕在传统矩阵分解上加入用户和项目的影响因子,进一步改善模型的推荐性能及准确性。

参 考 文 献

[1] ZHANG Z W. Summary of research on personalized recommendation algorithm [J]. Information and Computer (Theoretical Edition),2018(17):27-29.

[2] ZHANG M,DING B Y,MA W Z,et al. Hybrid recommendation method based on deep learning reinforcement[J]. Journal of Tsinghua University (Natural Science Edition),2017,57(10):1014-1021.

[3] LI F,XU G,CAO L. Coupled item-based matrix factorization [C]// International Conference on Web Information Systems Engineering. Cham:Springer,2014:1-14.

[4] WANG H,WANG N,YEUNG D Y. Collaborative deep learning for recommender systems[C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2015:1235-1244.

[5] GAO Q L,GAO L ,YANG J F,et al. Preference acquisition method based on user cognitive behavior in context-aware recommendation system [J]. Acta Computerica Sinica, 2015, 38(9):1767-1776.

[6] PENG Y,ZHU W,ZHAO Y,et al. Cross-media analysis and reasoning:advances and directions[J]. Frontiers of Information Technology &Electronic Engineering,2017,18(1):44-57.

[7] WER J,HE J,CHEN K,et al. Collaborative filtering and deep learning based recommendation system for cold start items[J]. Expert Systems with Applications,2016,69(10):1339-1351.

[8] CHEN P H,ZHU Y. A recommendation algorithm for fusion knowledge graph representation and matrix decomposition[J]. Computer engineering and design,2018,39(10):145-150.

[9] OKURA S,TAGAMI Y,ONO S,et al. Embedding-based news recommendation for millions of users[C]// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2017:1933-1942.

[10] VANDEN O A,DIELEMAN S,SCHRAUWEN B. Deep content-based music recommendation[C]// Advances in Neural Information Processing Systems. 2013:2643-2651.

[11] CHENG H T,KOC L,HARMSEN J,et al. Wide & deep learning for recommender systems[C]// Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. ACM, 2016:7-10.

[12] ZHENG L,NOROOZI V,YU P S. Joint deep modeling of users and items using reviewsfor recommendation[C]// Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. ACM,2017:425-434.

[13] RENDEL S. Factorization machines with libfm [J]. ACM Transactions on Intelligent Systems and Technology (TIST),2012, 3(3):57.

[14] WANG H,ZHANG F,XIE X,et al. DKN:Deep knowledge-aware network for news recommendation[C]// Proceedings of the 2018 World Wide Web Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2018:1835-1844.

[15] MNIH A,SALAKHUTDINOV R R. Probabilistic matrix factorization[C]// Advances in Neural Information Processing Systems. 2008:1257-1264.

[16] KIMD,PARK C,OH J,et al. Convolutional Matrix Factorization for Document Context-Aware Recommendation[C]// Acm Conference on Recommender Systems. ACM,2016:233-240.

[17] HARPER F M,KONSTAN J A. The movielens datasets:History and context[J]. Acm Transactions on Interactive Intelligent Systems (tiis),2016,5(4):19.

[18] DESHPANDEM,KARYPIS G. Item-based top- N recommendation algorithms[J]. ACM Transactions on Information Systems, 2004,22(1):143-177.

[19] LEE D D,SEUNG H S. Algorithms for non-negative matrix factorization[C]// Advances in Neural Information Processing Systems. 2001:556-562.

MA Hai-jiang, born in 1989, postgraduate. His main research interests include Intelligent data processing and analysis.