

融合多维语义表示的概率矩阵分解模型

任开旭,王玉龙,刘同存,李 炜

(北京邮电大学网络技术研究院,北京 100876)

摘 要: 协同过滤作为推荐系统核心技术,面临严重的评分数据稀疏性问题.融合物品文本信息可以有效的解决数据稀疏性问题,然而,目前的方法侧重于提取文本的单维特征,忽略了物品语义表示的多维特性.深度挖掘物品内容的多维特性可以更加精细化描述物品的语义信息,有助于提升推荐效果.为此,本文提出基于胶囊网络的概率生成模型.模型利用胶囊网络挖掘文本的多维语义特征,并以正则化方式融入概率矩阵分解框架,建立用户与物品之间的内在关系.实验结果表明本文提出的模型具有更高的评分预测精度.

关键词: 协同过滤; 概率矩阵分解; 胶囊网络; 多维语义特征; 正则化; 混合推荐

中图分类号: TP311 **文献标识码:** A **文章编号:** 0372-2112 (2019)09-1848-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2019.09.005

A Probabilistic Matrix Factorization Model Based on Multidimensional Semantic Representation Learning

REN Kai-xu, WANG Yu-long, LIU Tong-cun, LI Wei

(Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Collaborative filtering, as the core technology of recommendation systems, is currently facing the sparsity problem of rating data. This can be effectively solved through integrating item text information. However, current methods focus on extracting the one-dimensional features of the text, neglecting its multidimensional semantic features. Digging deeply into the multidimensional semantic features of the text can improve the recommendations. To help achieve this goal, a probabilistic matrix factorization model based on multidimensional semantic representation learning is proposed in the present study. The model uses a capsule network to mine the multidimensional semantic features of the text, and then integrates it into the probabilistic matrix decomposition framework using the regularization method to reveal hidden features linking users and items. Experimental results show that the proposed model has higher prediction accuracy.

Key words: collaborative filtering; probabilistic matrix factorization; capsule network; multidimensional semantic features; regularization; hybrid recommendation

1 引言

电商平台的爆发式发展使得用户陷入了商品选择困境.为此,推荐系统作为重要的工具成为各大电商平台的核心功能.近年来,协同过滤作为经典算法,已被成功应用于推荐系统中.然而该类算法却面临严重的数据稀疏性和冷启动问题.为了解决数据稀疏性和冷启动问题,常用的方法是向协同过滤融入辅助信息,如社交关系^[1,2]、用户评论和物品描述等^[3-12].

早期的研究工作,主要通过机器学习技术,利用物

品的内容特征学习用户-物品评分预测模型^[3],并作为用户-评分矩阵中的缺失值填充方法.该方法在一定程度上缓解了数据稀疏性问题,但仍然需要大量的用户-物品评分数据训练模型.为了进一步解决数据稀疏性问题,黄贤英等^[4]使用 TF-IDF 计算新闻内容的相似性,并采用加权方法与基于评分的相似性进行线性融合,该方法将物品的内容相似与行为相似进行了融合,有效改善了长尾物品相似度计算问题.这些方法虽然可以缓解数据稀疏性问题,但依赖统计的相似度计算仍然存在数据极端稀疏时置信度不足的问题.

收稿日期:2018-10-11;修回日期:2019-04-25;责任编辑:覃怀银

基金项目:国家自然科学基金(No. 61771068, No. 61671079);北京市自然科学基金(No. 4182041);教育部科学技术研究重点(重大)项目(No. MCM20130310)

近年来,随着物品内容信息的大量采集,向协同过滤融入物品内容特征挖掘算法成为解决数据稀疏性和物品冷启动推荐问题的有效方法.这些融合内容信息与协同过滤的混合推荐算法的核心是如何提取文本语义特征并巧妙的与协同过滤进行融合.主题模型和深度神经网络是最为常用的提取语义特征的方法,如潜语义主题模型(LDA)^[5]、多层感知机模型(MLP)^[6]、堆栈自编码模型(SDAE)^[7]等.然而,这些方法都是建立在词袋模型基础上,忽略了文本内在的语境和词序关系.为此, Kim 等^[8]使用卷积神经网络(CNN)学习文本内在的语境信息.为了处理文本的词序关系, Bansal 等^[9]使用循环神经网络学习文本的特征表示.这些模型将文本内容编码为由单维的语义特征所组成的向量,忽略了文本内容的多维语义特性.事实上,文本描述往往存在多个主题,每个主题均由不同的语义特征来表示,例如物品描述可能包含关于价格、关于商品质量和关于售后服务等主题,仅使用单维的语义特征表示缺乏对每个主题的深刻理解,无法充分提取文档中关于价格、商品质量和服务等主题的多方面的特征.

为了解决上述问题,本文提出了融合多维语义表示的概率矩阵分解模型(MFMSR)学习用户-物品之间的潜在关系.模型使用胶囊网络^[13]提取文档的多维语义特征表示,并进一步通过非线性变换转换为文档的内容特征向量,最后以正则化方式将提取的语义特征融入到概率矩阵分解^[14]框架.本文使用的胶囊网络模型利用胶囊向量代替 CNN 网络的标量输出,将文档编码为包含多个语义层面的特征表示,能够更全面的提取物品文本描述的多维语义信息;提取的多维语义特征在模型中搭建用户-物品关系数据与物品内容的桥梁,使得 MFMSR 可以缓解用户-物品交互行为数据的稀疏性,同时, MFMSR 可以利用物品内容特征向量预测用户对新物品偏好行为,可有效的解决物品冷启动推荐问题.

2 融合多维语义表示的概率矩阵分解模型

本节详细描述 MFMSR 模型,其主要包含两部分:(1)利用胶囊网络从物品的文本内容提取多维语义特征;(2)以正则化方式融入到概率矩阵分解模型.其总体架构如图 1 所示.为了便于表示,模型的主要符号及含义定义如表 1 所示.

设 N 个用户对 M 个物品的打分矩阵为 $\mathbf{R} \in \mathbb{R}^{N \times M}$,模型假定用户特征向量 \mathbf{u}_i 服从以 0 为均值, θ_u^2 为方差的正太分布;物品潜在特征向量 $\mathbf{\varepsilon}_j$ 服从以 0 为均值, θ_v^2 为方差的正太分布.为了解决用户-评分数据的稀疏性问题和物品冷启动推荐问题,模型假定物品的特征向量为胶囊网络提取的物品内容特征向量 $G(\mathbf{W}, \mathbf{X}_j)$ 与物品潜在特征向量 $\mathbf{\varepsilon}_j$ 之和:

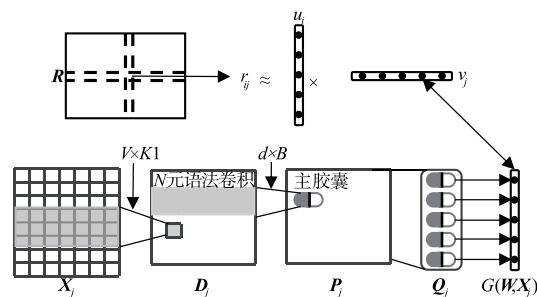


图1 融合多维语义表示的概率矩阵分解模型

表 1 主要符号及其含义

符号	含义
r_{ij}	用户 i 对物品 j 的真实评分
$\mathbf{u}_i, \mathbf{v}_j$	用户 i 和物品 j 的 k 维特征向量
\mathbf{X}_j	物品 j 的文本矩阵
\mathbf{O}_j	物品 j 的内容特征向量
\mathbf{W}	胶囊网络的所有权重矩阵和偏置向量.
θ_v^2	评分分布的方差
θ_u^2, θ_v^2	用户特征向量和物品特征向量分布的方差
k	潜在因子数目
L	物品文档中词的最大数量
V	词向量维度
N_r	已知评分数量
N, M	用户数量和物品数量

$$\mathbf{v}_j = G(\mathbf{W}, \mathbf{X}_j) + \mathbf{\varepsilon}_j \quad (1)$$

特别地,对于新物品 $\mathbf{\varepsilon}_j = \mathbf{0}$. 用户-物品评分矩阵中的任意元素均为用户与物品特征向量的内积:

$$\hat{r}_{ij} = \mathbf{u}_i^T \mathbf{v}_j \quad (2)$$

最后,本文采用坐标下降法学习模型中的参数.

2.1 多维语义特征表示

多维语义特征表示是指提取物品描述文档中的多维的语义特征,并将提取的多维语义特征映射到物品潜在特征向量 $\mathbf{\varepsilon}_j$ 相同的语义空间.在自然语言处理中,文本可以被映射到由 K_u 维的词向量所组成的文本矩阵 $\mathbf{X}_j = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_L]^T$. 基于胶囊网络的多维语义特征表示模型包含 N 元语法卷积、主胶囊、数字胶囊、转换等四步操作.

为了得到更丰富的文本特征,本文使用不同类型的卷积核提取文本特征.每个词向量视为不可分割的单位,卷积核的宽度是词向量的维度 V ,高度是可以设置的超参数,高度也被称为窗口,窗口大小表示周围几个词是相关的.那么第 a 个卷积核可以表示为一个宽度为 V ,窗口大小为 $K1$ 的共享矩阵 $\mathbf{W}_{CNN}^a \in \mathbb{R}^{V \times K1}$. 第 j 个文

档 X_j 经第 a 个卷积核提取的 N 元语法向量 D_j^a 可以表示为:

$$D_j^a = f(X_j * W_{\text{CNN}}^a + b_{\text{CNN}}^a) \quad (3)$$

这里 $*$ 表示卷积操作, b_{CNN}^a 表示偏置, f 表示 ReLu 激活函数. 假设卷积核的数量是 A , 则由 A 个 N 元语法向量组成的文本矩阵 D_j 表示为:

$$D_j = [D_j^1, D_j^2, \dots, D_j^A] \quad (4)$$

主胶囊层将 N 元语法卷积层检测的基本特征转化为胶囊向量. 胶囊可以同时表达词语的属性和词语之间的语义关系. 主胶囊层可分为卷积操作和维度变换两步操作, 其中卷积操作仍为一维卷积. 令 d 表示输出胶囊的维度, 令 B 表示主胶囊层的通道数, 令 $K2$ 表示卷积核的高度. 在卷积操作中, 卷积核的数量为 $d \times B$, 高度为 $K2$, 宽度为文本矩阵 M_j 的列数 A , 文本矩阵 D_j 经过卷积操作和维度变换得到维度为 $((L - K1 - K2 + 2) \times B) \times d$ 的文本矩阵 P_j , 表示由 $(L - K1 - K2 + 2) \times B$ 个胶囊组成的二维矩阵.

$$P_j = [p_1, p_1, \dots, p_{(L-K1-K2+2) \times B}] \quad (5)$$

数字胶囊层由非线性压缩函数和动态路由^[13]两部分构成. 数字胶囊层的输入来自主胶囊层的输出, 胶囊的方向编码了检测出的相应对象的一些内部特征, 经过特征映射函数 $p_{jli} = W_{ij} p_i$, 转化胶囊. W_{ij} 编码着底层特征和高层特征之间的空间关系和其他重要关系. p_{jli} 通过标量权重 c_{ij} 确定底层胶囊和上层胶囊之间的耦合系数, c_{ij} 通过动态路由算法学习得到, 这种方式避免了最大池化带来的信息损失, 更智能地将局部分配给整体. 上层胶囊由底层所有胶囊通过式(6)加权求和与式(7)压缩得到输出胶囊 q_j .

$$s_j = \sum_i c_{ij} p_{jli} \quad (6)$$

$$q_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (7)$$

假设数字胶囊层的输出胶囊数量为 k , 维度为 e . 经过以上计算, 最终将主胶囊层输入的 $(L - K1 - K2 + 2) \times B$ 个胶囊 p 映射为 k 个胶囊 q , 构成文本多维语义表示 Q_j .

$$Q_j = [q_1, q_2, \dots, q_k]^T \quad (8)$$

为了将由胶囊向量组成的多维语义表示转换为与 k 维潜在特征向量 ϵ_j 同维度的物品内容特征向量, 数字胶囊层输出胶囊的个数为 k , 计算每个胶囊的长度, 最终输出 k 维表示向量 O_j .

$$O_j = [\|q_1\|, \|q_2\|, \dots, \|q_k\|]^T \quad (9)$$

2.2 模型优化

将上述得到的物品内容特征向量, 以正则化方式融入到概率矩阵分解模型, 得到损失函数, 如式(10)

$$J = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (r_{ij} - u_i^T v_j)^2 + \frac{\lambda_u}{2} \sum_{i=1}^N \|u_i\|_2^2 + \frac{\lambda_v}{2} \sum_{j=1}^M \|v_j - G(W, X_j)\|_2^2 + \frac{\lambda_w}{2} \sum_{l=1}^{L'} \sum_{h=1}^{H_l} \|w_{l,h}\|_2^2 \quad (10)$$

其中, I_{ij} 是指示函数, 如果第 i 个用户对第 j 个物品打过分, 则 $I_{ij} = 1$, 否则 $I_{ij} = 0$, $\lambda_u = \theta_r^2 / \theta_u^2$, $\lambda_v = \theta_r^2 / \theta_v^2$, λ_w 为 $w_{l,h}$ 正则化系数. l 表示神经网络第 l 层, h 表示第 l 层网络第 h 个神经元, L' 表示神经网络层数, H_l 表示第 l 层网络的神经元个数, $w_{l,h}$ 表示第 l 层第 h 个神经元所对应的权重向量和偏置.

本文使用坐标下降法优化参数 U, V, W . 求损失函数 J 对 u_i 和 v_j 的偏导:

$$\frac{\partial J}{\partial u_i} = \sum_{j=1}^M (u_i^T v_j I_{ij} v_j^T - r_{ij} v_j^T) + \lambda_u u_i^T \quad (11)$$

$$\frac{\partial J}{\partial v_j} = \sum_{i=1}^N (u_i I_{ij} u_i^T v_j - u_i r_{ij}) + \lambda_v (v_j - G(W, X_j)) \quad (12)$$

令 u_i, v_j 的导数等于 0, 求解方程, 更新参数.

对于参数 W 的优化, 模型通过固定 u_i 和 v_j 得到胶囊网络模型的损失函数:

$$\ell(W) = \frac{\lambda_v}{2} \sum_{j=1}^M \|v_j - O_j\|_2^2 + \frac{\lambda_w}{2} \|W\|_2^2 \quad (13)$$

然后使用反向传播算法优化参数 W .

算法的时间复杂度主要由更新用户矩阵 U 、物品矩阵 V 和胶囊网络参数 W 所决定. 其中, 更新参数 U, V 的时间复杂度为 $O(k^2 N_r + k^3 N + k^3 M)$, 更新胶囊网络参数 W 的时间复杂度为 $O(ML(VK1A + AK2Bd + Bdek))$. 因此, 整个模型的时间复杂度为 $O(ML(VK1A + AK2Bd + Bdek) + k^2 N_r + k^3 N + k^3 M)$.

3 实验分析

3.1 数据集与评估标准

模型在 Yelp Dataset, MovieLens-1m 和 MovieLens-10m 三个真实的数据集进行实验分析. Yelp Dataset 选取商户最多的城市 Las Vegas, 包含用户对商户的评分和用户对商户的精简描述. MovieLens 数据集包含用户对电影的评分和电影情节等信息, 其中电影情节从 IMDb 得到.

对于评分数据, 我们去除没有文本描述的物品, 并将整个数据集划分训练集、验证集和测试集. 为确保三个数据集均含有同一用户的评分数据, 去除少于 3 条评分记录的用户. 对于物品的文本描述, 通过分词、去除停用词、高频词以及根据词的 TF-IDF 值进行过滤, 建立文本描述的词典. 最终数据情况如表 2 所示.

表 2 数据详细信息

数据集	用户量	物品量	评分数量	数据稀疏程度
Yelp (LasVegas)	14401	3904	370029	0.66%
ML-1m	6040	3544	993482	4.64%
ML-10m	69878	10073	9945875	1.41%

同时为了测试物品冷启动,本文使用类似 Schein 等人^[15]测试物品冷启动的方法,在确保训练集中的用户包含测试集所有用户前提下,从数据集的物品集合中随机选择一定数量物品,将与这些物品相关的所有评分均作为测试集,其余数据划分为训练集与验证集.经处理,ML-1m 测试集物品数量为 1888,评分数量为 99498;ML-10m 测试集物品数量为 5833,评分数量为 923619.

为了评估算法在三个真实的数据集上的效果,采用平均绝对误差(MAE)和均方根误差(RMSE)进行评估.其计算方法见式(14)、(15):

$$MAE = \frac{1}{N_r} \sum_{i,j} |r_{ij} - \hat{r}_{ij}| \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i,j} (r_{ij} - \hat{r}_{ij})^2}{N_r}} \quad (15)$$

为了验证提出模型的有效性,本文实验选择以下 5 种模型与本文模型 MFMSR 进行对比:

(1) PMF(Probabilistic Matrix Factorization)^[14]:评分矩阵的标准模型,该模型仅仅利用用户物品评分数据,没有加入任何辅助信息,使用随机梯度下降法优化模型参数.

(2) ALS(Alternating Least Squares)^[16]:为了解决传统矩阵分解在大数据量的计算问题,传统矩阵分解通常使用梯度下降的方法优化模型参数,ALS 使用交替最小二乘法优化模型参数.

(3) CTR(Collaborative Topic Regression)^[17]:将概率矩阵分解(PMF)和主题模型(LDA)结合的混合推荐

算法,该模型同时利用评分数据与物品文本信息.

(4) CDL(Collaborative Deep Learning)^[18]:使用堆栈式自编码(SDAE)提取物品文本特征,同时利用评分数据和物品文本信息的混合推荐算法.

(5) ConvMF(Convolutional Matrix Factorization)^[8]:使用卷积神经网络(CNN)提取物品文本特征,是当前效果最好的同时利用评分数据和物品文本信息的混合推荐算法.

3.2 性能对比

为了更全面的测试本文模型的效果,笔者在不同数据集下测试各模型的整体性能,同时在 ML-1m 和 ML-10m 数据集上测试各模型的物品冷启动的性能.实验结果分别如表 3 与表 4.从表 3 我们可以看出本文提出的 MFMSR 模型在不同数据集,其精度均高于对比模型;CTR、CDL、ConvMF 和 MFMSR 等模型均明显优于仅使用评分数据的 PMF 和 ALS 算法.实验结果表明利用物品文本信息能够有效提高推荐预测准确率,相比使用了其他文本算法的模型,本文提出的多维语义表示能够更有效地提取物品文本内容特征,融合概率矩阵分解模型,能有效提升推荐预测准确率,且针对不同数据集具有一定的鲁棒性.表 4 结果显示 MFMSR 模型的物品冷启动预测精度明显高于其他三种算法,在物品冷启动预测中,物品特征向量等于语言模型提取的物品内容特征向量,预测精度更加依赖物品内容特征,因而进一步表明本文提出的多维语义表示能够更有效地表达新物品与用户历史交互物品的内容相似性,有助于模型更精准的预测用户偏好.

表 3 不同推荐算法在评分预测上的性能对比

数据集	评测指标	PMF	ALS	CTR	CDL	ConvMF	MFMSR	误差降低(%)
Yelp (LasVegas)	MAE	0.95723	0.95919	0.87196	0.86284	0.84579	0.82913	1.97%
	RMSE	1.20982	1.19460	1.09980	1.08990	1.07809	1.06701	1.03%
ML-1m	MAE	0.73692	0.70682	0.70859	0.69926	0.67259	0.66388	1.29%
	RMSE	0.93339	0.90597	0.88905	0.87876	0.85553	0.84842	0.83%
ML-10m	MAE	0.83015	0.71006	0.65777	0.63046	0.61827	0.60209	2.62%
	RMSE	1.06365	0.93991	0.86635	0.82936	0.80691	0.78510	2.70%

3.3 参数分析

用户及物品特征向量的特征维度 k 决定特征向量承载信息上限,本文使用不同特征维度 k 在 Yelp (Las-

Vegas) 与 ML-1m 数据集上测试其对 MFMSR 模型性能的影响,如图 2 和图 3 所示, k 取值过小,特征向量不能充分表达用户和物品所具有的隐含特征,导致算法无

法精准预测用户对物品的评分; k 取值过大,特征向量 会包含过多的冗余特征,导致算法预测准确率下降.

表 4 不同推荐算法在物品冷启动评分预测上的性能对比

数据集	评测指标	CTR	CDL	ConvMF	MFMSR	误差降低(%)
ML-1m	MAE	1.24575	1.13983	0.96341	0.93646	2.80%
	RMSE	1.48145	1.36129	1.18572	1.1448	3.45%
ML-10m	MAE	1.03094	0.89397	0.86973	0.84194	3.20%
	RMSE	1.2298	1.09062	1.08984	1.03988	4.58%

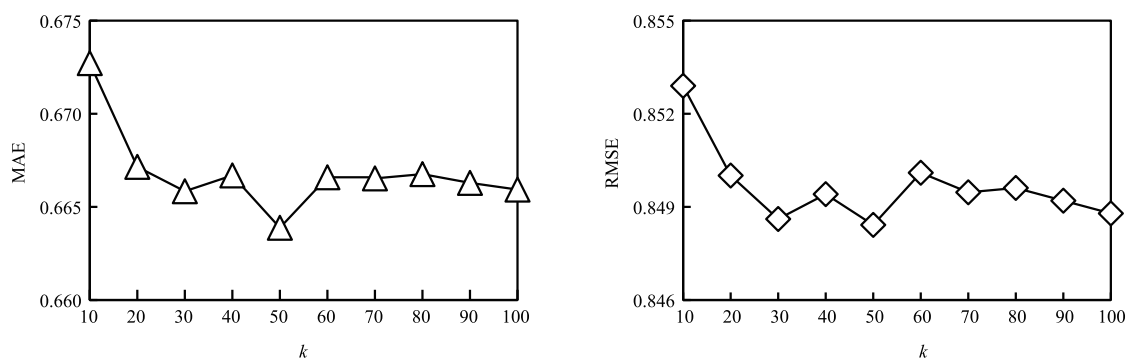


图2 MFMSR在Yelp (LasVegas)数据集上不同特征维度下的MAE与RMSE

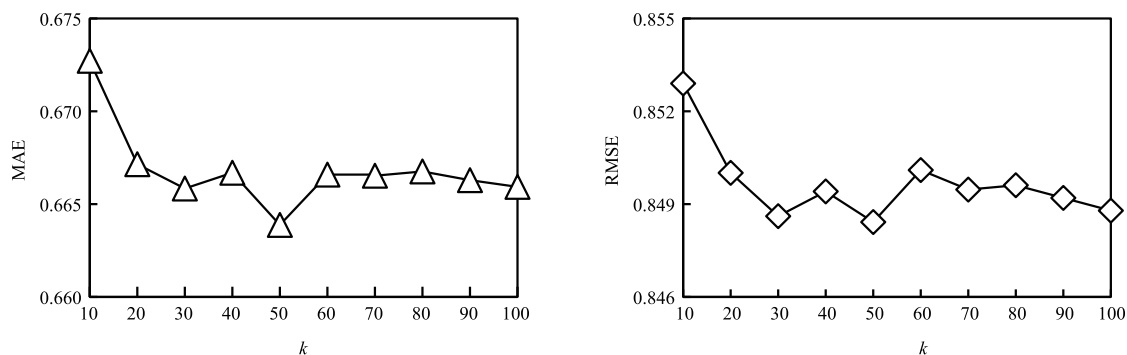


图3 MFMSR在ML-1m数据集上不同特征维度下的MAE与RMSE

正则化参数 λ_u, λ_v 用于约束模型参数,防止过拟合,由于两个参数的比例影响模型效果,因此本文 MFMSR 模型在 Yelp (Las Vegas) 与 ML-1m 数据集上通过交替改变 λ_u 和 λ_v 测试参数对模型的效果.如表 5 和表 6 所示, MFMSR 模型在 Yelp (Las Vegas) 与 ML-1m 数据集上正则化参数 λ_u, λ_v 的最优值分别为 10、100 与 1、100. 预测精度与正则化参数 λ_u, λ_v 呈现非线性关系,正则化参数 λ_u 与 λ_v 是为了调节评分数据与物品文本信息对算法的影响程度,对于不同的数据集,需要根据数据集中评分数据的稀疏程度与物品文本信息质量确定参数值,但为了减少人为主观性,常常使用网格搜索探索两个参数的最优值.

表 5 MFMSR 在 Yelp (Las Vegas) 数据集上不同 λ_u 和 λ_v 下的 MAE 与 RMSE

λ_u	评价指标	λ_v		
		1	10	100
1	MAE	0.98222	0.84026	0.86805
	RMSE	1.25623	1.07914	1.11800
10	MAE	0.93959	0.88476	0.82913
	RMSE	1.19114	1.13180	1.06701
100	MAE	0.88111	0.91017	0.94701
	RMSE	1.10341	1.12538	1.16321

表 6 MFMSR 在 ML-1m 数据集上不同 λ_u 和 λ_v 下的 MAE 与 RMSE

λ_u	评价指标	λ_v		
		1	10	100
1	MAE	0.69767	0.69167	0.66388
	RMSE	0.88833	0.88612	0.84842
10	MAE	0.74455	0.67652	0.66513
	RMSE	0.95258	0.86356	0.84843
100	MAE	0.67649	0.67154	0.70461
	RMSE	0.86234	0.85039	0.88854

表 7 模型训练时间(单位:s)

数据集	PMF	ALS	CTR	CDL	ConvMF	MFMSR
ML-1m	646.56	95.89	851.42	701.75	379.42	1254.71
ML-10m	3295.96	660.07	4270.5	2321.45	1264.88	4944.3
Yelp (Las Vegas)	2658.72	494.82	2761.55	1239.43	1069.78	2026.43

3.5 可解释性分析

为了说明本文模型的语义表达能力的效果,本文选择 ConvMF 模型与本文模型进行相似电影预测的对比实验.分别使用已训练好的最优模型学习 ML-1m 中电影的内容特征向量;实验设置为随机选择一部电影 M ,根据电影的内容特征向量为电影 M 计算最相似的 10 部电影,这里将包含电影 M 主题的相似电影作为正确预测;此实验执行 6 次,每次选择不同的电影.实验结果如图 4 所示, MFMSR 正确预测数量明显多于 ConvMF.结果表明 MFMSR 模型能够更有效地表达文本的主题信息,更准确地学习文本之间的相似性,促进模型为用户推荐与历史交互更为相似的物品,进而提升推荐预测精度.

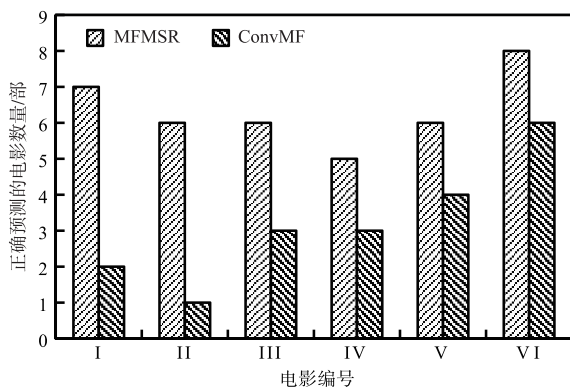


图 4 相似电影预测结果对比图

4 结论

针对协同过滤面临严重的数据稀疏性和物品冷启动问题,本文提出了融合多维语义表示的概率矩阵分解模型,该模型使用胶囊网络提取物品文本内容的多

3.4 时间消耗分析

本文选择三个公开数据集作为模型输入,得到各模型在模型训练时间.实验结果如表 7 所示,使用交替最小二乘法学习参数的 ALS 模型训练时间明显优于使用梯度下降的 PMF 模型;CTR、CDL、ConvMF 和 MFMSR 等模型训练时间明显大于 ALS 算法.虽然 MFMSR 模型相比其他模型训练时间均有所上升.但模型训练属于离线过程,在线预测只需训练得到的用户物品特征向量的点积运算,所有模型均相同;对于新物品,训练完成的 MFMSR 模型可直接预测物品向量,运算时间短.

维语义特征表示,通过非线性变换方式转换为物品的内容特征,并以正则化方式无缝融合到概率矩阵分解模型.通过在三个真实数据集上的大量实验,验证了本文模型的有效性.未来工作将围绕用户描述性资料 and 社交信息进一步完善推荐模型.

参考文献

- [1] Ma H, Yang H, Lyu M R, et al. Sorec: social recommendation using probabilistic matrix factorization [A]. Proceedings of the 17th ACM Conference on Information and Knowledge Management [C]. California: ACM, 2008. 931-940.
- [2] 陈克寒, 韩盼盼, 吴健. 基于用户聚类的异构社交网络推荐算法 [J]. 计算机学报, 2013, 36(2): 349-359. Chen Ke Han, Han Pan Pan, Wu Jian. User clustering based social network recommendation [J]. Chinese Journal of Computers, 2013, 36(2): 349-359. (in Chinese)
- [3] Melville P, Mooney R J, Nagarajan R. Content-boosted collaborative filtering for improved recommendations [A]. Proceeding Eighteenth National Conference on Artificial Intelligence [C]. Menlo Park: AAAI, 2002. 187-192.
- [4] 黄贤英, 熊李媛, 李沁东. 基于改进协同过滤算法的个性化新闻推荐技术 [J]. 四川大学学报(自然科学版), 2018, 55(01): 49-55. Huang X Y, Xiong L Y, Qin-Dong LI. Personalized news recommendation technology based on improved collaborative filtering algorithm [J]. Journal of Sichuan University, 2018, 55(01): 49-55. (in Chinese)
- [5] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(Jan): 993-1022.

- [6] Hsieh C K, Yang L, Cui Y, et al. Collaborative metric learning[A]. Proceedings of the 26th International Conference on World Wide Web[C]. Republic and Canton of Geneva; IW3C2, 2017. 193 – 201.
- [7] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11 (Dec): 3371 – 3408.
- [8] Kim D, Park C, Oh J, et al. Convolutional matrix factorization for document context-aware recommendation[A]. Proceedings of the 10th ACM Conference on Recommender Systems[C]. Boston: ACM, 2016. 233 – 240.
- [9] Bansal T, Belanger D, McCallum A. Ask the gru: Multi-task learning for deep text recommendations[A]. Proceedings of the 10th ACM Conference on Recommender Systems[C]. Boston: ACM, 2016. 107 – 114.
- [10] Li Z, Tang J. Weakly supervised deep matrix factorization for social image understanding[J]. IEEE Transactions on Image Processing, 2017, 26(1): 276 – 288.
- [11] Li Z, Tang J, Mei T. Deep collaborative embedding for social image understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 1(4): 1 – 7.
- [12] Liao J, Liu T, Liu M, et al. Multi-context integrated deep neural network model for next location prediction[J]. IEEE Access, 2018, (6): 21980 – 21990.
- [13] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[A]. Advances in Neural Information Processing Systems[C]. California: NIPS, 2017. 3856 – 3866.
- [14] Mnih A, Salakhutdinov R R. Probabilistic matrix factorization[A]. Advances in Neural Information Processing Systems[C]. California: NIPS, 2008. 1257 – 1264.
- [15] Schein A I, Popescul A, Ungar L H, et al. Methods and metrics for cold-start recommendations[A]. Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. Tampere: ACM, 2002. 253 – 260.
- [16] Zhou Y, Wilkinson D, Schreiber R, et al. Large-scale parallel collaborative filtering for the netflix prize[A]. International Conference on Algorithmic Applications in Management[C]. Shanghai: AAIM, 2008. 337 – 348.
- [17] Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles[A]. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. San Diego: ACM, 2011. 448 – 456.
- [18] Wang H, Wang N, Yeung D Y. Collaborative deep learning for recommender systems[A]. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. Sydney: ACM, 2015. 1235 – 1244.

作者简介



任开旭 (通信作者) 男, 1992 年生于河南信阳. 现为北京邮电大学硕士研究生. 主要研究方向为推荐算法.
E-mail: renkaixu@ebupt.com



王玉龙 男, 1968 年生于山东鱼台. 现为北京邮电大学副教授、硕士生导师. 主要研究方向为移动增值业务系统, 业务网络智能化.
E-mail: wangyulong@ebupt.com



刘同存 男, 1984 年生于山东临沂. 现为北京邮电大学博士研究生. 主要研究方向为推荐算法.
E-mail: liutongcun@ebupt.com



李 炜 男, 1973 年出生于天津, 现为北京邮电大学副教授、硕士生导师. 主要研究方向为业务网络智能化, 移动增值业务系统.
E-mail: liwei@ebupt.com