

基于深度学习的推荐算法研究综述

王俊淑^{1,2}, 张国明^{3,4}, 胡 斌^{1,2}

(1.南京师范大学地理科学学院,江苏 南京 210023)

(2.南京师范大学虚拟地理环境教育部重点实验室,江苏 南京 210023)

(3.南京大学计算机科学与技术系,江苏 南京 210023)

(4.江苏省卫生统计信息中心,江苏 南京 210008)

[摘要] 深度学习技术是机器学习领域的一个研究热点,已被深入研究并广泛应用于许多领域.推荐系统是缓解信息过载的重要技术,如何将深度学习融入推荐系统,利用深度学习的优势从各种复杂多维数据中学习用户和物品的内在本质特征,构建更加符合用户兴趣需求的模型,以提高推荐算法的性能和用户满意度,是深度学习应用于推荐系统的主要研究任务.对基于深度学习的推荐算法研究和应用现状进行了综述,讨论并展望了深度学习应用于推荐系统的研究发展趋势.

[关键词] 推荐系统,深度学习,协同过滤,内容推荐,动态推荐,标签推荐

[中图分类号] TP3-05 **[文献标志码]** A **[文章编号]** 1672-1292(2018)04-0033-11

A Survey of Deep Learning Based Recommendation Algorithms

Wang Junshu^{1,2}, Zhang Guoming^{3,4}, Hu Bin^{1,2}

(1.School of Geography, Nanjing Normal University, Nanjing 210023, China)

(2.Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Nanjing 210023, China)

(3.Department of Computer Science and Technology, Nanjing University, Nanjing 210023, China)

(4.Health Statistics and Information Center of Jiangsu Province, Nanjing 210008, China)

Abstract: Deep learning technology has recently become a very hot topic in the machine learning field, and has been thoroughly studied and widely used in many fields. The recommender system is an important technology to alleviate information overload. Deep learning has the advantages of learning the intrinsic characteristics of users and items from various complex multidimensional data. How to integrate deep learning into recommender systems to build a model that is more in line with the user preferences and improving recommendation performance is the main research task of deep learning based recommender systems. In this paper, the research and application of deep learning based recommender algorithms are reviewed, and the future research and development trends of integrating deep learning to recommender systems are discussed.

Key words: recommender systems, deep learning, collaborative filtering, content-based recommendation, dynamic recommendation, tag-based recommendation

近年来,深度学习(deep learning)技术成为机器学习领域理论及应用研究的热点,掀起人工智能和机器学习研究的又一次浪潮.与传统的浅层机器学习相比,深度学习具有优秀的自动提取抽象特征的能力,可从海量数据中学习更有效的特征.随着云计算和大数据技术的发展,深度学习在工程界得到了广泛应用,在图像识别、语音识别、自然语言处理等领域都取得了比传统机器学习更好的效果.另一方面,随着物联网、互联网、电子商务、社交网络等技术的迅猛发展,网络上所蕴含的信息量呈指数级增长^[1],推荐系统作为缓解信息过载的重要手段,得到了学术界和工业界的广泛研究和应用,并取得了一定成果.推荐系统本质上是从一堆看似杂乱无章的原始数据中,抽象出用户的兴趣特征,挖掘用户的偏好.把深度学习与推荐系统相结合已成为近年来推荐系统发展的一个新方向.事实上,Spotify、Netflix、谷歌、微软等公司已对如

收稿日期:2018-04-07.

基金项目:国家自然科学基金(41571389)、江苏省自然科学基金(BK20171037)、江苏省高校自然科学研究面上项目(17KJB420003).

通讯联系人:张国明,工程师,研究方向:推荐系统、健康大数据. E-mail: zgmiming@qq.com

何将深度学习应用到推荐系统中进行了深入研究,并在实际应用中取得了很好的效果^[2-5]。

1 推荐系统概述

随着上世纪 90 年代协同过滤(collaborative filtering, CF)算法被提出后,推荐系统就成为一门独立的学科被深入研究。近年来,互联网、电子商务的飞速发展,使得网络上各种各样的信息爆炸式增长,将用户淹没在信息的海洋中。推荐系统成为缓解信息过载的重要方法,可以为用户在选择物品和服务时提供决策帮助。传统的推荐算法主要分为 3 大类:协同过滤推荐算法、基于内容(content-based)的推荐算法、混合(Hybrid)推荐算法^[1]。

协同过滤算法是推荐系统应用最广泛的算法,主要分为基于内存(又称基于近邻^[6])的协同过滤算法(memory-based CF)和基于模型的协同过滤算法(model-based CF)^[7]。基于内存的协同过滤算法又分为基于用户的协同过滤算法(UserCF)和基于物品的协同过滤算法(ItemCF)^[8]。基于用户的协同过滤算法是最早出现的推荐算法,算法首先计算和目标用户兴趣相似的用户集合,然后为目标用户推荐该相似用户集合中用户喜欢且未接触过的物品。基于物品的协同过滤算法是目前业界应用最多的算法,该算法给用户推荐那些和他们之前喜欢的物品相似的物品。ItemCF 算法并不利用物品的内容属性计算物品之间的相似度,它主要通过分析用户的行为记录计算物品之间的相似度^[9]。

基于模型的协同过滤算法主要通过机器学习和数据挖掘模型,利用分类、回归、矩阵分解等算法提取用户和物品的隐含模式进行推荐。其中代表性的有基于贝叶斯信念网络的算法^[10-11]、基于聚类模型的算法^[12-13]、基于回归模型的算法^[14-16]、基于矩阵分解模型的算法^[17-18]等。与基于内存的协同过滤算法相比,基于模型的协同过滤算法具有较好的理论基础,且占用内存小,可缓解基于内存的方法中的数据稀疏和可扩展性弱等问题,同时提高了推荐算法的精度,但其无法提供推荐解释。

基于内容的推荐算法通过分析物品内容信息(如物品属性、描述等),提取用户对物品的兴趣偏好进行推荐。物品的内容可通过物品属性的向量空间模型表示,若物品内容是文本形式,则需要引入自然语言处理技术抽取关键词,生成关键词向量。根据物品向量计算物品的相似度,利用 ItemCF 算法的思想,给用户推荐和其历史上喜欢的物品内容相似的物品。一般认为协同过滤算法优于基于内容的推荐算法,但协同过滤存在冷启动和数据稀疏问题,基于内容的推荐算法常被用于评分数据非常稀疏的场景和解决物品的冷启动问题^[19]。此外,在物品内容特征非常强的情况下,基于内容的推荐算法更有优势^[9]。

基于内容的推荐算法和协同过滤推荐算法都存在各自的缺点,混合推荐算法通过组合多种推荐技术解决单一推荐模型的不足。根据组合方法的不同,文献[20]总结了 6 类混合推荐算法:加权混合推荐算法^[21]、交叉调和推荐算法^[22]、特征混合的推荐系统、瀑布型混合方法、特征扩充的推荐算法^[23]、元模型混合推荐算法^[24]。

2 深度学习概述

深度学习^[25-26](deep learning, DL)算法的概念源于神经网络的研究,是对神经网络的发展,与神经网络具有相似的分层结构。相比于传统神经网络,深度学习做出了重大的改进:通过构建具有多个隐层的非线性网络结构来实现复杂的函数逼近;通过海量的训练数据来学习更抽象的深层次特征提升分类或预测的准确性;通过逐层训练来有效降低训练难度(如梯度弥散问题)。深度学习可自动学习特征,不必请专家手工构造特征,极大地推进了智能化、自动化。通过对数据的逐层学习,深度学习将低层特征进行抽取或组合,形成更加抽象的高层特征,发现数据的内在规律,更能够刻画数据的本质和丰富的内在信息。实践表明,深度学习能提取高维数据的本质特征,具有很强的建模和推广能力。深度学习在图像识别^[27-29]、自然语言处理^[30-32]、语音识别^[33-34]等领域取得了很好的成绩,在很多其他领域也有非常优秀的表现^[35-37]。

深度学习作为机器学习的一个分支,其学习方法可以分为监督学习和无监督学习。监督学习模型包括多层感知机、卷积神经网络(convolutional neural networks, CNN)、循环神经网络(recurrent neural networks, RNN)等^[38]。无监督学习模型主要包括受限玻尔兹曼机(restricted Boltzmann machine, RBM)、深度置信网络(deep belief nets, DBN)、稀疏编码(sparse coding)、自动编码器(auto encoders)、降噪自动编码

器(denoising autoencoder)等^[38].

3 基于深度学习的推荐算法

推荐系统是人工智能和机器学习研究的一个分支. 由于深度学习在许多领域取得了巨大成功,许多研究人员开始将深度学习与推荐算法结合,解决推荐系统的各种问题(如冷启动、稀疏性等),提高推荐性能. 本文将当前基于深度学习的推荐算法研究分为4类:利用辅助信息的深度学习推荐算法;基于模型的深度学习推荐算法;动态深度学习推荐算法;基于标签的深度学习推荐算法. 其中,利用辅助信息的深度学习推荐算法又可分为3种类型:利用辅助信息仅提取用户(或物品)的特征表示;利用辅助信息分别提取用户和物品的特征表示;利用辅助信息提取用户和物品的共同特征表示.

3.1 利用辅助信息的深度学习推荐算法

深度学习在特征提取和特征表示方面与传统方法相比有很大优势,因此,利用辅助信息提取用户和物品的分布式特征表示并融合其他模型进行推荐,是深度学习在推荐领域应用最多的方法. 辅助信息主要包括用户的人口学特征、物品内容特征、上下文特征(时间上下文、位置上下文、环境上下文)等. 深度学习可以融合任意连续的特征和不同类别特征,自动提取用户和物品更高层的隐含特征表示,在很多数据集中的表现优于传统的利用辅助信息的推荐算法.

3.1.1 利用辅助信息提取物品(或用户)的特征表示

这类深度学习推荐算法仅利用辅助信息提取物品(或用户)的特征表示,然后融合近邻方法或概率分解进行推荐,一方面可以解决物品(或用户)冷启动问题,另一方面在很多应用领域提高了推荐性能. 其模型框架如图1所示.

部分论文提取物品特征表示后与基于物品近邻的协同过滤结合. 与传统基于近邻的算法区别在于:传统算法仅考虑用户对物品的反馈来计算物品的相似性,深度学习模型还用到了其他辅助信息计算相似度.

在音乐推荐中,一般认为协同过滤算法优于基于内容的推荐算法,但协同过滤存在冷启动问题. 为解决音乐推荐冷启动问题(如推荐新歌和非流行歌曲),Oord等人^[39]使用卷积神经网络从音乐文件的声音信号中提取时间-频率特征表示,结果优于传统的词袋模型. 文章认为卷积神经网络可以在不同的特征中共享中间特征,且池化层可操作多个时间尺度,因此非常适合提取音频的隐含特征.

在图像推荐中,McAuley^[40]和Zhou等人^[41]使用卷积神经网络,从ImageNet数据集中提取图像视觉特征向量,根据视觉特征所计算的物品之间的关联关系以及用户对物品的视觉风格偏好进行推荐.

在文章推荐中,Bansal等人^[42]使用可以挖掘单词序列的深度循环神经网络提取文本隐含特征向量,并结合多任务学习解决数据稀疏问题,在科学论文推荐中明显提升了推荐准确率,在冷启动问题上也取得了较好的效果.

在视频(电影)推荐中,Covington等人^[4]提出深度协同过滤模型,首先利用深度候选视频生成模型检索出候选集,然后利用深度排序模型对候选视频排序,优于YouTube之前基于矩阵分解的模型. 与矩阵分解模型相比,该模型可有效融合任意连续的特征和不同类别特征. 模型根据用户历史活动(观看记录、搜索记录)、上下文(所在地区、设备)以及人口学信息(年龄、性别等)利用多层全连接神经网络学习用户特征向量,模型学习时可看作一个极大多分类问题(每个视频作为一类),考虑到显示反馈(点赞、评分等)非常稀疏,采用隐式反馈训练网络,即用户看完的视频作为正样本. 推荐时根据学习的用户向量使用基于用户近邻的协同过滤筛选候选视频. 排序时,将视频的相关信息以及与视频关联的用户之间的关系共上百个特征作为输

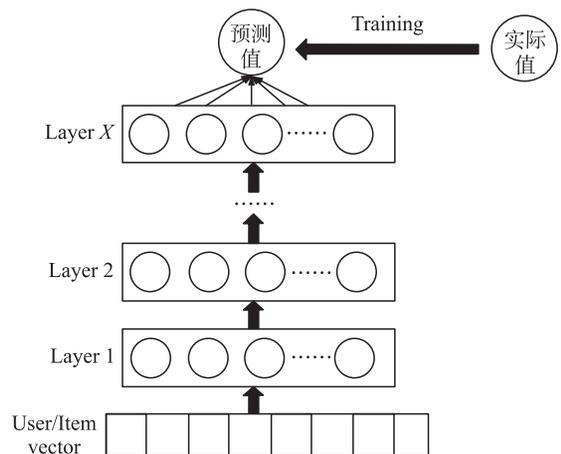


图1 仅利用辅助信息提取物品(或用户)的特征表示

Fig. 1 Extract feature of items(or users) by auxiliary information

入,与候选视频生成网络相似,构建多层全连接深度神经网络,并使用加权逻辑回归对每个候选视频基于观看时长打分并排序.与文献[4]类似,Zanotti 等人^[43]利用神经网络语言模型 CBOW (continuous bag of words) 和 Skip-gram^[44]学习电影多个来源的特征(如用户评分数据、标签数据、物品数据等),发现隐含的语义关系,提取用户和物品更丰富的分布式特征表示,并根据学习的用户(物品)特征使用传统的基于用户(物品)近邻的协同过滤进行评分预测.

部分论文提取物品(或用户)特征表示后,融合矩阵分解模型提取物品和用户的隐含语义表示进行推荐,在冷启动问题中有显著的效果.为处理数据稀疏问题,Wang 等人^[45]提出了一种通用的协同深度学习模型(collaborative deep learning, CDL),通过降噪自动编码器^[46]从用户评论中学习物品的深度特征表示,同时利用协同主题回归^[47]将学习的物品语义特征与标准的概率矩阵分解模型相结合进行预测.类似的,为解决冷启动问题,Wei 等人^[48]将物品内容信息用词袋模型转换成向量表示,然后使用堆栈降噪自动编码器学习内容特征,最后融合兼顾时间信息的矩阵分解模型 timeSVD++^[49]进行评分预测,取得了更好的效果.为解决文献[45]无法提取文本上下文信息的问题,Kim 等人^[50]利用 CNN 提取文档的上下文语义信息,融合概率矩阵分解(probabilistic matrix factorization, PMF)模型进行评分预测. Shen 等人^[51]提出了一种更复杂的融合深度特征提取与矩阵分解的框架,进行在线学习资源推荐.作者首先将历史评分矩阵用隐含语义模型(LFM)分解为用户偏好和学习资源特征,将学习资源特征作为卷积神经网络输出层,同时利用 LDA 主题模型提取学习资源的文本主题,作为卷积神经网络输入层,构建 4 层的卷积神经网络模型,并通过最小化卷积神经网络全连阶层输出与 LFM 分解得到的学习资源特征的均方误差来训练模型,实验结果表明该模型适用于推荐新物品和冷门物品.

3.1.2 利用辅助信息分别提取用户和物品的特征表示

仅利用辅助信息提取物品(或用户)的特征表示主要解决了物品(或用户)冷启动问题,同时提取物品和用户的特征进行推荐,可提高推荐性能.其模型框架如图 2 所示.

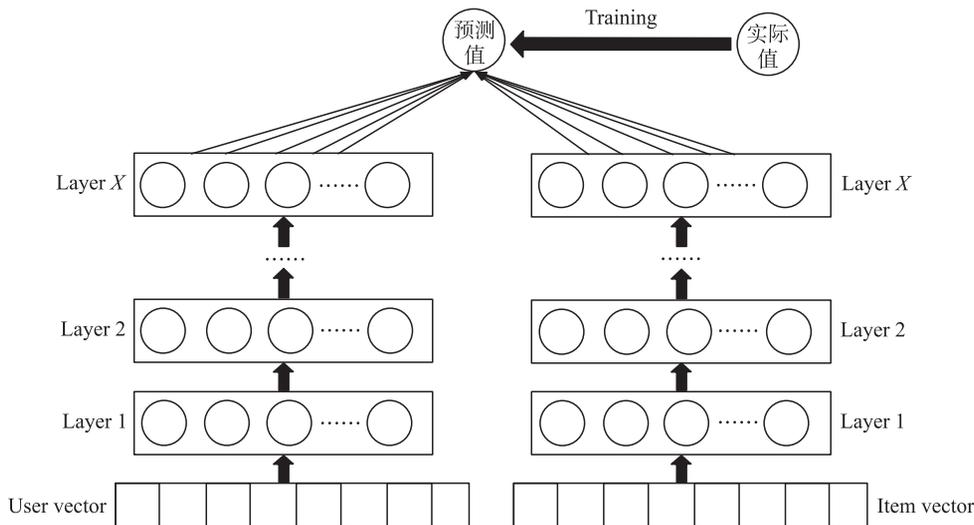


图 2 利用辅助信息分别提取用户和物品的特征表示

Fig. 2 Extract feature of users and items respectively by auxiliary information

Zheng^[8]使用两个卷积神经网络分别从用户评论中提取用户行为特征,从对物品的评论中提取物品属性特征,然后利用矩阵分解技术将提取的用户和物品隐含语义特征映射到公共特征空间进行评分预测.类似的,Li^[52]和 Shin 等人^[53]用深度学习分别提取用户和物品的特征,结合矩阵分解推荐,提高了推荐性能.

Lei 等人^[54]设计了 3 个深度神经网络,将图像和用户偏好映射到同一个隐含语义空间,提取图像与用户偏好的特征表示进行图像推荐.整个网络的输入是一个三元组 $(U, I+, I-)$, U 是用户信息,可将图像 $I+$ 看作是正例样本,图像 $I-$ 看作是负例样本,用户与图像 $I+$ 的距离大于用户与图像 $I-$ 的距离.两个卷积神经网络分别用于提取两个图像 $I+$ 和 $I-$ 的特征,一个全连接神经网络用于提取用户 U 的特征,分别计算提取的两个图像特征向量与用户特征向量的距离,并分别映射到两个全连接层,最后最小化交叉熵损失函数训

练模型.当推荐时仅需要两个深度神经网络.

Elkahky 等人^[3]在深度语义结构模型(deep semantic structured models, DSSM)^[55]基础上,提出一种基于内容的多视图深度学习推荐框架,既解决了新用户的冷启动问题,也显著提高了推荐性能,在多个数据集上取得了良好的结果.模型构建了包含4个深度神经网络的多视图深度学习模型,1个深度神经网络用于从用户的搜索和点击记录中提取用户兴趣特征,3个深度神经网络分别从3个领域(新闻、APP、电影)中提取各自领域的物品特征,通过最大化不同领域特征与用户特征的相似性之和求解.

3.1.3 利用辅助信息提取用户和物品的共同特征表示

此类模型用深度神经网络提取物品和用户共同的特征表示.其模型框架如图3所示.

不同于前面所述的使用已有的矩阵分解模型将用户与深度学习得到的物品特征进行简单的线性融合,He 等人^[56]提出一种更通用的神经协同过滤(neural collaborative filtering, NCF)框架,使用深度神经网络来学习用户和物品的非线性关联函数,在多个数据集中取得了明显提高.论文认为矩阵分解模型可作为NCF框架的特例,网络输入层是用户和物品的特征向量,输入层之上是全连接的嵌入层,嵌入层将原始的用户和物品的稀疏特征向量映射为稠密向量并作为NCF的输入,NCF利用多层感知机网络学习用户和物品之间的非线性关联函数进行预测,通过最小化预测值和实际值的平方差训练模型参数.论文比较了不同数量隐藏层情况下的实验结果,结果显示随着网络层数增多,预测效果提高.

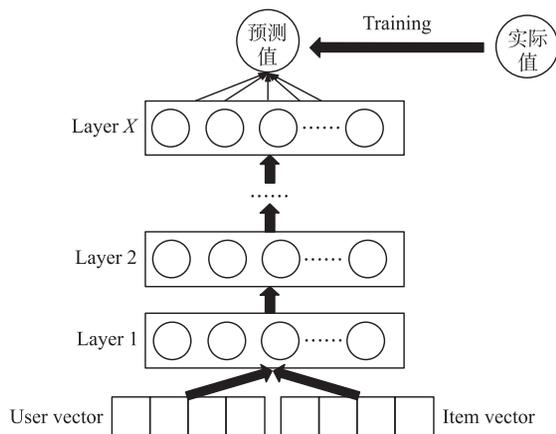


图3 利用辅助信息提取用户和物品的共同特征表示
Fig. 3 Extract common feature of users and items by auxiliary information

类似回归或分类问题,Cheng 等人^[5]提出组合线性模型和深度学习的特征提取方法用于APP推荐排序.模型使用的特征包含用户特征(如国家、语言、人口学特征等)、上下文特征(如所用设备、时间等)、印象特征(如APP年龄、APP历史统计数据等),线性模型输入为原始特征和转换特征,前馈深度神经网络用于学习特征组合的深度表示,输出层预测值是两个模型的加权和,通过逻辑回归损失函数联合训练网络,同时学习两个模型的参数.

在社交网络推荐中,Geng 等人^[57]提出一个基于社交网络的深度学习框架,以相互连接的“用户-图片”模型作为输入,并将大型稀疏的网络拓扑映射为一个分层树状结构,叶子节点表示用户和图像,非叶子节点表示对社交关系(如“用户-图片”、“用户-用户”、“图片-图片”)的编码,网络输出层为用户和图像的隐含特征表示.通过ImageNet pretrained卷积神经网络提取图像特征,用户特征和网络拓扑参数通过学习得到.该模型既学习了图像的视觉内容特征,同时也学习了图像和用户的拓扑关系,效果明显好于单纯使用图像视觉内容的推荐模型.

3.2 基于模型(model-based)的深度学习推荐算法

由于深度学习在特征学习方面的优势,基于辅助信息的深度学习推荐算法在推荐领域得到了较多的研究和应用,也有部分论文提出了基于模型的深度学习推荐算法,通过直接对评分矩阵进行学习提高推荐系统性能.

基于RBM的协同过滤算法,是深度学习最早在推荐系统的尝试.Salakhutdinov 等人^[58]使用两层RBM无向图模型自动抽取隐含的抽象特征,训练时将物品评分数据作为输入层,通过条件概率函数计算隐藏层的隐含向量表示,预测时利用隐藏层的向量表示反向求得评分.该方法在Netflix数据集上取得了较好的效果,被认为是最早基于深度学习的推荐模型.但严格来说,RBM并不属于深度神经网络,只是一个两层结构.受该RBM-CF模型启发,Zheng 等人^[59]采用RBM的替代模型(neural autoregressive distribution estimator, NADE)^[60]构造了神经自回归协同过滤模型,解决了RBM训练速度慢的问题,并将其扩展为具有多个隐层的深度模型.Georgiev 等人^[61]也对文献[58]提出的RBM模型进行了扩展,从用户维度和物品维度构建两个RBM模型,分别提取用户之间的关联特征和物品之间的关联特征,并将两个RBM模型组合为一

个混合模型求解。

基于自动编码器的推荐是另一类基于模型的深度学习推荐算法。Sedhain 等人^[62]提出使用基于物品(或用户)的自动编码器,将物品(或用户)的部分观察向量(评分)映射到低维隐含空间,然后在输出空间进行重建来评分预测。与矩阵分解模型相比,基于自动编码器的推荐算法仅使用了用户(或物品)的隐含特征,并可学习到非线性特征表示,与 RBM-CF 相比,需要学习的参数更少,结果优于概率分解、RBM-CF、近邻模型。与文献[62]类似,Strub 等人^[63]通过基于物品(或用户)的降噪自动编码器学习用户(物品)的非线性表示进行评分预测,同时将辅助信息融入自动编码器训练过程解决冷启动问题。Wu 等人^[64]提出了一个 top-N 推荐框架,使用三层(包含一个隐层)的降噪自动编码器从用户对物品的反馈信息中(输入层,不包含辅助信息)提取用户和物品的隐含分布式特征表示(隐藏层),然后在输出层将隐含特征表示映射到输入空间重建输入向量。

除上述两种类型外,Deng 等人^[65]提出了另一种新颖的基于矩阵分解的社交网络信任推荐模型。传统矩阵分解模型随机初始化参数,考虑到矩阵分解是非凸优化,该模型利用深度自动编码器从评分矩阵中学习矩阵分解模型的初始参数,可以达到更好的局部极小值,从而提高预测精度和学习效率。

3.3 动态深度学习推荐算法

用户兴趣和物品内容是随时间发生变化的,如何对用户和物品的动态特征建模是推荐算法研究的一个新方向。深度循环神经网络和卷积神经网络在捕捉时间序列特征和环境上下文特征中有很大优势。

一类模型是基于深度循环神经网络的动态推荐算法。基于会话的推荐系统中,用户的点击数据可看作是序列数据,用户的序列点击行为并不是孤立的,当前的点击行为往往是受到之前的结果影响,同理,当前的反馈也能够影响到今后的决策。Hidasi 等人^[66]提出了基于循环神经网络 GRU-RNN^[44]的方法训练历史会话点击(click-session)序列数据的深度模型,用户进入站点后点击的第一个物品作为 GRU-RNN 的初始输入,每一次点击后会在前面所有点击的基础上产生一个输出,利用排序学习损失函数训练整个网络。Tan 等人^[67]通过数据扩充(data augmentation)、时序适应等方法对文献[66]提出的深度循环神经网络推荐模型进行了改进。类似的,Ko 等人^[68]使用基于 GRU-RNN 的动态推荐模型,根据用户的上下文状态挖掘隐含的用户个性化特征向量。更一般地,Devooght 等人^[69]考虑用户消费物品的时间先后顺序,将长短记忆网络(long short-term memory, LSTM)应用于协同过滤推荐模型进行短期预测推荐,提高了推荐的多样性和覆盖率。基于自然语言处理单词预测的思想,该模型将每个物品看作一个单词,物品目录看做词汇表,每个用户消费物品的历史作为一个序列,通过最小化分类交叉熵(categorical cross-entropy)损失函数求解模型参数。

另一类动态模型通过组合循环神经网络 CNN、卷积神经网络 RNN、深度神经网络 DNN 等多种深度学习算法进行推荐。Wu 等人^[70]提出深度循环神经网络模型 DRNN,从用户一次会话(从用户进入网站到购买商品)的网页访问序列中学习用户浏览模式,同时利用用户的历史购买记录融合前馈神经网络(feedforward neural network, FNN)进行实时商品推荐,在考拉购物网站取得了明显的效果。Lee 等人^[71]通过组合循环神经网络和卷积神经网络从用户对话上下文中学习语义表示,推荐“名言”“谚语”的引用或常用回复。卷积神经网络用于提取会话上下文中每个句子的局部语义表示,循环神经网络 LSTM 用于对会话中的语句序列建模,通过最小化实际分布和预测分布的交叉熵训练模型,在 Twitter 会话中取得了较好的效果。Song 等人^[72]对深度语义结构模型(deep semantic structured models, DSSM)^[73]进行扩展,将长期的静态用户偏好和近期的用户兴趣时序变化相结合,提高推荐算法性能。物品特征和长期静态用户特征使用 DNN 提取,短期特征使用 RNN 提取。同时为了解决不同时间段粒度对用户短期兴趣的影响,模型使用两种时间段粒度,Fast-rate RNNs 提取近期用户兴趣,Slow-rate RNNs 提取季度用户兴趣变化,在新闻推荐中取得了显著效果。Dai 等人^[74-75]认为随着用户与物品之间的交互,用户的兴趣和物品的特征是相互影响并共同进化的,但传统的推荐算法仅考虑了用户与物品静态交互信息,基于此假设,作者提出了融合卷积神经网络与多元点过程(multivariate point process)的动态推荐算法,学习用户-商品隐含特征的共同进化模式进行推荐。

3.4 基于标签(tag-based)的深度学习推荐算法

用户用标签来描述对物品的看法,因此标签是联系用户和物品的纽带,也是反应用户兴趣的重要数据

源. 标签一方面描述了用户的兴趣,另一方面则表示了物品的语义,从而将用户和物品联系起来. 如何利用用户的标签数据提高个性化推荐结果的质量是推荐系统研究的重要课题^[9]. 与传统的二元评分矩阵不同,基于标签的推荐要学习三元组(用户、物品、标签)的关系,基于标签的深度学习算法主要利用深度学习从“物品-标签”、“用户-标签”和“用户-物品”矩阵中学习用户和物品的深度特征表示进行物品推荐.

Zuo 等人^[76]提出稀疏自动编码器深度学习模型进行基于标签的物品推荐. 首先将“用户-标签-物品”三元组映射为两个二维矩阵:“用户-标签”矩阵和“用户-物品”矩阵,将“用户-标签”矩阵作为深度神经网络输入层,输出层为学习到的用户深度特征表示,根据用户深度特征表示计算用户相似性,然后根据用户的相似性结合“用户-物品”矩阵推荐物品.

Xu 等人^[77]提出深度语义相似模型进行基于标签的物品推荐. 将与标签相关的用户信息(user-tags 向量)和物品信息(item-tags 向量)分别输入到两个深度神经网络,通过多个隐层将用户(物品)向量映射到抽象深度特征空间,计算用户和物品抽象特征的相似性,应用 softmax 函数对物品的相关性排序后进行物品推荐.

除物品推荐外,也有文献根据物品的内容特征为物品推荐标签. Wang 等人^[78]提出关系堆叠降噪自动编码器(relational stacked denoising autoencoders, RSDAE)的标签推荐模型,融合深度表示学习和关系学习,从物品内容信息和物品关系矩阵中提取物品的深度内容特征及关系表示,再结合矩阵(物品-标签矩阵)分解进行标签推荐,其中物品内容信息可以是论文摘要(论文推荐)、电影文本描述(电影推荐),物品关系矩阵通过 Matrix Variate Normal Distribution^[79]提取. Hamel 等人^[80]利用深度神经网络提取音频文件的信号特征,为音乐推荐标签. Rawat 等人^[81]使用深度神经网络将图像的上下文信息(时间、地理位置信息)和图像的视觉内容特征结合进行图像多标签分类推荐. 该模型使用卷积神经网络提取图像内容特征,使用两层全连接上下文网络 ContextNet 提取与图像关联的上下文信息,通过优化 Sigmoid 交叉熵损失函数训练网络.

3.5 基于深度学习的推荐算法性能讨论

不同算法是针对不同的推荐任务和数据集特点来设计的,难以用某一个或某几个数据集来衡量不同类型推荐算法的性能. 总体来看,在基于模型的深度学习推荐算法中,自动编码器算法具有更好的性能,优于 RBM-CF 和其他传统的推荐算法. 利用辅助信息的深度学习推荐算法,因为利用了用户和物品的额外信息,要优于单纯的基于模型的算法,其中利用辅助信息提取用户和物品的共同特征表示的算法效果最好. 动态深度学习推荐算法在利用静态辅助信息的同时,也考虑了动态时间信息,在很多推荐任务上要优于仅利用静态辅助信息的推荐算法. 基于标签的深度学习推荐算法在包含标签的推荐任务中具有更好的效果.

4 深度学习推荐算法研究趋势展望

尽管基于深度学习的推荐算法已取得了一定的研究成果,但深度学习在推荐系统的应用研究还处于初步阶段,仍有许多值得深入探索和研究的问题,本节将对相关研究方向和发展趋势进行讨论.

(1) 利用辅助信息

利用辅助信息的推荐模型仍是深度学习在推荐系统最主要的应用方向. 从特定领域的非结构化数据如文本、视频、音频、图像中提取高质量的特征表示,对用户偏好和物品特征进行建模,并融合或改进原有的推荐算法框架,是推荐系统应用深度学习最简单有效的方法,将会有越来越多的算法被提出.

(2) 用户动态行为建模

传统的推荐算法大都只对用户的静态特征进行建模,这在需求变化缓慢和偏好相关的场景下是有效的. 但有些领域需要更多地考虑用户和物品随时间的动态变化,综合考虑用户长期和短期兴趣偏好进行建模是非常重要的. 深度学习的序列模型(如循环神经网络)对解决动态建模问题有很大潜力和优势.

(3) 上下文感知的模型

随着物联网和智能终端的发展普及,越来越多的环境上下文数据可以被传感器获取,进而被用于推荐系统,上下文感知的推荐模型将会成为重要的发展方向. 大量的上下文感知数据需要强大的模式识别算法,传统的方法往往依赖专家去人为设计特征,效果未必好,深度神经网络将是解决该问题的重要手段.

(4) 多任务学习

数据稀疏性是推荐系统面临的重大问题之一,多任务学习是利用多个相关任务之间蕴含的有价值信息来提高学习能力,是缓解数据稀疏问题的重要方法.利用多任务学习可在多个推荐任务之中提取用户特征,提高推荐性能.深度学习在挖掘多个领域特征的复杂依赖关系中与传统算法相比有巨大优势,是未来应用的重要方向.

(5) 算法优化及可扩展性

尽管深度学习在推荐系统中有巨大潜力,并取得了一定的成果,但现有模型仍有很多改进的空间.基于模型的算法仅取得了与传统算法相当的效果,并无明显提高,仍需要进一步优化.在开发新模型时,模型训练速度、响应时间以及可扩展性等方面需要认真考虑.

5 结语

在信息爆炸的时代,推荐系统作为缓解信息过载问题的重要技术,得到了广泛的研究和应用,同时也面临很多挑战,在很多领域的推荐性能还很不理想.深度学习在许多领域已取得了巨大成功,其自动学习复杂高维数据特征的特性和优秀的泛化能力,有助于对各种类型的数据建模,更好地理解用户需求,进一步提高推荐质量.深度学习具有成为下一代推荐系统的核心技术的潜力,但目前深度学习在推荐系统的研究中还处于初步阶段.

[参考文献] (References)

- [1] 黄震华,张佳雯,田春岐,等.基于排序学习的推荐算法研究综述[J].软件学报,2016,27(3):691-713.
HUANG Z H,ZHANG J W,TIAN C Q, et al. Survey on learning-to-rank based recommendation algorithms[J]. Journal of software,2016,27(3):691-713.(in Chinese)
- [2] DIELEMAN S. Recommending music on spotify with deep learning[EB/OL]. [2014-08-05]. <http://benanne.github.io/2014/08/05/spotify-cnns.html>.
- [3] ELKAHKY A M,SONG Y,HE X. A multi-view deep learning approach for cross domain user modeling in recommendation systems[C]//Proceedings of the 24th International Conference on World Wide Web. Florence,Italy,2015.
- [4] COVINGTON P,ADAMS J,SARGIN E. Deep neural networks for youtube recommendations[C]//Proceedings of the 10th ACM Conference on Recommender Systems. Boston,USA,2016.
- [5] CHENG H T,KOC L,HARMSSEN J, et al. Wide and deep learning for recommender systems[C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston,USA,2016.
- [6] RICCI F,ROKACH L,SHAPIRA B, et al. Recommender systems handbook[M]. Berlin:Springer,2011.
- [7] 刘青文. 基于协同过滤的推荐算法研究[D]. 合肥:中国科学技术大学,2013.
LIU Q W. Research on recommender systems based on collaborative filtering[D]. Hefei:University of Science and Technology of China,2013.(in Chinese)
- [8] ZHENG L. A survey and critique of deep learning on recommender systems[EB/OL]. [2018-03-06]. <https://bdsc.lab.uic.edu/docs/survey-critique-deep.pdf>.
- [9] 项亮. 推荐系统实践[M]. 北京:人民邮电出版社,2012.
XIANG L. Recommended system practice[M]. Beijing:Posts and Telecom Press,2012.(in Chinese)
- [10] MIYAHARA K,PAZZANI M J. Collaborative filtering with the simple Bayesian classifier[M]//MIZOGUCHI R,SLANEY J. PRICAI 2000 topics in artificial intelligence. Berlin:Springer,2000:679-689.
- [11] SU X,KHOSHGOFTAAR T M. Collaborative filtering for multi-class data using belief nets algorithms[C]//IEEE International Conference on TOOLS with Artificial Intelligence. Arlington,USA,2006.
- [12] CHEE S H S,HAN J,WANG K. RecTree:an efficient collaborative filtering method[C]//Proceedings of Data Warehousing and Knowledge Discovery. Munich,Germany,2001.
- [13] UNGAR L H,FOSTER D P. Clustering methods for collaborative filtering[C]//Proceedings of AAAI Workshop on Recommendation Systems. Madison,USA,1998.
- [14] CANNY J. Collaborative filtering with privacy via factor analysis[C]//International ACM SIGIR Conference on Research and

- Development in Information Retrieval. Tampere, Finland, 2002.
- [15] VUCETIC S, OBRADOVIC Z. Collaborative filtering using a regression-based approach [J]. Knowledge and information systems, 2005, 7(1): 1–22.
- [16] LEMIRE D, MACLACHLAN A. Slope one predictors for online rating-based collaborative filtering [C]//Proceedings of the 2005 SIAM International Conference on Data Mining. Newport Beach, USA, 2005.
- [17] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of machine learning research, 2003, 3: 993–1022.
- [18] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model [C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, USA, 2008.
- [19] JANNACH D, ZANKER M, FELFERNIG A, et al. Recommender systems: an introduction [M]. Cambridge: Cambridge University Press, 2010.
- [20] 杨文龙. 基于动态集成方法的混合推荐系统研究 [D]. 济南: 山东大学, 2015.
YANG W L. Research on hybrid recommender systems on the dynamically integrated Methodologies [D]. Jinan: Shandong University, 2015. (in Chinese)
- [21] KOREN Y. The bellkor solution to the netflix grand prize [J]. Netflix prize documentation, 2009, 81: 1–10.
- [22] SMYTH B, COTTER P. Personalized electronic program guides for digital TV [J]. AI magazine, 2001, 22(2): 89–98.
- [23] SARWAR B M, KONSTAN J A, BORCHERS A, et al. Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system [C]//Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work. Seattle, USA, 1998.
- [24] CONDLIFF M K, LEWIS D D, MADIGAN D, et al. Bayesian mixed-effects models for recommender systems [C]//ACM SIGIR'99 Workshop on Recommender Systems: Algorithms and Evaluation. Berkeley, USA, 1999.
- [25] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436–444.
- [26] 孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展 [J]. 计算机科学, 2016, 43(2): 1–8.
SUN Z Y, LU C X, SHI Z Z, et al. Research and advances on deep learning [J]. Computer science, 2016, 43(2): 1–8. (in Chinese)
- [27] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]//Advances in neural information processing systems. Lake Tahoe, USA, 2012.
- [28] DONAHUE J, JIA Y, VINYALS O, et al. DeCAF: a deep convolutional activation feature for generic visual recognition [C]//Proceedings of the 31st International Conference on Machine Learning. Beijing, China, 2014.
- [29] KARPATY A, TODERICI G, SHETTY S, et al. Large-scale video classification with convolutional neural networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014.
- [30] LI J, MONROE W, RITTER A, et al. Deep reinforcement learning for dialogue generation [DB/OL]. [2018-03-06]. <https://arxiv.org/abs/1606.01541>.
- [31] BORDES A, CHOPRA S, WESTON J. Question answering with subgraph embeddings [DB/OL]. [2018-03-06]. <https://arxiv.org/abs/1406.3676>.
- [32] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks [C]//Advances in Neural Information Processing Systems. Montreal, Canada, 2014.
- [33] MIKOLOV T, DEORAS A, POVEY D, et al. Strategies for training large scale neural network language models [C]//IEEE Workshop on Automatic Speech Recognition and Understanding. Waikoloa, USA, 2011.
- [34] HINTON G, DENG L, YU D, et al. Deep neural networks for acoustic modeling in speech recognition; the shared views of four research groups [J]. IEEE signal processing magazine, 2012, 29(6): 82–97.
- [35] MA J, SHERIDAN R P, LIAW A, et al. Deep neural nets as a method for quantitative structure-activity relationships [J]. Journal of chemical information and modeling, 2015, 55(2): 263–274.
- [36] HELMSTAEDTER M, BRIGGMAN K L, TURAGA S C, et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina [J]. Nature, 2013, 500(7461): 168–174.
- [37] LEUNG M K K, XIONG H Y, LEE L J, et al. Deep learning of the tissue-regulated splicing code [J]. Bioinformatics, 2014, 30(12): 121–129.
- [38] 郭丽丽, 丁世飞. 深度学习研究进展 [J]. 计算机科学, 2015, 42(5): 28–33.
GUO L L, DING S F. Research progress on deep learning [J]. Computer science, 2015, 42(5): 28–33. (in Chinese)
- [39] OORD A V D, DIELEMAN S, SCHRAUWEN B. Deep content-based music recommendation [C]//Conference on Neural

- Information Processing Systems(NIPS 2013). Lake Tahoe, USA, 2013.
- [40] MCAULEY J, TARGETT C, SHI Q, et al. Image-based recommendations on styles and substitutes [C]//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. Santiago, Chile, 2015.
- [41] ZHOU J, ALBATAL R, GURRIN C. Applying visual user interest profiles for recommendation and personalisation [C]//International Conference on Multimedia Modeling. Miami, USA, 2016.
- [42] BANSAL T, BELANGER D, MCCALLUM A. Ask the GRU: multi-task learning for deep text recommendations [C]//Proceedings of the 10th ACM Conference on Recommender Systems. Boston, USA, 2016.
- [43] ZANOTTI G, HORVATH M, BARBOSA L N, et al. Infusing collaborative recommenders with distributed representations [C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, USA, 2016.
- [44] CHO K, VAN MERRIENBOER B, BAHDANAU D, et al. On the properties of neural machine translation: encoder-decoder approaches [DB/OL]. [2018-03-06]. <https://arxiv.org/abs/1409.1259>.
- [45] 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. Sydney, Australia, 2015.
- [46] 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: 3371-3408.
- [47] WANG C, BLEI D M. Collaborative topic modeling for recommending scientific articles [C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, USA, 2011.
- [48] WEI J, HE J, CHEN K, et al. Collaborative filtering and deep learning based recommendation system for cold start items [J]. *Expert systems with applications*, 2017, 69: 29-39.
- [49] KOREN Y. Collaborative filtering with temporal dynamics [J]. *Communications of the ACM*, 2010, 53(4): 89-97.
- [50] KIM D, PARK C, OH J, et al. Convolutional matrix factorization for document context-aware recommendation [C]//Proceedings of the 10th ACM Conference on Recommender Systems. Boston, USA, 2016.
- [51] SHEN X, YI B, ZHANG Z, et al. Automatic recommendation technology for learning resources with convolutional neural network [C]//2016 International Symposium on Educational Technology (ISET). Beijing, China, 2016.
- [52] LI S, KAWALE J, FU Y. Deep collaborative filtering via marginalized denoising auto-encoder [C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne, Australia, 2015.
- [53] SHIN D, CETINTAS S, LEE K C, et al. Tumblr blog recommendation with boosted inductive matrix completion [C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne, Australia, 2015.
- [54] LEI C, LIU D, LI W, et al. Comparative deep learning of hybrid representations for image recommendations [C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016.
- [55] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data [C]//Proceedings of the 22nd ACM International Conference on Information and Knowledge Management. San Francisco, USA, 2013.
- [56] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C]//Proceedings of the 26th International Conference on World Wide Web. Perth, Australia, 2017.
- [57] GENG X, ZHANG H, BIAN J, et al. Learning image and user features for recommendation in social networks [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015.
- [58] SALAKHUTDINOV R, MNIH A, HINTON G. Restricted Boltzmann machines for collaborative filtering [C]//Proceedings of the 24th International Conference on Machine Learning. Corralis, USA, 2007.
- [59] ZHENG Y, TANG B, DING W, et al. A neural autoregressive approach to collaborative filtering [C]//International Conference on Machine Learning. New York, USA, 2016.
- [60] LAROCHELLE H, MURRAY I. The neural autoregressive distribution estimator [C]//International Conference on Artificial Intelligence and Statistics. Lauderdale, USA, 2011.
- [61] GEORGIEV K, NAKOV P. A non-IID framework for collaborative filtering with restricted Boltzmann machines [C]//International Conference on Machine Learning. Atlanta, USA, 2013.
- [62] SEDHAIN S, MENON A K, SANNER S, et al. Autorec: autoencoders meet collaborative filtering [C]//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015.
- [63] STRUB F, GAUDEL R, MARY J. Hybrid recommender system based on autoencoders [C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, USA, 2016.

- [64] WU Y, DUBOIS C, ZHENG A X, et al. Collaborative denoising auto-encoders for top-n recommender systems[C]//Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. San Francisco, USA, 2016.
- [65] DENG S, HUANG L, XU G, et al. On deep learning for trust-aware recommendations in social networks[J]. IEEE Transactions on neural networks and learning systems, 2017, 28(5):1164–1177.
- [66] HIDASI B, KARATZOGLU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks[C]//International Conference on Learning Representations. San Juan, Puerto Rico, 2016.
- [67] TAN Y K, XU X, LIU Y. Improved recurrent neural networks for session-based recommendations[C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, USA, 2016.
- [68] KO Y J, MAYSTRE L, GROSSGLAUSER M. Collaborative recurrent neural networks for dynamic recommender systems[C]//Proceedings of the 8th Asian conference on Machine Learning. Hamiton, New Zealand, 2016.
- [69] DEVOOGHT R, BERSINI H. Collaborative filtering with recurrent neural networks[DB/OL]. [2018-03-06]. <https://arxiv.org/abs/1608.07400>.
- [70] WU S, REN W, YU C, et al. Personal recommendation using deep recurrent neural networks in NetEase[C]//IEEE International Conference on Data Engineering. Paris, France, 2016.
- [71] LEE H, AHN Y, LEE H, et al. Quote recommendation in dialogue using deep neural network[C]//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa, Italy, 2016.
- [72] SONG Y, ELKAHKY A M, HE X. Multi-rate deep learning for temporal recommendation[C]//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa, Italy, 2016.
- [73] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data[C]//ACM International Conference on Information and Knowledge Management. San Francisco, USA, 2013.
- [74] DAI H, WANG Y, TRIVEDI R, et al. Recurrent coevolutionary feature embedding processes for recommendation[DB/OL]. [2018-03-06]. <https://arxiv.org/abs/1609.03675>.
- [75] DAI H, WANG Y, TRIVEDI R, et al. Recurrent coevolutionary latent feature processes for continuous-time recommendation[C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, USA, 2016.
- [76] ZUO Y, ZENG J, GONG M, et al. Tag-aware recommender systems based on deep neural networks[J]. Neurocomputing, 2016, 204:51–60.
- [77] XU Z, CHEN C, LUKASIEWICZ T, et al. Tag-aware personalized recommendation using a deep-semantic similarity model with negative sampling[C]//Proceedings of the 25th ACM International Conference on Information and Knowledge Management. Singapore, 2016.
- [78] WANG H, SHI X, YEUNG D Y. Relational stacked denoising autoencoder for tag recommendation[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA, 2015.
- [79] GUPTA A K, NAGAR D K. Matrix variate normal distribution[M]. Boca Raton; CRC Press, 1999.
- [80] HAMEL P, LEMIEUX S, BENGIO Y, et al. Temporal pooling and multiscale learning for automatic annotation and ranking of music audio[C]//International Society for Music Information Retrieval Conference. Miami, USA, 2011.
- [81] RAWAT Y S, KANKANHALLI M S. ConTagNet: exploiting user context for image tag recommendation[C]//Proceedings of the 2016 ACM on Multimedia Conference. Amsterdam, Netherlands, 2016.

[责任编辑:严海琳]