

基于深度学习的推荐系统研究综述

黄立威¹⁾, 江碧涛¹⁾, 吕守业¹⁾, 刘艳博¹⁾, 李德毅²⁾

¹⁾(北京市遥感信息研究所 北京 100192)

²⁾(清华大学计算机科学与技术系 北京 100084)

摘 要 深度学习是机器学习领域一个重要研究方向,近年来在图像处理、自然语言理解、语音识别和在线广告等领域取得了突破性进展。将深度学习融入推荐系统中,研究如何整合海量的多源异构数据,构建更加贴合用户偏好需求的用户模型,以提高推荐系统的性能和用户满意度,成为基于深度学习的推荐系统的主要任务。本文对近几年基于深度学习的推荐系统研究进展进行综述,分析其与传统推荐系统的区别以及优势,并对其主要的研究方向、应用进展等进行概括、比较和分析。最后,对基于深度学习的推荐系统的未来发展趋势进行分析和展望。

关键词 推荐系统;深度学习;协同过滤;个性化服务;数据挖掘;多源异构数据;综述

中图法分类号 TP18

Survey on Deep Learning Based Recommender Systems

HUANG Li-Wei¹⁾ JIANG Bi-Tao¹⁾ LV Shou-Ye¹⁾ LIU Yan-Bo¹⁾ LI De-Yi²⁾

¹⁾(Beijing Institute of Remote Sensing, Beijing 100192)

²⁾(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

Abstract With the ever-growing volume, complexity and dynamicity of online information, recommender systems have been an effective key solution to handle the increasing information overload problem by retrieving the most relevant information and services from a huge amount of data, and providing personalized recommendation. In recent years, deep learning technology has become an important research direction in the field of machine learning, which has been widely applied in the image processing, natural language understanding, speech recognition and online advertising. Meanwhile, recent studies also demonstrate its effectiveness in coping with information retrieval and recommendation tasks. Applying deep learning techniques into recommender systems has been gaining momentum due to its state-of-the-art performances and high-quality recommendations. In this paper, we investigate the deep learning based recommender systems, for which the main tasks are how to organize the massive multi-source heterogeneous data, build more suitable user models according to user preferences requirements, and improve the performance and user satisfaction. For specific, we first introduce the basic concepts and methods of traditional recommendation systems, including content-based recommendation method, collaborative filtering and hybrid recommendation method, and then we give an overview of the main deep learning techniques and briefly introduce their applications in the recommender systems. And secondly, we provide a comprehensive summary of current research on deep learning based recommender systems. According to the data sources used in recommender systems and the classification of the traditional recommender systems, we categorize the current research into five main directions: the application of

本课题得到国家重点基础研究发展计划(973)(Nos. 2014CB340401);国家自然科学基金重大研究计划(Nos. 91638301);国家自然科学基金(Nos. 61272111, 61273216, 6160011950)资助。黄立威(通信作者),男,1985年生,博士,中国计算机学会会员,主要研究方向为机器学习、推荐系统.E-mail: dr_huanglw@163.com. 江碧涛,女,1967年生,博士,研究员,主要研究方向为数据挖掘。吕守业,男,1979年生,博士,研究员,主要研究方向为数据挖掘.E-mail: lvshouye@bit.edu.cn. 刘艳博,女,1988年生,硕士,主要研究方向为图像处理、机器学习.E-mail: liuyanbonudt@163.com. 李德毅,男,1944年生,博士,研究员,中国工程院院士,主要研究方向为人工智能。

deep learning in content-based recommender systems, the application of deep learning in collaborative filtering, the application of deep learning in hybrid recommender systems, the application of deep learning in social network-based recommender systems, and the application of deep learning in context-aware recommender systems. Then, we analyze the differences and advantages of deep learning based recommender systems compared with the traditional recommender systems. First, by using deep learning, complex feature engineering can be avoided, especially when faced with unstructured data such as image and video. Second, deep learning can learn the multi-level and abstract feature representation of users and items, and is able to effectively capture the non-linear and non-trivial user-item interactions. Third, deep learning can incorporate various multi-source heterogeneous data into recommender systems, and help to mitigate the data sparseness problem considerably. Finally, this paper summarizes the future development trend of deep learning based recommender systems, e.g., the combination of deep learning and traditional recommendation methods, the application of deep learning in cross domain recommendation, the combination of the attention mechanism and deep learning based recommender systems, new deep learning recommendation architectures and the interpretability of deep learning based recommender systems. In short, deep learning has become popular in the recommender systems community both in academia and in industry. Meanwhile, this area of research is very young, there is much room for improvement in the aforementioned research directions, but we also believe that deep learning will revolutionize the recommender systems dramatically and bring more opportunities in reinventing the user experiences for better customer satisfaction in the near future.

Key words recommender systems; deep learning; collaborative filtering; personalized services; data mining; multi-source heterogeneous data; survey

1 引言

近年来,随着云计算、大数据、物联网等技术的迅猛发展,互联网空间中各类应用的层出不穷引发了数据规模的爆炸式增长^[1]。根据国际数据集团(IDC)2012年的一份报告显示:到2020年,预计全球数据总量是2011年的22倍,将达到35.2ZB^[2]。大数据中蕴含着丰富的价值与巨大的潜力,将给人类社会带来变革性的发展,但同时也带来了严重的“信息过载”问题,如何快速有效地从纷繁复杂的数据中获取有价值的信息成为了当前大数据发展的关键难题。推荐系统作为解决“信息过载”问题的有效方法^[3],已经成为学术界和工业界的关注热点并得到了广泛应用,形成了众多相关研究成果。推荐系统根据用户需求、兴趣等,通过推荐算法从海量数据中挖掘出用户感兴趣的项目(如信息、服务、物品等),并将结果以个性化列表的形式推荐给用户。目前,推荐系统在很多领域得到了成功应用,包括电子商务(如Amazon、eBay、Netflix、阿里巴巴等)、信息检索(如iGoogle、MyYahoo、百度等)、社交网络(Facebook、Twitter、腾讯等)、位置服务(如Foursquare、Yelp、大众点评等)、新闻推

送(如Google News、GroupLens、今日头条等)等各个领域。

传统的推荐方法主要包括协同过滤、基于内容的推荐方法和混合推荐方法。其中,最经典的算法是协同过滤,如矩阵因子分解,其利用用户与项目之间的交互信息为用户产生推荐,协同过滤是目前应用最为广泛的推荐算法,近年来在Netflix大奖赛中屡获大奖,但是同时也遭遇到了严重的数据稀疏(一个用户评分过的项目仅仅占总项目数量的极少部分)和冷启动(新的用户和新的项目往往没有评分数据)问题。此外,经典的协同过滤方法采用浅层模型无法学习到用户和项目的深层次特征。基于内容的推荐方法利用用户已选择的项目来寻找其他类似属性的项目进行推荐,但是这种方法需要有效的特征提取,传统的浅层模型依赖于人工设计特征,其有效性及可扩展性非常有限,制约了基于内容的推荐方法的性能。随着互联网中越来越多的数据能够被感知获取,包括图像、文本、标签在内的多源异构数据蕴含着丰富的用户行为信息及个性化需求信息,融合多源异构辅助信息(side information)的混合推荐方法由于能够缓解传统推荐系统中的数据稀疏和冷启动问题,而越来越受到重视,但是由于辅助信息往往具有多模态、数据异

构、大规模、数据稀疏和分布不均匀等复杂特征，融合多源异构数据的混合推荐方法研究依然面临着严峻的挑战^[4-5]。

近年来，深度学习在图像处理、自然语言理解和语音识别等领域取得了突破性进展^[6]，已经成为人工智能的一个热潮，为推荐系统的研究带来了新的机遇。一方面，深度学习可通过学习一种深层次非线性网络结构，表征用户和项目相关的海量数据，具有强大的从样本中学习数据集本质特征的能力，能够获取用户和项目的深层次特征表示。另一方面，深度学习通过从多源异构数据中进行自动特征学习，从而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^[7]，在此基础上融合传统推荐方法进行推荐，能够有效利用多源异构数据，缓解传统推荐系统中的数据稀疏和冷启动问题。近三年来，基于深度学习的推荐系统研究开始受到国际学术界和工业界越来越多的关注，ACM推荐系统年会（ACM RecSys）在2016年专门召开了第一届基于深度学习的推荐系统研究专题研讨会（DLRS'16），研讨会指出深度学习将是推荐系统的下一个重要方向，DLRS'17也已经在意大利的科莫举行。计算机领域的数据挖掘和机器学习顶级会议（SIGKDD, NIPS, SIGIR, WWW, AAAI等）中，基于深度学习的推荐系统研究的文章逐年增加，国内外许多大学和研究机构也对基于深度学习的推荐系统开展了广泛研究^[8-20]。基于深度学习的推荐系统研究目前已经成为推荐系统领域的研究热点之一。

本文主要对基于深度学习的推荐系统的研究与应用进展进行综述。第2节简要介绍传统的推荐算法。第3节介绍了深度学习的主要方法。第4节重点分析基于深度学习的推荐系统的研究进展。第5节展望基于深度学习的推荐系统的未来研究趋势。第6节是全文总结。

2 传统推荐系统

上世纪90年代，协同过滤技术^[21]的首次提出，标志着推荐系统成为了一门独立的学科而受到广泛关注。推荐系统的核心是推荐算法，它利用用户与项目之间的二元关系，基于用户历史行为记录或相似性关系帮助用户发现用户可能感兴趣的项目。文献[3]给出了推荐算法的形式化定义：用 U 表示所有用户（user）的集合，用 I 表示所有项目（item）的集

合。在实际系统中， U 和 I 具有非常大的规模。定义一个效用函数 s ，用来计算项目 i 对用户 u 的推荐度，即 $s:U \times I \rightarrow R$ ，其中 R 是一个全序集合（在一定范围内非负的整数或实数），推荐算法的研究问题就是通过计算推荐度为每一个用户 $u \in U$ 找到其最感兴趣的项目 $i' \in I$ ，如下：

$$\forall u \in U, i'_u = \arg \max_{i \in I} s(u, i) \quad (1)$$

推荐系统中的用户和项目都可以通过一组不同的属性或特征来进行表示。推荐系统面临的一个关键问题是效用函数 s 通常定义在 $U \times I$ 的一个子空间上，推荐算法必须将 s 外推到整个 $U \times I$ 空间。例如，我们通常将推荐度定义为用户对项目的评价，但真实的推荐系统中，用户仅仅评价了一小部分项目，因此在选择推荐度最高的项目推荐给用户之前，必须先根据已知的评分来实现对未知评分的预测，这就是外推的过程。推荐算法对未知评分的预测能够采用不同的方法，包括近似理论、机器学习和各种启发式方法等。但是传统的推荐方法主要可以分为以下三种^[22]：基于内容的推荐（content-based recommendation）^[23]、协同过滤推荐（collaborative filtering recommendation）^[24]和混合推荐（hybrid recommendation）^[25]。

（1）基于内容的推荐：主要根据用户已经选择或者评分的项目，挖掘其他内容上相似的项目作为推荐，属于Schafer划分^[26]中的项目到项目关联（Item-to-Item Correlation）的方法。首先通过显式反馈（例如评分、喜欢/不喜欢）或隐式反馈（例如观看、搜索、点击、购买等行为）的方式获取用户交互过的项目，然后从这些项目的特征中学习用户的偏好并表示为特征，就能计算用户与待预测项目在内容（由特征刻画）上的匹配度（或相似度），最后根据匹配度对所有待预测项目进行排序，从而为用户推荐潜在感兴趣的项目。基于内容的推荐方法依赖于关于用户偏好和项目的特征信息，不需要大量的评分记录，因此不存在评分数据稀疏的问题。同时，对于新项目，只需要进行特征提取就可以向用户进行推荐，解决了新项目的冷启动问题，但常常会遭遇到特征提取困难的问题。

（2）协同过滤推荐：源于现实生活中口碑相传（word-of-mouth）的过程，协同过滤利用相似用户之间具有相似兴趣偏好的方法，来发现用户对项目的潜在偏好。主要包括启发式和基于模型两种类型，启发式方法首先通过用户的历史评分差异计算

用户（或者项目）之间的相似度，然后根据用户的历史评分和用户之间的相似度计算效用值，基于模型的方法主要通过构建一个用户偏好模型预测用户对项目的潜在偏好。协同过滤仅仅需要利用用户的历史评分数据，因此简单有效，是目前应用最为成功的推荐方法。但是，由于用户对项目的评分数据相对项目的总数量非常少，常常遭遇数据稀疏的问题，此外，对于新的用户或项目，由于没有评分数据而无法进行推荐，存在冷启动问题。

（3）混合推荐：考虑到单一推荐方法都存在各自的不足，通过组合不同的推荐算法进行混合推荐，往往能够产生更好的推荐性能。常见的组合策略主要包括三种，分别为：后融合、中融合和前融合。后融合是指将两种或两种以上的推荐算法产生的推荐结果，以投票机制、线性组合或者可信度选择组合等方式来产生最终的推荐结果，本质上是决策层面上的混合。中融合的基本框架是以一种推荐算法为基础，同时融合另一种推荐算法，例如：以协同过滤算法为框架，融入基于内容的推荐算法可以有效缓解数据稀疏问题，本质上是模型层面上的混合。前融合则是直接将多种推荐算法融合到统一的模型中，然后从各类数据中提取的特征作为模型的输入，由统一的模型产生推荐结果，例如，将所有用户属性、用户行为等数据作为输入，通过训练一个统一的分类器产生推荐结果，本质上是特征层面的融合。

需要注意的是，随着社交媒体的迅速发展，大量的用户生成内容（如社会化关系、标签、评论、位置信息）能够被获取，推荐系统的内容更加多样，包括商品、朋友、标签、音乐、视频、新闻等；同时，除了传统的推荐方法，也催生了大量新的推荐方法，例如，基于社交网络的推荐方法、情境感知的推荐方法等；推荐的对象也更加多样，不仅仅针对单个用户，还可以针对一群用户，产生了组推荐；随着移动互联网的发展，位置社交网络的推荐系统也受到越来越多的关注。因此可以发现，推荐系统在推荐内容、推荐方法、推荐对象等各个方面都朝着越来越多元的方向发展。

3 深度学习技术

深度学习已经成为互联网大数据和人工智能的一个热潮^[27]。深度学习通过组合低层特征形成更加稠密的高层语义抽象，从而自动发现数据的分布

式特征表示，解决了传统机器学习中需要人工设计特征的问题，在图像识别、机器翻译、语音识别和在线广告等领域取得了突破性进展。图像识别领域，在 2016 年的 ImageNet 图像分类竞赛中，深度学习的准确率超过了 97%。在机器翻译领域，基于深度学习的 Google 神经机器翻译系统（GNMT）在英语与西班牙语和英语与法语的翻译中都取得了接近于人类的翻译水平^[28]。在语音识别领域，2016 年年底百度、科大讯飞和搜狗都宣布，他们基于深度学习的中文语音识别准确率都超过了 97%。在线广告领域，深度学习广泛应用于广告点击率预测，在 Google^[9,29]、微软^[30,31]、华为^[32]、阿里巴巴^[33]等企业的应用中取得了很大成功。深度学习涉及相当广泛的机器学习技术和结构，下面我们对常用的深度学习模型和方法进行介绍

3.1 自编码器

1986 年 Rumelhart 等人^[34]提出了自编码器（Autoencoder, AE）的概念，并将其用于高维复杂数据处理。自编码器通过一个编码和一个解码过程来重构输入数据，学习数据的隐层表示。基本的自编码器可视为一个三层神经网络结构：一个输入层 \mathbf{x} 、一个隐层 \mathbf{h} 和一个输出层 \mathbf{y} ，其中输出层和输入层具有相同的规模，结构如图 1 所示。

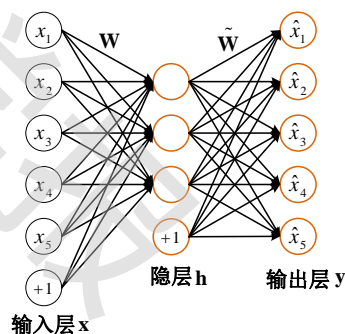


图 1 自编码器结构示意图

自编码器的目的是使得输入 \mathbf{x} 与输出 \mathbf{y} 尽可能接近，这种接近程度通过重构误差表示，根据数据的不同形式，通常重构误差有均方误差和交叉熵两种定义方式。

如果仅仅通过最小化输入输出之间的误差来实现对模型的训练，自编码器很容易学习到一个恒等函数。为了解决这个问题，研究者提出了一系列自编码器的变种，其中比较经典的包括稀疏自编码器^[35]和降噪自编码器^[35]。通过在损失函数中加入 L1 正则项，便可以训练得到稀疏自编码器，其目的是对过大的权重进行惩罚，使隐层表示中的大量节

点为 0，从而确保隐层表示尽量稀疏。降噪自编码器则是通过在自动编码器的输入数据中加入噪声得到，这样降噪自编码器在重构输入数据时，就被迫去除这种噪声来学习到更加鲁棒的输入数据的表达，降噪自编码器通过这种方式提升了泛化能力。2007 年，Bengio 等人^[36]通过堆叠多个降噪自编码器，提出了栈式降噪自动编码器（Stacked Denoising Autoencoder, SDAE）的概念，其是一种深度神经网络结构，通过逐层非监督学习的预训练可以学习多层次的数据抽象表示。

自编码器，尤其是栈式降噪自编码器，在推荐系统中主要被应用于学习用户和项目的隐层特征表示^[4,5,17,20,37-42]，其通过对用户或项目相关的信息（包括评分数据和文本、图像等信息）进行重构学习到用户或项目的隐表示，然后基于这种隐表示预测用户对项目的偏好。应用场景主要包括评分预测、文本推荐、图像推荐等。

3.2 受限玻尔兹曼机

玻尔兹曼机（Boltzmann machine, BM）^[43]是一种生成式随机神经网络，由Hinton和Sejnowski在1986年提出，BM由一些可见单元（对应可见变量，亦即数据样本）和一些隐层单元（对应隐层变量）构成，可见变量和隐层变量都是二元变量，其状态取0-1，状态0表示该神经元处于抑制状态，状态1代表该神经元处于激活状态。BM能够学习数据中复杂的规则，具有强大的无监督学习能力。但是，玻尔兹曼机的训练过程非常耗时。为此，Sejnowski 等人^[44]进一步提出了一种受限玻尔兹曼机（Restricted Boltzmann Machine, RBM），其在玻尔兹曼机的基础上，通过去除同层变量之间的所有连接极大地提高了学习效率。受限玻尔兹曼机的结构如图2所示，包括可见层 v 以及隐层 h ，两层之间的节点是全连接的，同层节点间是互不连接的。

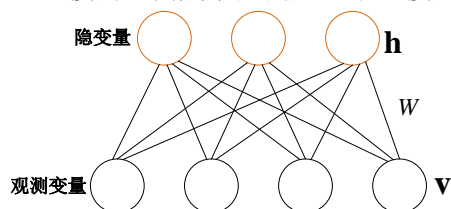


图2 受限玻尔兹曼机结构示意图

从RBM的结构可以发现，在给定隐层单元的状态时，可见层单元之间是条件独立的；反之，在给定可见层单元的状态时，各隐层单元之间也条件独立。因此，尽管无法有效计算RBM所表示的分布，但是通过Gibbs采样能够得到RBM所表示的分布的

随机样本。Gibbs采样的问题是需要使用较大的采样步数，使得RBM的训练效率仍不高。考虑到这种情况，Hinton^[45]提出了一种对比散度（contrastive divergence, CD）快速学习算法，CD算法同样利用Gibbs采样过程（即每次迭代包括从可见层更新隐层，以及从隐层更新可见层）来获得随机样本，但是只需迭代 k （通常 $k=1$ ）次就可获得对模型的估计，而不需要像Gibbs采样一样直到可见层和隐层达到平稳分布。

RBM是推荐系统中最早被应用的神经网络模型^[46-50]，当前的应用主要是通过对用户的评分数据进行重构学习到用户的隐表示，从而实现未知评分的预测。应用场景主要是用户评分预测。

3.3 深度信念网络

Hinton 等人在 2006 年提出了一种深度信念网络（Deep Belief Network, DBN）^[51]，其是一种由多层非线性变量连接组成的生成式模型。在深度信念网络中，靠近可见层的部分是多个贝叶斯信念网络，最远离可见层的部分则是一个 RBM，其结构如图3所示。DBN 的结构可以看作由多个受限玻尔兹曼机层叠构成，网络中前一个 RBM 的隐层视为下一个 RBM 的可见层。这样，在 DBN 的训练过程中，每一个 RBM 都可以使用上一个 RBM 的输出单独训练，因此与传统的神经网络相比，DBN 的训练更加简单。同时，通过这种训练方法，DBN 也能够从无标记数据获取深层次的特征表示。

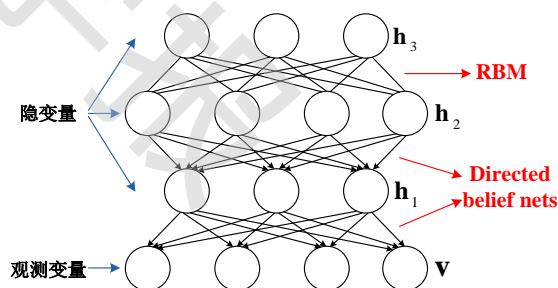


图3 深度信念网络结构示意图

DBN 网络的训练可采用一种贪婪逐层算法^[52]。首先，由最底层 RBM 开始，通过对比散度算法从原始观测数据中学习第一层隐层单元的状态，然后将参数保存，将隐层单元的状态作为下一层 RBM 的输入，按照这种方式继续训练，直到整个深层结构训练完成。

深度信念网络当前在推荐系统中应用较少。由于DBN在建模一维数据上比较有效，因此被应用于提取音乐的特征表示，从而进行音乐推荐^[53]。当前的应用场景仅限于音乐推荐。

3.4 卷积神经网络

卷积神经网络 (Convolutional Neural Network, CNN) 已成为当前图像理解领域的研究热点^[54-56]。卷积神经网络是一种多层感知机, 主要被用来处理二维图像数据。相比传统的多层感知机, CNN 使用池化操作减少了模型中的神经元数量, 同时对输入空间的平移不变性具有更高的鲁棒性。另外, CNN 的权值共享网络结构能够减少模型中参数数量, 降低了网络模型的复杂度, 提升了模型的泛化能力。尤其是当网络的输入是多维图像时, 通过将图像直接作为网络的输入, 从而避免了传统图像处理算法中复杂的特征提取和数据重建过程。卷积神经网络的基本结构由输入层、卷积层、下采样层(池化层)、全连接层和输出层构成, 如图 4 所示。

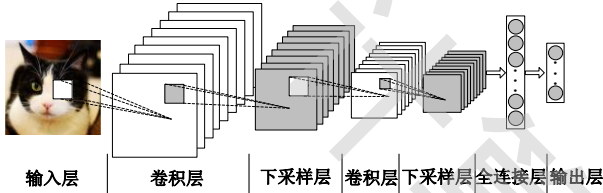


图 4 卷积神经网络结构示意图

卷积神经网络在推荐系统中应用较为广泛^[5,57-62], 主要被用于从图像、文本、音频等内容中提取项目的隐藏特征, 从而获取项目的低维向量表示, 并结合用户隐表示为用户产生推荐。当前的应用场景主要包括图像推荐、音乐推荐、文本推荐等。

3.5 循环神经网络

1986 年 Rumelhart 等人^[34]提出循环神经网络 (Recurrent Neural Network, RNN) 的概念。普通的全连接网络或卷积神经网络, 是从输入层到隐层再到输出层的结构, 层与层之间是全连接的, 每层之间的节点是无连接的。这种神经网络结构在面对序列数据建模时往往显得无能无力。例如, 当需要预测句子中下一个单词是什么, 一般需要依据前面的单词。RNN 因为能够建模序列数据中不同时刻数据之间的依赖关系, 在机器翻译^[63]、语音识别^[64]、图标注生成^[65]等领域取得了广泛应用。

RNN 的最大特点在于神经网络各隐层之间的节点是具有连接的, 它能够通过获取输入层的输出和前一时刻的隐层状态来计算当前时刻隐层的输出, 也就是说 RNN 能够对过去的信息进行记忆。理论上来说, RNN 能够对任意长度的序列数据进行建模, 但在实际应用中往往假设当前状态仅与前几个时刻的历史状态相关, 从而帮助降低模型的复杂

度, 图 5 是一个典型的 RNN 结构, 包含输入单元、输出单元和隐层单元。

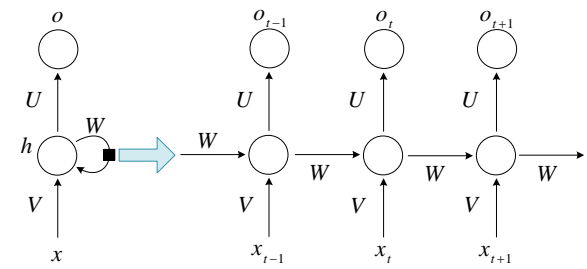


图 5 循环神经网络结构示意图

将 RNN 展开之后发现, 它是一类所有层共享相同权值的深度前馈神经网络。但是, 普通的 RNN 结构存在梯度消失问题, 很难学习数据之间的长程依赖关系。针对这个问题, 研究者相继提出了一些 RNN 的变种, 其中最著名的包括 Hochreiter 等人^[66]提出的长短时记忆网络 (Long Short-Term Memory, LSTM) 和 Cho 等人^[67]提出的门限循环单元 (Gated Recurrent Unit, GRU)。LSTM 和 GRU 采用了特殊的隐层结构, 通过增加保存长期状态的隐层单元, 能够更加有效地建模长程依赖关系, 是目前应用最为广泛的循环神经网络模型。近几年, 随着深度学习的不断发展, 通过增加更加广泛的记忆模块, 研究者提出了记忆网络 (Memory Network)^[68]、栈式增强循环网络 (Stack-augmented Recurrent Net)^[69]、神经图灵机 (Neural Turing Machines, NTM)^[70]和可微分神经计算机 (Differentiable Neural Computer, DNC)^[71]等模型来建模数据之间长程依赖关系。尤其是 NTM 等工作, 通过将 RNN 与注意力机制进行结合, 极大地发展了循环神经网络的研究与应用。

循环神经网络在推荐系统中的应用主要是用来建模数据之间的序列影响, 从而帮助获取更有效的用户和项目隐表示。主要包括两个方面: 首先是被应用于建模推荐系统中用户行为的序列模式^[10,12,72-78], 其次是在获取用户和项目隐表示的过程中, 循环神经网络被应用于建模用户和项目相关的文本信息中词语之间序列影响^[4,38]。当前的应用场景主要包括评分预测、图像推荐、文本推荐、基于位置社交网络中的兴趣点推荐等。

4 基于深度学习的推荐系统

基于深度学习的推荐系统通常将各类用户和项目相关的数据作为输入, 利用深度学习模型学习到用户和项目的隐表示, 并基于这种隐表示为用户

产生项目推荐。一个基本的架构如图6所示，包含输入层、模型层和输出层。输入层的数据主要包括：用户的显式反馈（评分、喜欢\不喜欢）或隐式反馈数据（浏览、点击等行为数据）、用户画像（性别、年龄、喜好等）和项目内容（文本、图像等描述或内容）数据、用户生成内容（社会化关系、标注、评论等辅助数据）。在模型层，使用的深度学习模型比较广泛，包括自编码器、受限玻尔兹曼机、卷积神经网络、循环神经网络等。在输出层，通过利用学习到的用户和项目隐表示，通过内积、Softmax、相似度计算等方法产生项目的推荐列表。

本文通过充分调研当前深度学习在推荐系统研究中的应用情况，根据推荐系统中利用的数据类型并结合传统推荐系统的分类，将当前的研究主要分为五个方向：

（1）深度学习在基于内容的推荐系统中的应用。利用用户的显式反馈或隐式反馈数据、用户画像和项目内容数据，以及各种类型的用户生成内容，采用深度学习方法来学习用户与项目的隐向量，并将与用户访问过的项目相似的项目推荐给用户；

（2）深度学习在协同过滤中的应用。利用用户的显式反馈或隐式反馈数据，采用深度学习学习方法学习用户或项目的隐向量，从而基于隐向量预测用户对项目的评分或偏好；

（3）深度学习在混合推荐系统中的应用。利用用户的显式反馈或隐式反馈数据、用户画像和项目内容数据，以及各种类型的用户生成内容产生推荐，模型层面主要是基于内容的推荐方法与协同过滤方法的组合；

（4）深度学习在基于社交网络的推荐系统中的应用。利用用户的显式反馈或隐式反馈数据、用户的社会化关系等各类数据，采用深度学习模型重点建模用户之间的社会关系影响，更好地发现用户对项目的偏好；

（5）深度学习在情景感知的推荐系统中的应用。利用用户的显式反馈或隐式反馈数据，以及用户的情境信息等各类数据，采用深度学习模型对用户情境进行建模，发现用户在特定情境下的偏好。

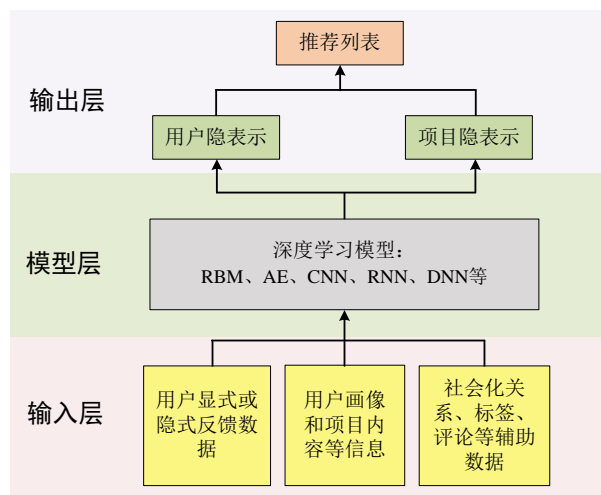


图6 基于深度学习的推荐系统框架

4.1 深度学习在基于内容的推荐系统中的应用

基于内容的推荐方法的性能严重依赖于有效的数据特征提取。深度学习的最大优势是能够通过一种通用的端到端的过程学习到数据的特征，自动获取到数据的高层次表示，而不依赖于人工设计特征。因此，深度学习在基于内容的推荐中主要被用于从项目的内容信息中提取项目的隐表示，以及从用户的画像信息以及历史行为数据中获取用户的隐表示，然后基于隐表示通过计算用户和项目的匹配度来产生推荐。在假设用户和项目携带辅助信息的情况下，深度神经网络模型被作为有效的特征提取工具。

4.1.1 基于多层感知机的方法

深度结构化语义模型。Elkahky等人^[10]考虑到传统的基于内容的推荐系统中，用户特征难以获取的问题，通过分析用户的浏览记录和搜索记录提取用户的特征，丰富用户的特征表示。作者将深度结构化语义模型（DSSM, Deep Structured Semantic Models）^[79]进行扩展，提出了一种多视角深度神经网络模型（Multi-View Deep Neural Network, Multi-View DNN），该模型通过用户和项目两种信息实体的语义匹配来实现用户的项目推荐，是一种实用性非常强的基于内容的推荐方法。其基本思想是设置两类映射通路，分别通过深度学习模型将两类信息实体映射到同一个隐空间，在这个隐含空间中通过余弦相似度计算两个实体的匹配度，然后根据匹配度产生推荐。图7展示了一个Multi-View DNN的通用示意图，在用户视角上，通过利用用户的搜索、浏览、下载、视频观看等历史记录作为输入 x_u ，通过深度学习模型学习用户的隐表示 y_u ，

在项目视角上, 通过利用项目的标题、类别、描述等信息作为输入 x_i , 通过深度学习模型学习项目的隐表示 y_i , 模型共包括一个用户视角和 N 个项目视角, 其中 N 为所有项目的数量, 用户视角的深度神经网络模型为 $f_U(x_U, W_U)$, 第 i 个项目视角的深度神经网络模型为 $f_i(x_i, W_i)$ 。假设有 M 个样本 $\{(x_{u,j}, x_{a,j})\}_{0 \leq j \leq M}$, $(x_{u,j}, x_{a,j})$, 其表示用户 u 和项目 a 之间的一次交互, 通过拟合用户与项目的交互历史进行参数学习:

$$\arg \max_{W_U, W_1, \dots, W_N} \sum_{j=1}^M \sum_{1 \leq a \leq N} \frac{e^{\cos(f_U(x_{u,j}, W_U), f_a(x_{a,j}, W_a))}}{e^{\cos(f_U(x_{u,j}, W_U), f_a(x_{a,j}, W_a))}} \quad (2)$$

在模型训练完成之后, 基于模型学习到的用户隐表示 y_U 和项目隐表示 y_i , 通过在隐空间中计算用户和项目的相似度, 选择相似度最高的 k 个项目产生推荐。

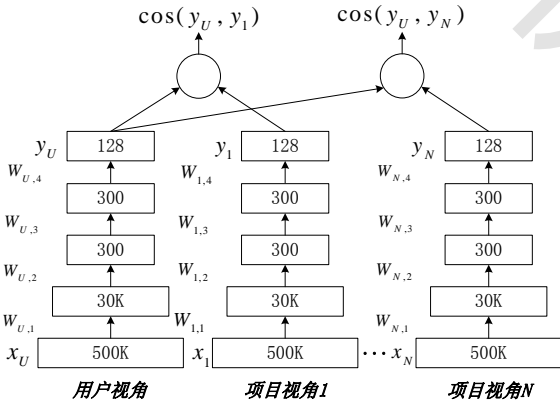


图7 Multi-View DNN 的模型结构

Zheng等人^[18]通过采用与DSSM相似的结构, 考虑将评论信息融入到推荐系统中以缓解推荐系统的数据稀疏问题, 并提高推荐系统的质量, 提出了一种深度协作神经网络模型 (Deep Cooperative Neural Network, DeepCoNN)。其主要思路是利用两个并行的神经网络模型学习用户和项目的隐特征, 一个网络通过用户的所有评论数据建模了用户的偏好, 另一个网络通过项目的所有评论信息建模项目的特征, 然后在两个神经网络上面构建一个交互层来预测用户对项目的评分。该方法通过利用用户评论的文本内容信息, 有效提升了推荐的质量。

Xu等人^[80]基于DSSM模型研究标签感知的个性化推荐问题, 分别利用用户的所有标签和项目的所有标签定义用户和项目的输入特征, 从而学习用户和项目的隐表示, 通过计算用户隐表示和项目隐表示的相似度来产生推荐。Chen等人^[81]研究了位置感知的个性化新闻推荐问题。通过DSSM中增加一个位置通道, 利用MLP从用户信息、项目信息和位置局部主题分布中学习用户、项目和位置的隐表示, 最后联合三个方面的信息计算特定位置下用户兴趣与新闻内容之间的关联度来产生新闻推荐。

深广学习模型。Cheng 等人^[9]通过利用用户特征、情境特征和项目特征等多源异构数据, 提出了一种深广学习 (Wide&Deep Learning) 模型, 用于手机 APP 推荐, 模型同时具有了高的记忆 (memorization) 能力和泛化 (generalization) 能力。记忆主要依靠统计方法, 通过关联学习分析历史记录中的共现现象, 主要利用了相关性, 不具有泛化性且需要手动特征工程。泛化需要研究关联的传递性, 通过探索更多的信息能够提升推荐的多样性。图8展示了一个深广学习模型的通用示意图, 这种模型联合训练一个宽广线性模型 (图中左侧) 和一个深度神经网络 (图中右侧) 来确保模型记忆能力和泛化能力的均衡。

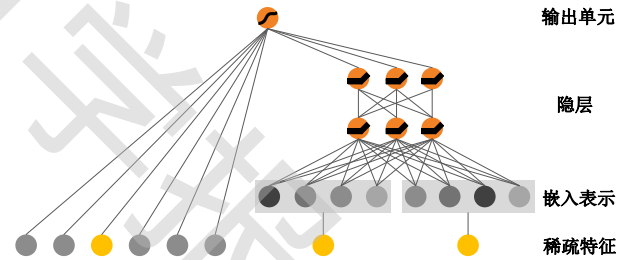


图8 深广学习模型的模型结构

类似于深广学习模型, He等人^[82]利用深度神经网络采用非线性方式建模用户与项目之间的复杂交互, 提出一种神经协同过滤方法 (Neural Collaborative Filtering, NCF), 将用户和项目的特征作为输入, 然后利用多层神经网络学习用户与项目之间的交互函数。通过将NCF结构进行实例化, 得到一种矩阵因子分解的泛化结构和一种多层感知机结构, 分别可以建模用户与项目之间交互的线性和非线性特征, 最后在NCF框架下组合以上两种结构, 提出一种神经矩阵因子分解模型 (Neural Matrix Factorization Model, NeuMF), 在建模用户与项目的交互中组合了矩阵因子分解的线性特征和深度神经网络的非线性特征。

基于MLP的广告点击率预测。多层感知机模型目前在广告点击率预测问题中得到了广泛应用。Web领域中的输入特征通常是离散和稀疏的,为了有效建模这类数据,学习特征之间的交互至关重要^[81,83]。其中一种思路是利用MLP直接学习特征之间的交互,例如Wide&Deep^[9]、Deep Crossing^[30]、Deep&Cross^[29]、DEF^[31]和DIN^[33]。另一种思路是结合因子化机^[83]与MLP,首先利用因子化机建模特征之间的成对交互,然后通过增加全连接层来进一步建模高阶的特征交互,例如PNN^[84]、DeepFM^[32]、NFM^[85]和AFM^[86]。

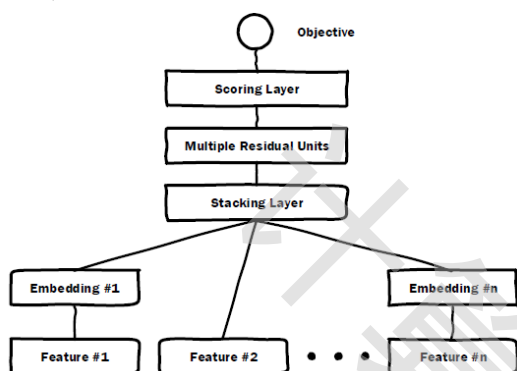


图9 Deep Crossing 模型的模型结构

Shan等人^[30]提出了一种Deep Crossing模型用于广告点击率预测。Deep Crossing的模型结构如图9所示,是由一个嵌入层、一个堆栈层、一个残差单元和一个评分层组成的多层感知机。Wang等人^[29]基于Deep Crossing,提出了一个Deep&Cross网络模型用户广告点击率预测。Deep&Cross同时在输入特征上采用一个深度网络（Deep Network）和交叉网络（Cross Network），然后组合两部分输出进行预测。Deep&Cross同时集成了深度神经网络（DNN）和Deep Crossing的优点。Zhu等人^[31]基于Deep Crossing提出了一个深度嵌入森林模型（Deep Embedding Forest, DEF），将Deep Crossing中的残差单元替换为森林层,通过预训练能够降低模型的在线预测时间。Zhou等人^[33]考虑到大部分点击率预测研究中缺乏对用户行为多样性和局部激活（Local Activation）现象的建模,提出了一种深度兴趣网络（Deep Interest Network, DIN），DIN基于一种多层感知机结构,通过引入用户兴趣分布和注意力机制实现对用户行为多样性和局部激活现象的建模。

Qu等人^[84]基于因子化机模型和MLP提出了一个基于产品的神经网络模型（Product-based Neural Network, PNN）。模型结构如图10所示,由一个特征嵌入层、一个成对特征交互层和两个全连接层

组成。Guo等人^[32]基于Wide&Deep,结合因子化机和深度学习,提出了一种基于深度网络的因子化机模型（Factorization-Machine based Neural Network, DeepFM）进行点击率预测,分别采用因子化机和深度神经网络建模低层次和高层次的特征交互,相比Wide&Deep,DeepFM不需要进行人工的特征工程处理。He等人^[85]研究了稀疏数据输入情况下的推荐问题,基于因子化机模型提出了一种神经因子化机模型（Neural Factorization Machine, NFM），能够实现特征之间的高层次非线性交互,从而提升预测能力。该模型与深广模型的最大区别是,其在特征嵌入层之上增加了一个双线性交互池化操作。相比深度学习模型,神经因子化机模型在使用更少隐层的情况下获得了更好的预测性能,而且因为参数更少,训练更加容易。Xiao等人^[86]通过扩展NFM模型,提出了一种注意力因子化机模型（Attentional Factorization Machine, AFM），通过将注意力机制引入双线性交互池化操作中,提升了NFM的表示能力和可解释性。

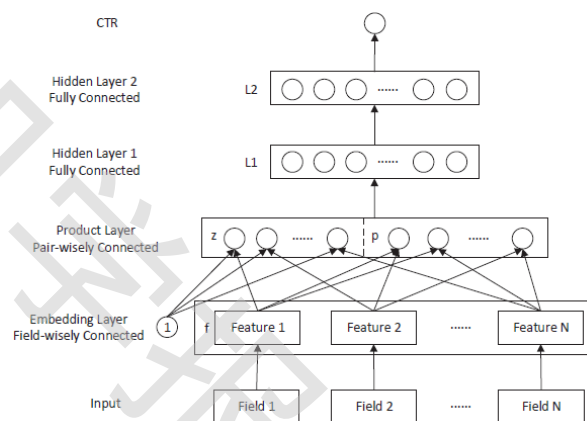


图10 PNN 模型的模型结构

基于 MLP 的 YouTube 视频推荐方法。

Covington等人^[8]通过利用用户信息、情境信息、历史行为数据和项目的特征信息等多源异构数据,提出了一种深度神经网络模型用于YouTube视频推荐。YouTube视频推荐主要面临三个方面的挑战:可扩展性、新鲜度和数据噪声问题。为了克服这三个挑战,该研究将深度神经网络模型应用到视频推荐系统的两个关键过程:候选集生成和排序,系统架构如图11所示。候选集生成的目的是从海量视频库中筛选出和用户相关的几百个视频,主要利用用户在YouTube上的历史行为数据、用户特征和情境信息建模用户对视频的个性化偏好,其核心方法是将推荐问题转化为一个基于深度神经网络的分

类问题, 寻找与用户向量(神经网络变换后的特征向量)距离最近的 N 个视频。排序过程是通过进一步考虑更多的视频特征, 利用神经网络和逻辑回归模型对每个候选视频进行打分, 并根据打分值对视频进行排序。

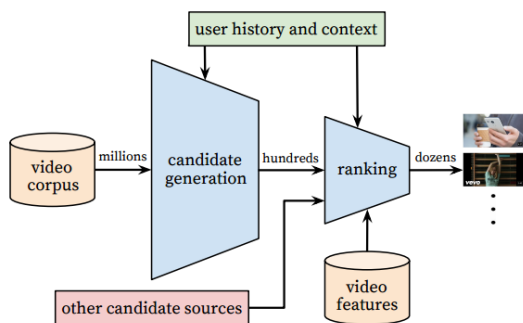


图 11 Covington 等人^[8]提出的方法的模型架构

4.1.2 基于卷积神经网络的方法

基于注意力的 CNN。Gong 等人^[87]提出了一种基于注意力的卷积神经网络(CNN)来进行微博中的 Hashtag 推荐。总的来说, 作者将 Hashtag 推荐作为一个多标记分类问题, CNN 被作为一种特征提取手段来获取微博的特征。提出的模型包括一个全局通道和一个局部注意力通道。全局通道由一个卷积层和一个 Pooling 层组成, 局部注意力通道由一个注意力层和一个 Pooling 层组成, 模型的架构如图 12 所示。

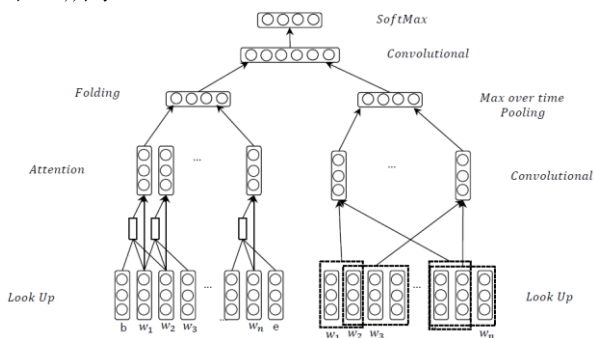


图 12 基于注意力的 CNN 模型架构

之后, Zhang 等人^[88]通过利用多模信息来进行微博的 hashtag 推荐。该工作考虑了文本和图像, 分别采用 CNN 和 RNN 从图像和文本中提取特征, 然后组合两个方面的特征进行标签推荐。同时, 考虑到标签仅仅与图像和文本中的部分信息存在关联, 该工作采用注意力机制来建模这种局部关联性。Seo 等人^[89]研究了如何利用评论信息来进行推荐的问题。模型是一个基于注意力的 CNN 模型, 其包含一个用户网络和一个项目网络, 分别采用 CNN 从用户的所有评论和项目的所有评论中学习用户和项目

的隐表示, 同时采用注意力机制来建模评论中的不同部分与用户偏好和项目特征的关联度。Wang 等人^[90]提出了一种动态注意力深度模型 (Dynamic Attention Deep Model, DADM) 来研究编辑者的文章推荐问题, DADM 利用 CNN 来学习文章的语义信息, 同时利用注意力机制来抓住编辑者选择文章行为的动态性。

比较深度学习方法。Lei 等人^[61]基于深度学习方法研究了图像推荐的问题。该研究指出图像推荐最重要的是需要在图像的语义理解与用户对图像的偏好或意图之间建立桥梁, 因此学习到的图像表示不仅仅需要具有高的表达性和可分类性, 更重要的是需要反映用户对图像的偏好。针对这个问题, 该研究提出了一种比较深度学习方法 (Comparative Deep Learning, CDL), 其主要思路是利用 MLP 和 CNN 分别从用户的多源异构数据(包括用户画像、标签信息等)和图像的视觉信息中学习用户和图像的隐表示, 并将用户和图像映射到同一隐空间中。模型的训练过程中, 利用了比较学习的思想, 即同时利用正反馈图像和负反馈图像, 比较它们与用户之间距离(即正反馈图像与用户的距离应该比负反馈图像与用户的距离要小), 并采用交叉熵损失函数进行参数学习。最后通过计算用户和图像之间的距离来产生图像推荐。

基于 CNN 的音乐推荐。Oord 等人^[57]研究了如何利用深度学习模型来解决音乐推荐系统中的冷启动问题。在音乐推荐中, 协同过滤通常面临冷启动问题, 即对于一些没有用户数据的音乐, 往往不能够被推荐给用户。作者首先利用用户的历史收听数据和音乐的音频信号数据, 通过组合加权矩阵因子分解和卷积神经网络, 将用户和音乐投影到一个共享的隐空间, 从而能够学习到用户和歌曲的隐表示。对于新的歌曲, 可以通过训练好的卷积神经网络从自身的音频信号中提取出歌曲的隐表示, 从而能够在共享隐空间中通过计算用户与新音乐之间的相似性来为用户推荐音乐, 帮助解决新项目的冷启动问题。

4.1.3 基于循环神经网络的方法

基于注意力的 RNN。前面我们讨论过基于注意力的 CNN 模型。与其类似, 注意力机制也被用于基于 RNN 的推荐方法中, Li 等人^[91]提出了一种基于注意力的 LSTM 来进行微博中的 hashtag 推荐, 注意力机制与 RNN 结合的优势是能够抓住文本的序列特征, 同时能够从微博中识别最具有信息量的词。模

型首先利用 LSTM 来学习微博的隐状态 (h_1, h_2, \dots, h_N) , 同时采用主题模型来学习微博的主题分布。隐状态的注意力权值 a_j 通过微博第 j 个位置附近的词和微博的主题分布来计算。注意力层的输出 $vec = \sum_{j=1}^N a_j h_j$ 。模型的架构如图 13 所示。

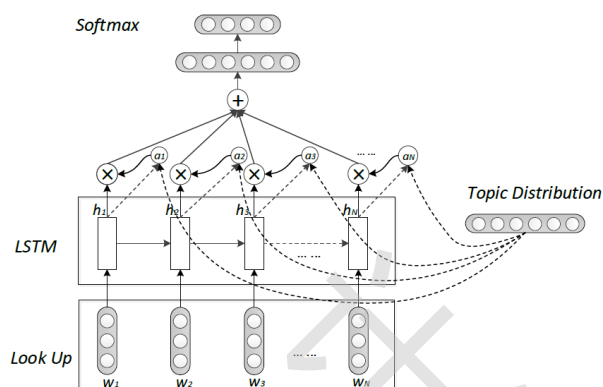


图 13 基于注意力的 RNN 模型架构

类似于[87,91]的工作, Huang等人^[92]使用记忆网络(Memory Networks)代替CNN和LSTM, 提出了一个基于注意力的记忆网络来进行hashtag推荐。此外, Huang等人^[93]还提出了一个基于注意力的记忆网络来进行微博的提示推荐。考虑到微博长度通常很短, 存在单词稀疏和单词同义问题, 仅仅依靠语言模型(如词嵌入模型)获得的推荐性能往往非常有限。作者通过利用用户的历史微博作为外部记忆单元来建模用户的兴趣, 提升了提示推荐的准确性。具体地, 作者利用两个记忆网络分别从作者的历史微博和目标用户的历史微博中发现作者和目标用户的兴趣, 最后联合微博的内容、作者的兴趣和目标用户的兴趣实现微博用户的提示推荐。

基于RNN的新闻推荐。Okura等人^[94]采用深度学习研究方法研究了新闻推荐问题。首先为了抓住文章的语义信息, 采用降噪自编码器(DAE)从新闻中提取文章的隐表示; 然后为了学习用户的偏好, 采用RNN从用户的历史行为列表中学习用户的隐表示; 最后, 为了利用用户与新闻之间的关联, 基于新闻和用户的隐表示采用点乘的方式为用户产生新闻推荐列表。

4.1.4 基于深度信念网络的方法

基于 DBN 的音乐推荐。传统的基于内容的音乐推荐将音乐内容特征提取与音乐推荐分为两个独立过程, 这可能导致推荐性能的不足, Wang 等人^[53]通过深度信念网络和概率矩阵分解

(Probabilistic Matrix Factorization, PMF) 将两个过程组合到一个统一框架中, 提升了音乐推荐的性能。具体地, 首先学习两种用户和项目的隐表示, 一种是利用概率矩阵分解学习到用户和项目表示, 第二种是采用深度信念网络从音乐中提取的音乐特征表示以及由用户偏好得到的用户表示, 然后利用两种隐表示分别做内积, 并组合两种内积通过拟合用户-项目评分矩阵进行模型训练, 最后基于学习到的两种表示对未知评分进行预测。

总的来说, 深度学习能够有效缓解用户和项目特征提取困难的问题, 以及新项目的冷启动问题, 同时能够将用户和项目特征提取与推荐过程融合到统一的框架中, 但是并不能够解决基于内容的推荐方法自身存在的新用户问题, 也不能为用户发现新的感兴趣的资源, 只能发现与用户已有兴趣相似的资源。

4.2 深度学习在协同过滤中的应用

协同过滤, 如矩阵因子分解, 通过学习用户和向量的低维向量表示来实现推荐, 通常面临可扩展性不足的问题^[46]。深度学习由于能够适应于大规模数据处理, 目前被广泛应用于协同过滤推荐问题中。基于深度学习的协同过滤方法利用用户对项目的显式反馈或隐式反馈数据, 采用深度学习训练一个推荐模型, 是一类基于模型的协同过滤推荐方法。其主要思路是将用户的评分向量或项目的被评分向量作为输入, 利用深度学习模型学习用户或项目的隐表示, 然后利用逐点损失(Point-wise Loss)和成对损失(Pair-wise Loss)等类型的损失函数构建目标优化函数对深度学习模型的参数进行优化, 最后利用学习到的隐表示进行项目推荐。根据深度学习模型的不同, 本文将深度学习应用到协同过滤中的研究分为五类。

4.2.1 基于受限玻尔兹曼机的协同过滤方法

基于RBM的协同过滤。2007年Salakhutdinov等人^[46]首次将深度学习应用于解决推荐问题, 提出一种基于受限玻尔兹曼机的协同过滤推荐模型。假设有 m 部电影, 则使用 m 个Softmax单元来作为可见单元构造RBM。每个用户都有一个单独的RBM, 对于不同的RBM仅仅是可见单元不同, 因为不同的用户会对相同的电影打分, 因此所有这些RBM的可见单元共享相同的偏置以及可见单元与隐层单元的连接权值 W 。Salakhutdinov等人对传统的RBM模型进行了改进, 一是在可见层, 评分数据通过一个固定长度的0-1向量进行表示; 二是考虑到用户只对

很少的项目进行了评分,使用一种不与任何隐层单元连接的Missing单元表示未评分的项目。为了考虑用户未评分的电影的信息,Salakhutdinov等人采用一个0-1向量 \mathbf{r} 表示电影是否被评分过,通过融入这种辅助信息提出了一种条件RBM。模型的结构如图14所示。

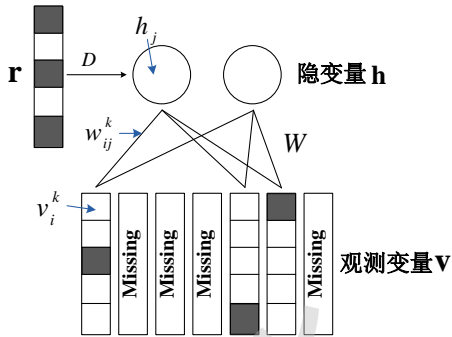


图14 基于RBM的协同过滤方法的模型结构

根据RBM模型可见层单元之间和隐层单元之间条件独立的性质,当给定可见单元状态时,可见单元 \mathbf{v} 与隐层单元 \mathbf{h} 的条件概率可以表示为:

$$p(v_i^k = 1 | \mathbf{h}) = \frac{\exp(b_i^k + \sum_{j=1}^F h_j w_{ij}^k)}{\sum_{l=1}^K \exp(b_i^l + \sum_{j=1}^F h_j w_{ij}^l)} \quad (4)$$

$$p(h_j = 1 | \mathbf{v}, \mathbf{r}) = \sigma(b_j + \sum_{i=1}^m \sum_{k=1}^K v_i^k w_{ij}^k + \sum_{i=1}^m r_i D_{ij}) \quad (5)$$

其中, K 和 F 分别表示用户和隐变量的数量,模型需要训练的参数包括权值 $\{w_{ij}^k\}$,隐层节点偏置 $\{b_j\}$ 和可视节点偏置 $\{b_i^k\}$ 。参数训练采用2002年Hinton提出的对比散度算法^[45]。

Phung等人^[47]通过将Salakhutdinov等人的工作进行扩展,用于建模用户评分的序数特征,采用一种统一的方式同时抓住了相似性和共现性。此外,考虑到RBM模型仅仅利用了项目之间的关联,Georgiev等人^[48]通过增加用户之间的关联,对RBM模型进行了扩展,并且对模型的训练和预测过程进行了简化,同时还使得模型能够直接处理实值评分数据。何洁月等人^[49]将RBM模型进行扩展,提出一种基于实值状态的玻尔兹曼机,该模型从三个方面对RBM进行改进,一是能够直接将评分数据作为可见单元的状态,不再需要转化为 K 维的0-1向量表示,二是在训练数据中增加使用了未评分信息,三是将好友信任关系融入到该模型之中,能够有效缓

解模型的稀疏性。

基于受限玻尔兹曼机的协同过滤方法的最大问题是连接隐层和可见层的权重参数规模过大,同时由于受限玻尔兹曼机的训练过程往往依靠变分推理和蒙特卡罗采样等近似优化方法,训练的时间过长,导致基于受限玻尔兹曼机的协同过滤方法在实际应用中受到很大限制。

4.2.2 基于自编码器的协同过滤方法

最近,自编码器也被应用于协同过滤中,其通过对用户或项目的显式或隐式反馈数据进行重构,能够学习用户或项目的隐表示,并基于这种隐表示预测用户对项目的偏好。图15给出了一个基本的模型结构,给定一个用户的评分向量 R_i ,其中包含评分数据(图中用灰色圆圈表示)和未评分数据(图中用白色圆圈表示),通过最小化自编码器的重构误差 $\ell(R_i, R'_i)$ 训练得到模型的参数,最后对未知评分进行预测:

$$\hat{R}_i = g(W'_i f(W_i R_i + \mu) + b) \quad (6)$$

其中, g 和 f 为激励函数, W_i 和 W'_i 为权值矩阵, μ 和 b 为偏置向量。

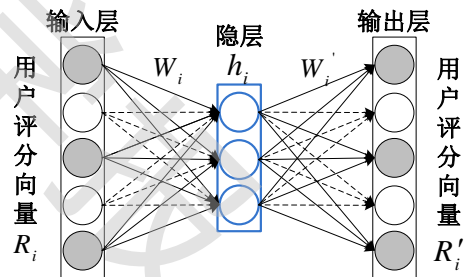


图15 基于自编码器的协同过滤方法的模型架构

基于自编码器的协同过滤方法。Sedhain等人^[37]提出了一种基于自编码器的协同过滤方法(AutoRec),该模型的输入为评分矩阵 R 中的一行(User-based)或者一列(Item-based),利用一个编码过程和一个解码过程产生输出,通过最小化重构误差进行模型参数优化。Strub等人^[95]采用两个栈式降噪自编码器(SDAE),将用户和项目的评分向量分别作为输入,分别学习用户和项目的隐表示,然后通过隐表示对缺失评分进行预测。该方法与AutoRec略微不同,其针对评分矩阵的数据稀疏问题,在训练过程中通过将评分矩阵中的缺失值直接归零,从而减少了网络的连接数量,但同时这种

方式也导致未评分数据的信息被忽略。

协同降噪自编码器模型。 Wu等人^[20]利用降噪自编码器来解决top- N 推荐问题,提出了一种协同降噪自编码器模型(Collaborative Denoising Auto-Encoders, CDAE)。CDAE与AutoRec方法的结构类似,通过将用户的评分向量作为输入,学习用户的低维向量表示来进行推荐。但是与AutoRec相比存在一些差异:一是该方法不是评分预测,而是top- N 推荐;二是该方法通过在评分向量中加入了噪声数据,提升了模型的鲁棒性;三是考虑不同用户的个性化因素,为每个用户引入了一个用户因子提升了推荐的准确性。

基于自编码器的表示学习。 Wu等人在[20]中指出,自编码器在最小重构误差时既能采用逐点损失(Point-wise Loss)也能采用成对损失(Pair-wise Loss),但具体选择需要针对特定的任务来决定,可是作者并没有给出选择的依据。Zhuang等人^[96]在后来的研究中指出,推荐问题中每个用户的评分标准存在差异,例如有的用户对自己不是特别满意的项目可能给出高分,而一些标准苛刻的用户可能对满足自己要求的项目给出低分,因此,推荐系统的评分预测不仅仅需要使预测的评分数据与真实评分数据一致,而且还需要使预测的评分数据之间的相对排序与真实评分数据排序一致。针对这个问题,作者提出在推荐问题中融入成对排序损失(Pair-wise Ranking Loss)来确保预测数据与真实数据的排序保持一致。模型利用自编码器分别学习用户和项目的隐表示,再通过组合逐点损失和成对排序损失构建目标函数,采用梯度下降方法进行参数优化,最后基于学习到的用户和项目隐表示,通过内积方式进行评分预测。

基于自编码器的协同过滤方法通常来说简单有效,尤其是利用栈式降噪自编码器,通过提高模型的深度和增加噪声,推荐的有效性和鲁棒性都得到了提升。

4.2.3 基于分布式表示技术的协同过滤方法

传统的协同过滤(如矩阵因子分解)忽略了用户行为中的序列模式,但是很多推荐场景中,序列模式会对推荐的结果起到至关重要的作用,例如基于会话的推荐(Session-based Recommendation)问题中,一个会话中之前的行为对后面的行为具有决定性作用。传统的序列模式建模方法(例如马尔科夫模型)往往存在计算复杂且准确性不高的问题。

最近几年,浅层神经网络模型得到了广泛应

用,特别是词分布式表示模型(Embedding Models)。分布式表示模型最早应用在自然语言处理领域中,利用背景信息构建词汇的分布式表示。分布式表示模型往往简单而且比较有效,因此很快被用于建模推荐系统中的用户行为序列模式。其核心思想是利用用户对项目的访问序列同时构建用户和项目(或者其他背景信息)的分布式表示,并将多种实体的分布式表示映射到同一个隐空间中,进而通过计算两个实体之间的相似性实现项目的推荐。图 16 是一个基本的分布式表示模型,给定一个用户的访问序列 $S_j = \{v_{i-k}, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_{i+k}\}$, 定义一个目标项目(Target Item) $v_i \in V$, 其中 V 是所有项目的集合,序列中其他所有项目为 v_i 的情境项目(Context Items),给定 S_j , 能够通过最大化序列的似然 $\ell(S_j)$ 求目标序列 v_i 的分布式表示:

$$\ell(S_j) = \frac{1}{|S_j|} \sum_{v_i \in S_j} \sum_{-k \leq c \leq k} \log p(v_{i+c} | v_i) \quad (7)$$

其中 $p(v_{i+c} | v_i)$ 为给定目标项目 v_i , 生成情境项

目 v_{i+c} 的概率,通常采用 Softmax 函数进行定义:

$$p(v_{i+c} | v_i) = \frac{\exp(\mathbf{v}_{i+c} \cdot \mathbf{v}_i)}{\sum_{v_t \in V} \exp(\mathbf{v}_t \cdot \mathbf{v}_i)} \quad (8)$$

对模型的学习通常采用 Mikolov 等人^[97]提出的层次 Softmax 和 Negative Sampling 方法。最终,通过学习到的项目分布式表示,能够利用基于用户的协同过滤、最近邻等方法来产生推荐。

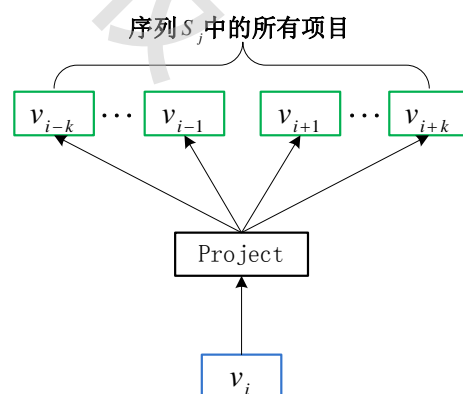


图 16 分布式表示模型结构

基于分布式表示技术的下一次购物篮推荐。

Wang 等人^[98]将分布式表示技术用于解决下一次购物篮推荐问题。购物篮推荐问题是给定用户的购物历史(通常是交易数据的序列),预测用户下一次

购物行为要选择的物品。该研究考虑建模用户的历史序列行为和用户的长时偏好进行购物篮推荐,提出了一种层次表示模型(Hierarchical Representation Model, HRM)。HRM是一种两层结构的混合表示模型,第一层通过聚合最后一次交易中的所有项目的隐向量形成交易表示(建模序列模式),第二层通过聚合用户的隐向量(建模用户偏好)和交易表示构建混合表示,最终利用混合表示进行下一次购物篮预测。模型结构如图17所示, \mathbf{v}_u^U 和 \mathbf{v}_i^I 分别表示用户 u 和项目 i 的隐向量,对于用户 u , 给定其上一次交易历史序列 T_{t-1}^u , 需要预测下一次购物序列 T_t^u , HRM定义用户 u 在下一次购物中选择项目 i 的概率为:

$$p(i \in T_t^u | u, T_{t-1}^u) = \frac{\exp(\mathbf{v}_i^I \cdot \mathbf{v}_{u,t-1}^{Hybrid})}{\sum_{j=1}^{|I|} \exp(\mathbf{v}_j^I \cdot \mathbf{v}_{u,t-1}^{Hybrid})} \quad (9)$$

其中 $|I|$ 表示所有项目的数量, $\mathbf{v}_{u,t-1}^{Hybrid}$ 是第二层通过聚合用户隐向量 \mathbf{v}_u^U 和交易表示 $f_1(\mathbf{v}_i^I \in T_{t-1}^u)$ 构建的混合表示, 即:

$$\mathbf{v}_{u,t-1}^{Hybrid} = f_2(\mathbf{v}_u^U, f_1(\mathbf{v}_i^I \in T_{t-1}^u)) \quad (10)$$

其中 $f_1(\cdot)$ 和 $f_2(\cdot)$ 都是聚合函数, 可以是最大池化(Max Pooling)和平均池化(Average Pooling)等形式。最后通过对所有用户的历史交易数据进行拟合, 采用梯度下降方法进行模型的训练, 进而利用训练好的HRM进行下一次购物篮预测。

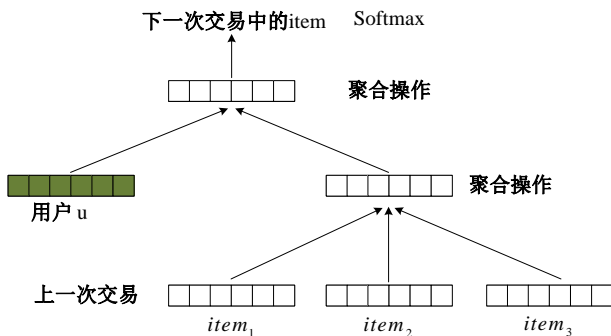


图17 HRM的模型结构

基于分布式表示技术的广告推荐。Grbovic等人^[14]利用分布式表示技术研究邮件系统的广告推荐

问题, 利用用户与不同电子商务网站之间的邮件往来获取到的用户历史购买行为数据, 预测用户对产品的兴趣。首先, 作者提出了两种产品分布式表示方法获取用户或项目的分布式表示, 一种是Prod2Vec方法, 所有产品被独立考虑, 其在产品层次上建模, 采用分布式表示技术将所有产品嵌入到一个连续的、低维的向量空间; 另一种方法是Bagged-Prod2Vec, 其考虑一封邮件中可能包含多个产品的购买行为, 在邮件层次上进行建模, 采用分布式表示技术获取到更加有效的产品向量表示。获取项目的分布式表示之后, 作者提出了两种推荐模型, 一种是产品到产品的推荐模型, 其在嵌入式空间中计算产品之间的相似度, 然后利用基于用户的协同过滤产生广告推荐; 另一种是用户到产品的推荐模型, 其通过将用户和产品嵌入到统一的低维空间, 从而通过计算用户与产品的相似度产生推荐。

基于分布式表示技术的协同过滤方法, 能够建模用户行为的序列影响。同时, 由于其简单而且训练高效, 目前在推荐系统研究中引起了大量的关注^[98]。但是, 这种方法仅仅建模了用户行为的局部情境信息, 并不能完全有效地抓住用户行为的序列模式。

4.2.4 基于循环神经网络的协同过滤方法

循环神经网络模型正是针对序列数据建模而提出, 因此很快被引入到用户行为序列模式建模的研究中。基于循环神经网络的协同过滤与基于分布式表示技术的协同过滤类似, 都能用来建模用户行为的序列模式, 但区别在于分布式表示技术仅仅抓住了用户行为的局部情境信息, 而RNN能够建模用户行为之间的相互依赖关系。基于循环神经网络的协同过滤的主要思路是利用循环神经网络建模用户历史序列行为对当前时刻用户行为的影响, 从而实现用户的项目推荐和行为预测。图18是一个基本的基于RNN的协同过滤方法框架, 已知一个用户的行为序列 $S = \{x_1, x_2, \dots, x_t\}$, 首先, 对其进行嵌入式表示并作为循环神经网络在每一时刻的输入, 时刻 t 的隐向量 $h_t = f(Vx_t + Wh_{t-1})$, 其中 f 为激励函数。输出 o_t 为时刻 t 选择特定项目的概率, 通常基于 t 时刻的隐向量采用 Softmax 等方法进行计算。根据应用场景不同, 基于循环神经网络的协同过滤模型主要被用于基于会话的推荐、融入时间序

列信息的协同过滤等应用中。

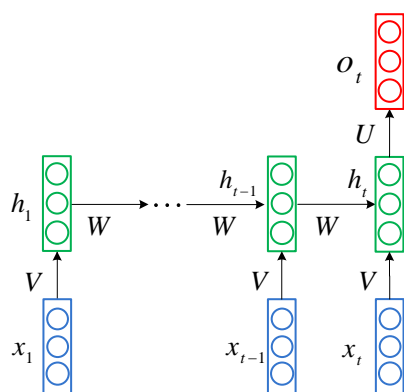
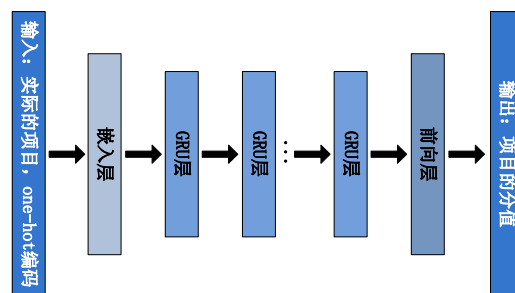


图 18 基于循环神经网络的推荐方法的模型架构

基于会话的推荐。基于会话的推荐系统主要利用当前会话中的历史行为记录来预测下一步点击每个项目的概率，其最大特点是仅仅考虑了一个会话期间的行为数据。Hidasi 等人^[75]采用 GRU 来抓住会话中行为之间的依赖关系，基本架构如图 19 所示，在每一个时间点上，模型的输入是当前点击的项目的 one-hot 编码，然后通过一个嵌入层压缩为低维连续向量，中间是多个基本的 GRU 层和一个前向层，输出层利用 Softmax 等方法计算的每个项目的点击概率。为了实现并行计算，该研究通过把不同的会话拼接起来，然后采用 mini-batch 处理。此外，在输出阶段，考虑到项目数量太多会导致计算量过大，该研究通过对所有项目进行抽样，仅仅预测部分项目被点击的概率。

在[76]中，Tan 等人对 Hidasi 等人^[75]的工作提出了一系列改进，一是数据增强（data augmentation），首先是通过将每一个长度的序列都作为一个训练样本来增加训练样本的数量，然后采用嵌入式 Dropout 随机去除一些序列中的节点，帮助缓解由于噪声数据导致的过拟合；二是预训练，考虑到用户偏好会随时间变化，通过利用所有历史数据对模型进行预训练，然后仅仅利用最近的数据进行更精细粒度的训练，这样同时抓住了用户的长时偏好和短时偏好；三是利用特权信息（Privileged Information）进行训练，在模型训练过程中，通过融入用户行为序列中预测时间点之后的数据，提升了模型的预测精度。



融入时间序列信息的协同过滤。传统的协同过滤方法并没有考虑用户行为的时间信息，但是时间信息反映了用户行为的时间序列模式，有效地利用时间信息有利于提升推荐系统的性能。Song 等人^[12]通过融入时间信息并在多种粒度上建模用户的兴趣偏好，提出一种多等级时间深度语义结构化模型（Mutli-Rate TDSSM）。Liu 等人^[73]考虑到推荐系统中的用户行为往往存在多种类型，采用循环神经网络模型和 Log 双线性模型（Log-BiLinear, LBL）^[99]分别建模用户行为之间的长程依赖关系和短时情境信息，从而提出了一种循环 Log 双线性模型（Recurrent Log-BiLinear, RLBL），实现用户在下一时刻的行为类型预测。Wu 等人^[77]将循环神经网络用于建模用户的时间序列行为，该研究区分了不同类型的显式反馈和隐式反馈数据，并组合一个循环部分和一个非循环部分来对用户进行推荐。循环部分是一个 RNN 结构，其通过区分不同的行为类型抓住所有历史反馈对当前用户行为的影响。非循环部分是一个全连接神经网络结构，其主要建模了用户基本的偏好。

以上研究实际上仅仅利用了用户的序列行为建模用户偏好的演化，同时假设项目的特征保持不变。但是在实际的推荐系统中，项目的特征可能会发生变化，例如一部电影的受欢迎程度或者受众群体可能会随着时间发生改变。Dai 等人^[78]考虑到用户偏好和项目特征会因为用户交互而随时间演化，基于循环神经网络和多维时间点过程模型，提出了一种循环共演化特征嵌入过程模型（Recurrent Coevolutionary Feature Embedding Processes）实现用户偏好和项目特征的演化跟踪，并学习用户和项目在每一时刻的隐表示，最后通过对用户和项目的隐表示进行内积来产生项目推荐。Wu 等人^[10]通过利用循环神经网络建模用户偏好和项目特征的演化，提出了一种循环推荐网络（Recurrent Recommender Network, RRN），能够预测用户未来

的行为轨迹。具体地, RRN 首先利用低维矩阵因子分解学习用户和项目的静态隐表示, 同时将用户的历史评分数据作为输入, 采用 LSTM 学习用户和项目在每一时刻的动态隐表示, 最后通过聚合两类隐表示的内积实现单一时刻的评分预测。

基于循环神经网络的协同过滤方法, 由于其能够有效建模用户行为中的序列模式。同时通过改变循环神经网络的输入^[77]和定义不同的权值矩阵^[73], 还能够融入时间等情境信息^[10], 以及各种类型的辅助数据来提升推荐的质量^[72], 模型具有高的适用性, 在当前的推荐系统中得到了广泛应用。

4.2.5 基于生成对抗网络的协同过滤方法

Wang 等人^[100]提出的 IRGAN 首次将 GAN 引入到信息检索领域, 实现了信息检索建模中两种思维流派(生成检索模型和判别检索模型)的统一描述。生成检索模型聚焦于给定一个查询 q , 生成相关的文档 d , 判别检索模型聚焦于给定一个查询和文档 (q, d) , 预测二者之间的相关性。IRGAN 的目的是借鉴 GAN 中生成器和判别器相互对抗的思想, 采用一个博弈理论式中的极小化极大算法来将生成检索模型和判别检索模型集成到一个统一的框架中。

形式上讲, 假设 $p_{true}(d|q_n, r)$ 是用户真实的偏好分布, 生成检索模型 $p_\theta(d|q_n, r)$ 需要尽力接近用户真实的偏好分布, 判别检索模型 $f_\phi(q, d)$ 需要尽力区分相关文档和不相关文档。类似于 GAN, 可以定义 IRGAN 的目标函数:

$$J^{G^*, D^*} = \min_q \max_f \sum_{n=1}^N \left\{ E_{d \sim p_{true}(d|q_n, r)} [\log D(d|q_n)] + E_{d \sim p_\theta(d|q_n, r)} [\log(1 - D(d|q_n))] \right\} \quad (11)$$

其中, $D(d|q_n) = \sigma(f(d, q_n))$, σ 是一个 sigmoid 函数, θ 和 ϕ 分别是生成检索模型和判别检索模型的参数, 能够通过采用梯度下降法进行迭代学习而获取到。以上目标函数是通过采用成对排序损失来构建的, 假设 $p_\theta(d|q_n, r)$ 能够通过一个 Softmax 函数进行定义:

$$p_\theta(d_i|q, r) = \frac{\exp(g_\theta(q, d_i))}{\sum_{d_i} \exp(g_\theta(q, d_i))} \quad (12)$$

$g_\theta(q, d)$ 是给定查询 q , 生成文档 d_i 的概率,

$g_\theta(q, d)$ 和 $f_\phi(q, d)$ 通常根据具体的任务进行定义, 可以是相同的形式或不同的形式。在文献[100]中, 作者将其定义为相同的形式: $g_\theta(q, d) = s(q, d)$ 和 $f_\phi(q, d) = s(q, d)$ 。在项目推荐任务中, 作者采用了矩阵因子分解方法来定义 $s(q, d)$:

$$s(u, i) = b_i + U_u^T V_i \quad (12)$$

其中, b_i 是项目 i 的偏置, U_u 和 V_i 分别为用户 u 和项目 i 的隐向量。 $s(u, i)$ 也可以通过因子化机或者神经网络进行定义。

生成对抗网络由于引入了对抗机制, 在推荐系统中能够取得不错的效果。但目前, 生成对抗网络在推荐系统中的应用还处于探索阶段, 更深入的研究有待于进一步展开。

4.2.6 基于其他深度学习模型的协同过滤方法

除了 RBM、AE、分布式表示技术、RNN 和对抗神经网络之外, 目前被用于协同过滤的深度学习模型还包括卷积神经网络、神经自回归分布估计模型 (Neural Autoregressive Distribution Estimator, NADE)^[101]。

Geng 等人^[58]针对社交内容网络中用户与图像之间的连接稀疏以及图像内容多样的问题, 提出了一种神经网络模型来学习社交内容网络中用户和图像的统一表示。该模型利用用户和图像之间由于历史交互形成的二部图网络结构, 采用卷积神经网络模型通过最大化模块度^[102]的方法学习用户与图像的深层次统一表示, 从而在同一空间中通过计算用户和图像的相似度为用户进行图像推荐。Zheng 等人^[103]考虑到基于 RBM 的协同过滤存在优化困难的问题, 使用 NADE 替代 RBM 来进行协同过滤推荐, NADE 不需要融入任何隐变量, 因此避免了复杂的隐变量推理过程, 从而降低了模型复杂度。

总的来说, 基于深度学习的协同过滤能够看做是传统隐因子模型的一种非线性泛化^[20]。其最大的优点是在学习用户和项目隐表示的过程中引入了非线性的特征变换, 相比传统的协同过滤方法(如矩阵因子分解)具有更好的性能^[82, 96]。但同时这类方法仍然无法改变传统协同过滤方法存在的数据

稀疏问题，以及新用户和新项目的冷启动问题。

4.3 深度学习在混合推荐系统中的应用

传统的协同过滤方法仅仅利用了用户的显式反馈或隐式反馈数据，面临数据稀疏问题。通过融入用户画像数据、项目内容数据、社会化标注、评论等辅助数据，混合推荐方法能够有效缓解数据稀疏问题，但是这种方法面临最大的难题是辅助数据的表示问题，经典的方法，如协同主题回归（Collaborative Topic Regression, CTR）^[104]，并不能够获取有效的辅助数据表示^[4]。深度学习通过自动特征提取，能够从辅助数据中学习有效的用户和项目隐表示。

基于深度学习的推荐方法的主要思路是组合基于内容的推荐方法与协同过滤，将用户或项目的特征学习与项目推荐过程集成到一个统一的框架中，首先利用各类深度学习模型学习用户或项目的隐特征，并结合传统的协同过滤方法构建统一的优化函数进行参数训练，然后利用训练出来的模型获取用户和项目最终的隐向量，进而实现用户的项目推荐。根据深度学习模型的不同，本文将基于深度学习的推荐方法分为以下两类。

4.3.1 基于自编码器的混合推荐方法

当前，自编码器在基于深度学习的混合推荐方法中应用最为广泛。基于自编码器的混合推荐方法的基本架构如图 20 所示，自编码器由于具有强的表示学习能力，很自然地用来从用户特征 X 或项目特征 Y 中学习用户隐表示 U 或项目隐表示 V ，然后将隐表示融入到隐因子模型中拟合用户-项目交互矩阵 R （如评分矩阵），最后联合自编码器的重构误差 $\ell(X, U)$ 和 $\ell(Y, V)$ ，以及拟合交互矩阵的误差 $\ell(R, U, V)$ 构建统一的损失函数，通过梯度下降等方法学习到最终的用户和项目隐表示，从而对用户进行推荐。在不同研究中，可能同时利用了项目特征信息和用户特征信息，也可能只利用了其中一种。

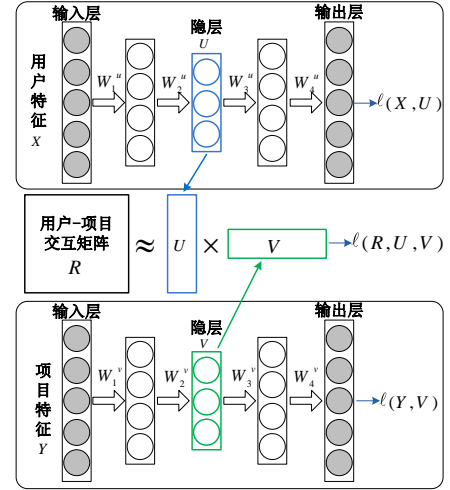


图 20 基于自编码器的混合推荐方法的模型架构

协同深度学习模型。Wang 等人在 KDD2015 年的论文^[4]中提出了一种贝叶斯版本的降噪自编码器模型，即贝叶斯栈式降噪自编码器模型（Bayesian SDAE），然后组合 Bayesian SDAE 和概率矩阵分解，提出了一种协同深度学习（Collaborative Deep Learning, CDL）混合推荐方法。该方法主要利用项目的文本类辅助数据，采用 Bayesian SDAE 学习项目的隐表示。CDL 模型的图表示如图 21 所示，假设系统有 I 个用户和 J 个项目，对应一个评分矩阵 $\mathbf{R} = [R_{ij}]_{I \times J}$ ，文本数据用词袋模型表示，所有项目的文本数据对应一个特征矩阵 \mathbf{X}_c （每一行代表一个项目的词频向量）。通过在 \mathbf{X}_c 中加入噪声得到自编码器的输入 \mathbf{X}_0 ，自编码器第 l 层的输出为 \mathbf{X}_l ，相应权值矩阵和偏置为 \mathbf{W}_l 和 \mathbf{b}_l 。CDL 的生成过程如下：

1、对 SDAE 网络的每一层 l

(a) 对权值矩阵 \mathbf{W}_l 的每一列 n ，提取

$$\mathbf{W}_{l, *n} \square N(\mathbf{0}, \lambda_w^{-1} \mathbf{I}_{K_l})$$

(b) 提取偏置向量 $\mathbf{b}_l \square N(\mathbf{0}, \lambda_w^{-1} \mathbf{I}_{K_l})$

(c) 对 \mathbf{X}_l 的每一行 j ，提取

$$\mathbf{X}_{l, j*} \square N(\sigma(\mathbf{X}_{l-1, j*} \mathbf{W}_l + \mathbf{b}_l), \lambda_s^{-1} \mathbf{I}_{K_l})$$

2、对每个项目 j

(a) 提取一个干净的输入

$$\mathbf{X}_{c,j^*} \sim N(\mathbf{X}_{L,j^*}, \lambda_n^{-1} \mathbf{I}_J)$$

(b) 提取一个隐项目偏移向量

$$\varepsilon_j \sim N(\mathbf{0}, \lambda_v^{-1} \mathbf{I}_K), \text{ 然后设置隐项目向量为:}$$

$$\mathbf{v}_j = \varepsilon_j + \mathbf{X}_{L/2,j^*}^T$$

3、对每个用户 i , 提取一个隐用户向量:

$$\mathbf{u}_i \sim N(\mathbf{0}, \lambda_u^{-1} \mathbf{I}_K)$$

4、对每对用户和项目 (i, j) , 提取一个评分 R_{ij}

$$R_{ij} \sim N(\mathbf{u}_i^T \mathbf{v}_j, \mathbf{C}_{ij}^{-1})$$

其中 λ_w 、 λ_n 、 λ_u 、 λ_s 和 λ_v 是超参数, \mathbf{C} 是一个置信度参数矩阵。

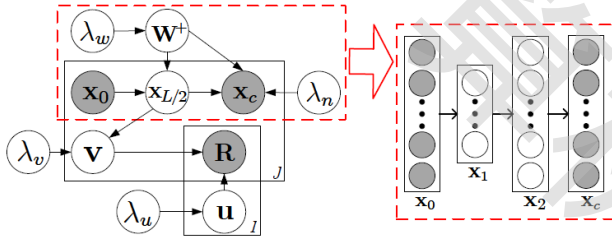


图 21 CDL 的模型结构

之后, 有研究者对 CDL 模型进行了一系列扩展。文献[39]中, Wang 等人将 CDL 扩展应用到标签推荐问题中, 提出一种关系栈式降噪自编码器模型 (Relational Stacked Denoising Autoencoder, RSDAE)。该研究采用 SDAE 从项目的内容信息中学习项目的隐表示, 利用概率矩阵分解建模标签与项目之间的共现信息, 同时融入了标签系统中可能存在的各类关系数据。CDL 模型中, 项目的文本辅助信息是由词袋模型表示, 这种方式忽略了文本内容中词语序列所包含的信息。针对这个问题, Wang 等人^[38]提出了一种协同循环自编码器模型 (Denoising Recurrent Autoencoder, CRAE), 用来代替 SDAE 进行项目推荐。CRAE 拥有一种自编码器结构, 在编码器和解码器中, 分别利用一个循环神经网络建模文本序列的生成, 从而抓住文本中词语的序列信息。另外一个 CDL 的扩展是协同变分自编码器方法 (Collaborative Variational Autoencoder, CVAE)^[42], CVAE 使用变分自编码器代替 SDAE 从项目的内容中学习项目的隐表示。相比 CDL, 推理

的过程更加简单, 且不需要在输入中加入噪声, 能够从图像、文本等项目内容中提取特征。此外, Ying 等人^[105]将 CDL 模型进行扩展, 提出一种协同深度排序模型 (Collaborative Deep Ranking, CDR), CDR 与 CDL 的区别在于, CDR 利用了成对损失函数代替 CDL 使用的逐点损失函数进行模型优化。

协同知识库嵌入模型。Zhang 等人^[5]利用一个与 CDL 相似的模型架构, 融入项目之间的结构化信息 (项目之间的各类关系) 和非结构化信息 (文本数据和图像数据) 到推荐系统中, 通过组合深度学习方法和概率矩阵分解模型, 提出了一种协同知识库嵌入模型 (Collaborative Knowledge Base Embedding, CKE)。CKE 利用了三种非监督表示学习方法从项目相关的结构化和非结构化数据中学习项目的隐表示。具体来说, 为了利用项目相关的结构化信息, 通过采用一种知识库嵌入式方法 TransR^[106]将项目嵌入到隐空间来学习项目的结构化向量, 利用项目相关的文本信息通过采用 SDAE 提取文本的分布式表示来得到项目的文本向量, 利用项目相关的图像信息通过提出一种栈式卷积自编码器 (Stacked Convolutional Autoencoders, SCAE) 模型提取视觉图像的分布式表示来得到项目的视觉向量, 并通过融合项目的三类向量得到项目的隐表示, 最后结合 PMF 学习到用户和项目最终的隐表示。与 CDL 相比, CKE 集成了更多的辅助信息来学习项目的隐表示, 另外, CKE 中的协同过滤利用了成对排序损失。

基于 SDAE 的方法。Wei 等人^[41]通过组合协同过滤方法 TimeSVD++^[107]和栈式降噪自编码器 SDAE, 提出了一种混合推荐方法。该研究利用 SDAE 从项目的辅助信息中学习项目的隐表示, 利用 TimeSVD++ 拟合用户和项目之间的评分矩阵。与 CDL 相比, 该模型利用的 TimeSVD++ 是一类融合了时间因素的隐因子模型, 因此建模了用户偏好和项目特征随时间的变化, 有利于提升推荐系统的性能。

基于 mDA 的方法。Li 等人^[17]通过将边缘降噪自编码器 (Marginalized Denoising Auto-encoder, mDA)^[108]和概率矩阵分解方法进行组合, 提出一种混合推荐方法。mDA 是降噪自编码器的一类扩展, 其通过对降噪自编码器的输入噪声进行边缘化, 避免了训练降噪自编码器所需要的大的计算开销, 提升了模型的可扩展性。与 CDL 和 CRAE 的不同之处在于, 该模型不仅考虑了项目相关的辅助

信息来学习项目的隐表示，还融入了用户相关的辅助数据来学习用户的隐表示，此外该模型并没有像 CDL 和 CRAE 一样采用贝叶斯形式，而是利用边缘降噪自编码器，具有更少的模型参数，具有更高的可扩展性。

基于 aSDAE 的方法。Dong 等人^[40]提出了一种基于附加栈式降噪自编码器（Additional Stacked Denoising Autoencoder, aSDAE）的混合推荐方法，aSDAE 是 SDAE 的扩展，其输入是用户或者项目的评分向量，aSDAE 在 SDAE 的每个隐层都加入用户或者项目的辅助信息，在输出端同时重构输入数据和辅助数据。该研究利用 aSDAE 学习用户和项目的隐表示，然后结合矩阵因子分解（MF）和 aSDAE 构建联合目标优化函数，采用梯度下降方法进行参数优化，进而对未知评分进行预测。相比 CDL 方法，该模型能够融入更加精细的辅助数据，且具有更少的模型参数。

4.3.2 基于其他深度学习模型的混合推荐方法

除了自编码器之外，目前被用于混合推荐的深度学习模型还包括多层感知机、卷积神经网络和循环神经网络等。虽然基于深度学习的混合推荐方法都采用深度学习加协同过滤的模型框架，但是自编码器从辅助信息中学习用户和项目隐表示的过程是一个非监督过程，因此在模型最终的优化过程中除了拟合用户与项目的显式反馈或隐式反馈数据之外，还融入了自编码器的重构误差，而其他监督深度学习方法将从辅助信息中学习用户和项目隐表示的过程和拟合协同信息的过程集成到一个统一的监督学习框架中，从而在最终的模型优化过程中只拟合用户与项目的显式反馈或隐式反馈数据。因此，与基于自编码器的混合推荐方法的最大区别在于最终优化目标函数的构建。图 22 给出了一个基本的模型架构，深度学习模型用来从用户特征 X 和项目特征 Y 中学习用户隐表示 U 或项目隐表示 V （此处并没有自编码器的输入数据重构过程），然后将隐表示融入到隐因子模型中拟合用户-项目交互矩阵 R （如评分矩阵），通过拟合交互矩阵的误差 $\ell(R, U, V)$ 构建统一的目标优化函数，利用最终学习到的用户和项目隐表示对用户产生推荐。

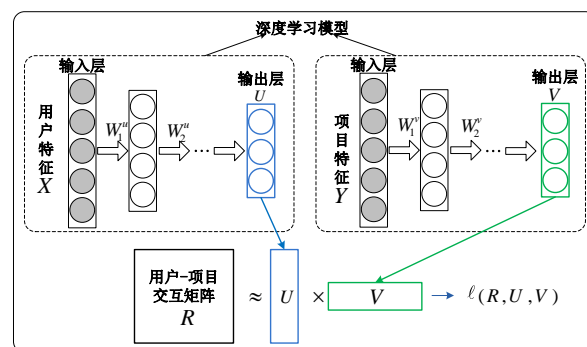


图 22 基于其他深度学习模型的混合推荐方法模型架构

基于 MLP 的混合推荐方法。传统矩阵分解方法通常采用内积的方式来衡量用户向量和项目向量的距离，Hsieh 等人^[15]利用测度学习来学习数据之间的相似度，并将测度学习（Metric Learning, ML）和协同过滤结合，提出了一种协同测度学习（Collaborative Metric Learning, CML）方法。该研究指出内积操作由于不具有距离相似度的传递性，并不是一个合理的距离测度，会带来相似度学习不准确的问题。因此，CML 通过大边界最近邻方法（Large Margin Nearest Neighbor, LMNN）进行测度学习来学习数据之间的相似度。最后，在测度学习损失函数的基础上，通过 MLP 融入项目相关的特征数据，同时将加权近似排序成对损失^[109]融入用户的隐式反馈数据，并加入正则化因子，构建统一的目标优化函数，最后学习到用户和项目隐向量，进而实现 Top-K 推荐。

Chen 等人^[110]在多媒体推荐中，通过引入两种层次的注意力机制到隐因子模型中，提出了一个注意力协同过滤模型（Attentive Collaborative Filtering, AFC）。AFC 是一个多层感知机，包括两种层次的注意力机制，一个项目层次的注意力机制被用来刻画不同项目匹配用户兴趣的程度，一个部件层次的注意力机制被用来抓住多媒体辅助信息中对用户最具有代表性的特征。

基于 CNN 的混合推荐方法。Kim 等人^[59]利用卷积神经网络将项目相关的文本辅助数据融入到推荐系统中，提出了一种卷积矩阵因子分解模型（Convolutional Matrix Factorization, ConvMF）。考虑到 CDL 等方法采用词袋模型对文本进行表示时，无法有效利用文本内部的情境信息，ConvMF 利用 CNN 的多层卷积操作抓住文本中词语之间的相互关联，并帮助学习用户和项目的隐表示，提升推荐系统评分预测的准确性。ConvMF 首先对文档中的每个词做嵌入式表示，然后将所有嵌入式向量拼接

成一个矩阵, 这样每篇文档就可用一个二维矩阵表示, 然后在该矩阵上进行卷积、池化以及映射, 从而获取到项目的隐向量。ConvM 的目标函数由矩阵因子分解和 CNN 的损失函数共同组成。

基于 RNN 的混合推荐方法。与 ConvMF 相同, 为了抓住文本辅助数据中词语之间的序列关系, Bansal 等人^[19]基于循环神经网络模型提出了一种文本推荐方法, 在利用评分数据的同时, 还利用了文本内容信息和文本标签数据进行推荐。具体来说, 为了更好地抓住文本中词语的序列模式, 该模型利用 GRU 学习文本内容的向量表示, 然后基于隐因子模型采用一个多任务学习框架(包括文本推荐和标签预测两个任务)构建一个联合优化目标函数, 在监督学习框架下实现对模型参数的优化。与 CDL 和 CRAE 相比, 虽然该模型也利用了项目相关的文本数据学习项目的隐向量表示, 但是该模型是一个完全端到端的监督学习架构。

Li 等人^[11]利用深度学习模型研究了 App 的评分预测与抽象 tips 生成问题, 通过一个多任务学习框架同时进行评分预测与抽象 tips 生成。具体地, 在评分预测任务中, 作者利用用户和项目的隐向量作为输入, 利用一个多层感知机 (MLP) 来进行评分预测。在抽象 tips 生成中, 将用户和项目的隐向量作为输入, 利用一个 GRU 来进行文本生成。最后在一个多任务框架中进行模型训练。

基于深度学习的混合推荐中, 深度学习主要用于从辅助数据中学习用户和项目的隐表示。当前的大部分研究都采用深度学习加隐因子模型的推荐框架, 各种研究的区别主要体现在辅助数据的类型、深度学习的方法和隐因子方法三个方面的选择不同, 表 1 总结了相关的代表性研究。

总的来说, 基于深度学习的混合推荐通过融入辅助信息能够有效缓解协同过滤推荐中存在的稀疏和冷启动问题。但是, 目前的大部分研究都是针对具体的辅助数据采用不同的深度学习模型, 构建一个针对所有数据的、统一的混合推荐框架是下一步的重要方向。

表 1 基于深度学习的混合推荐方法比较

研究	辅助数据类型	深度学习模型	隐因子模型
Wang <i>et al</i> [4]	文本	SDAE	PMF
Wang <i>et al</i> [38]	文本、关系	RSDAE	PMF
Wang <i>et al</i> [39]	文本	CRAE	PMF
Ying <i>et al</i> [105]	文本	SDAE	PMF
Li <i>et al</i> [42]	文本、图像等	VAE	PMF

Zhang <i>et al</i> [5]	文本、图像、关系等	SDAE、SCAE、TransR	PMF
Li <i>et al</i> [17]	用户和项目特征	mDA	PMF
Wei <i>et al</i> [41]	用户特征	SDAE	TimeSVD++
Dong <i>et al</i> [40]	用户和项目特征	aSDAE	MF
Hsieh <i>et al</i> [15]	图像、文本、标签等	MLP	ML
Chen <i>et al</i> [110]	图像、视频、音频等	Attention-based MLP	SVD
Kim <i>et al</i> [59]	文本	CNN	MF
Bansal <i>et al</i> [19]	文本	GRU	LFM
Li <i>et al</i> [111]	文本	MLP、GRU	MF

4.4 深度学习在基于社交网络的推荐系统中的应用

当前, 社交网络作为一种开放的公众参与的信息交流与业务服务平台, 迅速进入了人们的日常工作和生活中。同时, 随着移动互联网的发展, 位置社交网络也已经迅速普及。基于社交网络的推荐系统旨在解决社交媒体平台中“信息过载”的问题, 已经成为社交媒体挖掘领域的热门话题。目前, 深度学习与基于社交网络的推荐系统的结合也引发了一系列的研究成果, 主要可以分为两个方向: (1) 基于深度学习的社交网络社会化关系影响建模;

(2) 基于深度学习的位置社交网络序列模式建模。

4.4.1 基于深度学习的社交网络社会化关系影响建模

社交网络中用户之间存在各种类型的社会化关系(例如朋友关系、关注关系等), 用户之间通过社会化关系会产生相互影响, 并具有相似的兴趣偏好。基于社交网络的推荐系统最重要的是需要通过建模用户之间的社会化关系影响来提升推荐系统的质量。通常, 社会化关系通过图结构进行表示, 因此传统的方法一般采用图模型或正则化技术来建模用户之间的社会化关系影响^[112], 但是这些方法容易受到图结构的稀疏性以及高的计算复杂性的影响。当前, 针对这些问题, 研究者通过利用深度学习开展了一些针对性工作。

基于 MLP 的推荐方法。Wang 等人^[113]研究了跨信息网络和社交网络的社会化推荐问题, 即如何利用信息网络中用户与项目之间的交互和社交网络中用户之间的连接, 将信息网络中的项目推荐给社

交网络中的潜在用户。该研究通过分别在两个领域进行嵌入式学习，同时让桥接用户（同时出现在两个网络中的用户）共享相同的嵌入式表示，来学习用户和项目一致的隐表示。在信息网络领域，作者基于MLP，提出了一种属性感知的深度协同过滤模型（Attributed-aware Deep CF Model），模型架构如图23所示。模型将用户和项目的ID号以及它们的属性信息作为输入，通过多层神经网络预测用户对项目的偏好。在社交网络领域，基于信息网络中桥接用户的隐表示，通过假设朋友之间具有相近的兴趣来学习其他用户的隐表示，最后通过组合两个领域的损失函数构建统一的优化函数进行模型训练。

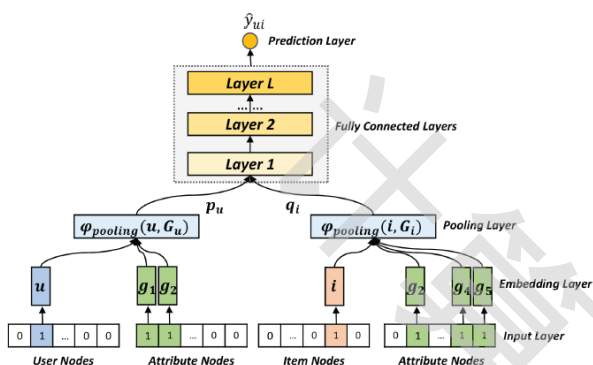


图 23 属性感知的深度协同过滤模型架构

基于分布式表示技术的推荐方法。Yang 等人^[74]为了融入社会化影响到推荐系统中，采用了基于深度学习的网络表示技术建模用户之间的社会化影响。基于用户的社交网络结构，通过网络表示技术将用户节点嵌入到一个共享的隐空间，每个用户用一个低维、稠密的向量进行表示。通过使用网络表示技术建模社交网络的生成，能够缓解社交网络图结构的稀疏性问题，同时基于深度学习的网络表示技术具有高的可扩展性，能够适应于大规模数据集。Zhao 等人^[114]通过组合深度学习与矩阵因子分解，研究如何利用社交网络数据为社交网络中的冷启动用户（在电子商务网站中没有购买记录的用户）推荐电子商务网站中的产品。首先，通过利用神经网络模型（即 Word2Vec 和 Para2Vec）从用户购买行为数据中学习用户和项目的嵌入式表示，同时为了融入社交网络信息，利用改进的梯度 Boost 树方法从社交媒体用户数据中学习用户的嵌入表示，最后，利用基于特征的矩阵因子分解联合以上两个方面的隐表示进行模型训练和预测。

基于自编码器的推荐方法。Deng 等人^[115]针对传统矩阵因子分解中用户和项目的隐特征初始化困难以及用户信任关系多样化的问题，采用深度自

编码器模型初始化用户和项目的隐特征向量，通过正则化技术融入用户信任关系的影响到推荐系统中，有效提升了推荐系统的性能。Pan 等人^[116]研究了如何融入信任关系到评分预测问题中，作者采用降噪自编码器分布通过重构用户的评分向量和信任关系向量学习评分隐向量和信任隐向量，中间通过一个加权隐层来平衡这两类隐向量的重要性，基于学习到的隐向量来进行 top-N 推荐。此外，文章还利用了一个关联正则项来避免由于数据稀疏所导致的过拟合问题

基于 RBM 的推荐方法。Nguyen 等人^[50]在 Salakhutdinov 等人^[46]的工作的基础上，通过两种方式建模社会化关系对推荐结果的影响，一种是将社会化关系和用户评分一样，都作为 RBM 的观测变量，通过同时拟合社会化关系和用户评分实现社会化影响建模；另一种是在 RBM 中增加一个隐含层，在这个隐层上通过朋友间的权值共享实现社会化影响建模。

通过深度学习方法将社交网络中的社会化关系影响融入到推荐系统中，有利于缓解推荐系统的数据稀疏和冷启动问题，从而提升推荐系统的质量。但总的来说，深度学习在基于社交网络的推荐系统中的应用还非常少，诸如如何区分关系的类型和强度、如何有效建模结构洞等社会理论到推荐系统中，都还有待深入的研究。

4.4.2 基于深度学习的位置社交网络序列模式建模

位置社交网络中，所有项目（也就是兴趣点，Point-of-interest, POI）都具有位置属性，用户行为具有时间和空间上的序列模式，对这种时空序列模式的建模有利于提升 POI 推荐的准确率。因此，位置社交网络的推荐系统除了需要关注用户之间的社会化关系，还需要抓住用户的位置影响、序列移动模式等因素。

基于MLP的推荐方法。Yang 等人^[117]指出当前的 POI 推荐方法都是针对特定的数据和问题而设计，提出了一种通用的半监督学习模型，即偏好与情境嵌入模型 (Preference and Context Embedding, PACE)，能够利用相邻用户和位置的信息来缓解推荐系统的数据稀疏问题。PACE 首先构建情境图作为观测数据，包括利用位置间的距离构建的位置图，以及利用用户间的朋友关系构建的用户图。PACE 的模型架构如图24所示，其输入是用户和项目的 one-hot 向量，输出是用户和项目的情境信息以及用户对 POI 的访问次数，分别建模了用户与用户

之间的情境、项目与项目之间的情境,以及用户对 POI 的偏好。最后,PACE 联合三个方面的损失建立优化函数,通过学习到的模型进行 POI 推荐。

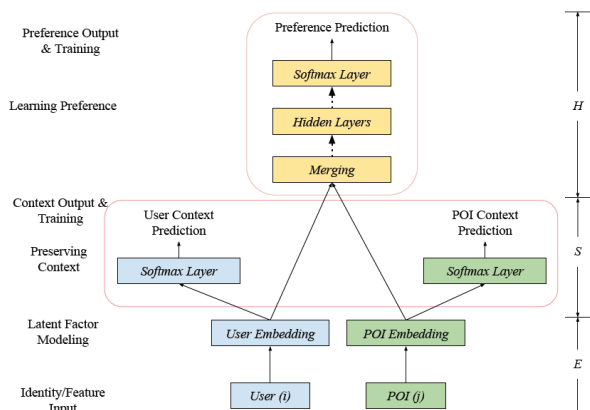


图 24 偏好与情境嵌入模型架构

基于 CNN 的推荐方法。Wang 等人^[62]研究了如何融入图像内容信息来提升 POI 推荐的问题,作者基于 CNN 和概率矩阵因子分解,提出了一种视觉内容增强的 POI 推荐模型 (Visual Content Enhanced POI Recommendation, VPOI)。VPOI 利用 CNN 从图像内容中提取特征,通过用户-图像关系、POI-图像关系和用户-POI 关系三类关系,构建了一个概率主题模型,最后将图像的特征提取和概率主题模型融合到一个统一框架中构建优化函数,采用 Negative Sampling 方法进行参数优化。

基于 RNN 的推荐方法。Liu 等人^[72]研究了位置社交网络中的行为预测问题,通过利用循环神经网络抓住序列行为之间的依赖关系,从而基于用户的历史行为序列帮助预测下一时刻的行为。Yang 等人^[74]利用循环神经网络建模位置社交网络中的用户序列移动模式,同时融合用户的社会化关系影响进行位置推荐。具体地,模型由两个部分组成:社会网络的构建和移动轨迹的生成。首先采用网络嵌入方法构建社会网络,再利用用户的访问偏好、用户朋友的影响、用户的短序列情境和长序列情境共同建模行为序列的生成。为了刻画两种情境信息,在短时和长时两个情境下分别采用 RNN 和 GRU 抓住用户移动模式中的序列关联性。最后组合社会网络的构建和移动轨迹的生成建立目标函数,采用随机梯度下降法进行模型训练。

基于分布式表示的推荐方法。Ozsoy 等人^[118]利用用户与项目之间的共现信息,采用 Word2Vec 技术开展位置社交网络推荐问题研究。首先,通过 Word2Vec 技术利用用户的历史签到行为将用户和项目嵌入到一个共享的低维空间,然后,借鉴基于

内容的推荐方法和基于用户的协同过滤方法,采用最近邻算法找到用户最感兴趣的项目,并产生用户的项目推荐。Zhao 等人^[119]考虑建模用户的兴趣偏好和移动模式来进行兴趣点推荐,提出了一种序列嵌入排序模型 (SEquential Embedding Rank, SEER) 模型。具体来说,SEER 利用分布式表示技术学习用户的嵌入式表示,然后将用户嵌入作为约束加入到成对排序模型中来抓住用户行为的序列模式,同时 SEER 还融入了时间和空间信息。

通过深度学习方法将用户行为序列模式融入到推荐系统中,有利于抓住用户行为之间的依赖关系,从而提升推荐系统的质量。深度学习在位置社交网络推荐系统中的应用目前主要聚焦在序列模式建模,其研究还有待进一步拓展,例如,如何通过深度学习融合各类异构的时空数据,以及如何集成社会化关系影响、位置影响、时间影响和序列模式等要素构建统一的推荐框架等。

4.5 深度学习在情境感知的推荐系统中的应用

情境感知的推荐系统主要通过集成用户的情境信息到推荐过程中,从而改善推荐效果。在情境感知的推荐系统中,因为考虑了情境信息,传统的推荐函数从 $s: U \times I \rightarrow R$ 变为 $s: U \times I \times C \rightarrow R$, 其中 C 表示所有情境的集合。当前,深度学习在情境感知的推荐系统中的应用主要集中在如何采用深度学习方法对情境信息进行有效建模,主要表现在两个方面:(1) 如何采用深度学习方法融入情境信息到推荐系统中,即基于深度学习的情境感知的推荐;(2) 如何采用深度学习方法实现对情境信息的有效表示,即基于深度学习的情境信息表示。

基于深度学习的情境感知的推荐。传统的情境感知的推荐系统通过将问题转化为一个三维矩阵上的矩阵补全问题,能够简单地融入情境信息到推荐系统中。但是对于很多复杂场景下的推荐问题(例如序列推荐问题),融入情境信息到推荐系统中是很困难的^[120]。Liu 等人^[72,120]考虑到传统的序列推荐除了需要考虑用户的序列影响,还需要考虑用户行为的情境信息,通过改进传统的循环神经网络结构融入了情境信息到序列推荐中。在[72]中,作者提出了一种空时循环神经网络模型 (Spatial Temporal Recurrent Neural Network, ST-RNN), 针对不同粒度的时间间隔和距离长度,通过在循环神经网络模型中定义不同的时间转换矩阵和位置转换矩阵,有效地抓住了连续的时间和空间情境信息对用户行为的影响,提升了位置推荐的质量。在

[120]中,作者将情境信息分为输入情境(如位置、天气等)和转换情境(时间间隔长度),通过在循环神经网络中采用可适应的、情境具体的输入矩阵和转换矩阵对两种情境信息进行建模,帮助提升了序列推荐的质量。

基于深度学习的情境信息表示。传统的情境感知的推荐系统通常采用图模型、主题模型和张量分解技术等,但是,当情境数据大量增加时,这些技术会因为情境集合维度过大而遭遇到严重的数据稀疏问题。深度学习通过将各种类型的情境数据进行建模,并表示为一个低维、稠密的隐向量,能够有效缓解这种数据稀疏问题^[121]。

Zhou等人^[122]将分布式表示技术用于情境信息建模,提出一种多情境轨迹嵌入模型(Multi-Context Trajectory Embedding Model, MC-TEM),通过将包括用户、轨迹、临近的地点、类别、时间、区域在内的情境信息嵌入到一个统一的隐空间中,能够抓住多种类型的情境信息用于轨迹数据挖掘,此外,分布式表示技术通过刻画相关因素之间的距离以及联系,可以增加模型的可解释性。Huang等人^[123]提出了一种神经概率模型用于基于情境的引用推荐。首先用一个词语序列来表示情境,然后利用词分布式表示学习方法学习情境和文档的向量表示,最后通过神经概率模型联合情境和文档的向量表示来产生文章推荐。Fang等人^[124]提出了一种Encoder-Decoder结构来解决情境感知的引用推荐问题,在解码器中,作者采用了一个基于CNN的深度模型来学习情境数据的隐表示,在解码器中,作者利用一个RNN来生成引用的文章标题,同时利用作者的元数据构建了一个作者网络来抓住作者的自身偏好,采用注意力机制抓住不同部分输入对生成文章标题的重要性。Unger等人^[121]将情境信息分为清晰情境信息和隐情境信息,首先采用深度自编码器从各种类型的、粗糙的情境数据中提取隐情境表示,然后采用隐因子模型融合隐情境表示、清晰的情境信息和用户对项目评分数据构建一个评分模型来进行评分预测,通过融入隐情境信息提升了推荐系统的质量。Rawat等人^[60]利用深度神经网络模型研究了融合用户情境信息的图像标签推荐问题,同时考虑了用户情境信息和图像语义信息。主要思路是采用深度卷积神经网络学习图像的语义表示,采用一个深度神经网络学习用户的情境表示,最后联合两方面的表示,将标签推荐问题作为

一个多标记分类问题进行标签推荐。

总的来说,深度学习应用于情境感知的推荐系统中。一方面,在很多复杂推荐场景中,通过深度学习方法能够有效融入情境信息到推荐系统中;另一方面,通过深度学习获取情境信息的隐表示,并在此基础上进行情境感知的推荐,能够有效融入各类粗糙的情境数据,缓解情境感知的推荐系统中的数据稀疏问题。但同时,由于情境数据被表示为隐情境,模型面临可解释性不足的问题。

4.6 应用比较

基于深度学习的推荐方法能够融入多源异构数据进行推荐,包括用户的显式反馈或隐式反馈数据、用户画像和项目内容数据、用户生成内容等。深度学习方法通过将多源异构数据作为输入从而采用一个端到端的模式自动训练预测模型,能够有效融入多源异构数据到推荐系统中,从而缓解传统推荐系统中面临的数据稀疏和冷启动问题,并提升推荐系统的能力。其主要的优势表现在:(1)可以避免复杂的人工特征工程^[5,53,61],尤其是面对图像、视频等非结构化数据,深度学习的表示学习能力能够帮助从非结构化数据中提取特征信息;(2)可以学习非线性的多层次抽象特征表示,而且获取的特征通常是稠密和低维的,这是传统浅层模型所不具备的^[9,29-33];(3)可以克服不同数据之间的异构性,将各类粗糙的原始数据作为输入来学习用户和项目的隐表示^[5,8,17,80]。但同时也存在一些不足:

(1)可扩展性问题。当前,虽然深度学习在Youtube视频推荐^[8]、Google的App推荐^[9]和Yahoo的新闻推荐^[94]等实际场景中得到了应用,但由于其模型复杂往往需要长的训练时间,如何平衡模型可扩展性和复杂度仍然是一个大的问题;(2)可解释性问题。深度学习模型往往类似于一个黑盒,容易带来推荐系统可解释性不足的问题,如何在增强推荐能力的同时增强可解释性仍然有待深入研究。

总的来说,基于深度学习的推荐系统利用深度学习方法学习用户和项目的隐表示,从而实现项目推荐,但是不同类型的方法在深度学习模型、数据类型以及推荐对象等方面存在着差异。表2列出了5个主要研究方向在深度学习模型、数据类型方面的区别,以及各自的优点和面临的难点问题。

表2 深度学习在不同推荐系统中的应用比较

应用方向	深度学习模型	数据类型	主要优点	主要难点
深度学习在基于内容的推荐系统中的应用	MLP、CNN、RNN、DBN等	用户的显式反馈或隐式反馈数据、用户画像和项目内容数据,以及各种类型的用户生成内容	1、自动的用户和项目特征提取 2、不存在新项目的冷启动问题 3、高的可扩展性	1、新用户冷启动问题 2、不能为用户发现多样的感兴趣的资源
深度学习在协同过滤中的应用	RBM、AE、CNN、NADE、RNN、GAN等	用户与项目之间的显式反馈或隐式反馈数据(如评分矩阵)	1、系统自动化程度高 2、推荐结果丰富 3、能够实现非线性特征建模	1、数据稀疏问题 2、新用户和新项目冷启动问题
深度学习在混合推荐中的应用	AE、CNN、RNN、MLP等	用户的显式反馈或隐式反馈数据、用户画像和项目内容数据,以及各种类型的用户生成内容	1、不存在冷启动问题 2、不受数据稀疏问题约束 3、有效利用辅助数据提升推荐性能	1、异构数据的有效特征表达 2、各类推荐算法的混合
深度学习在基于社交网络的推荐系统中的应用	RBM、AE、分布式表示技术、CNN、RNN、MLP等	用户与项目之间的显式反馈或隐式反馈数据、用户的社会化关系数据,以及时间、位置等情境数据	1、缓解推荐系统的数据稀疏问题 2、缓解推荐系统的冷启动问题 3、能够建模社会化关系影响和用户行为序列模式	1、探索更多基于深度学习的模型架构来有效建模社会化关系影响 2、位置社交网络中,如何同时融合用户移动模式、社会化影响、时间影响、位置影响等多种要素
深度学习在情境感知的推荐系统中的应用	分布式表示技术、AE、RNN、MLP等	用户与项目之间的显式反馈或隐式反馈数据,用户的情境信息(时间、地点、天气等)	1、能够融合各种类型的情境数据 2、能够实现情境数据的有效表示	1、确定不同情境信息的作用程度 2、提升情境模型的可解释性

5 基于深度学习的推荐系统研究趋势展望

随着大数据时代的不断深入,深度学习在推荐系统中的应用已经受到学术界和工业界越来越多的重视,基于深度学习的推荐系统研究已经成为当前的研究热点。但是基于上面的讨论可以看到,目前深度学习在推荐系统中的应用仍处于起步阶段,在未来必将会有更多、更广泛的尝试^[125]。以下总结了五个可能的研究方向。

(1) 深度学习与现有推荐方法的结合

传统的推荐方法,包括基于内容的方法和协同过滤方法,都采用浅层模型进行预测,依赖于人工特征提取,很难有效学习到深层次的用户和项目隐表示。通过利用深度学习模型融合广泛的多源异构数据,包括社会化关系、用户或项目属性、以及用户的评论和标签信息等,能够学习到更加抽象、更加稠密的用户和项目的深层次表示,同时采用深层神经网络结构构建预测模型也能够更好地抓住用户和项目之间交互的非线性结构特征^[15]。但同时,传统的推荐方法,具有简单、可解释性强等优势。

因此,将深度学习与现有推荐方法结合,能够融合两种方法的优势,虽然目前已有相关研究出现,但这个方向仍值得更多的研究者加以关注。

(2) 基于深度学习的跨领域信息融合的推荐

随着数据获取能力的不断提升,用户在不同领域的历史记录或项目在不同领域的信息能够被获取。例如,一个用户可能在多个社交媒体平台上注册账号,融合用户在不同平台上的数据能够进行跨领域信息融合的推荐,帮助克服单一领域信息的不足,从而有效缓解传统推荐系统中的数据稀疏和冷启动问题,同时利用多个领域数据能够更好地发现用户的个性化偏好。针对跨领域推荐问题的研究,当前最主要的研究方法包括基于协同过滤的方法^[126]、基于迁移学习的方法^[127]和基于张量分解的方法^[128]等。但是,这些方法都只针对不同领域中特定类型的信息进行融合,适应性非常有限。当前,利用深度学习技术,通过将各类数据通过嵌入式表示等方法作为统一输入,构建深层预测模型能够有效融合各种不同类型的、跨平台的异构数据进行推荐,已经在Google^[8]和微软^[10]等互联网公司的实际系统中被应用。未来,通过构建深度学习模型来实现跨领域信息融合的推荐将是学术界和工业界研究的重点方向。

(3) 注意力机制与基于深度学习的推荐系统的结合

基于注意力机制的深度学习是人类视觉中的选择注意力机制与深度神经网络的结合，目前在计算机视觉、自然语言处理等领域取得了巨大成功。当前，注意力机制已经被应用于MLP、RNN、CNN和其他深度学习模型，其中最为引人关注的是基于注意力的RNN和基于注意力的CNN。基于注意力的RNN能够更好地建模序列数据中的长期记忆，基于注意力的CNN能够从输入中识别与问题最具有信息量的部分。将注意力机制应用到在推荐系统中，能够帮助推荐系统抓住项目中最具有信息量的特征，推荐最具有代表性的项目，同时增强模型的可解释性^[110]。当前，注意力机制已经在hashtag推荐^[87,88,91,92]、文章推荐^[90]、多媒体推荐^[110]、引用推荐^[124]等问题中得到了应用。但是总的来说，目前的研究还比较少，未来还待更深入和更广泛的研究。

(4) 新的深度学习推荐系统架构

对于推荐系统来说，涉及到不同的推荐对象和推荐场景，例如电影推荐、音乐推荐、图像推荐、商品推荐、地理位置推荐等等。一方面，针对所有任务构建统一的深度学习推荐模型几乎是不可能的，需要根据不同的推荐场景考虑不同的数据构建新的深度学习框架来产生推荐，包括推荐项目的具体内容信息，推荐系统中涉及的辅助数据（评论、标签、用户画像信息、用户的社会化关系等），以及推荐的情景信息（时间、位置等）等，因此，面对新的推荐场景需要设计新的深度学习推荐系统架构。另一方面，当前的推荐系统需要建模的要素众多，不仅仅包括用户与项目之间的交互数据，还涉及到用户行为的时空序列模式、社会化关系影响、用户偏好的动态演化和项目特征的动态变化等，建模更多的要素能够提升推荐系统的性能。因此，研究能够表达和融合多种要素的、新的深度学习架构也是未来的研究方向之一。

(5) 基于深度学习的推荐系统的可解释性

推荐系统除了直接展示推荐结果之外，往往还要展示恰当的推荐理由来告诉用户为什么系统认为这样的推荐是合理的。提升推荐系统的可解释性可以提高用户对推荐结果的接受度，同时也可以提高用户在系统透明度、可信度、可辨性、有效性和满意度等方面的体验。已有的推荐方法使用主题模型学习到的话题^[104]以及显式的物品特征^[129]来加强可解释性。但是，基于深度学习推荐系统采用端到

端的模型直接将多源异构数据作为输入预测用户对项目的偏好，模型训练的结果是给出深度神经网络的结构和神经元之间的连接权重，很难对推荐结果直接给出合理的解释。因此，有必要从数据、模型和经济意义等层面上进行研究，提升基于深度学习的推荐系统的可解释性。

6 结束语

在互联网迅猛发展的今天，随着互联网用户对信息需求的日益膨胀，“信息过载”问题逐年升温，推荐系统在各个领域的数字化进程中扮演着越来越重要的角色。将深度学习技术融入推荐系统中，开展基于深度学习的推荐系统研究，通过从海量数据中学习用户和项目的隐表示，然后构建推荐模型，最终向用户产生有效的推荐列表。与传统的推荐系统相比，基于深度学习的推荐系统能够利用深度学习技术通过融合各种类型的多源异构数据，自动学习用户和项目抽象的隐藏特征，建模用户行为中的序列模式，能够更有效地反映用户的不同偏好以及提高推荐的准确性。本文在分析传统推荐算法所存在问题的基础上，介绍和分析了基于深度学习的推荐系统的研究现状和进展，并讨论了今后的发展方向，希望能对相关领域的研究人员和工程技术人员提供有益的帮助。

参考文献

- [1] Marz N, Warren J. Big Data: Principles and best practices of scalable realtime data systems. Greenwich, USA: Manning Publications Co, 2015
- [2] Gantz J, Reinsel D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future, 2012, 2007(2012): 1-16
- [3] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749
- [4] Wang H, Wang N, Yeung D Y. Collaborative deep learning for recommender systems//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015: 1235-1244
- [5] Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems// Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 353-362

- [6] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436-444
- [7] Peng Y, Zhu W, Zhao Y, et al. Cross-media analysis and reasoning: advances and directions. *Frontiers of Information Technology & Electronic Engineering*, 2017, 18(1): 44-57
- [8] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations//*Proceedings of the 10th ACM Conference on Recommender Systems*. Boston, USA, 2016: 191-198
- [9] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems//*Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. Boston, USA, 2016: 7-10
- [10] Wu C Y, Ahmed A, Beutel A, et al. Recurrent recommender networks// *Proceedings of the 10th ACM International Conference on Web Search and Data Mining*. Cambridge, UK, 2017: 495-503
- [11] Elkahky A M, Song Y, He X. A multi-view deep learning approach for cross domain user modeling in recommendation systems//*Proceedings of the 24th International Conference on World Wide Web*. Florence, Italy, 2015: 278-288
- [12] Song Y, Elkahky A M, He X. Multi-rate deep learning for temporal recommendation//*Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. Pisa, Italy, 2016: 909-912
- [13] Vasile F, Smirnova E, Conneau A. Meta-Prod2Vec: Product embeddings using side-information for recommendation// *Proceedings of the ACM Conference on Recommender Systems*. Boston, USA, 2016: 225-232
- [14] Grbovic M, Radosavljevic V, Djuric N, et al. E-commerce in your inbox: product recommendations at scale//*Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Sydney, Australia, 2015: 1809-1818
- [15] Hsieh C K, Yang L, Cui Y, et al. Collaborative metric learning//*Proceedings of the 26th International Conference on World Wide Web*. Perth, Australia, 2017: 193-201
- [16] Roy S, Guntuku S C. Latent factor representations for cold-start video recommendation//*Proceedings of the 10th ACM Conference on Recommender Systems*. Boston, USA, 2016: 99-106
- [17] Li S, Kawale J, Fu Y. Deep collaborative filtering via marginalized denoising auto-encoder//*Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. Melbourne, Australia, 2015: 811-820
- [18] Zheng L, Noroozi V, Yu P S. Joint deep modeling of users and items using reviews for recommendation//*Proceedings of the 10th ACM International Conference on Web Search and Data Mining*. Cambridge, UK, 2017: 425-434
- [19] Bansal T, Belanger D, McCallum A. Ask the GRU: Multi-task learning for deep text recommendations//*Proceedings of the 10th ACM Conference on Recommender Systems*. Boston, USA, 2016: 107-114
- [20] Wu Y, DuBois C, Zheng A X, et al. Collaborative denoising auto-encoders for top-n recommender systems//*Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. San Francisco, USA, 2016: 153-162
- [21] Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 2009: 4
- [22] Verbert K, Manouselis N, Ochoa X, et al. Context-aware recommender systems for learning: a survey and future challenges. *IEEE Transactions on Learning Technologies*, 2012, 5(4): 318-335
- [23] Mooney R J, Roy L. Content-based book recommending using learning for text categorization//*Proceedings of the 5th ACM Conference on Digital libraries*. San Antonio, USA, 2000: 195-204
- [24] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering//*Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*. Madison, USA, 1998: 43-52
- [25] Balabanović M, Shoham Y. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 1997, 40(3): 66-72
- [26] Schafer J B, Konstan J, Riedl J. Recommender systems in e-commerce//*Proceedings of the 1st ACM Conference on Electronic Commerce*. Denver, USA, 1999: 158-166
- [27] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*, 2016, 529(7587): 484-489
- [28] Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016
- [29] Wang R, Fu B, Fu G, et al. Deep & cross network for ad click predictions. *arXiv preprint arXiv:1708.05123*, 2017
- [30] Shan Y, Hoens T R, Jiao J, et al. Deep Crossing: Web-scale modeling without manually crafted combinatorial features//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, USA, 2016: 255-262
- [31] Zhu J, Shan Y, Mao J C, et al. Deep embedding forest: forest-based serving with deep embedding features. *arXiv preprint arXiv:1703.05291*, 2017
- [32] Guo H, Tang R, Ye Y, et al. DeepFM: A factorization-machine based neural network for ctr prediction//*Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, Australia, 2017:1725-1731
- [33] Zhou G, Song C, Zhu X, et al. Deep interest network for click-through rate prediction. *arXiv preprint arXiv:1706.06978*, 2017
- [34] Williams D, Hinton G. Learning representations by back-propagating errors. *Nature*, 1986, 323(6088): 533-538
- [35] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks//*Proceedings of the Advances in Neural Information Processing Systems*. Vancouver, Canada, 2007, 19: 153
- [36] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders//*Proceedings of the 25th International Conference on Machine Learning*. Helsinki, Finland,

- 2008: 1096-1103
- [37] Sedhain S, Menon A K, Sanner S, et al. Autorec: Autoencoders meet collaborative filtering//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015: 111-112
- [38] Wang H, Shi X, Yeung D Y. Collaborative recurrent autoencoder: Recommend while learning to fill in the blanks// Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 415-423
- [39] Wang H, Shi X, Yeung D Y. Relational stacked denoising autoencoder for tag recommendation//Proceedings of the 29th Conference on Artificial Intelligence. Austin, USA, 2015: 3052-3058
- [40] Dong X, Yu L, Wu Z, et al. A hybrid collaborative filtering model with deep structure for recommender systems//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017: 1309-1315
- [41] Wei J, He J, Chen K, et al. Collaborative filtering and deep learning based recommendation system for cold start items. Expert Systems with Applications, 2017, 69: 29-39
- [42] Li X, She J. Collaborative Variational Autoencoder for Recommender Systems//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017: 305-314
- [43] Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines. Massachusetts, USA: MIT Press, 1986
- [44] Smolensky P. Information processing in dynamical systems: Foundations of harmony theory[Ph.D. dissertation]. Colorado University at Boulder Department of Computer science, Colorado, USA, 1986
- [45] Hinton G E. A practical guide to training restricted Boltzmann machines. Momentum, 2010, 9(1): 926
- [46] Salakhutdinov R, Mnih A, Hinton G. Restricted boltzmann machines for collaborative filtering//Proceedings of the 24th International Conference on Machine Learning. Corvallis, USA, 2007: 791-798
- [47] Phung D Q, Venkatesh S. Ordinal Boltzmann machines for collaborative filtering//Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial Intelligence. Virginia, USA, 2009: 548-556
- [48] Georgiev K, Nakov P. A non-iid framework for collaborative filtering with restricted boltzmann machines//Proceedings of the 30th International Conference on Machine Learning. Atlanta, USA, 2013: 1148-1156
- [49] HE Jie-Yue, Ma Be. Based-on real-valued conditional Restricted Boltzmann Machine and social network for collaborative filtering. Chinese Journal of Computers, 2016, 39(1): 183-195(in Chinese)
(何洁月, 马贝. 利用社交关系的实值条件受限玻尔兹曼机协同过滤推荐算法. 计算机学报, 2016, 39(1): 183-195)
- [50] Nguyen T T, Lauw H W. Representation learning for homophilic preferences//Proceedings of the 10th ACM Conference on Recommender Systems. Boston, USA, 2016: 317-324
- [51] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. Neural Computation, 2006, 18(7): 1527-1554
- [52] Bengio Y. Learning deep architectures for AI. Foundations and Trends® in Machine Learning, 2009, 2(1): 1-127
- [53] Wang X, Wang Y. Improving content-based and hybrid music recommendation using deep learning//Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA, 2014: 627-636
- [54] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition. Neural Computation, 1989, 1(4): 541-551
- [55] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324
- [56] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097-1105
- [57] Van den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2013: 2643-2651
- [58] Geng X, Zhang H, Bian J, et al. Learning image and user features for recommendation in social networks//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 4274-4282
- [59] Kim D, Park C, Oh J, et al. Convolutional matrix factorization for document context-aware recommendation//Proceedings of the 10th ACM Conference on Recommender Systems. Boston, USA, 2016: 233-240
- [60] Rawat Y S, Kankanhalli M S. ConTagNet: Exploiting User Context for Image Tag Recommendation//Proceedings of the 2016 ACM on Multimedia Conference. Amsterdam, Netherland, 2016:1102-1106
- [61] Lei C, Liu D, Li W, et al. Comparative deep learning of hybrid representations for image recommendations//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2545-2553
- [62] Wang S, Wang Y, Tang J, et al. What your images reveal: Exploiting visual contents for point-of-interest recommendation//Proceedings of the 26th International Conference on World Wide Web. Western Australia, 2017: 391-400
- [63] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks// Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014: 3104-3112
- [64] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks//Proceedings of the 31th International Conference on Machine Learning. Beijing, China, 2014:1764-1772
- [65] Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions//Proceedings of the IEEE Conference

- on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3128-3137
- [66] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735-1780
- [67] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014
- [68] Weston J, Chopra S, Bordes A. Memory networks. *arXiv preprint arXiv:1410.3916*, 2014
- [69] Joulin A, Mikolov T. Inferring algorithmic patterns with stack-augmented recurrent nets//*Proceedings of the Advances in Neural Information Processing Systems*. Montreal, Canada, 2015: 190-198
- [70] Graves A, Wayne G, Danihelka I. Neural Turing machines. *arXiv preprint arXiv:1410.5401*, 2014
- [71] Graves A, Wayne G, Reynolds M, et al. Hybrid computing using a neural network with dynamic external memory. *Nature*, 2016, 538(7626): 471-476
- [72] Liu Q, Wu S, Wang L, et al. Predicting the next location: A recurrent model with spatial and temporal contexts//*Proceedings of the 30th AAAI Conference on Artificial Intelligence*. Phoenix, USA, 2016: 194-200
- [73] Liu Q, Wu S, Wang L. Multi-behavioral sequential prediction with recurrent log-bilinear model. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(6): 1254-1267
- [74] Yang C, Sun M, Zhao W X, et al. A neural network approach to jointly modeling social networks and mobile trajectories. *ACM Transactions on Information Systems (TOIS)*, 2017, 35(4): 36
- [75] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*, 2015
- [76] Tan Y K, Xu X, Liu Y. Improved recurrent neural networks for session-based recommendations//*Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. Boston, USA, 2016: 17-22
- [77] Wu C, Wang J, Liu J, et al. Recurrent neural network based recommendation for time heterogeneous feedback. *Knowledge-Based Systems*, 2016, 109: 90-103
- [78] Dai H, Wang Y, Trivedi R, et al. Recurrent coevolutionary feature embedding processes for recommendation//*Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. Boston, USA, 2016: 29-34
- [79] Huang P S, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data//*Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*. San Francisco, USA, 2013: 2333-2338
- [80] Xu Z, Chen C, Lukasiewicz T, et al. Tag-aware personalized recommendation using a deep-semantic similarity model with negative sampling//*Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. Indianapolis, USA, 2016: 1921-1924
- [81] Chen C, Meng X, Xu Z, et al. Location-aware personalized news recommendation with deep semantic analysis. *IEEE Access*, 2017, 5(99): 1624-1638
- [82] He X, Liao L, Zhang H, et al. Neural collaborative filtering//*Proceedings of the International Conference on World Wide Web*. Perth, Australia, 2017: 173-182
- [83] Rendle S. Factorization machines//*Proceedings of the 2010 IEEE 10th International Conference on Data Mining*. Sydney, Australia, 2010: 995-1000
- [84] Qu Y, Cai H, Ren K, et al. Product-based neural networks for user response prediction//*Proceedings of the 2016 IEEE 16th International Conference on Data Mining*. Barcelona, Spain, 2016: 1149-1154
- [85] He X, Chua T S. Neural Factorization Machines for Sparse Predictive Analytics//*Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Tokyo, Japan, 2017: 355-364
- [86] Xiao J, Ye H, He X, et al. Attentional factorization machines: Learning the weight of feature interactions via attention networks//*Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, Australia, 2017: 3119-3125
- [87] Gong Y, Zhang Q. Hashtag recommendation using attention-based convolutional neural network//*Proceedings of the 25th International Joint Conference on Artificial Intelligence*. New York, USA, 2016: 2782-2788
- [88] Zhang Q, Wang J, Huang H, et al. Hashtag recommendation for multimodal microblog using co-attention network//*Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, Australia, 2017: 3420-3426
- [89] Seo S, Huang J, Yang H, et al. Representation Learning of Users and Items for Review Rating Prediction Using Attention-based Convolutional Neural Network//*Proceedings of the 3rd International Workshop on Machine Learning Methods for Recommender Systems (SDM'17)*. Huston, USA, 2017
- [90] Wang X, Yu L, Ren K, et al. Dynamic Attention Deep Model for Article Recommendation by Learning Human Editors' Demonstration//*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Halifax, Canada, 2017: 2051-2059
- [91] Li Y, Liu T, Jiang J, et al. Hashtag recommendation with topical attention-based LSTM//*Proceedings of the 26th International Conference on Computational Linguistics*. Osaka, Japan, 2016: 943-952
- [92] Huang H, Zhang Q, Gong Y, et al. Hashtag Recommendation Using End-To-End Memory Networks with Hierarchical Attention//*Proceedings of the 26th International Conference on Computational Linguistics*. Osaka, Japan, 2016: 943-952
- [93] Huang H, Zhang Q, Huang X, et al. Mention recommendation for

- twitter with end-to-end memory network//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia, 2017: 1872-1878
- [94] Okura S, Tagami Y, Ono S, et al. Embedding-based News Recommendation for Millions of Users//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017: 1933-1942
- [95] Strub F, Mary J. Collaborative Filtering with Stacked Denoising AutoEncoders and Sparse Inputs//Proceedings of the NIPS Workshop on Machine Learning for eCommerce. Montreal, Canada, 2015
- [96] Zhuang F, Luo D, Yuan N J, et al. Representation learning with pair-wise constraints for collaborative ranking//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. Cambridge, UK, 2017: 567-575
- [97] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2013: 3111-3119
- [98] Wang P, Guo J, Lan Y, et al. Learning hierarchical representation model for next basket recommendation//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. Santiago, Chile, 2015: 403-412
- [99] Mnih A, Hinton G. Three new graphical models for statistical language modelling//Proceedings of the 24th International Conference on Machine Learning. Corvallis, USA, 2007: 641-648
- [100] Wang J, Yu L, Zhang W, et al. IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan, 2017: 515-524
- [101] Larochelle H, Murray I. The Neural Autoregressive Distribution Estimator. *Journal of Machine Learning Research*, 2011, 15:29-37
- [102] Clauset A, Newman M E J, Moore C. Finding community structure in very large networks. *Physical Review E*, 2004, 70(6): 066111
- [103] Zheng Y, Tang B, Ding W, et al. A neural autoregressive approach to collaborative filtering//Proceedings of the 33rd International Conference on Machine Learning. New York, USA, 2016: 764-773
- [104] Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, USA, 2011: 448-456
- [105] Ying H, Chen L, Xiong Y, et al. Collaborative deep ranking: a hybrid pair-wise recommendation algorithm with implicit feedback//Proceedings of the 20th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Auckland, New Zealand, 2016: 555-567
- [106] Lin Y, Liu Z, Zhu X, et al. Learning entity and relation embeddings for knowledge graph completion//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA, 2015:2181-2187
- [107] Koren Y. Collaborative filtering with temporal dynamics. *Communications of the ACM*, 2010, 53(4): 89-97
- [108] Chen M, Xu Z, Weinberger K, et al. Marginalized denoising autoencoders for domain adaptation. *arXiv preprint arXiv:1206.4683*, 2012
- [109] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada, 2009: 452-461
- [110] Chen J, Zhang H, He X, et al. Attentive collaborative filtering: multimedia recommendation with item-and component-level attention//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan, 2017: 335-344
- [111] Li P, Wang Z, Ren Z, et al. Neural rating regression with abstractive tips generation for recommendation//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan, 2017:345-354
- [112] Ma H, Zhou D, Liu C, et al. Recommender systems with social regularization//Proceedings of the 4th ACM International Conference on Web Search and Data Mining. Hong Kong, China, 2011: 287-296
- [113] Wang X, He X, Nie L, et al. Item silk road: recommending items from information domains to social users//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan, 2017:185-194
- [114] Zhao W X, Li S, He Y, et al. Connecting social media to e-commerce: Cold-start product recommendation using microblogging information. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(5): 1147-1159
- [115] Deng S, Huang L, Xu G, et al. On deep learning for trust-aware recommendations in social networks. *IEEE Transactions on Neural Networks & Learning Systems*, 2017, 28(5):1164
- [116] Pan Y, He F, Yu H. Trust-aware collaborative denoising auto-encoder for top-N recommendation. *arXiv preprint arXiv:1703.01760*, 2017
- [117] Yang C, Bai L, Zhang C, et al. Bridging Collaborative Filtering and Semi-Supervised Learning: A Neural Approach for POI Recommendation//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017: 1245-1254
- [118] Ozsoy M G. From word embeddings to item recommendation. *arXiv preprint arXiv:1601.01356*, 2016
- [119] Zhao S, Zhao T, King I, et al. Geo-Teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation//Proceedings of the 26th International Conference on World Wide Web Companion. Perth, Australia, 2017: 153-162
- [120] Liu Q, Wu S, Wang D, et al. Context-aware sequential recommendation//Proceedings of the 2016 IEEE 16th International Conference on Data Mining. Barcelona, Spain, 2016: 1053-1058
- [121] Unger M, Bar A, Shapira B, et al. Towards latent context-aware recommendation systems. *Knowledge-Based Systems*, 2016, 104:

- 165-178
- [122] Zhou N, Zhao W X, Zhang X, et al. A general multi-context embedding model for mining human trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(8): 1945-1958
- [123] Huang W, Wu Z, Liang C, et al. A neural probabilistic model for context based citation recommendation//*Proceedings of the 29th AAAI Conference on Artificial Intelligence*. Austin, USA, 2015:2404-2410
- [124] Fang Y, Fang Y. Neural citation network for context-aware citation recommendation//*Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Tokyo, Japan, 2017:1093-1096
- [125] Zhang S, Yao L, Sun A. Deep Learning based Recommender System: A Survey and New Perspectives. *arXiv preprint arXiv:1707.07435*, 2017
- [126] Singh A P, Gordon G J. Relational learning via collective matrix factorization// *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Las Vegas, USA, 2008: 650-658
- [127] Li B, Yang Q, Xue X. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction// *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, Pasadena, USA, 2009:2052-2057
- [128] Hu L, Cao J, Xu G, et al. Personalized recommendation via cross-domain triadic factorization//*Proceedings of the 22nd International Conference on World Wide Web*. Rio de Janeiro, Brazil, 2013: 595-606
- [129] Zhao W X, Wang J, He Y, et al. Mining product adopter information from online reviews for improving product recommendation. *ACM Transactions on Knowledge Discovery from Data*, 2016, 10(3): 29



HUANG Li-Wei, born in 1985, Ph.D. His main research interests include machine learning and recommender systems.

JIANG Bi-Tao, born in 1967, Ph.D., professor. Her main research interests focus on data mining.

LV Shou-Ye, born in 1979, Ph.D., professor. His main research interests focus on data mining.

LIU Yan-Bo, born in 1988, engineer. Her main research interests include image processing and machine learning.

LI De-Yi, born in 1944, Ph.D., professor. His main research interests focus on artificial intelligence.

Background

Deep Learning is one of the next big things in recommendation systems technology. The past few years have seen the tremendous success of deep neural networks in a number of complex tasks such as computer vision, natural language processing and speech recognition. After its relatively slow uptake by the recommender systems community, deep learning for recommender systems became widely popular in past three years. This paper gives a comprehensive overview of the current state of the art and provides a thorough analysis and comparison of the latest studies on deep learning based recommendation systems, including the comparison with the traditional recommender systems, the main research direction and application progress. In addition, the paper discusses the

future development trend of deep learning based recommendation systems.

The work is supported by the National Basic Research Program of China (973 Program) under Grant No. 2014CB340404, the Major Research Plan of the National Natural Science Foundation of China under Grant Nos. 91638301 and the National Natural Science Foundation of China under Grant Nos. 61272111, 61273216 and 6160011950. All of these projects are around the problem of data mining, deep learning, remote sensing information processing, and the research of these topics has been studied by our group for many years.