



# 当推荐系统邂逅深度学习

## 当推荐系统邂逅深度学习



张小磊

专注于机器学习、推荐系统方面的小白，公众号：机器学习游记

已关注

84 人赞同了该文章

### 前言

这是一篇关于讲述推荐系统邂逅深度学习故事的文章。

推荐系统之于用户的角色，有时更像是无微不至的男朋友，你口渴时递给你符合口味的饮料，你饥饿时还你以常吃的披萨，你无聊时帮你推荐有趣的音乐亦或带你欣赏感兴趣的电影。

但男友也会有手足无措、不知如何是好的境地。由于深度学习这个大家伙阅女无数，鬼点子太多，因此推荐系统不得不去拜见深度学习，希望能够助之一臂之力，以期提高女友的幸福感。

注：经历秋招的我，拖延了半年，终于又写了一篇，此文写的仓促，若有不妥，欢迎指教。最后提前祝大家元旦快乐呀，最后的最后，提前给大家拜个早年吧。

### 摘要

本文围绕深度学习技术在推荐系统方面的研究展开介绍，首先介绍推荐系统的相关内容以及面临的主要挑战，然后介绍目前主流的解决方案以及应用深度学习技术的必要性；第二节介绍深度学习技术的相关概念以及它的适用场景及限制条件；最后总结深度学习技术与推荐系统相结合的具体案例，希望对大家有所帮助。更多关于推荐系统相关的内容，请移步张小磊：推荐系统干货总结。

本文首发于专栏社会化推荐那些事。

▲ 赞同 84 ▼

● 18 条评论

➤ 分享

★ 4

知乎



首发于

社会化推荐那些事

推荐系统对于我们来说并不陌生，已经渗透到我们生活的方方面面，比如网易云的音乐推荐、淘宝的商品推荐、美团的餐厅推荐、抖音的短视频推荐等等。之所以推荐系统无处不在，是因为它可以提高用户的驻留时间、增加网站主的效益，以此实现互利共赢。数据显示，零售业巨头亚马逊

(Amazon) 35%的收益来自于它的推荐引擎；视频网站网飞 (Netflix) 中75%的用户都是基于推荐算法选择影片。**这也充分说明了推荐系统是整个生态系统中不可或缺的组件，可谓流量担当、变现神器。**

那么推荐系统是如何产生推荐的呢？宽泛来讲，就是系统基于用户的历史行为信息（评分、评论、浏览、点赞等）来训练用户的画像，进而对每个用户进行个性化推荐。精细阐述的话，主要分为以下两步：召回和精排。召回的作用是从海量物品中过滤出候选集合，过滤策略可以是基于内容的推荐算法或者是基于矩阵分解的算法等，算法细节详见拙作[张小磊：推荐系统之矩阵分解家族](#)。精排的作用是从召回层获取的候选集合中结合用户的兴趣来进行排序，以此来给用户产生最可能感兴趣的推荐列表。

推荐系统之所以可以产生个性化推荐列表，是基于用户的历史行为信息，但往往用户的历史行为微乎其微，弥足珍贵。**因此数据稀疏一直以来都是推荐系统所面临的挑战。目前主流的缓解方案是结合除用户-项目的交互矩阵之外的辅助边信息**，比如用户维度的社交信息，物品维度的文档信息、图像信息，以及用户-物品的上下文信息（时间、地理信息）等。对于边信息的利用，可以利用人工特征工程来提取想要的特征，但由于结合的信息越多，人工付出的代价往往过高。**好在深度学习在各类边信息中都有其发挥作用的场景，使得系统可以将复杂耗时的特征工程交给深度学习来自动提取，而推荐系统则可以更专注于如何提供更好的推荐策略，正所谓术业有专攻。**更多关于推进系统的入门可参见拙作[张小磊：推荐系统从入门到接着入门](#)。

## 深度学习

深度学习，顾名思义，具有深层结构的特征学习技术。它是建立在人工神经网络基础上发展而来的表示学习方法，又叫做表示学习。它通过构建多层表示学习结构，组合原始数据中的简单特征，从而得到更高层、更抽象的分布式表示。其中的分布式表示是指语义概念到神经元是一个多对多映射，直观来讲，即**每个语义概念由许多分布在不同神经元中被激活的模式表示；而每个神经元又可以参与到许多不同语义概念的表示中去。**

传统的机器学习技术，是一个pipeline的作业方式，需要首先进行复杂的人工特征工程任务，同时要求人们需要掌握相对较多的领域知识，进而将整理好的特征喂给分类器。而深度学习是一个端到端的学习技术，它将特征工程与特定任务整合到同一框架，可以在完成特定任务的过程中完成自动的特征提取，使得人们从复杂的人工特征工程中摆脱出来，进而将更多的精力专注于更高层的任务。

当然深度学习并不像媒体所吹嘘的那样，是所有场景。**首先，在数据量偏小的情况下，并不适合**

▲ 赞同 84 ▼

● 18 条评论

➤ 分享

★ 4



放异彩，有很大原因是图像和文本数据都具有局部相关性。比如图像中，相邻的点组成边，相连的边组成轮廓，可谓一个像素不能表达足够的信息，但一堆像素就能表示这是人脸还是狗头；语言中由单词组成句子，相邻的单词间存在相近的表达，同时单词间的次序一旦被打乱，那么整体所表达的信息也同时被打乱了。这些具有局部相关特性的数据，配合以特定的网络结构可以提取出其中的局部相关特性，同时配合深度架构达到层次特征的提取，从而达到令人满意的效果。而对于不具有局部相关特性的数据，没法用特定的网络拓扑来捕捉它们的信息，在深度学习中就只能用MLP来完成模型的训练，而通常MLP的效果通常要弱于GDBT, RF等传统集成树模型。**再者，不同的数据形式需要设计特定的网络结构，不存在通用的深度学习架构。**图像数据被表示为二维矩阵的形式（当然三通道的图像被表示为三维矩阵），为了获取图像的局部特征，人们基于平移不变性设计出了参数共享的卷积神经网络（CNN），利用滑动窗口可以很好的捕捉局部特征，并且缩小了参数空间；文本数据被表示为序列的形式，人们基于分布式假设设计出了带滑动窗口的词嵌入技术

（Word2vec），使得学到的词向量具有推理的作用；对于带有时间序列的数据，人们设计出了循环神经网络（RNN），使得参数可以受上一时刻的影响；对于图结构的数据，最近人们又设计出了图卷积神经网络（GCN）来更好的获取图结构上的特征；对于没有特殊形式的数据，深度学习不见得能更胜一筹，当人工特征工程做到一定程度后，传统模型是可以超越深度学习的。**另外，在需要给出较强业务理解的时候，不适合应用深度学习。**因为深度学习的黑盒性质，使得结果要么惊人的好，要么吓人的差，使得人们不理解其中的原理，难以指导工程师进一步调整方案。另外即使通过调参达到了模型上线的标准，工程师也不敢轻易尝试，毕竟用户需要的是合理的解释，工程师需要的是稳定的性能和知根知底的算法。

虽然深度学习技术需要前提条件，但不能否认的是它确实给各领域带来了技术上的革新，这也得益于大数据时代的**高算力（硬件层面）与强算法（软件层面）**的强大支撑。深度学习技术由于出色的拟合能力，现已广泛的应用于计算机视觉、语音识别以及自然语言处理任务中，并且都已取得了领先水平。推荐系统，作为机器学习中主要的应用之一，同样也借势深度学习，逐渐取得了令人振奋的结果。

## 插上深度学习翅膀的推荐系统

对于推荐系统来说存在两大场景即评分预测与Top-N推荐。因此接下来的介绍中主要涉及以上两方面内容。另外，由于结合的信息不同，所利用的深度学习模型也因此而异。比如对于文本信息的利用更侧重于Word2vec模型，图像信息更侧重于CNN模型，序列化的数据侧重于RNN模型等。接下来将一一进行介绍。

### 1.与MLP相关的推荐系统

#### · NeuMF

He, Xiangnan et al. Neural collaborative fil

▲ 赞同 84 ▼

● 18 条评论

➤ 分享

★ 4

不能很好的保持，因此破坏了数据的原有结构。

通过利用神经网络框架来取代内积操作，这样可以从中学习任意的函数，可以很好的保持数据的原始特性，同时该框架可以用来实现矩阵分解。另外为了能让该模型捕捉非线性因素，提出了利用多层感知机（MLP）来建模用户-项目交互矩阵。该算法同时结合了GMF层和MLP层来得到更好的特征表示，具体框架见下图。

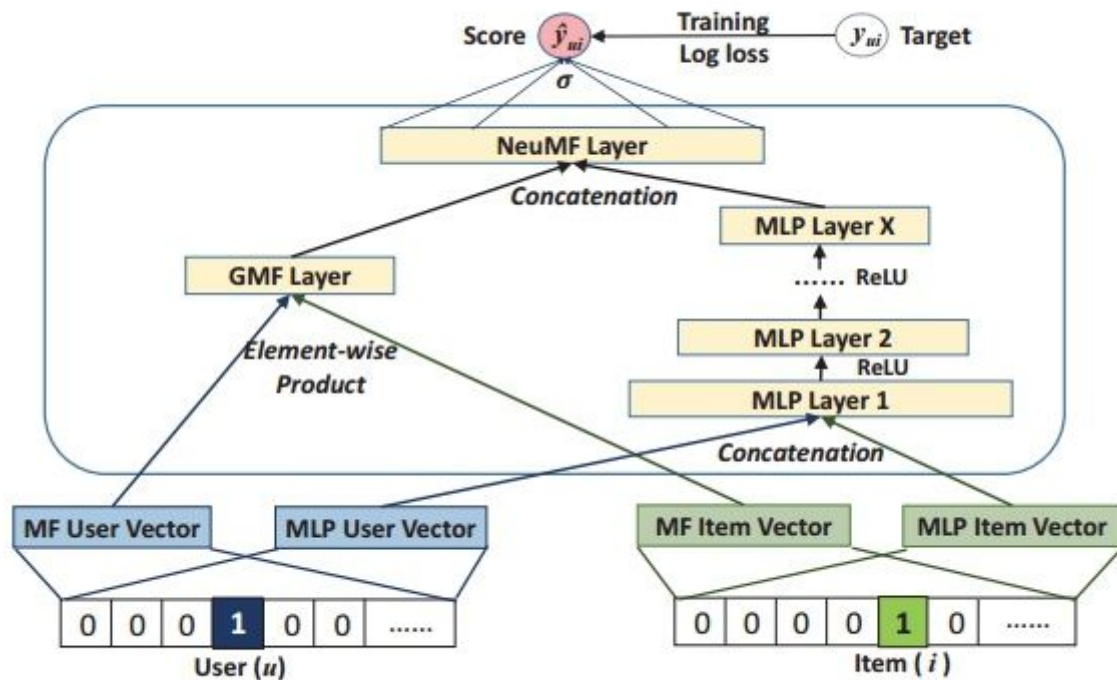


Figure 3: Neural matrix factorization model

NCF

## · EMCDR

Man, Tong et al. Cross-domain recommendation: an embedding and mapping approach. IJCAI, 2017.

该文是一篇利用跨域思想来缓解推荐系统中冷启动和数据稀疏问题的文章，思路很清晰。

数据稀疏一直以来是推荐系统中最具挑战性的问题之一。其中解决这一问题的有效思路是跨域推荐，即利用来自多个域的反馈或评分信息以众包的方式来改进推荐性能。本文提出了一种跨域推荐的嵌入式映射框架，称为EMCDR。提出的EMCDR框架从两个方面区别于现有的跨域推荐模型。首先，在每个域中利用隐因子模型来进行Embedding学习，以此来学习每个域中实体的特定特征。第二，在域间利用Mapping技术来补充不同域的数据稀疏。其中涉及的Mapping技术主要是线性映射和多层感知机映射（MLP），值得注意的是，由于MLP可以捕捉非线性因素以及出色的拟合能力，性能要优于线性映射。具体模型见下图。其（BP），算法思想详见拙作[张小磊：反向传播](#)

▲ 赞同 84 ▼

18 条评论

分享

★ 4

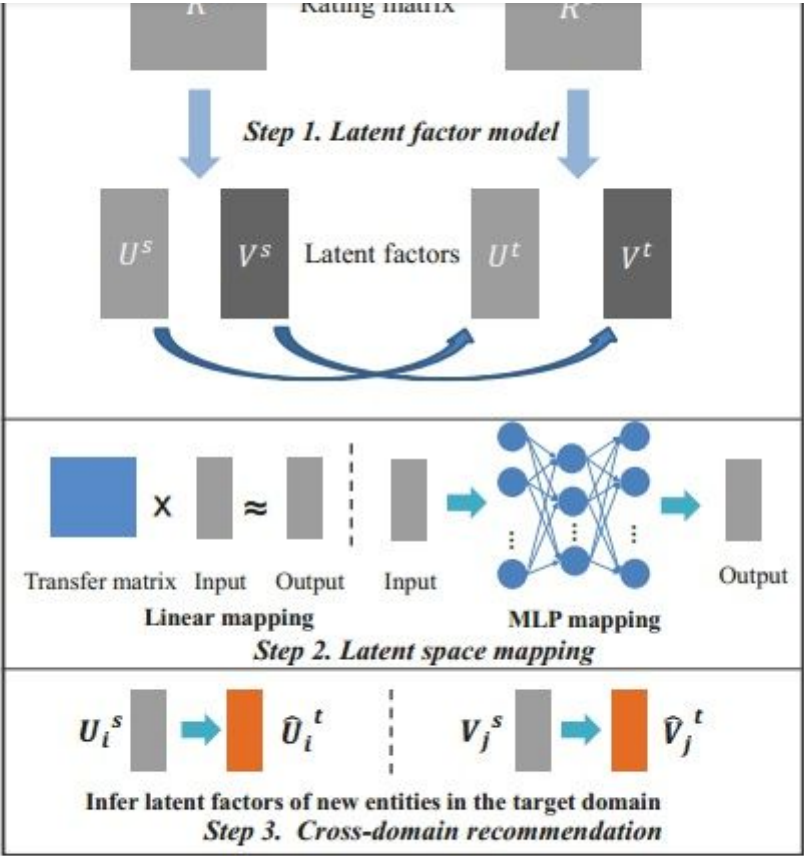


Figure 1: Illustrative diagram of the EMCDR framework. 知乎 @张小猫

EMCDR

2.与AE相关的推荐系统

· AutoRec

Sedhain et al. Autorec: Autoencoders meet collaborative filtering. WWW, 2015.

该文采用无监督学习的思想来进行推荐，利用AutoEncoder来进行用户或者项目的隐特征构建，而不像矩阵分解那样同时去分解用户和项目到同一隐空间中，通过最小化重构损失的同时完成预测任务，利用RMSE来指导参数优化。下图为Item-based AutoRec。



知乎



首发于

社会化推荐那些事

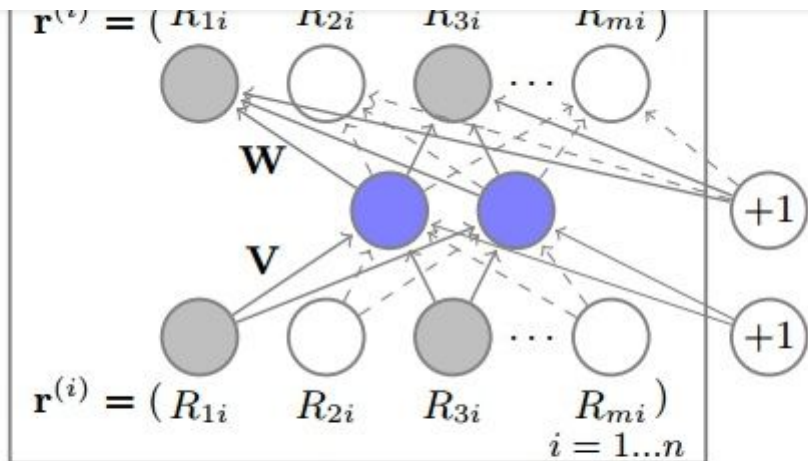


Figure 1: Item-based AutoRec model. We use plate notation to indicate that there are  $n$  copies of the neural network (one for each item), where  $\mathbf{W}$  and  $\mathbf{V}$  are tied across all copies.

Item-based AutoRec

当然，由于AutoRec的抗噪声能力以及对于数据稀疏和冷启动问题不能很好解决，人们基于此也提出了一些扩展版本，在此不再一一介绍。

① *Collaborative denoising auto-encoders for top-n recommender systems*, WSDM16.

② *Variational Autoencoders for Collaborative Filtering*, WWW, 2018.

### 3.与CNN相关的推荐系统

#### · DeepCoNN

Zheng et al. Joint deep modeling of users and items using reviews for recommendation. WSDM, 2017.

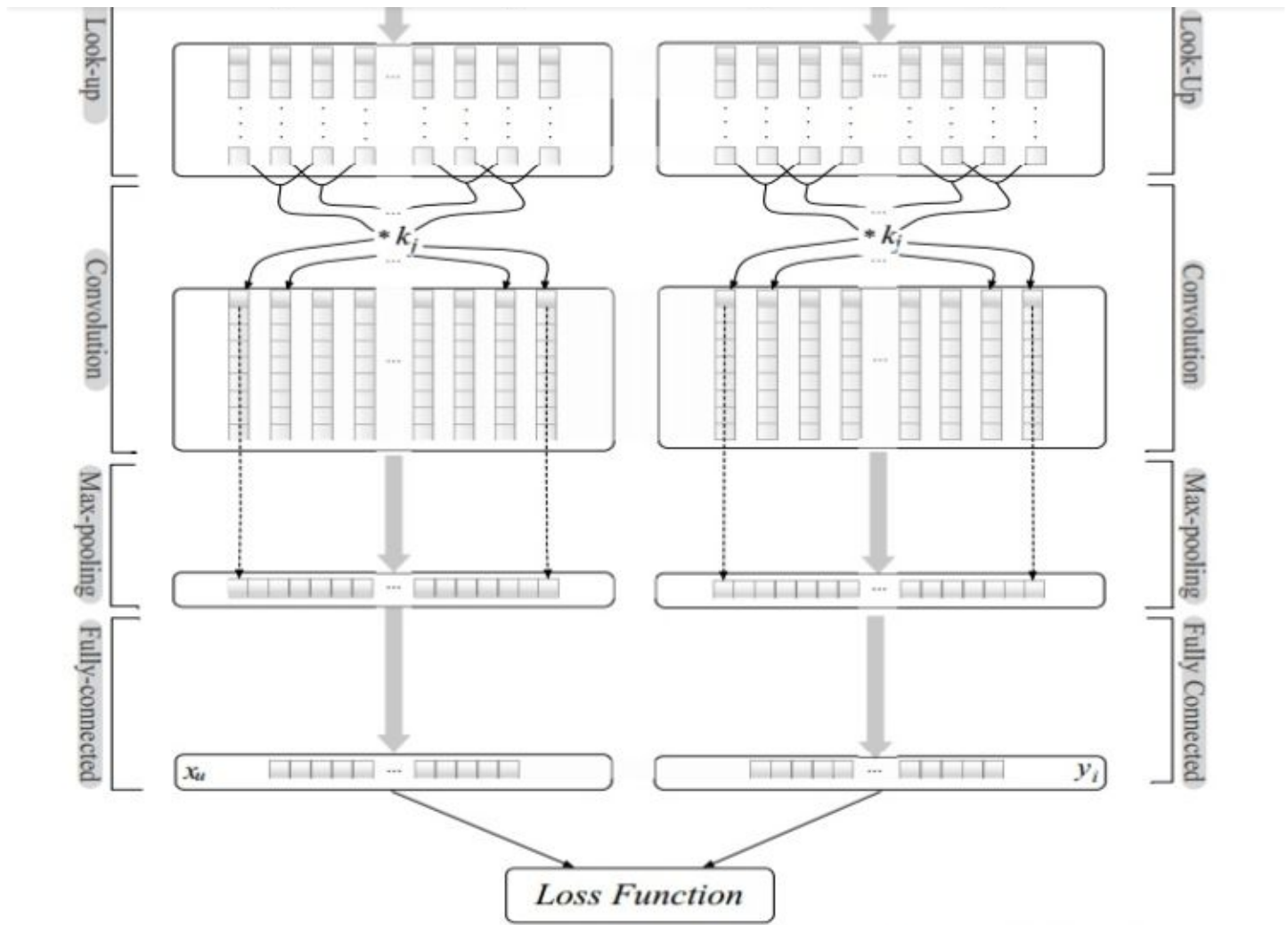
该文提出了利用两个并行的CNN来对文本信息进行处理，分别对应于用户层面的评论以及项目层面的评论，以此来挖掘用户的行为特点与项目的属性特征。最后利用FM来作为共享层，以达到利用评分标签来监督用户隐特征以及项目隐特征的训练。整体的算法框架如下图。

▲ 赞同 84 ▼

● 18 条评论

➤ 分享

★ 4

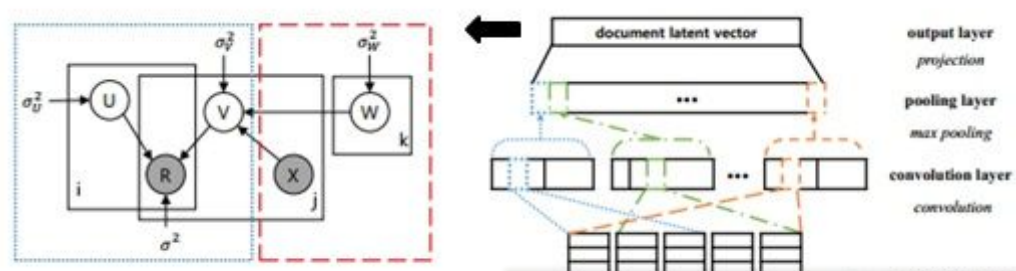


知乎 @张小磊

· ConvMF

Kim et al. Convolutional matrix factorization for document context-aware recommendation. RecSys, 2016.

该文同样是利用的文本信息，通过将文本信息作为辅助边信息来缓解评分矩阵的稀疏问题。与上文 DeepCoNN模型不同的是，文本将CNN学得文本特征作为正则项来对待，使得在拟合评分矩阵的同时能够兼顾项目维度的文本信息。算法的框架图如下。

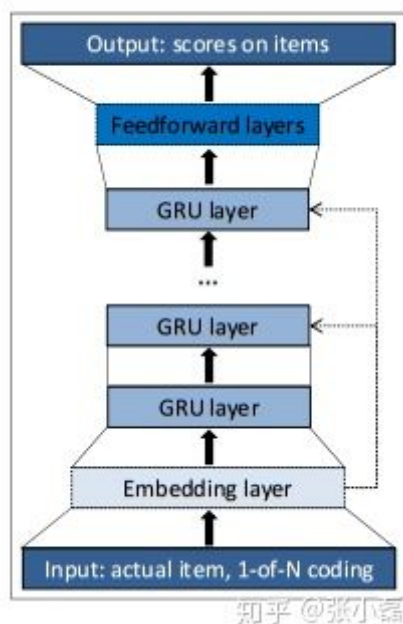


## 4.与RNN相关的推荐系统

### · SR-RNN

Hidasi et al. Session-based recommendations with recurrent neural networks. ICLR, 2016.

对于基于会话的推荐问题，较之传统推荐问题，不同点在于如何利用用户的短期会话交互数据来预测用户可能感兴趣的点击行为。基于会话的推荐可以建模为序列化问题，即基于用户的短期历史点击日志来预测下一时刻可能感兴趣的点击。深度学习中的循环神经网络模型（RNN）正是一类用于处理序列数据的神经网络，不同于MLP，在标准的RNN结构中，隐层的神经元之间也是带有权值的。也就是说，随着序列的不断推进，前面的隐层将会影响后面的隐层。于是将用户的历史交互数据作为输入，经过如下多层神经网络，达到预测用户兴趣的目的。整体框架如下图。



Session-based RNN

基于以上描述，后面也有许多这方面的扩展工作，参考文献如下。

- ① Tan et al. Improved recurrent neural networks for session-based recommendations. Workshop on Deep Learning for Recommender Systems, 2016.
- ② Hidasi et al. Recurrent neural networks with top-k gains for session-based recommendations. CIKM, 2018.

## 5.与NRL相关的推荐系统

### · Item2vec

▲ 赞同 84 ▼

💬 18 条评论

➦ 分享

★ 4



受自然语言处理中的Word2vec思想的影响，人们将用户对于项目的点击记录序列看做是自然语言处理领域中的句子，项目看做是单词，然后将此作为模型的输入，最终给出每个项目的低维表示，然后利用近邻方法来进行推荐。

## · CUNE

Zhang et al. Collaborative User Network Embedding for Social Recommender Systems.SDM, 2017.

对于社交网络数据的挖掘，传统方法可以利用降维方法如局部线性嵌入算法（LLE），特征值分解算法（LE）或者图分解算法（GF）等，但随着近年来深度学习技术的发展，受自然语言处理中Word2vec的思想影响，也提出了许多基于深度学习的网络表示学习方法（NRL），如Deepwalk，node2vec等，更多关于该算法的介绍详见拙作[张小磊：网络表示学习概述](#)。因此CUNE根据用户-项目二部图数据抽取出用户的隐式社交网络，然后经过类似于Word2vec的网络结构来提取每个用户的低维嵌入表示，最后将此作为正则项来约束最终的优化目标，框架图如下。

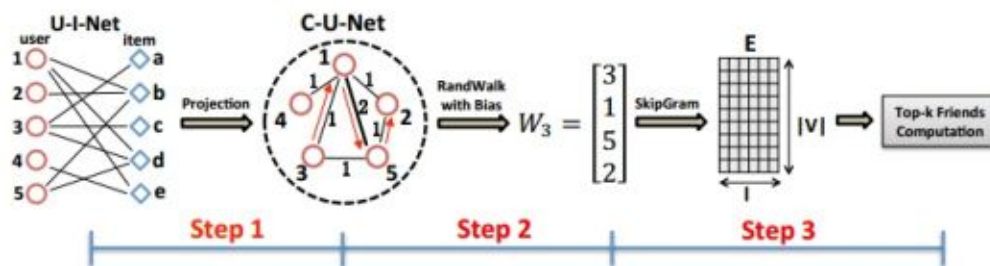


Figure 3: The illustration of CUNE method. The first step is to construct collaborative user network (C-U-Net) via one-mode projection onto users of user-item bipartite network (U-I-Net). The second step is to generate semantic social corpus (e.g.,  $W_3$ ) via random walk with bias. The third step is to learn users' latent representations (E) via SkipGram using semantic social corpus.

CUNE

## 6.与GNN相关的推荐系统

### · SR-GNN

Wu et al. Session-based Recommendation with Graph Neural Networks. AAAI, 2019.

对于基于会话的推荐问题，刚才提到，可以利用循环神经网络（RNN）来进行建模。但本文作者认为之前的工作不能够得到用户的精确表示以及忽略了项目中负责的过渡特性，因此本文工作的创新点是将序列化问题转换为图的问题，然后经过图神经网络（GNN）来学习每个项目的低维表示，同时经过注意力网络（Attention Network）来捕捉用户的短期兴趣，以达到捕获长期与短期兴趣共存的向量表示，算法框架图如下。

知乎

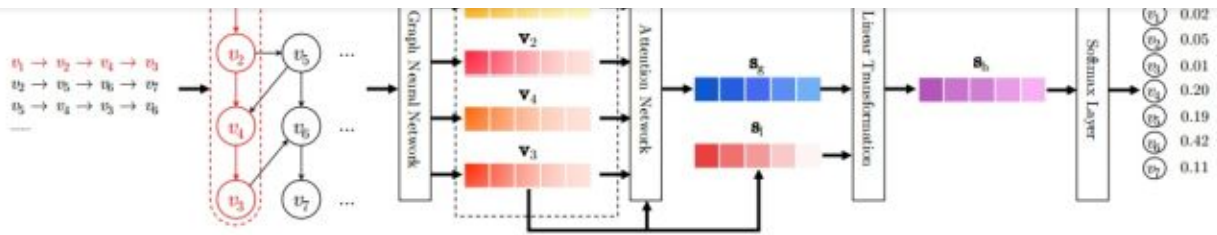
首发于  
社会化推荐那些事

Figure 1: The workflow of the proposed SR-GNN method. We model all session sequences as session graphs. Then, each session graph is proceeded one by one and the resulting node vectors can be obtained through a gated graph neural network. After that, each session is represented as the combination of the global preference and current interests of this session using an attention net. Finally, we predict the probability of each item that will appear to be the next-click one for each session.

SR-GNN

## 7.与深度特征提取相关的推荐系统

基于特征提取的方法，最基本的模型是逻辑回归（LR），即可以通过模型学习每个特征的权重进而起到特征提取的作用，然后LR只能够捕捉一阶特征的信息，后来人们发明了因子分解机（FM），以此来学习高阶特征组合的信息，并且通过矩阵分解的技巧做到了线性时间的训练，更多关于LR展开的模型，可参见拙作[张小磊：由Logistic Regression所联想到的...](#)

### · Wide & Deep Model

Cheng et al. Wide & deep learning for recommender systems. Workshop on RecSys, 2016.

刚才提到LR可以提取一阶特征的信息（Wide Model），具有很好的解释能力，另外，由于前馈深度神经网络（DNN）可以挖掘深层次的特征表示，拟合能力更强，于是Google提出了Wide & Deep learning框架，将LR和DNN网络结合起来，这样既发挥LR模型的优势，又利用DNN和Embedding的自动特征组合学习来构造了统一的学习框架。算法框架见图中间部分。

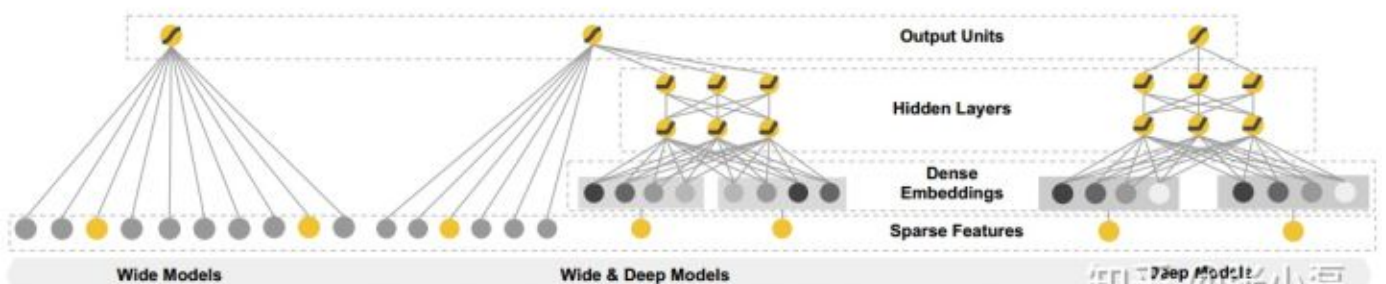


Figure 1: The spectrum of Wide & Deep models.

### · DeepFM

Huifeng et al. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. IJCAI, 2017.

▲ 赞同 84 ▼

18 条评论

分享

★ 4

取最多二阶的组合特征，因此结合了深度模型来获取更高阶复杂的特征。本文对于高阶特征与低阶特征的组合方式给出了两种方案：并行与串行的方案。整体框架如下。

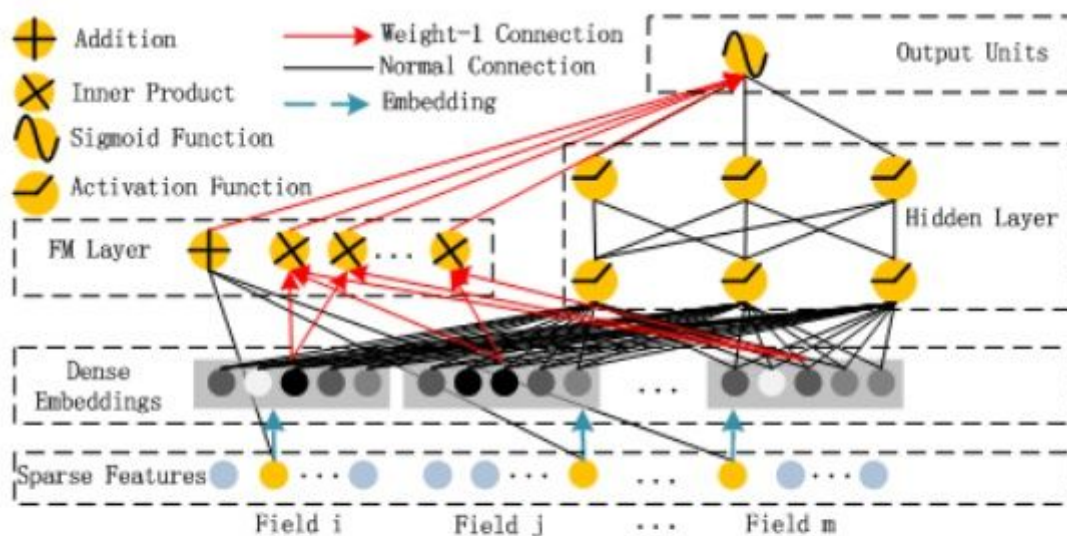


Figure 1: Wide & deep architecture of DeepFM. The wide and deep component share the same input raw feature vector, which enables DeepFM to learn low- and high-order feature interactions simultaneously from the input raw features.

知乎 @张小磊

DeepFM

最近，微软亚洲研究院对于DeepFM提出了改进，提出了极深因子分解机模型（xDeepFM），不仅能同时以显式和隐式的方式自动学习高阶的特征交互，使特征交互发生在向量级，还兼具记忆与泛化的学习能力。

## 8. 工业界利用深度学习进行推荐的范例

工业界相比于学术界对于学术研究来说，我觉得更有优势。因为可以拿到一线的数据，并且拥有足够的场景，因此基于以上可以设计出更灵活的算法。同时由于工业界的真实环境，可以通过A/B测试以及其他用户问卷来进行算法效果的反馈，而对于学术界，只能根据基本的评价指标来进行测量。另外，工业界更侧重于算法的效用、实时性以及用户体验，而学术界有时更侧重于简单指标的绝对提升。

### · Airbnb

Grbovic et al. Real-time Personalization using Embeddings for Search Ranking at Airbnb. KDD, 2018.

该文体现了业务理解与模型算法的完美融合，在真实环境中，没有普适的算法，只有结合自身业务理解设计的算法才能算得上是完美的算法。该技术创新，更多的是在Word2vec模型的基础上：

赞同 84

18 条评论

分享

★ 4



## kdd最佳论文?

· 京东

Zhou et al. Micro behaviors: A new perspective in e-commerce recommender systems. *WSDM*, 2018.

当前大多数推荐系统更注重用户和商品之间的宏观交互（如用户-商品评分矩阵），很少有人会结合用户的微观行为数据（如浏览商品的时长、对商品的阅读和评论）进行推荐。本文从微观行为的角度对推荐系统进行改进，作者将用户的固有数据视为用户和商品之间的宏观交互，并保留了宏观交互的顺序信息，同时，每个宏观交互都包含一系列微观行为。具体来说，论文提出了一个全新模型——RIB，它由输入层、Embedding层（解决数据稀疏和数据高维的问题）、RNN层（建模时序信息）、Attention层（捕捉各种微观行为影响）和输出层组成。整体框架如下图。

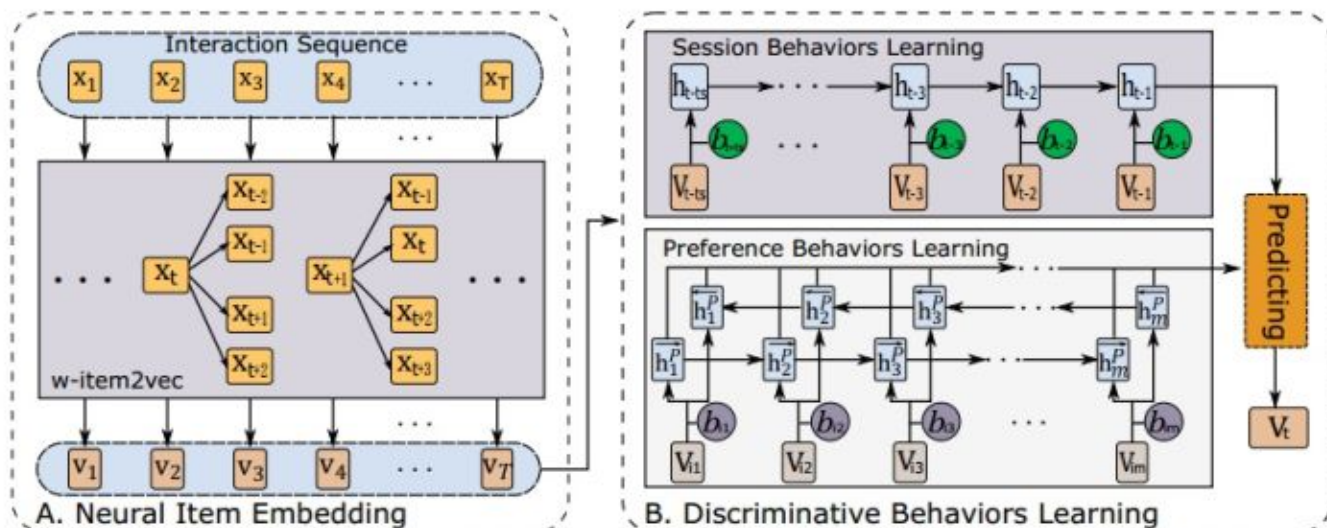


Figure 2: The overview of Behavior-Intensive Neural Network (BINN). A. Neural Item Embedding converts sequential items into a unified embedding space by w-item2vec. B. Discriminative Behaviors Learning constructs two alignments of user behaviors and discriminatively learns behavior information based on two LSTM-based architectures.

RIB

当然，还有许多公司将深度学习应用于推荐，在此不再一一简介，给出参考文献如下。

① **Youtube** Covington et al. Deep neural networks for youtube recommendations. *RecSys*, 2016.

② **雅虎** Okura et al. Embedding-based news recommendation for millions of users. *KDD*, 2017.

③ **阿里** Wang et al. Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba. *KDD*, 2018.

▲ 赞同 84 ▼

● 18 条评论

➤ 分享

★ 4

知乎



首发于  
社会化推荐那些事

基于深度学习技术的推荐系统的总结到此也就结束了，希望能对大家有所帮助。另外，虽然深度学习技术在当下甚是流行，但我们也应该保持冷静，毕竟许多东西高潮过后很少有人问津，高潮过后依然坚守在这方土地的人才是最热爱的人吧。

欢迎关注我的公众号【机器学习游记】，回复关键字“推荐系统论文”获取超百篇推荐系统论文链接下载地址，后期更多好玩的文章将在此发布，感恩。



机器学习游记

编辑于昨天 08:36

「一碗热腾腾的红烧牛肉面」

赞赏

还没有人赞赏，快来当第一个赞赏的人吧！

数据挖掘   机器学习   推荐系统

文章被以下专栏收录



社会化推荐那些事  
社会化推荐，又名社交推荐，旨在融合社交媒

▲ 赞同 84   ▼   18 条评论   分享   ★ 4