

基于 SDAE-WDL 的推荐策略优化研究

张静怡, 邝坚

(北京邮电大学计算机学院, 北京市 100089)

摘要: 本文针对传统的推荐模型方法在处理海量高维度数据时效果不佳的问题, 提出了一种融合栈式去噪自编码器 (SDAE) 和 WDL (Wide & Deep Learning for Recommender Systems) 的推荐系统。首先, 利用栈式去噪自编码器对数据进行特征降维, 实现从高维度到低维度的非线性转换; 然后利用 WDL 模型对数据进行分类推荐, 利用 Movielen 数据集的实验结果表明, 与其他推荐模型方法相比, SDAE-WDL 模型性能要优于其他方法, 取得了更好的推荐效果。

关键词: 神经网络; 推荐系统; 模型融合

中图分类号: TP391.1

Research on Optimization of Recommendation Strategy Based on SDAE-WDL

Zhang Jingyi, KUANG Jian

(Computer Science School, Beijing University of Posts and Telecommunications, Beijing city, 100089)

Abstract: Aiming at the problem that the traditional recommendation model method is not effective when processing massive high-dimensional data, this paper proposes a recommendation system that combines stacked denoising autoencoder (SDAE) and WDL (Wide & Deep Learning for Recommender Systems). First, use the stacked denoising autoencoder to reduce the feature dimension of the data to realize the non-linear conversion from high dimension to low dimension; then use the WDL model to classify and recommend the data, and the experimental results of the Movielen dataset show that it is different from other Compared with the recommended model method, the performance of SDAE-WDL model is better than other methods, and a better recommendation effect is achieved.

Key words: Neural network; recommendation system; model fusion

0 引言

互联网技术的飞速发展极大地改变民众的生活,与此同时也带来数据规模的激增,提高用户准确快速地找到目标信息的难度。搜索引擎在一定程度上缓解了“信息过载”的问题,但是依旧无法满足民众变幻无常的需求,因此便产生推荐系统。推荐系统已普遍应用于各行各业,一个好的推荐算法可以极大地提高企业的收入,提升用户的满意度。然而,现有的推荐算法饱受数据稀疏性、冷启动等问题的困扰,导致推荐精度较低,用户体验度较差。随着深度学习技术的蓬勃发展,将推荐算法与深度学习技术^[1]融合解决上述问题成为新的发展趋势。

如何从海量数据中快速精准地筛选出满足用户需求的信息是一件极富挑战性的任务。推荐系统的迅猛发展为解决这些问题提供了可能。推荐算法常借鉴机器学习^[2]、深度学习领域的方法,获得了快速发展,同时也促进了其他领域的发展。本文对推荐算法及应用进行研究,将堆栈去噪自编码器 SDAE (stacked Denoised Autoencoder)^[3]进行扩展,并且与深度学习模型 WDL(Wide and Deep Learning)^[4]相融合,提供自动化交叉特征^[5]的学习。并将特征间的结构学

作者简介: 张静怡 (1996), 女, 主要研究方向: 机器学习

通信联系人: 邝坚, 男, 教授, 嵌入式智能化. E-mail: jkuang@bupt.edu.cn

习^[6]引入到 WDL 模型中,建立一个混合推荐^[7]算法 SADE-WDL 模型,并且最终将模型在 MovieLens 数据集上进行了实验分析。与相关代表性算法相比,从 ACU^[8]方面对于模型进行测评。

1 背景知识

对于推荐算法,主要分为传统的协同过滤以及基于深度学习的推荐算法模型,对于深度模型,主要分为基于 representation learning 的深度模型以及 match function learning 的深度模型。基于 representation learning 的深度模型学习的是用户和物品的表示,然后通过匹配函数来计算,这里重点在与 representation learning 阶段,可以通过 CNN 网络, auto-encoder, 知识图谱等模型结构来学习。对于 match function learning 的深度模型,也分为基于协同过滤的模型和基于特征的模型。前者与传统 CF 模型一样,不同在于后面接入了 MLP 模型来增强非线性表达,目的是为了使得 user 和 item 的 vector 尽可能接近,这种方法就是基于 NCF 的模型;也有通过引入 relation vector 来是的 user vector 加上 relation vector 后接近 item vector,这种方法是基于翻译的模型。对于 match function learning 另一种模型框架,是基于特征层面的,有基于 fm 模型的,基于 attention 的,以及高阶特征捕捉的,另外还有基于时间序列的模型。下文将介绍 SDAE 和基于 feature-based 的深度模型-WDL。

1.1 栈式去噪自编码器

简单的自编码器是一类在半监督学习和非监督学习中使用的人工神经网络(Artificial Neural Networks, ANNs),其功能是通过将输入信息作为学习目标,对输入信息进行表征学习(representation learning),包含数据输入层、隐藏层和输出重构层,一个编码器主要分为两部分:编码器 encoder 和解码器 decoder,其中输入层到隐藏层成为编码过程,隐藏层到输出层称为解码过程。

编码器:把输入数据的特征向量进行压缩变换,构成隐含层的神经单元,隐含层的神经单元激活公式:

$$H_i = f_q(x) = s\left(\sum_{j=1}^n W_{ij}^{input} X_j + j_i^{input}\right) \quad (1)$$

其中 X 是输入数据的特征向量, W 表示权重矩阵。此过程中,输入矢量被编码为较低维向量。

解码器:对于编码器隐含层输出的特征进行解码重构回到原始输入数据的特征向量大小。映射函数如下:

$$Y_i = g_q(H) = s\left(\sum_{j=1}^n W_{ij}^{hidden} H_j + j_i^{hidden}\right) \quad (2)$$

之后,对编码器和解码器参数进行优化、调整,来尽量见效输入数据和重构数据之间的误差,损失函数一般使用均方差函数表示,公式如下:

$$J_E(W, j) = \frac{1}{2} \sum_{r=1}^N \|Y^{(r)} - X^{(r)}\|^2 \quad (3)$$

SDAE 是由多个 DAE 堆叠形成的。这时的 SDAE 并不能进行模式识别,因为它只是一

个特征提取器，并不具有分类功能。为了使 SDAE 具有分类功能，需在其顶层添加分类器，如 SVM、softmax 等，并使用带标签的数据对 SDAE 进行有监督训练，最后使用利用 BP 算法对整个网络参数进行微调，便得到具有分类功能的 SDAE。

SDAE 的思想就是将多个 DAE 堆叠在一起形成一个深度的架构。只有在训练的时候才会对输入进行腐蚀(加噪)，训练完成就不需要在进行腐蚀。栈式自编码器结构如下图 1 所示，可以看出 SDAE 采用逐层贪婪训练：每层自编码层都单独进行非监督训练，以最小化输入（输入为前一层的隐层输出）与重构结果之间的误差为训练目标。前 K 层训练好了，就可以训练 $K+1$ 层，因为已经前向传播求出 K 层的输出，再用 K 层的输出当作 $K+1$ 的输入训练 $K+1$ 层。

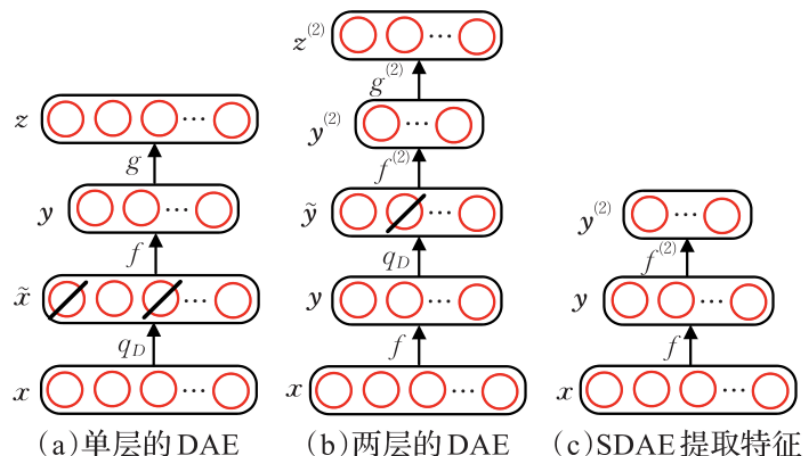


图 1 栈式自编码器结构

Fig. 1 Stacked autoencoder structure

一旦 SDAE 训练完成，其高层的特征就可以用做传统的监督算法的输入，也可以在最顶层添加一层 logistic regression layer (softmax 层)，本文通过将 SDAE 预处理后的数据输入给 WDL 的 embedding 层，使模型具有更好的鲁棒性。

1.2 WDL 深度学习模型

WDL (Wide and deep) 模型是 TensorFlow 在 2016 年 6 月左右发布的一类用于分类和回归的模型，并应用到了 Google Play 的应用推荐中。wide and deep 模型的核心思想是结合线性模型的记忆能力 (memorization) 和 DNN 模型的泛化能力 (generalization)，在训练过程中同时优化 2 个模型的参数，从而达到整体模型的预测能力最优。wide and deep 模型部分的输出是线性模型输出与 DNN 模型输出的叠加。模型训练采用的是联合训练 (joint training)，模型的训练误差会同时反馈到线性模型和 DNN 模型中进行参数更新。相比于 ensemble learning 中单个模型进行独立训练，模型的融合仅在最终做预测阶段进行，joint training 中模型的融合是在训练阶段进行的，单个模型的权重更新会受到 wide 端和 deep 端对模型训练误差的共同影响。因此在模型的特征设计阶段，wide 端模型和 deep 端模型只需要分别专注于擅长的方面，wide 端模型通过离散特征的交叉组合进行 memorization，deep 端模型通过特征的 embedding 进行 generalization，这样单个模型的大小和复杂度也能得到控制，而整体模型的性能仍能得到提高。

WDL 模型的整体结构如下图所示：

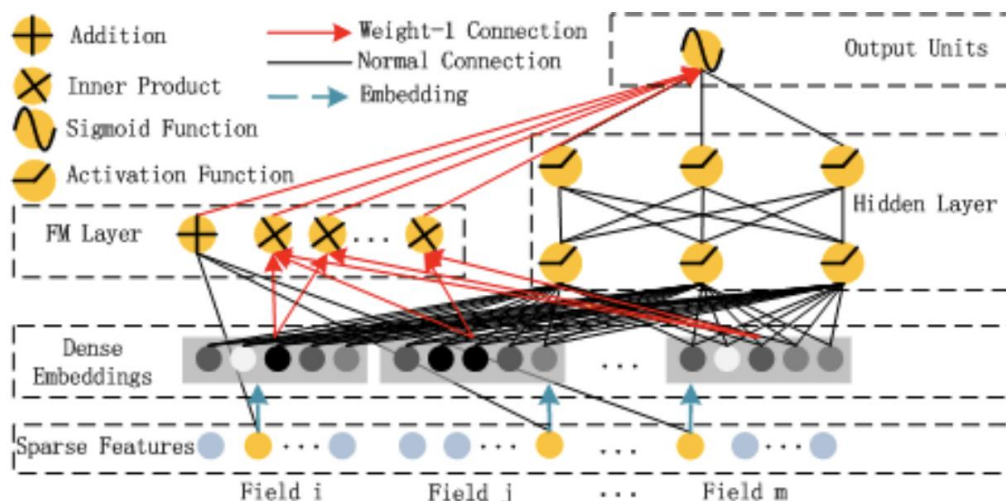


图 2 WDL 模型整体结构

Fig. 2 The overall structure of the WDL model

2 基于 SDAE-WDL 的算法推荐模型

基于以上两种算法, 本文提出一种基于深度学习方法 WDL 的神经网络预测模型: SDAE-WDL, 该模型由两部分组成, 利用 SDAE 从原始输入中提取到的特征信息作为第二层 WDL 算法输入, 对用户的多纬度信息进行特征交叉, 来预测用户对物品的喜好程度。

2.1 SDAE-WDL 模型整体结构

第一部分是深层网络采用 SDAE 自动去噪方法从原始的输入向量中提取到特征, 第二部分是用于预测 user-item 评分的 WDL 算法, 堆栈自动编码最后一层的隐藏层表示, 将作为 user 属性以及行为信息分概括和表示输入到 WDL 模型。

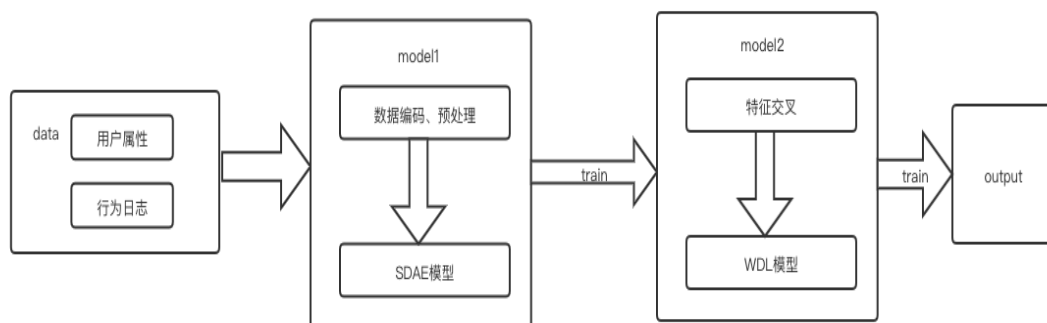


图 3 WDL 模型整体结构

Fig. 3 WDL model overall structure

2.2 WDL 模型部分算法优化

如图 3 所示, 传统的 WDL 模型在 embedding 层之后到 MLP 之间, 都是将 embedding 做 concat 的。这些 concat 后的信息其实能够表达的特征交叉信息其实是非常有限的, 仅靠

神经网络想完全捕捉到特征的有效交叉其实是非常困难的。embedding layer 进入 MLP 之前，引入了 product layer 来显式的学习每个 field 的 embedding 向量之间的两两交叉，如图 4 所示。

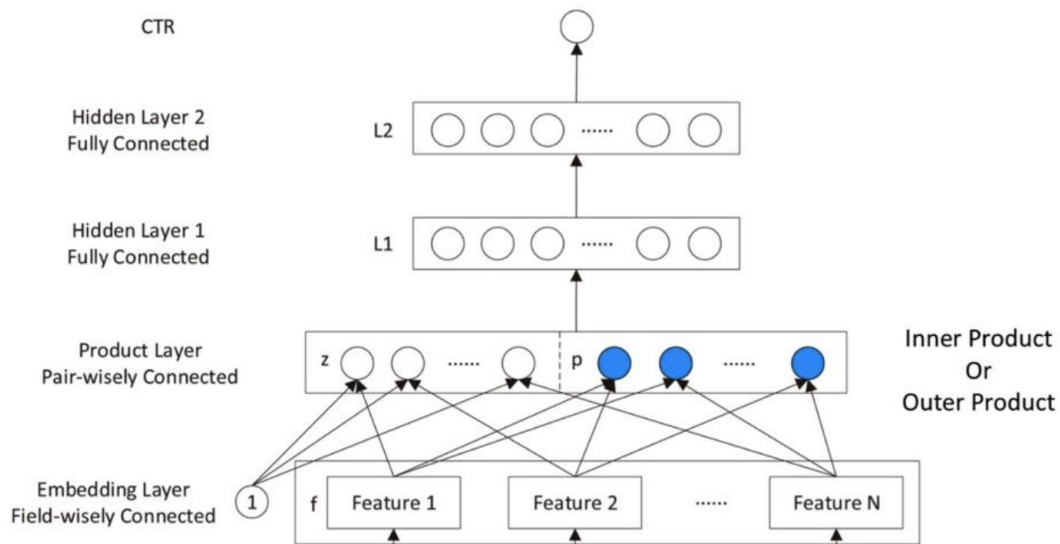


图 4 改进 WDL 模型结构图

Fig. 3 Improve WDL model structure diagram

左边 z 为 embedding 层的线性部分，右边为 embedding 层的特征交叉部分。 z 这样可以很好的挖掘推荐系统中“且”的关系。例如，职业为医生且喜欢恐怖片的人群，比起职业为医和喜欢恐怖片的人群，前者的组合比后者更能体现特征交叉的意义，对模型优化有很大的研究意义。

2.3 模型训练流程

步骤一：数据预处理，使用 one-hot 编码对数据集进行数值化，即数据的归一化处理，将所有数据特征转化到的范围。步骤二：将预处理后的数据输入给 SDAE，利用编码器解码器对数据进行重构，同时通过训练调整优化模型参数。步骤三：将 SDAE 解码器将获得的数据加上用户属性输入给 WDL 中，利用反向传播算法对模型进行有监督训练，进而优化模型参数。最后将测试集输入到训练好的模型进行最后的模型评估。

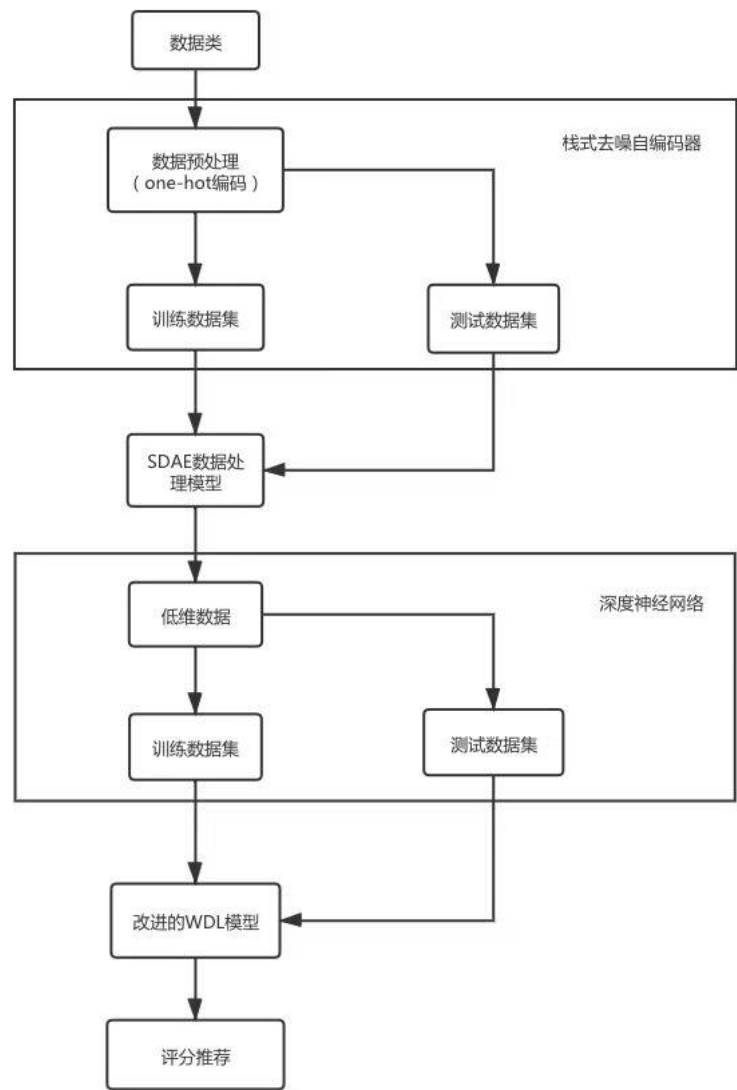


图5 模型训练流程

Fig. 5 Model training process

150

3 实验设计

本实验主要针主要任务是根据用户信息和 user-item 的交互信息，针对用户进行 item 评分预测，主要任务是推荐用户可能感兴趣的 item 列表给用户。

155

3.1 实验参数

MovieLens 数据集由 GroupLens 研究组在 University of Minnesota — 明尼苏达大学中组织的。MovieLens 是电影评分的集合，有各种大小。数据集命名为 1M, 10M 和 20M，是因为它们包含 1,10 和 20 万个评分。MovieLens 数据集中，用户对自己看过的电影进行评分，分

值为 1~5。MovieLens 包括两个不同大小的库，适用于不同规模的算法。小规模库是 943 个独立用户对 1682 部电影作的 10000 次评分的数据；大规模库是 6040 个独立用户对 3900 部电影作的大约 100 万次评分。

本实验采用 20M 数据库进行模型训练。

3.2 实验结果

如表 1 所示，实验采取了对比的方式。可以清楚地看到，线性模型的表现不佳。在引入深度模型后，效果有了较大提升，比较传统的 WDL 模型，融合模型在 AUC 上略有提升。

表 1 不同模型的 AUC 比较
Tab. 1 AUC comparison of different models

Model	AUC
FM	0.782
PNN	0.863
WDL	0.864
SDAE-WDL	0.870

4 结论

本文提出了一种基于栈式去噪自编码器和深度神经网络的推荐算法模型，利用 SDAE 算法提取数据特征，输入给 WDL 模型，同时通过 WDL 模型中特征工程的优化，进一步提高推荐准确值，实验结果表明本模型在 AUC 上优于其他传统的推荐模型。

[参考文献] (References)

- [1] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]. Advances in neural information processing systems. 2014: 3104-3112.
- [2] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [3] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [4] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016: 1480-1489.
- [5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Advances in neural information processing systems. 2017: 5998-6008.
- [6] Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[C]. Twenty-ninth AAAI 280 conference on artificial intelligence. 2015.
- [7] Yang Z, Dai Z, Yang Y, et al. Xlnet: Generalized autoregressive pretraining for language understanding[C]. Advances in neural information processing systems. 2019: 5754-5764.
- [8] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014. AUC comparison of different models.