论文中文版：

1. 摘要  
   学习用户行为背后的复杂功能交互对于最大化推荐系统的点击率至关重要。尽管取得了巨大的进步，现有的方法似乎对低端或高阶互动有着强烈的偏见，或者需要专门的技术特征工程。在本文中，我们表明可以推导出一种强调低阶和高阶特征相互作用的端对端学习模型。提出的模型DeepFM将功能学习的推理和深度学习的分解机的功能结合在一起新的神经网络架构。与谷歌最新的Wide＆Deep模型相比，DeepFM具有“宽”和“深”部分的共享输入，除了原始功能之外，不需要特征工程。进行综合实验，以证明DeepFM与现有的CTR预测模型在基准数据和商业数据上的有效性和效率。

1．介绍

点击率（CTR）的预测在推荐系统中至关重要，其中任务是估计用户点击推荐项目的概率。在许多推荐系统中，目标是最大限度地提高点击次数，因此返回给用户的项目可以按预计的点击率排序;而在其他应用场景（如在线广告）中，提高收入也很重要，因此排名策略可以根据所有候选人的点击率出价进行调整，其中“出价”是系统收到的项目，如果项目被点击用户。在任一情况下，很明显，关键在于正确估算CTR。 CTR预测重要的是学习用户点击行为背后的隐性特征交互。通过我们在主流应用程序市场的研究，我们发现人们经常在用餐时下载应用程序进行食物交付，这表明应用程序类别和时间戳之间的（第2阶段）交互可以用作CTR的信号。作为第二个观察，男性青少年喜欢射击游戏和RPG游戏，这意味着应用程序类别，用户性别和年龄的（3级）互动是CTR的另一个信号。一般来说，用户点击行为背后的功能的这种交互可能是非常复杂的，低和高阶特征交互都应该发挥重要作用。根据谷歌的“宽＆深”模型[Cheng et al。，2016]的深入分析，考虑到低阶和高阶特征交互同时对单独考虑的情况带来了额外的改进。

关键的挑战是有效地建模特征交互。一些特征交互可以很容易理解，因此可以由专家设计（如上面的实例）。然而，大多数其他功能交互都隐藏在数据中，难以识别（例如，经典的关联规则“尿布和啤酒”是从数据挖掘而不是由专家发现的），这只能通过机器学习自动捕获。[靠FNN模型]即使是易于理解的互动，专家们似乎不太可能对其进行全面的建模，特别是当功能数量较大时。

尽管简单，但广义线性模型，如FTRL [McMahan等，2013]在实践中表现出了良好的表现。然而，线性模型缺乏学习特征交互的能力，通常的做法是在其特征向量中手动包括成对特征交互。这种方法难以推广，以模拟高阶特征相互作用或从未或很少出现在训练数据中的方法[Rendle，2010]。因子分解机（FM）[Rendle，2010]模型成对特征相互作用作为特征之间潜在向量的内在产物，并显示出非常有希望的结果。原则上，FM可以模拟高阶特征交互，实际上通常只考虑2阶特征相互作用，因为复杂度高。

作为学习特征表征的强大方法，深层神经网络有潜力学习复杂的特征交互。一些想法延续了CNN和RNN的CTR优势[Liu et al。，2015; Zhang等，2014]，但是基于CNN的模型偏向相邻特征之间的相互作用，而基于RNN的模型更适合于具有顺序依赖性的点击数据。 [Zhang et al。，2016]研究特征表征，并提出了分解机器支持的神经网络（FNN）。该模型在应用DNN之前对FM进行预先训练，因此受到FM的能力的限制。特征相互作用在[Qu etal。 2016]，通过在嵌入层和完全连接层之间引入产品层，并提出了基于产品的神经网络（PNN）。如[Cheng et al。，2016]所述，PNN和FNN与其他深层模型一样，捕获了一些低阶特征相互作用，这也是CTR预测所必需的。为了模拟低阶和高阶特征相互作用，[Cheng et al。，2016]提出了一种结合线性（“宽”）模型和深度模型的有趣的混合网络结构（Wide＆Deep）。在这个模型中，“广泛部分”和“深部分”分别需要两种不同的输入，“广泛部分”的输入依赖于专业技能特征工程。

可以看出，现有的模型偏向于低阶或高阶特征交互，或依赖于特征工程。在本文中，我们可以得出一个学习模型能够以端到端的方式学习所有订单的功能交互，除了原始功能之外没有任何功能工程。我们的主要贡献总结如下：

我们提出了一种新的神经网络模型DeepFM（图1），它集成了FM和深层神经网络（DNN）的架构。它可以模拟低阶功能交互，如FM和高阶功能 互动，如DNN。不同于广深模型[Cheng et al。，2016]，DeepFM可以在没有任何特征工程的情况下进行端到端的训练。

DeepFM可以有效地进行培训，因为与[Cheng等人，2016]不同，它的广泛部分和深度部分共享相同的输入和嵌入向量。在[Cheng等人，2016]中，输入向量可以是巨大的，因为它包括在其广泛的输入向量中的手动设计的成对特征交互，这也大大增加了其复杂性。

我们对基准数据和商业数据评估DeepFM，这显示了与现有的CTR预测模型的一致性改进

2我们的方法

假设用于训练的数据集由n个实例（; y）组成，其中是通常涉及一对用户和项目的m场数据记录，y 2 f0; 1g是指示用户点击行为的相关标签（y = 1表示用户点击了该项目，否则为y = 0）。可以包括分类字段（例如，性别，位置）和连续字段（例如，年龄）。每个分类字段被表示为单热编码的向量，并且每个连续字段被表示为值本身，或离散化之后的一热编码的向量。然后，每个实例被转换为（x; y），其中x = [xfield1; xfield2; :::; xfiledj :::; xfieldm]是一个d- 其中xfieldj是第j个字段的向量表示。通常，x是高维度和非常稀疏的。 CTR预测的任务是建立一个预测模型^ y = CTR模型（x）来估计用户在给定上下文中点击特定应用程序的概率

2.1 DeepFM

我们的目标是学习低阶和高阶特征交互。为此，我们提出了一种基于分解机的神经网络（DeepFM）。如图11所示，DeepFM由共享相同输入的两个组件FM组件和深组件组成。对于特征i，使用标量wi来衡量其顺序1的重要性，潜伏向量Vi用于测量其与其他特征的相互作用的影响。 Vi被馈送到FM组件中以对2阶特征相互作用进行模型化，并将其加入深部分以建模高阶特征交互。所有参数，包括wi，Vi和网络参数（W（1），b（l））均为联合预测模型进行训练：^ y = sigmoid（yFM + yDNN）; （1）其中^ y 2（0; 1）是预测的CTR，yFM是FM分量的输出，yDNN是深分量的输出。

FM组件是一个分解机，在[Rendle，2010]中提出了学习功能交互的推荐。除了特征之间的线性（顺序-1）相互作用之外，FM模型成对（order-2）将相互作用作为各个特征潜向向量的内积。

它可以比以前的方法更有效地捕获2阶特征交互，特别是当数据集issparse时。在以前的方法中，只有当特征i和特征j都出现在同一数据记录中时，才能训练特征i和j的交互参数。在FM中，通过它们的潜伏矢量Vi和Vj的内积来测量。由于这种灵活的设计，每当i（或j）出现在数据记录中时，FM可以训练潜伏矢量Vi（Vj）。因此，训练数据中永远不会或很少出现的特征相互作用，更好地被FM学习。

如图2所示，FM的输出是加法单位和内产品单位数的总和：yFM = hw; xi + Xd j1 = 1 Xd j2 = j1 + 1hVi; Vji xj1 xj2; （2）其中w 2 Rd和Vi 2 Rk（k给出）2。加法单元（hw; xi）反映了1阶特征的重要性，内部乘积单位代表了2阶特征相互作用的影响。

深层组件是前馈神经网络，用于学习高阶特征交互。如图3所示，将数据记录（矢量）送入神经网络。与图像[Heet al。，2016]或音频[Boulanger-Lewandowski等人，2013]数据作为纯粹连续和密集的输入的神经网络相比，CTR预测的输入是完全不同的，这需要一个新的网络结构设计。具体来说，用于CTR预测的原始特征输入向量通常是高度稀疏3，超高维数4，分类连续混合，并且分组在字段（例如，性别，位置，年龄）中。这表明嵌入层在进一步馈送到第一隐藏层之前将输入向量压缩为低维密集实数值向量，否则网络可能爆炸式地训练

**Zhang et al., 2016] studies feature representations and proposes Factorization-machine supported Neural Network (FNN). This model pre-trains FM before applying DNN, thus limited by the capability of FM.**  
图4突出显示了从输入层到嵌入层的子网结构。我们想指出这个网络结构的两个有趣的特征：1）虽然不同输入场矢量的长度可以不同，但​​它们的嵌入是相同的大小（k）; 2）FM现在服务器中的潜在特征向量（V）作为网络权重，其被学习并用于将输入域向量压缩到嵌入向量。在[Zhang et al。，2016]中，V被FM预先训练，并用作初始化。在这项工作中，除了使用FM的潜在特征向量来初始化网络，如[Zhang et al。，2016]，除了其他DNN模型之外，我们还将FM模型作为我们整体学习架构的一部分。因此，我们消除了FM的预训练需要，而是以端对端的方式共同训练整个网络。将嵌入层的输出表示为：a（0）= [e1; E2; :::;的EM; （3）其中ei是第i个场的嵌入，m是场的数量。然后，a（0）被馈送到深层神经网络，前向过程是：（1 + 1）= \_（W（1）a（1）+ b（1））; （4）其中l是层深度，\_是激活函数。 a（1），W（1），b（l）是第l层的输出，模型权重和偏置。之后，生成一个密集的实值特征向量，最终被馈送到用于CTR预测的S形函数中：yDNN = \_（WjHj + 1 \_ aH + bjHj + 1），其中jHj是隐藏层数。