

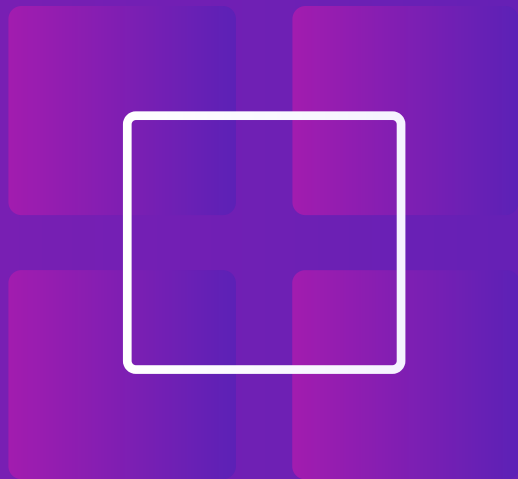


Feature Generation by Convolutional Neural Network for Click-Through Rate Prediction

WWW'19, 华为诺亚方舟实验室
乔梁 2022.11.11



- 一、摘要
- 二、引言
- 三、FGCNN
- 四、结论
- 五、学习收获







CTR

背景：点击率预测是推荐系统中的一项重要任务，其目的是估计用户点击给定项目的概率。最近，已经提出了许多深度模型来从原始特征中学习低阶和高阶特征交互。但是，由于有用的交互总是稀疏的，因此DNN很难在大量参数下有效地学习它们。

之前的方法：在实际场景中，人工特征能够提高深度模型的性能，但是特征工程成本很高，需要领域知识，在不同的场景中是不切实际的。因此，有必要自动扩展特征空间。

本文方法：在本文中，我们提出了一种卷积神经网络 (FGCNN) 模型的特征生成，该模型具有两个组件：特征生成和深度分类器。特征生成利用CNN的优势来生成局部模式并将其重新组合以生成新特征。深度分类器采用IPNN的结构，从增强特征空间学习交互。

在三个大规模数据集上的实验结果表明，FGCNN的性能明显优于九个最先进的模型。此外，当应用一些最先进的模型作为深度分类器时，总是可以获得更好的性能，这表明我们的FGCNN模型具有很大的兼容性。这项工作探索了CTR预测的新方向：**通过自动识别重要特征来减少DNN的学习困难非常有用。**





CTR任务的挑战

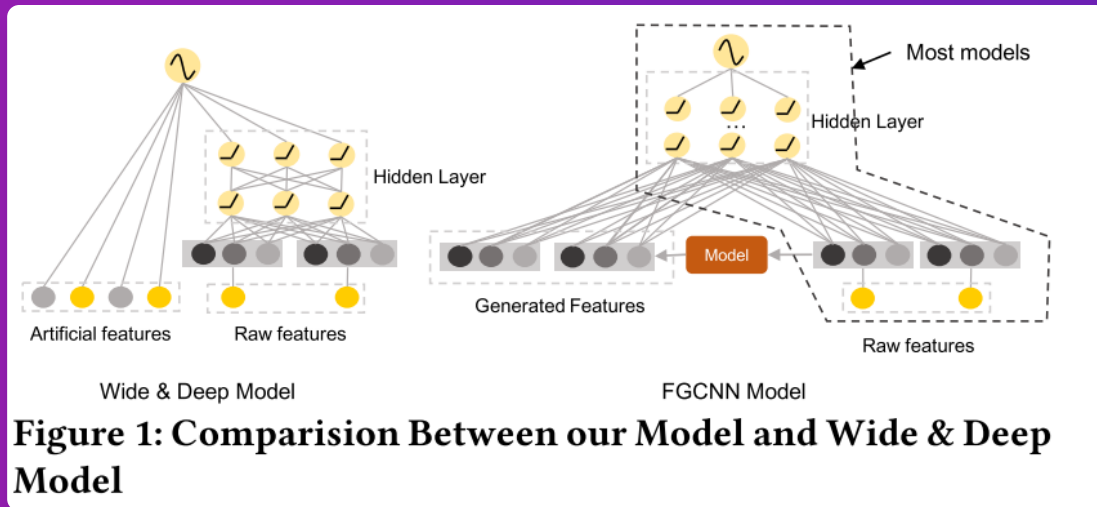
CTR预测任务的关键挑战是有效地对特征交互进行建模。

- 从理论上讲，DNN能够从原始特征中学习任意特征交互。但是，由于与原始特征的组合空间相比，有用的交互通常是稀疏的，因此很难从大量参数中有效地学习它们。
- 在人工特征的帮助下，深度组件的性能显著提高。但是，特征工程可能很昂贵，并且需要领域知识。如果我们可以通过机器学习模型自动生成复杂的特征交互，它将更加实用和健壮。

因此，如图1所示，我们提出了一个用于自动特征生成的通用框架。将原始特征输入到机器学习模型 (由图1中的红色框表示) 中，以识别并生成新特征。之后，将原始特征和生成的新特征组合并馈送到深度神经网络中。生成的特征能够通过预先捕获稀疏但重要的特征交互来减少深度模型的学习困难。



图1: 本文模型与Wide&Deep模型的比较





自动特征生成

自动特征生成最直接的方法是使用MLP，并使用隐藏的神经元作为生成的特征。但是，由于有用的特征交互通常是稀疏的，因此MLP很难从巨大的参数空间中学习这种交互。

- 例如，假设我们有四个用户特征: 姓名、年龄、身高、性别来预测用户是否会下载在线游戏。假设年龄和性别之间的特征交互是唯一重要的信号，因此，最佳模型应该识别这种且仅识别这种特征交互。当仅使用一个隐藏层执行MLP时，与Name和Height的嵌入相关联的最佳权重应全为0，这很难实现。



自动特征生成

作为一种先进的神经网络结构，卷积神经网络 (CNN) 在计算机视觉和自然语言处理领域取得了巨大的成功。在CNN中，共享权重和池化机制的设计大大减少了查找重要局部模式所需的参数数量，并减轻了后期MLP结构的优化困难。因此，CNN提供了一个潜在的很好的解决方案来实现我们的想法 (识别稀疏但重要的特征交互)。但是，直接应用CNN可能会导致性能不理想。

在CTR预测中，原始特征的不同排列顺序没有不同的含义。

- 例如，特征的排列顺序是 (姓名，年龄，身高，性别) 还是 (年龄，姓名，身高，性别) 对描述样本的语义没有任何区别，这与图像和句子的情况完全不同。如果我们仅使用CNN提取的邻居模式，许多有用的全局特征交互将会丢失。这也是CNN模型在CTR预测任务中表现不佳的原因。为了克服这一限制，我们将CNN和MLP相互补充，以学习用于特征生成的全局-局部特征交互。



本文工作

在本文中，我们提出了一种用于CTR预测任务的新模型，即通过卷积神经网络 (FGCNN) 生成特征，该模型由两个组件组成: 特征生成和深度分类器。

- 在特征生成中，CNN MLP结构被设计用于从原始特征中识别和生成新的重要特征。更具体地说，执行CNN以学习邻居特征交互，而应用MLP重新组合它们以提取全局特征交互。特征生成后，可以通过结合原始特征和新特征来扩大特征空间。
- 在深度分类器中，几乎所有最先进的网络结构 (例如PIN, xDeepFM, DeepFM) 都可以采用。因此，我们的模型与推荐系统中的最新模型具有良好的兼容性。为了便于说明，由于其在模型复杂性和准确性之间的良好权衡，我们将采用IPNN模型 作为FGCNN中的深度分类器。

在三个大规模数据集中的实验结果表明，FGCNN的性能明显优于九个最先进的模型，证明了FGCNN的有效性。在深度分类器中采用其他模型的同时，始终可以实现更好的性能，这表明了所生成特征的有用性。逐步分析表明，FGCNN中的每个组件都有助于最终性能。与传统的CNN结构相比，当原始特征的顺序发生变化时，我们的CNN MLP结构表现得更好，更稳定，这证明了FGCNN的鲁棒性。



本文贡献

本文的主要贡献如下:

- 确定了CTR预测的重要方向: 通过提前自动生成重要特征来减少深度学习模型的优化困难很有必要, 也很有用。
- 我们提出了一种用于自动生成和分类特征的新模型——FGCNN, 该模型由两个部分组成: 特征生成和深度分类器。特征生成利用CNN和MLP相互补充, 来识别重要但稀疏的特征。此外, 几乎所有其他CTR模型都可以应用于深度分类器中, 以基于生成的和原始特征进行学习和预测。
- 在三个大规模数据集上的实验证明了FGCNN模型的整体有效性。当生成的特征用于其他模型时, 总是可以实现更好的性能, 这表明我们的FGCNN模型具有极大的兼容性和鲁棒性。



三、FGCNN



FGCNN模型

如图2所示，FGCNNmodel由两个组件组成: 特征生成和深度分类器。更具体地说，特征生成侧重于识别有用的局部和全局模式，以生成新的特征作为原始特征的补充，而深度分类器通过深度学习模型基于增强的特征空间进行学习和预测。

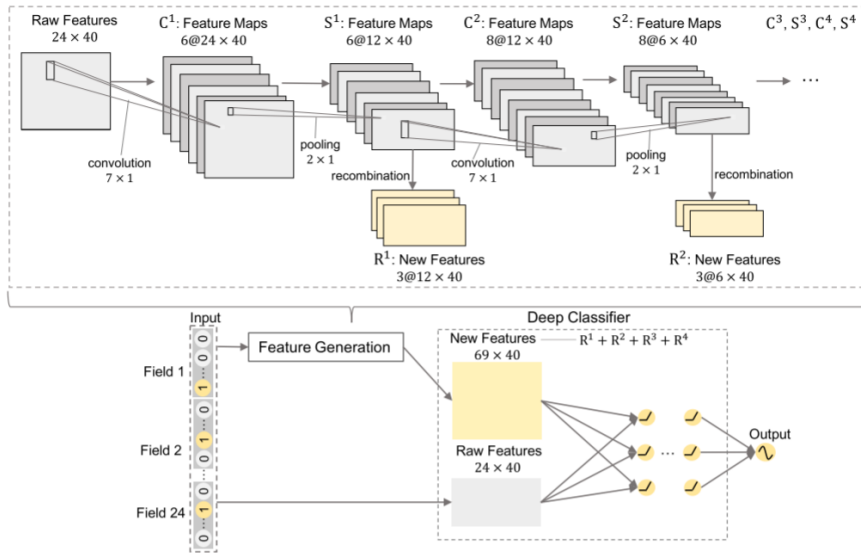


Figure 2: An overview of Feature Generation by Convolutional Neural Network Model (The hyper-parameters in the figure are the best setting of FGCNN on Avazu Dataset)



特征生成

从原始特征生成新特征有助于提高深度学习模型的性能。为了实现此目标，特征生成设计了适当的神经网络结构以识别有用的特征交互，然后自动生成新的特征。

仅使用MLP或CNN无法从原始特征生成有效的特征交互：

1. 有用的特征交互在原始特征的组合空间中总是稀疏的。因此，MLP很难从大量的参数中学习它们。
2. 尽管CNN可以通过减少参数的数量来减轻MLP的优化困难，但它仅生成邻居特征交互，而邻居特征交互会丢失许多有用的全局特征交互。

为了克服单独应用MLP或CNN的弱点，我们联合使用CNN和MLP作为相互补充的特征生成。图3显示了CNN重组结构以捕获全局特征交互的示例。可以观察到，CNN学习具有有限数量参数的有用的邻居特征模式，而重组层（其是全连接层）基于CNN提供的邻居模式生成全局特征交互。因此，可以通过这种神经网络结构有效地生成重要的特征，该结构比直接应用MLP进行特征生成具有更少的参数。在以下部分中，我们详细介绍了特征生成的CNN重组结构，即卷积层，池化层和重组层。



图3: CNN重组结构能够捕获全局非邻居特征交互以生成新特征

CNN由卷积层和池化层组成，而重组由全连接层组成。

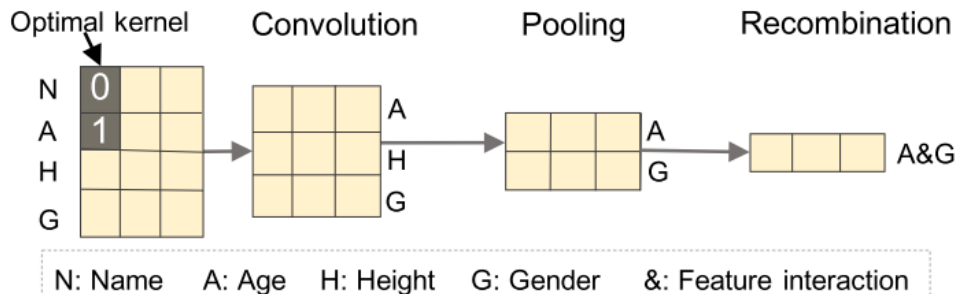


Figure 3: CNN+Recombination structure is able to capture global non-neighbor feature interactions to generate new features. CNN consists of convolutional layer and pooling layer, while Recombination consists of a fully connected layer.



卷积层

每个实例通过特征嵌入表示为嵌入矩阵 $E \in R^{n_f \times k}$, 其中 n_f 是字段数, k 是嵌入大小。为方便起见, 将嵌入矩阵重塑为 $E_1 \in R^{n_f \times k \times 1}$ 作为第一卷积层的输入矩阵。为了捕获邻居特征相互作用, 通过用非线性激活函数 (其中 h^1 是第一卷积权重矩阵的高度, m_c^1 是第一卷积层中的特征图的数量) 对矩阵 $\mathbb{W}C^1 \in R^{h^1 \times 1 \times 1 \times m_c^1}$ 进行卷积得到卷积层。假设第一卷积层的输出表示为 $C^1 \in R^{n_f \times k \times m_c^1}$, 我们可以将卷积层公式化为:

$$C_{p,q,i}^1 = \tanh \left(\sum_{m=1}^1 \sum_{j=1}^{h^1} E_{p+j-1,q,m}^1 \mathbb{W}_{j,1,1,i}^1 \right)$$
$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$

其中 $C_{:, :, i}^1$ 表示第一卷积层中的第 i 个特征图, p, q 是第 i 个特征图的行和列索引。



池化层

在第一卷积层之后，应用maxpooling层以抓住最重要的特征交互并减少参数数量。我们将 h_p 称为池化层的高度 (宽度 = 1)。在第一池化层中的输出为 $S^1 \in R^{(n_f/h_p) \times k \times m_c^1}$ ：

$$S_{p,q,i}^1 = \max \left(C_{p \cdot h_p, q, i}^1, \dots, C_{p \cdot h_p + h_p - 1, q, i}^1 \right)$$

第 i 个池化层的池化结果将是第 $i + 1$ 个卷积层的输入: $E^{i+1} = S^i$



重组层

在第一卷积层和池化层之后, $S^1 \in R^{(n_f/h_p) \times k \times m_c^1}$ 包含邻居特征的模式。由于CNN的性质, 如果将 S^1 视为生成的新特征, 则全局非邻居特征交互将被忽略。因此, 我们引入了一个全连接层来重新组合局部邻居特征模式并生成重要的新特征。我们将权重矩阵表示为 $\mathbb{W}R^1 \in R^{(n_f/h_p k m_c^1) \times (n_f/h_p k m_r^1)}$, 偏差表示为 $\mathbb{B}R^1 \in R^{(n_f/h_p k m_r^1)}$, 其中, m_c^1 是第一卷积层中的特征图的数量, m_r^1 是第一重组层中的新特征图的数量。因此, 在第 i 个重组层中, 产生 $n_f/h_p^i m_r^i$ 特征:

$$R^1 = \tanh(S^1 \cdot \mathbb{W}R^1 + \mathbb{B}R^1)$$



拼接

可以通过多次执行CNN重组来生成新特征。假设存在 n_c 个卷积层，池化层和重组层，并且 $N_i = n_f/h_p^i m_r^i$ 特征域由第 i 轮生成，表示为 R^i 。整体新特征 $\mathcal{R} \in R^{N \times k}$ (其中 $N = \sum_{i=1}^{n_c} N_i$)：

$$\mathcal{R} = (R^1, R^2, \dots, R^{n_c})$$

然后，原始特征和新特征拼接为

$$\mathbb{E} = (E'^T, \mathcal{R}^T)^T$$

其中 E' 是深度分类器的原始特征的嵌入矩阵。原始特征和新特征都用于深度分类器中的CTR预测。



深度分类器

如上所述，原始特征和新特征被拼接为增广嵌入矩阵 $\mathbb{E} \in R^{(N+n_f) \times k}$ ，其中 N 和 n_f 分别是新特征和原始特征的字段数。 \mathbb{E} 被输入到深度分类器中，该分类器旨在了解原始特征与新生成的特征之间的进一步交互。

在本小节中，为了便于展示，我们采用IPNN模型作为深度分类器中的网络结构，因为它在模型复杂性和准确性之间有很好的权衡。实际上，可以采用任何高级网络结构，这表明了FGCNN与现有作品的兼容性。



网络结构

IPNN模型结合了FM和MLP的学习能力。它利用FM层通过内积操作从嵌入向量中提取成对特征交互。之后，将输入特征的嵌入和FM层的结果拼接并馈送到MLP进行学习。

如图4所示，通过FM层对增强嵌入矩阵 $\mathbb{E} \in R^{(N+n_f) \times k}$ 的成对特征相互作用进行建模，如下

$$R_{fm} = \left(\langle \mathbb{E}_1, \mathbb{E}_2 \rangle, \dots, \langle \mathbb{E}_{N+n_f-1}, \mathbb{E}_{N+n_f} \rangle \right)$$

其中 \mathbb{E}_i 是第 i 个字段的嵌入， $\langle a, b \rangle$ 表示 a 和 b 的内积。FM层中的成对特征交互次数为 $\frac{(N+n_f)(N+n_f-1)}{2}$



IPNN模型结构

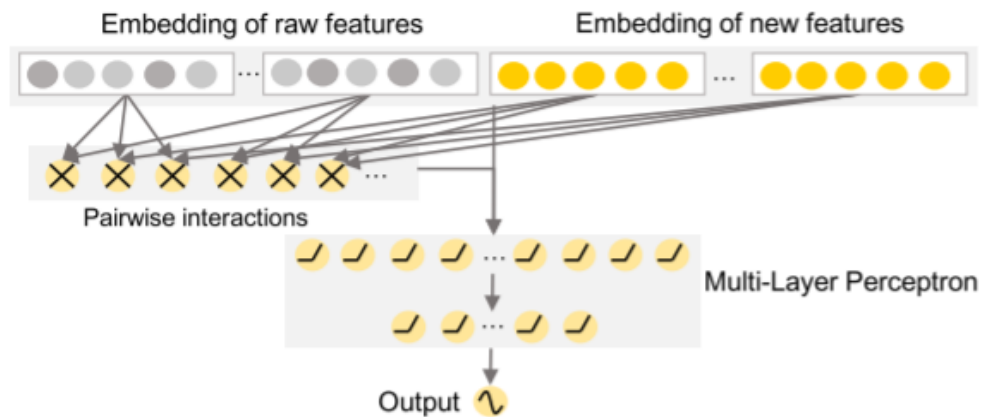


Figure 4: Structure of IPNN Model



网络结构

在FM层之后，将 R_{fm} 与增强的嵌入矩阵 \mathbb{E} 连接，将其馈入具有 n_h 隐藏层的MLP中，以学习隐式特征交互。我们将第 i 个隐藏层的输入称为 I_i ，将第 i 个隐藏层的输出称为 O_i 。MLP公式化为：

$$\begin{aligned} I_1 &= (R_{fm}, \text{flatten}(\mathbb{E})) \\ O_i &= \text{relu}(I_i W^i + B^i) \\ I_{i+1} &= O_i \end{aligned}$$

其中 W^i 和 B^i 是MLP中第 i 个隐藏层的权重矩阵和偏置。对于最后的隐藏层，进行最终预测：

$$\hat{y} = \text{sigmoid}(O_{n_h} W^{n_h+1} + b^{n_h+1})$$





结论

本文提出了一种用于CTR预测的FGCNN模型，该模型旨在通过预先识别重要特征来减少DNN模型的学习困难。该模型由两个部分组成: 特征生成和深度分类器。

- 特征生成利用CNN来识别有用的局部模式，并通过引入重组层从局部模式的重组生成全局新特征来缓解CNN的弱点。
- 在深度分类器中，大多数现有的深度模型都可以应用于增强特征空间。

在三个大规模数据集上进行了广泛的实验，结果表明FGCNN的性能优于九个最先进的模型。此外，在深度分类器中应用其他模型时，与未生成特征的原始模型相比，始终可以获得更好的性能，这证明了生成特征的有效性。逐步实验表明，FGCNN中的每个组件都有助于最终性能。此外，与传统的CNN结构相比，在调整原始特征的排列顺序时，我们在特征生成中的CNN重组结构始终表现得更好，更稳定。这项工作探索了CTR预测的新方向，即首先自动生成重要特征是有效的，而不是直接将原始嵌入提供给深度学习模型。





学习收获

- 如何进行自动特征交互;
- 通过引入CNN进行创新。

Thank you