

# 揭秘华为应用市场APP推荐中的深度学习算法：DeepFM

雷锋网 2017-09-28

IJCAI (International Joint Conference on Artificial Intelligence, 人工智能国际联合大会) 是人工智能领域的顶级会议，每年接收来自学术界及工业界在机器学习和人工智能领域的最新研究进展，历年在IJCAI发表的论文在机器学习和人工智能技术的发展上作出了积极的贡献。

今年8月下旬，在澳大利亚墨尔本召开的IJCAI2017会议上，来自华为伏羲推荐团队的专家发表了他们在深度学习推荐算法方面的最新成果。

**伏羲推荐引擎是华为应用市场联合华为诺亚方舟实验室开发的一款推荐系统。**针对华为应用市场的业务特点和数据特征，伏羲推荐算法团队提出的端到端的深度学习推荐模型DeepFM，助力华为应用市场更加精准、个性化的推荐体验。



华为诺亚方舟实验室算法专家在IJCAI2017演讲

特征工程 (Feature Engineering) 是影响推荐系统的重要因素。相比于其他机器学习系统，推荐系统更依赖于特征工程。

传统的大规模线性模型（如逻辑回归），通常依赖于人工方式设计特征。同时由于线性模型无法对特征间的非线性关系进行自动建模，通常依赖于特征组合 (Feature combination) 来实现二阶或高阶特征的构造。例如，在应用市场场景下：

- 一些流行的应用比较容易被用户下载，说明应用的标识会影响用户下载（**一阶特征**）
- 吃饭时间一些外卖类的应用比较容易被用户下载，说明应用的类型和时间综合起来会影响用户下载（**二阶特征**）
- 年轻男性喜欢下载射击类和角色扮演类的游戏，说明应用类型、用户年龄和用户性别综合起来会影响用户下载（**三阶特征**）

**人工方式的特征工程，通常有两个问题：**

- 一个是特征爆炸。以通常使用的Poly-2模型为例，该模型采用直接对2阶特征组合建模来学习它们的权重，这种方式构造的特征数量跟特征个数乘积相关，例如：加入某类特征有1万个可能的取值（如APP），另一类特征也有1万个可能的取值（如用户），那么理论上这两个特征组合就会产生1亿个可能的特征项，引起特征爆炸的问题；如果要考虑更高阶的特征，如3阶特征，则会引入更高的特征维度，比如第三个特征也有1万个（如用户最近一次下载记录），则三个特征的组合可能产生10000亿个可能的特征项，这样高阶特征基本上无法有效学习。
- 另一个问题是大量重要的特征组合都隐藏在数据中，无法被专家识别和设计（关于这个的一个有名的例子是啤酒和尿片的故事）。依赖人工方式进行特征设计，存在大量有效的特征组合无法被专家识别的问题。实现特征的自动组合的挖掘，就成为推荐系统技术的一个热点研究方向，深度学习作为一种先进的非线性模型技术在特征组合挖掘方面具有很大的优势。

针对上述两个问题，广度模型和深度模型提供了不同的解决思路。其中广度模型包括FM/FFM等大规模低秩 (Low-Rank) 模型，FM/FFM通过对特征的低秩展开，为每个特征构建隐式向量，并通过隐式向量的点乘结果来建模两个特征的组合关系实现对二阶特征组合的自动学习。

**作为另外一种模型，Poly-2模型则直接对2阶特征组合建模来学习它们的权重。FM/FFM相比于Poly-2模型，优势为以下两点。**

- 第一，FM/FFM模型所需要的参数个数远少于Poly-2模型：FM/FFM模型为每个特征构建一个隐式向量，所需要的参数个数为 $O(km)$ ，其中 $k$ 为隐式向量维度， $m$ 为特征个数；Poly-2模型为每个2阶特征组合设定一个参数来表示这个2阶特征组合的权重，所需要的参数个数为 $O(m^2)$ 。

- 第二，相比于Poly-2模型，FM/FFM模型能更有效地学习参数：当一个2阶特征组合没有出现在训练集时，Poly-2模型则无法学习该特征组合的权重；但是FM/FFM却依然可以学习，因为该特征组合的权重是由这2个特征的隐式向量点乘得到的，而这2个特征的隐式向量可以由别的特征组合学习得到。

总体来说，FM/FFM是一种非常有效地对二阶特征组合进行自动学习的模型。

深度学习是通过神经网络结构和非线性激活函数，自动学习特征之间复杂的组合关系。目前在APP推荐领域中比较流行的深度模型有FNN/PNN/Wide & Deep。FNN模型是用FM模型来对Embedding层进行初始化的全连接神经网络。PNN模型则是在Embedding层和全连接层之间引入了内积/外积层，来学习特征之间的交互关系。Wide & Deep模型由谷歌提出，将LR和DNN联合训练，在Google Play取得了线上效果的提升。

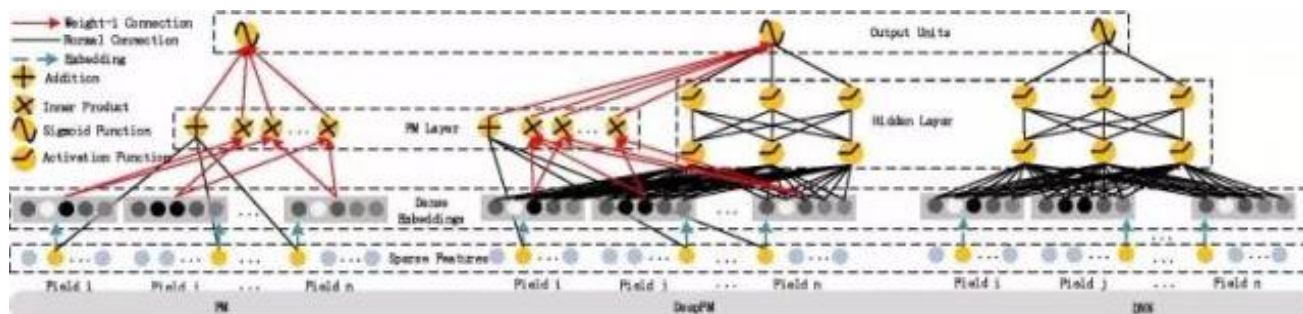
但目前的广度模型和深度模型都有各自的局限。广度模型(LR/FM/FFM)一般只能学习1阶和2阶特征组合；而深度模型(FNN/PNN)一般学习的是高阶特征组合。在之前的举例中可以看到无论是低阶特征组合还是高阶特征组合，对推荐效果都是非常重要的。Wide & Deep模型依然需要人工特征工程来为Wide模型选取输入特征。

DeepFM模型结合了广度和深度模型的有点，联合训练FM模型和DNN模型，来同时学习低阶特征组合和高阶特征组合。

此外，DeepFM模型的Deep component和FM component从Embedding层共享数据输入，这样做的好处是Embedding层的隐式向量在(残差反向传播)训练时可以同时接受到Deep component和FM component的信息，从而使Embedding层的信息表达更加准确而最终提升推荐效果。

DeepFM相对于现有的广度模型、深度模型以及Wide & Deep模型的优势在于：

- (1) DeepFM模型同时对低阶特征组合和高阶特征组合建模，从而能够学习到各阶特征之间的组合关系；
- (2) DeepFM模型是一个端到端的模型，不需要任何的人工特征工程。



## DeepFM算法架构

为了验证DeepFM模型的精度，分别在Criteo Kaggle的竞赛数据集和华为应用市场的数据集上进行了实验。

Criteo Kaggle的数据集按照9:1的比例划分训练集和测试集。华为应用市场数据集由8天连续数据构成：前7天数据为训练集，后1天数据为测试集。实验结果表明：相比较业界最新的推荐算法，DeepFM模型在华为数据集上，AUC提升0.36%~0.86%，LogLoss提升0.34%~1.1%。

	Company*		Criteo	
	AUC	LogLoss	AUC	LogLoss
LR	0.8641	0.02648	0.7804	0.46782
FM	0.8679	0.02632	0.7894	0.46059
FNN	0.8684	0.02628	0.7959	0.46350
IPNN	0.8662	0.02639	0.7971	0.45347
OPNN	0.8657	0.02640	0.7981	0.45293
PNN*	0.8663	0.02638	0.7983	0.45330
LR & DNN	0.8671	0.02635	0.7858	0.46596
FM & DNN	0.8658	0.02639	0.7980	0.45343
DeepFM	<b>0.8715</b>	<b>0.02619</b>	<b>0.8016</b>	<b>0.44985</b>

DeepFM算法效果对比

**DeepFM算法是华为伏羲推荐算法经过1年多的研究，并同业界深入合作的成果。**团队还在深度学习推荐算法方面持续投入，研究更好的深度学习网络架构，实现更好的特征表达和特征交互方式，持续改进深度学习推荐模型，为华为应用市场和华为游戏中心的用户提供更好的个性化、精准化的用户体验，保持华为应用市场推荐系统的技术领先。