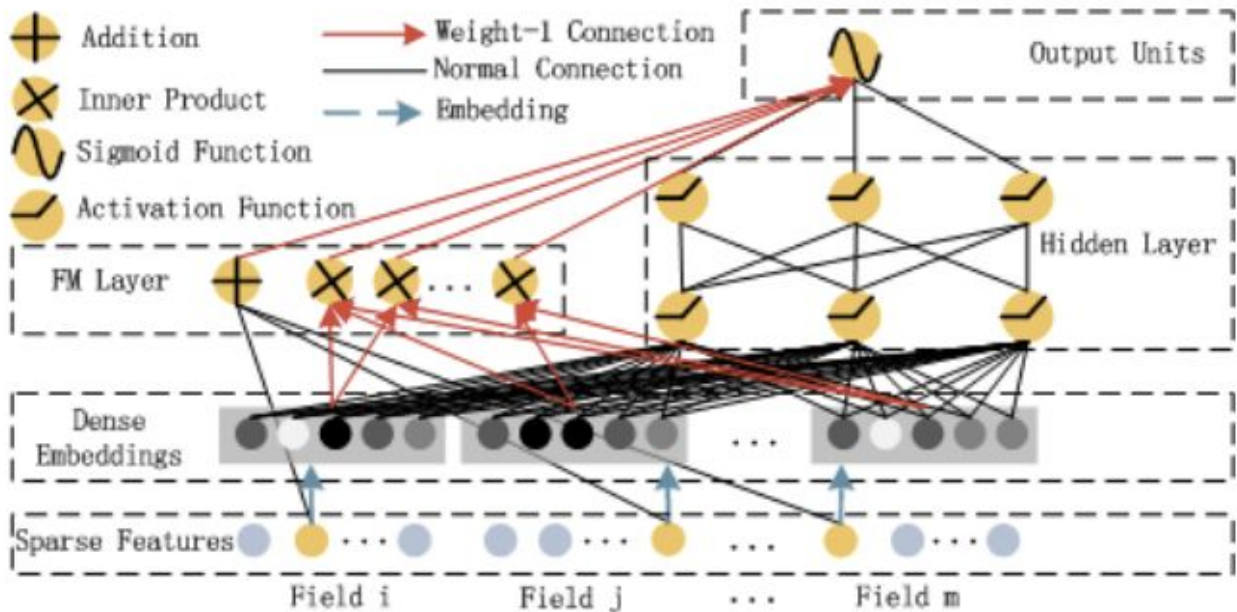


# 推荐算法之DeepFM

原创 易尔山 机器学习简明教程 2019-05-06



论文地址: <https://paperswithcode.com/paper/deepfm-a-factorization-machine-based-neural>

DeepFM由华为诺亚方舟实验室的几个算法研究员在2017年3月联合发表。它融合了FM模型与Deep模型，兼具记忆性与泛化性，省去了繁杂的特征工程部分。从效果看，DeepFM在AUC和LOGLOSS两项评估指标评估上都领先于其他模型。

DeepFM有5个模块：Sparse Features，Dense Embeddings，FM Layer，Deep Layer以及Output Units。

下面看一下每个模块到底做了什么？

## 一、Sparse Features

DeepFM是一个多输入模型，输入多个稀疏特征，后续需要针对每个特征单独做处理。

## 二、Dense Embeddings

该模块对每个稀疏特征做embedding，学习到他们的embedding向量。需要注意的一点是每个稀疏特征对应的embedding向量的维度是一样的，因为后续这些embedding向量需要送到FM层做内积。

## 三、FM Layer

FM层由两个部分组成，一部分是原始特征的相加，另一部分原始特征embedding后的embedding向量两两内积。这两部分的结果级联后，作为FM层的输出。FM层跟机器学习算法FM的做法是一样的，既需要考虑每个特征的信息，又要考虑特征之间的交叉信息。

### 三、Deep Layer

Deep层是将每个embedding向量做级联，之后做多层的全连接，学习更深的信息。

### 四、Output Units

FM层的输出级联Deep层的输出，接一个dense，作为最后的输出。

### 五、Evaluation

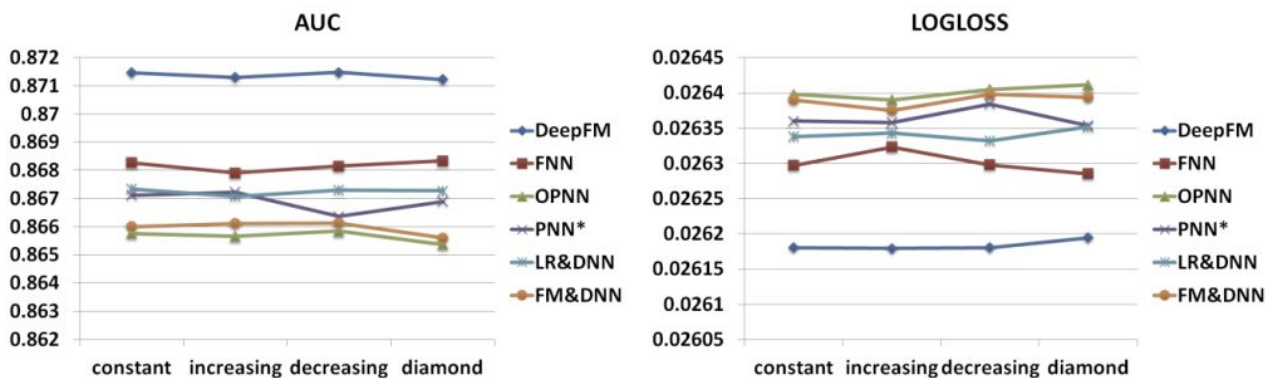


Figure 11: AUC and Logloss comparison of network shape.

我们可以看一下原论文中各模型的AUC，LOGLOSS这两项指标。DeepFM的AUC在0.8715左右,LOGLOSS在0.02615左右，远好于其他模型。