[深度模型] 谷歌多任务学习模型MMoE

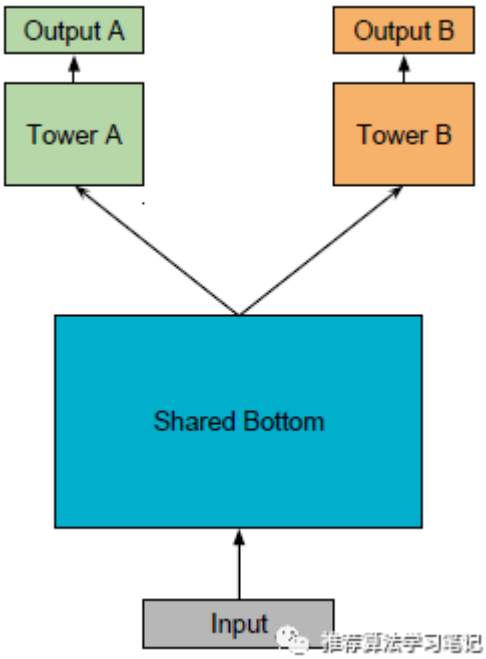
原创 xxxhuang 推荐算法学习笔记 5月20日

在前面的文章CVR深度预估模型ESMM：阿里是怎么做点击后的转化率预测的中，我们曾经介绍过阿里的ESMM模型是怎么通过构造多任务模型来做转化率预估的。而在本文介绍的google发表在kdd2018的paper《Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts》中，作者给我们介绍了一种基于Multi-gate Mixture-of-Experts（MMoE）的多任务学习模型。

一. 背景

在推荐系统当中，除了想要预估item的点击率的同时，我们可能还想要预估其他目标，例如点赞，评论，时长等等。我们希望在一个模型里面可以同时学习到多个目标。这就是多任务学习。

二. Shared-bottom多任务模型



在多任务学习模型当中，最常见的一种模型就是Shared-bottom模型。首先每个任务共享底部的network，然后再根据任务的个数在上面划分出多个tower network来分别学习不同的目标。转换成公式是

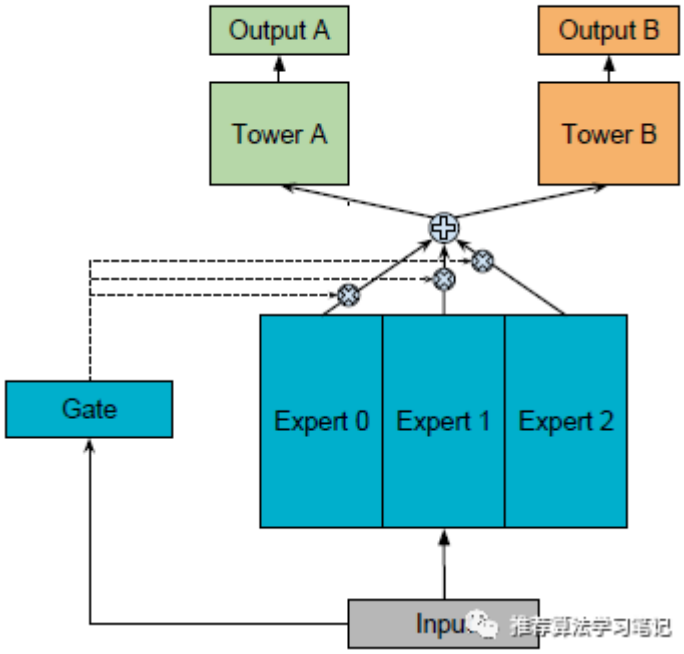
$$y_k = h^k(f(x)).$$

推荐算法学习笔记

其中f表示shared-bottom network， **h**k表示第k个tower network，**y**k是最终的输出。

对于Shared-bottom多任务模型，它的优缺点如下
优点：降低overfit风险，利用任务之间的关联性使模型学习效果更强
缺点：任务之间的相关性将严重影响模型效果。假如任务之间相关性较低，模型的效果相对会较差。

三. Mixture-of-Experts(MoE)多任务模型



在MoE中，我们引入了一个叫Expert的概念，每个Expert其实就是一个feed forward network。Gate为每个Expert输出一个标量。因此整个模型就像一种Ensemble的方式，由专家给出自己的“建议”，gate根据input去给每个专家一定的权重，然后最终根据加权后的专家意见给出最终的结果。
转换成公式如下所示

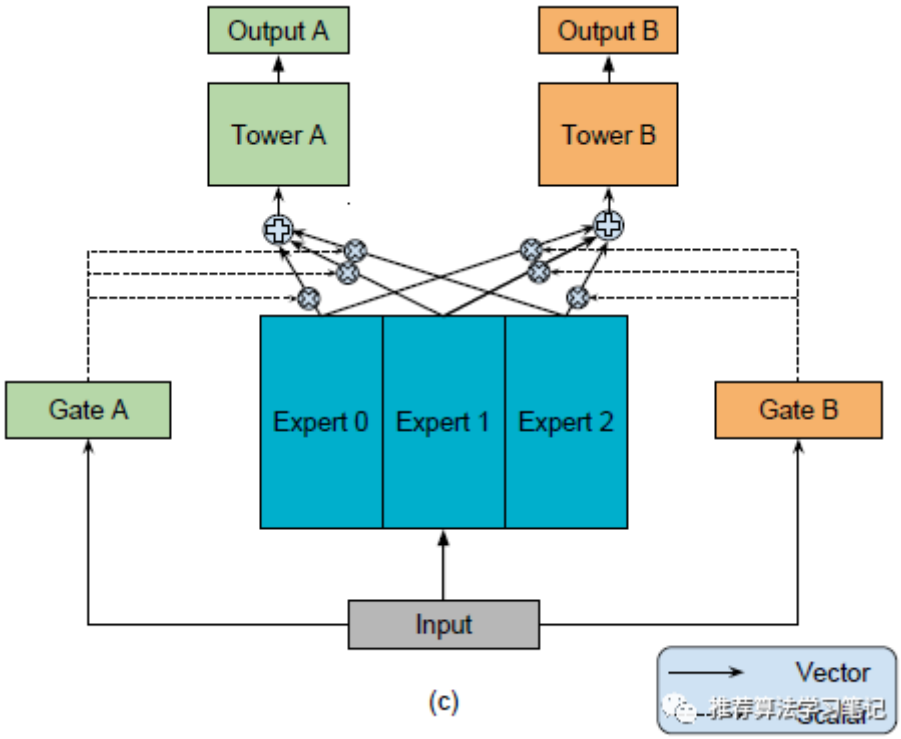
$$y = \sum_{i=1}^{n} g(x)_i f_i(x),$$

推荐算法学习笔记

其中**f**i表示第i个专家，**g**i表示第i个专家的权重

四. Multi-gate Mixture-of-Experts(MMoE)多任务模式

为了解决任务之间相关度降低导致模型效果下降的问题。paper作者在MoE的基础上进行了改进，引入了多个gate的模式来控制不同任务不同专家的权重，如下图所示



(c)

转换成公式如下所示

$$y_k = h^k(f^k(x)),$$

$$\text{where } f^k(x) = \sum_{i=1}^{n} g^k(x)_i f_i(x).$$

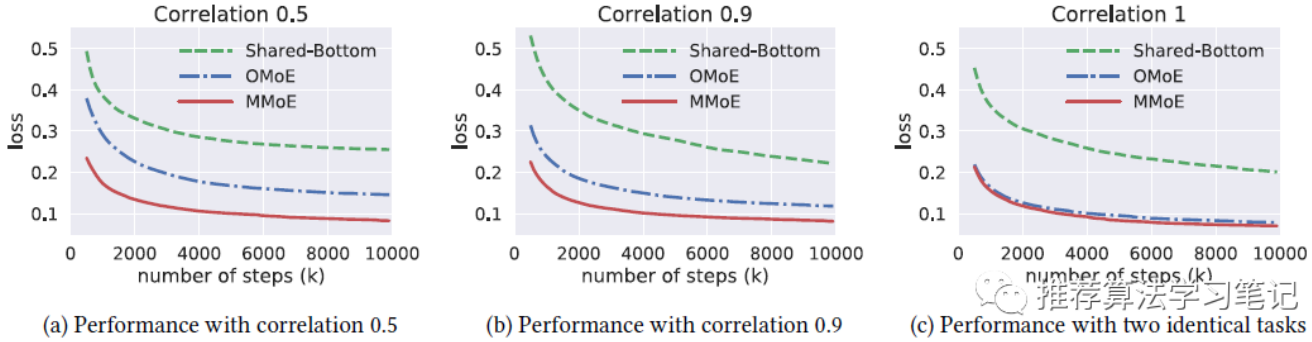
其中**f**i(x)表示第i个专家，**g**k(x)i表示第k个任务中第i个专家的权重，**h**k是第k个任务的tower network

gk(x)i公式就是一个线性模型+softmax，如下所示

$$g^k(x) = \text{softmax}(W_{gk}x),$$

五. 实验结果

在paper中作者针对多任务不同的相关度分别对这三个模型进行了实验，如下图所示



(a) Performance with correlation 0.5 (b) Performance with correlation 0.9 (c) Performance with two identical tasks

可以看到，不管多任务之间的相关度是0.5，0.9和1，MMoE的表现都要好于MoE，MoE的表现要好于Shared-bottom

六. 总结

以上就是MMoE的全部内容，如果有问题，欢迎和我联系。在下一篇中，我将会分享MMoE在Youtube相关推荐上的应用，敬请期待~