

【面试版】CTR预估模型-DeepMCP

原创 小橙子爱人工智能 小橙子爱人工智能 2019-11-01

这一周下班后有一点沉迷下棋，耽误了更新。另外一点原因是一直没有太理解论文里面的关联子网络的输入是什么，这一点论文里面讲得不是很清楚，结合代码来看终于明白了。这是阿里最新发表在IJCAI2019的一篇关于CTR预估的论文。网络结构虽然不复杂，但是思想精妙。下面依旧采取面试提问形式讲解。

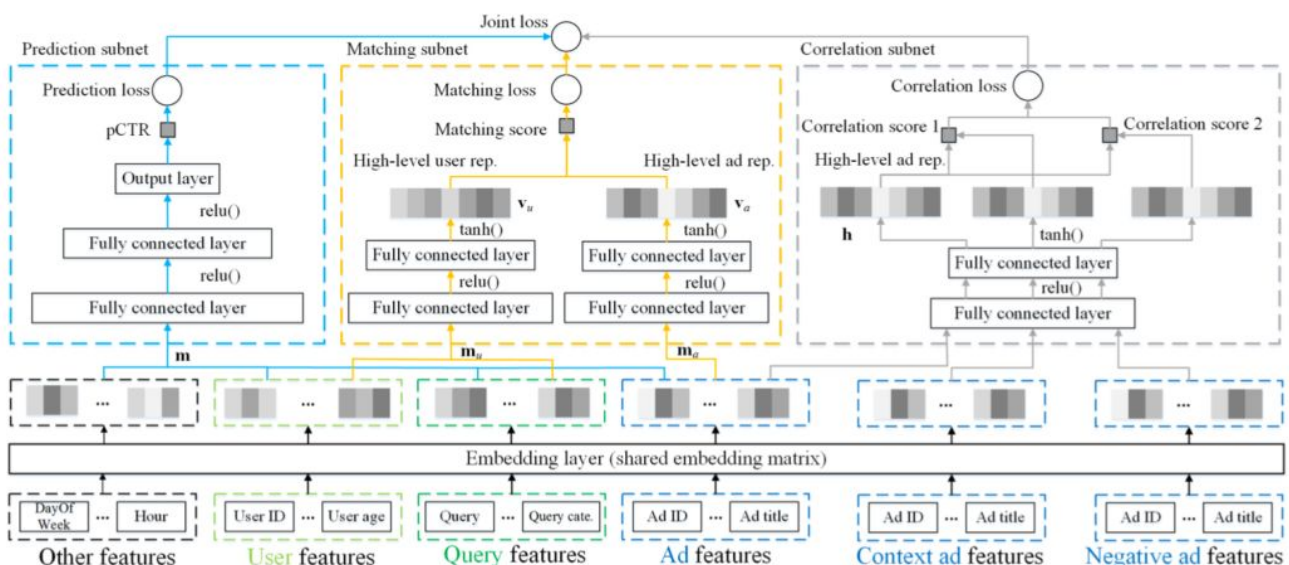
1. 什么叫DeepMCP?

这个模型由三部分模型组成，匹配（Matching），关联（Correlation）和预测（Prediction）网络，每一部分网络都是深度学习的模型。

2. 三个子网络分别的作用是什么?

Prediction网络就是常规的CTR预估的深度学习模型，可以是已经非常熟悉的DNN，Wide&Deep，DeeFM，DIN等网络，输入的特征是用户维度特征，query维度特征，广告维度特征和其他特征。Matching网络是用来建模用户与广告之间的关系，并且在用户的embedding和广告的embedding上体现出来，该子网络的输入是用户维度特征，query维度特征和广告维度特征。Correlation网络是用来建模广告与广告之间的关系，并能在广告的embedding上体现出来，该子网络的输入是三种类型的广告的特征。

3. DeepMCP的网络结构是怎样的?



前面已经讲了各个子网络的输入特征情况。三个子网络共用同一个embedding层。在Prediction网络中，各embedding拼接后输入一个DNN网络，label为是否点击广告。

Matching网络中，将用户维度和query维度特征拼接输入一个DNN网络，广告维度特征输入另外一个DNN网络，可以理解为双塔模型，DNN网络作为表示层，DNN后的激活函数用tanh，然后再进行计算匹配分，label也为是否点击广告。Correlation网络中，输入比较特别，与前两个子网络的输入不是同一份数据，而是单独的数据，每一条输入包括一个目标广告，一个上下文广告，以及若干个负采样广告，所有的广告特征embedding之后共用同一个DNN网络作为表示层得到高阶特征，与Matching网络一样DNN网络之后用tanh作为激活函数，然后分别计算目标广告与上下文广告以及负采样广告的关联分，上下文广告label为1，负采样广告label为0，Correlation网络实际上可以理解成skip-gram模型计算广告的embedding。最后将三个子网络计算得到的loss相加。

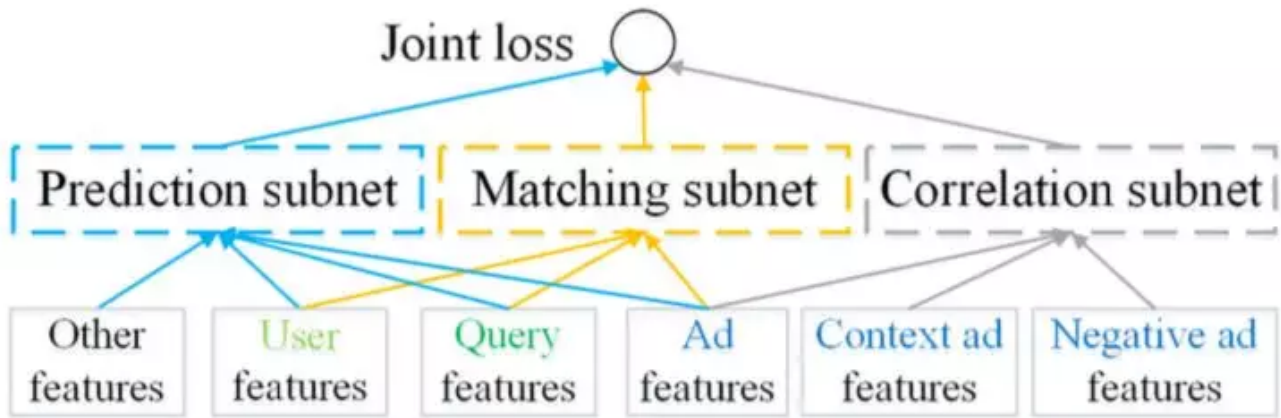
4. 在Matching子网络和Correlation子网络中，DNN的表示层之后为什么用tanh做激活函数而不用relu？

在全连接层之后，如果用relu做激活，会出现很多0。在接下来的匹配操作中，向量的内积容易计算得到0，而tanh不会出现这种情况。

5. 如何理解Correlation子网络的输入？

借助word2vec的思想。找到用户的点击广告列表，选定一个固定大小的窗口，每次选一个点击广告列表中的一条广告作为目标广告，在窗口内找一条也点过的作为上下文广告，再找出给用户展现过的但是没有被点击的广告作为负采样广告。

6. 如何理解这样的网络设计？



(a) DeepMCP - Training



(b) DeepMCP - Testing

Matching子网络和Correlation子网络只在训练的时候起作用，在预测时不计算。结合三个网络共用同一个embedding层，可以理解为，Matching子网络和Correlation子网络是在训练的时候，同时训练更精确的embedding层，服务于预测阶段。如果没有Matching子网络和Correlation子网络，训练阶段就无法学习到用户和广告以及广告和广告的关系，无法在embedding层中体现，则预测网络就获取不到这类信息。

7. 评估模型的指标是什么？

CTR预估常用的AUC和Logloss。

刚看这篇论文仿佛发现了新大陆，感慨于网络设计之巧妙，用两个辅助网络训练embedding层服务预测网络。正好工作中也有过类似的疑惑，现在就能完全明白了。关注后在公众号后台发送“DeepMCP代码”可直接获取开源代码链接。下一篇接着写让我有发现新大陆感觉的论文，同样来自阿里的最新CTR预估论文。