

2020年6月22日多目标模型之MMoE

原创 mrchor 软客圈 6月22日



前言

随着深度学习的不断发展，其在推荐场景下的应用也越来越广泛了。在业界研发人员不断追求模型CTR效果的同时，深度学习模型复杂度也在急剧上升，然而由于推荐业务场景本身的低响应要求，导致我们希望在为线上用户推荐TA喜欢的物品的同时，还要保证模型足够简洁，以达到线上算力的要求。

按照当前DL-Serving的性能，深度学习的CTR模型几乎已经达到了复杂度的瓶颈，这就要求我们需要换一种思维去提升业务效果，于是我们另辟蹊径，从同时预测用户多种行为的角度上思考推荐。本期开始讲解多任务模型在推荐业务场景的演化，以及多目标任务可以取得的效果。

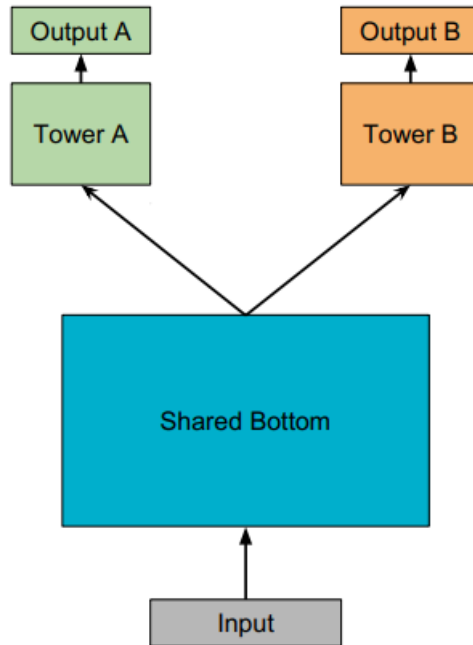
多任务优化

何为多任务优化？顾名思义，多任务优化，即一个模型的产出可以同时多个目标或任务进行推理，多目标之所以现在走红，是因为它一定程度上简化了我们对业务的细化分。例如，在电商场景中，传统的CTR模型，只会推理出用户要不要点，而不会对后续用户的是否下单有任何的关联性推理（如需推理则需要再做一个下单的模型），而多任务则不同，它不仅仅要推理出用户是否点击，还要推理出用户点击后是否会对商品下单，从而增加用户对商品的购买意愿，这样我们可以在同一个模型中推理出用户在点击后的并行行为，而不需要对点击场景和购买场景分开做处理，从而简化了业务场景分类。

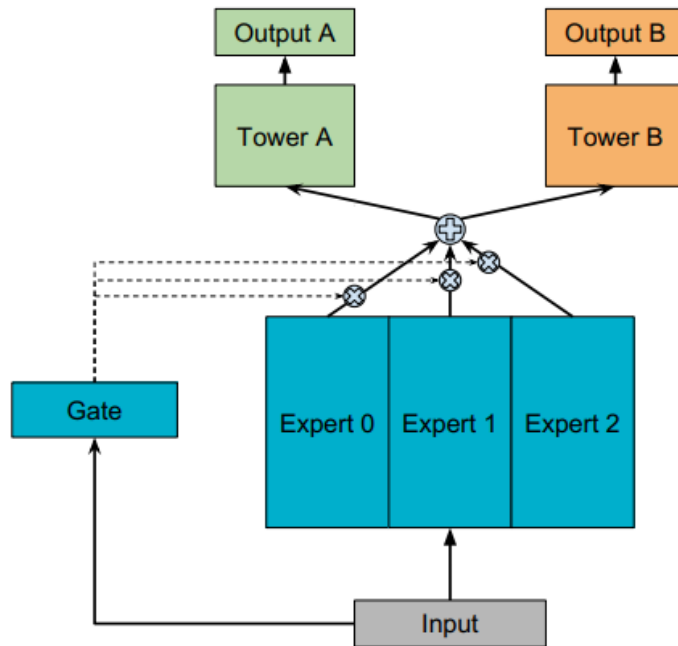
本期主要介绍由Google在KDD 2018中提出的Multi-gate Mixture-of-Experts，简称MMoE。我们从目的、模型架构和实验等方面阐述。

MMoE主要解决什么问题？

在MMoE提出之前，多任务模型已经有许多经典架构被提出，其中绝大多数的优化都基于share-bottom架构，即不同的任务共享相同的feature或feature_map。



然而，这种架构极大地限制了模型表达的能力，为什么这么说？因为我们在共享特征的上层直接接入了多个目标的输出，而由于多个任务各自有不同的数据分布，也就是说我们对不同任务的输出具有一定的差异性，而相同的特征输入会极大地削弱模型的多任务输出表达而在某种程度上降低了多目标模型的泛化能力。

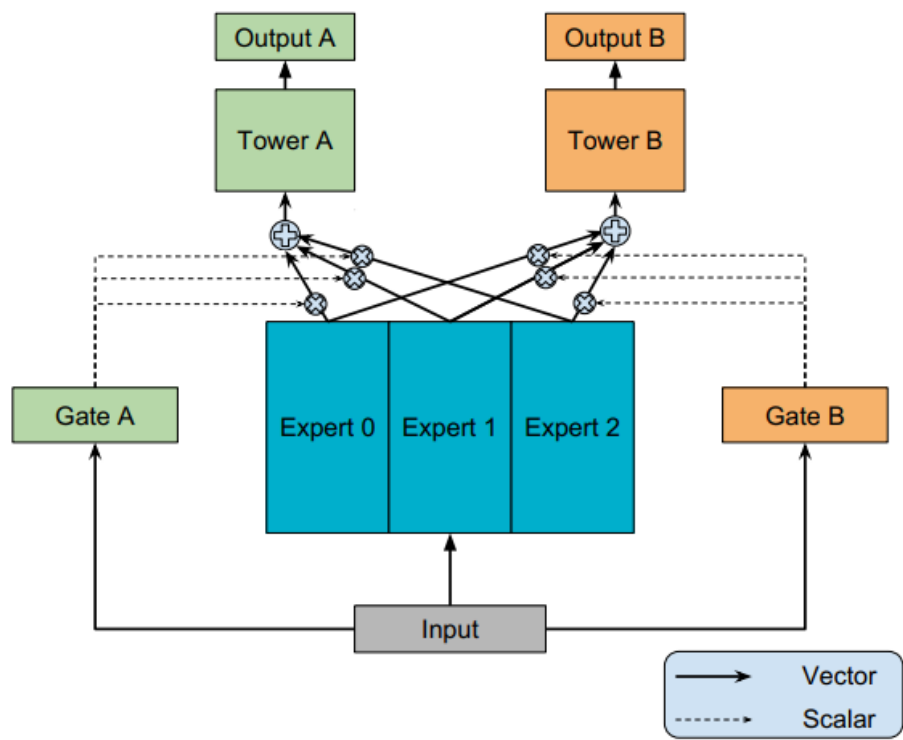


那么，如何去降低这种架构带来的影响，作者首先提出了One-gate MoE model，这种模型架构就在一定程度上解决了上述问题，即虽然说多个模型还是共享相同的输入特征，但是每个任务都利用“gate network”来区分特征表达的权重，从而提高了模型的表达能力。但是这种模型架构的“区分”还不是很大，毕竟输入的特征还是只有一个，于是作者受集成学习（ensemble learning）思想的影响，提出了multi-experts，即我们可以把单个的共享特征看做是一个弱学习器的输入，那么，根据集成思想，若干弱学习器的组合可以作为一个强学习器来对结果进行推理，再通过“gate network”就可以极大地提高多任务模型的表达能力了。

说模型，不如说是一种框架！

根据上一小节问题的提出以及解决方案的设计，作者提出了MMoE模型，如图是模型的整体架构。由图可以看出，我们在定义两个任务优化模型时，在特征输入阶段，进行特征转换，分别产生若干expert，作为我们模型的基学习器，然后在每个任务对应的输入，分别使用“gate network”来表征最后的结果输出。从某种角度上来讲，MMoE的厉害之处在于它的expert可以定义成任何一种单独的模型，可以说，MMoE是一个框架，而非简简单单的一个模型！这就使得我们在expert上可以灵活自由地实现自己想要的模型设

计。



实验对比

在本部分，作者主要做了两个实验，一个是UCI的人口收入普查，另外一个Google的大规模内容推荐。第一个实验我们暂且不谈，我们主要是说一下Google的内容推荐（YouTube）。在这个实验中，作者使用了超过300亿的用户行为数据作为模型训练的输入，并且在100万用户行为进行离线实验评估，而所优化的任务分别是交互和满意度。以下是与基线的离线实验结果评估对比：

Table 3: Engagement performance on the real large-scale recommendation system.

Metric	AUC@2M	AUC@4M	AUC@6M	R2@2M	R2@4M	R2@6M
Shared-Bottom	0.6879	0.6888	0.6900	0.08812	0.09159	0.09287
L2-Constrained	0.6866	0.6881	0.6895	0.08668	0.09030	0.09213
Cross-Stitch	0.6880	0.6885	0.6899	0.08949	0.09112	0.09332
OMoE	0.6876	0.6891	0.6893	0.08749	0.09085	0.09230
MMoE	0.6894	0.6897	0.6908	0.08978	0.09263	0.09362

实验结果表明，MMoE在不同的参数设置上的实验AUC均超过了其他基线模型，这就证明了MMoE模型的有效性。并且作者对MMoE做了上线，也得到了比较不错的结果：

Table 4: Live experiment results

Live experiment	Engagement Metric	Satisfaction Metric
Shared-Bottom Improvement over Single-Task	-0.22% *	19.72% **
MMoE Improvement over Shared-Bottom	0.25% **	2.65% **

* indicates confidence interval level 90%

** indicates confidence interval level 95%

总结

笔者认为，随着在线推理服务瓶颈的到来，多任务模型是未来的一个趋势，因为在推荐场景中，人的交互行为是并发甚至是并行的，所以我们单单去预测用户的某一种行为，可能会在某种角度上降低了用户其他行为的体验，而随着多目标模型的发展，我们可以同时对用户的不同行为做判断，那么这会在一定程度上提高用户的体验，因此，多任务模型，未来可期！

相关文献：<https://dl.acm.org/doi/10.1145/3219819.3220007>

如果你喜欢我的文章，欢迎关注我的微信公众号【软客圈】（ID: recoquan）
纯手工打造，实属不易，欢迎大家分享和转发~
原创内容，转载需注明出处，否则视为侵权并将被追诉！

