MoSE理论与实践

# 背景

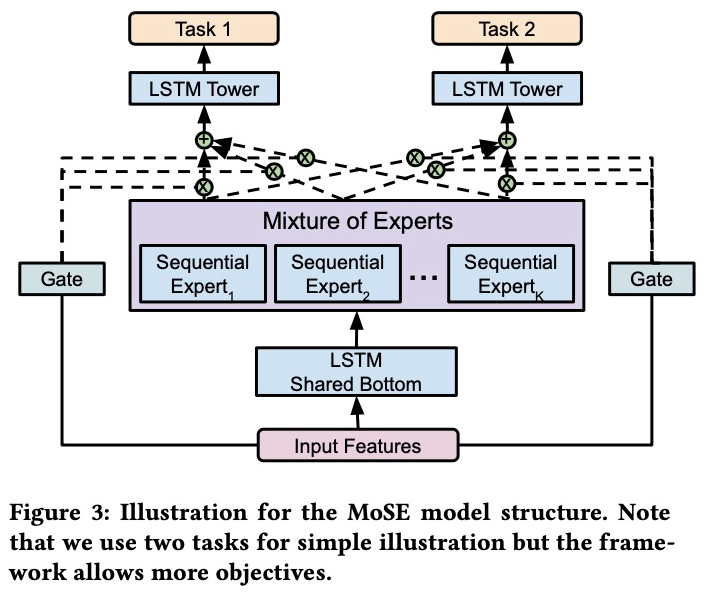
MMoE多任务学习框架由于其模型的简单，已经在工业界进行了大量的应用。然而，因为MMoE中experts的足够“简单”，使其在表达各个任务的同时，不能很好地对用户的行为序列进行解析，因此就会导致MMoE模型在进行多任务学习时，不能很好地对序列行为数据进行有效表达。

基于此，Google研究者提出了MoSE模型，在expert中集成了Long Short-Term Memory (LSTM)，旨在充分利用用户的序列行为数据，更好的达到业务需求。

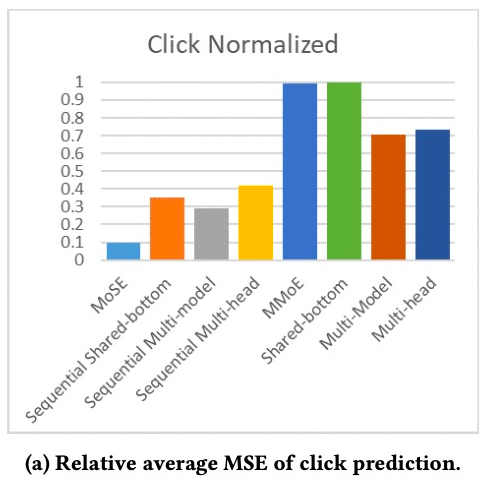
# 模型

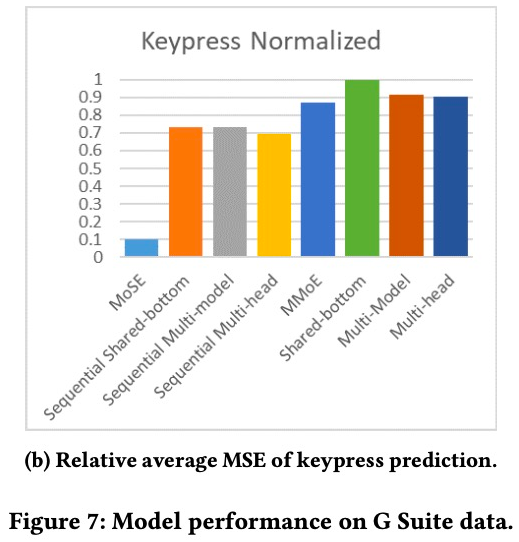
MoSE模型结构如下：

* 模型采用LSTM做share bottom，充分利用user的行为序列数据；
* 顺序expert层的混合，通过使用LSTM代替全连接网络，进一步增加了MoE层，更好地处理序列数据。
* 用门控网络对expert的输出进行门控。每个门控网络可以学习“选择”一个子集的expert使用条件输入的例子，这允许在不同的变量之间建模复杂的交互作用。
* 每个任务的Tower网络采用LSTM机制；



# 实验结果





论文比较了MoSE和七个备选模型在G Suite数据集上的效果，结论如下：

* MoSE的模型性能优于其他模型，尤其是在复杂的真实数据集上。
* 序列模型的性能优于非序列模型，这说明了在用户活动流中显式地建模序列依赖关系的必要性。
* MoSE性能优于其他序列模型。这说明MoSE模块能够有效地处理用户活动流中的稀疏变量和异构数据源之间复杂的交互等问题。
* 与其他非序列模型相比，MMoE本身并没有显示出显著的优势。当在MoSE中使用顺序expert时，expert混合框架是最有益的，因为用户活动流中的大多数复杂性似乎源于顺序复杂性和稀疏性。

# 缺陷

MoSE模型在序列化建模中可以取得不错的效果，但是采用LSTM建模，模型的运行时间长、速度慢，线上响应也会是个问题。

参考：

<https://github.com/ShaoQiBNU/Google_MTL>