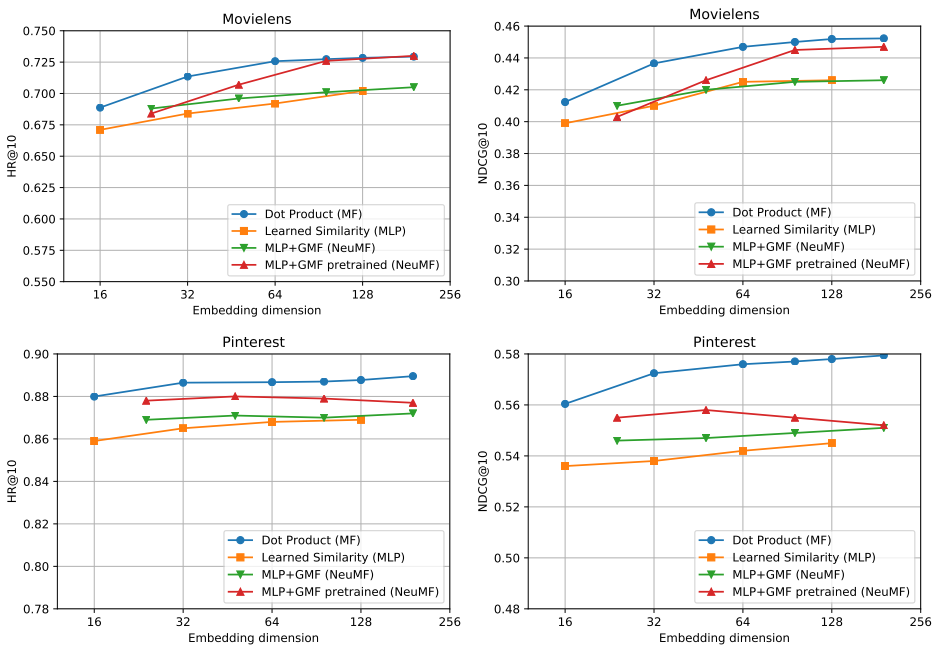
1. **NeuMF**

近年来，在推荐系统领域，许多工作使用学习到的相似度代替点积，比如：在神经协同过滤算法中使用MLP完成相似度学习。针对于此，Steffen Rendle等人，重新审视神经协同过滤与矩阵分解方法，指出：**经过适当的超参选择，点积实质上优于所提出的学习相似度。同时，MLP虽然理论上能近似任何函数，但其在学习点积的过程中存在某种程度的困难**[[1]](#footnote-0)。

具体而言，在使用相同损失和负采样方法的前提下，Steffen Rendle等人分别在数据集MovieLens和Pinterest下，采用HR和NDCG作为性能度量指标，比较经过适宜调参的MF与MLP、NeuMF及Pretrained NeuMF等方法间的推荐性能。具体实验结果，如下图所示。从中可以看出，相较于MLP，结合GMF和MLP的NeuMF在性能上仅有微小改善，总体推荐质量尚不如采用点积方法计算相似性的MF。此外，结合预训练的GMF和MLP，进行微调学习，在推荐性能上具有更为直观的改善效果，甚至在某些设置下效果同MF相当。

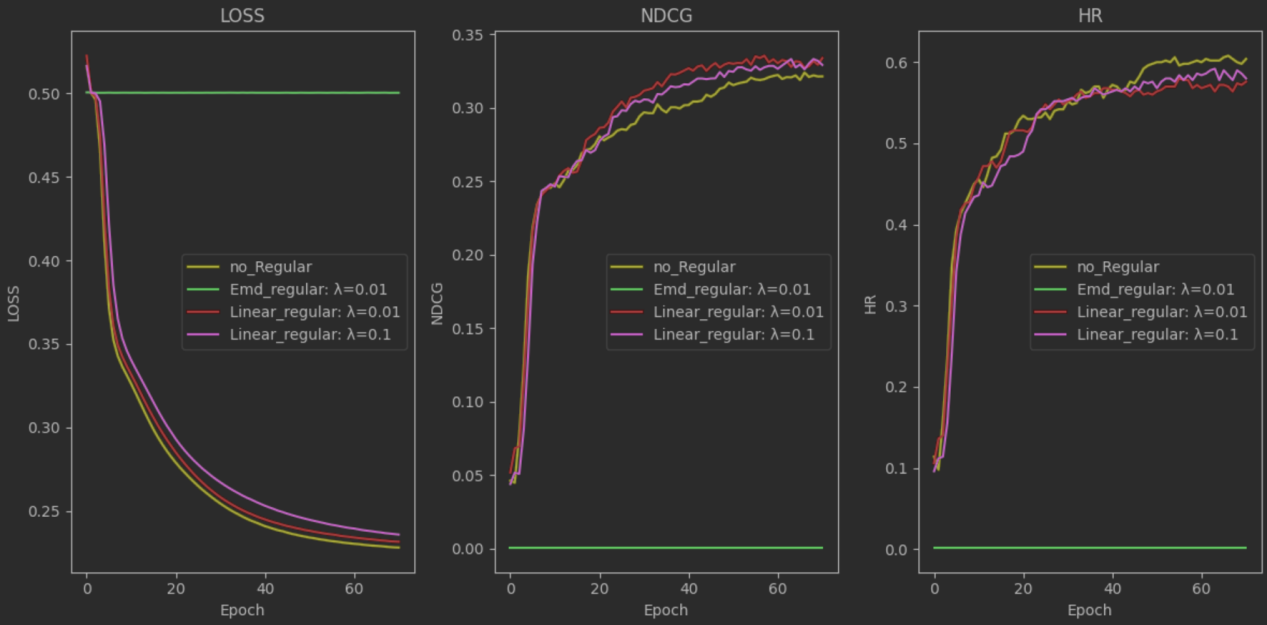


1. **关于GMF**

**问题**在于：虽然NeuMF通过额外引入GMF模块的方式，一定程度上提高了推荐性能。但是同MF相比较，为什么NeuMF的推荐性能总体上提升有限？换而言之，相较于MF，GMF在使用加权点积计算相似性的过程中，如何理解这种关于“额外引入可学习权重W”的差异，对于GMF模块性能表现的影响。

参考Steffen Rendle等人的观点，本节实验选择从正则化的角度出发，**探究**GMF由于前述差异所造成的不利影响（调参复杂性和过拟合风险），以及不同正则化措施的实际效果。具体而言，本次实验在维数为（500，3706）的用户-物品交互数据集上，选用NDCG和HR作为性能度量指标，比较无正则化、仅嵌入层正则化（λ=0.01）、仅点积权重正则化（λ=0.01）和仅点积权重正则化（λ=0.1）对照组的迭代训练效果。

**实验结果**，如下图所示。其中，在最小化损失的过程中，考虑对嵌入进行L2正则化，将导致GMF模型失去迭代训练效果（图中绿线部分）。同时，对于点积权重的正则化处理，虽然某种程度上能够提升模型在NDCG上的性能表现，但是这种措施在HR上效果相反。此外，也可以发现对点积权重设置不同程度的正则约束，能够影响模型的推荐性能。



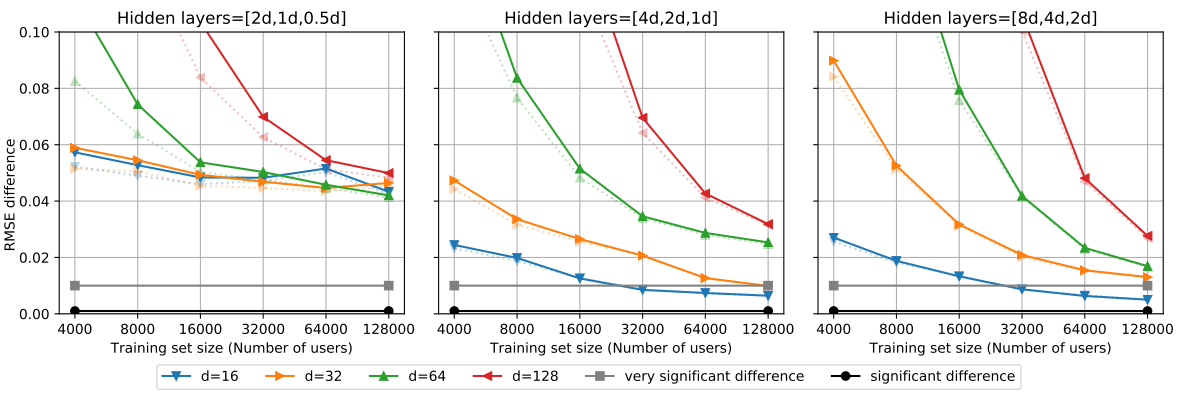
针对实验结果，可进行如下**结果分析**。其一，在使用小型数据集的情况下，对嵌入进行L2正则化将导致嵌入权重的取值更加平滑，这使得嵌入层学习特征表示的能力明显受限，甚至出现训练失效现象。其二，针对点积权重的正则处理，能提升GMF模型在NDCG上的性能表现。这表明相较于MF，GMF关于“额外引入可学习权重W”的差异，确实在提高GMF模型复杂性的同时又带来了过拟合风险。其三，虽然正则化能通过缓解过拟合的方式，提升推荐的排序性能。但是这种措施可能会导致推荐策略相对保守，忽略少数用户的个性化需求，从而使得考虑对点积权重正则化处理的GMF在HR上（更关注在Top-N推荐列表中的覆盖率）表现不佳。这表明，在具体应用情景中，对推荐模型的不恰当正则处理，将对其性能表现造成不利影响。

1. **关于MLP**

如前所述，至少在输入中仅考虑用户或物品ID的情况下，GMF关于额外引入点积权重的差异，将导致其在Top-N推荐任务中的性能表现，反而不如使用点积计算相似性的MF。至此，另一个值得探究的**问题**在于：相较于MF，若忽略NeuMF中GMF导致的性能差异，为什么NeuMF通过额外结合MLP模块学习相似性的方式，未取得显著的性能提升效果。换而言之，针对用户-物品交互数据，为什么用MLP近似点积，存在某种程度的困难？

为了**探究**MLP对于点积结构的拟合学习能力，Steffen Rendle等人采用RMSE作为性能度量指标，来比较MLP在不同训练数据量级和隐藏层宽度的情况下，同点积结果间的差异。

**实验结果**，如下图所示。其中，当训练数据量相同时，MLP采用较小的嵌入尺寸（d）可以更好地近似点积结果；MLP采用更宽的隐藏层，也能一定程度上提高对点积的近似效果。然而，针对采用更大嵌入尺寸的MLP而言（图中红线），有时候即使投喂更多的训练数据，也难以取得相对理想的点积近似效果。



根据前述实验结果，可得出如下**结论**。其一，MLP理论上确实可以学习拟合任何函数，至少在使用足够多的训练数据和足够宽的隐藏层的情况下，MLP是可以近似点积的。其二，实际上MLP在近似学习点积结构的过程中，有时候即使采用更复杂的MLP模型结构，也很难显著地拟合学习到目标水平。这表明，在组合嵌入的过程中，需要谨慎考虑采用MLP完成相似性的计算。除非数据集很大或嵌入维数很小，否则MLP关于归纳偏置方面的非线性能力，反而在相似性学习的过程中属于累赘。相反，这种情况下，点积则是一个简单而又高效的选择。

1. **关于NeuMF的讨论**

3.1 争议

实际上，很多人质疑NeuMF工作中关于MF和NeuMF性能对比结果的说服力，比如：原文中NeuMF采用的损失函数为log loss，而进行比较的基准模型MF则采用eALS和BPR loss作为损失函数。一般情况下，即使是相同的模型，不同损失函数的优化结果差异也很大。因此，Steffen Rendle等人重新审视神经协同过滤与矩阵分解方法，指出：在初始输入中仅考虑用户或物品ID的情况下，MF通过点积计算相似性的方式，性能表现上优于GMF加权点积计算相似性的方式，甚至优于通过MLP学习相似性的方式。只有结合预训练MLP和GMF模块的NeuMF，在特定情况下性能表现才能同MF相当。

然而，有人认为，由于Rendle是自己手算微分进行梯度下降实现的MF，而NCF则是采用keras自动微分的框架和分批次更新的方法，不排除它们之间的实现差异，从而导致MF在Top-N推荐任务上表现更好[[2]](#footnote-1)。

3.2 共识

考虑到计算资源的限制，NeuMF使用**负样本采样**的方式完成测试，存在一定程度上的偏差，并不能完全反映Top-K推荐的结果。同时，虽然MLP理论上具备更强的表达能力，但是NeuMF模型确实没有通过考虑适宜正则化或dropout的措施，来应对神经网络在学习交互函数过程中的**过拟合问题**。此外，在稠密数据上采用较大的嵌入尺寸，矩阵分解等线性方法效果更好，神经网络则难以发挥其非线性学习交互函数的优势。这是因为，线性模型能更准确地拟合稠密**数据**的整体分布，不需要过度的非线性拟合；相比之下，稀疏数据则需要更多的非线性建模能力，用于捕获数据中的模式。而且，采用较大的**嵌入尺寸**时，线性模型能够更好地捕获特征间的复杂关系，而神经网络则可能会由于过度拟合或参数量过多而难以训练。实际上，以前基于skopt自动调参NeuMF模型的实验结果也恰好映证了这点，即：相较于GMF的最佳嵌入尺寸，MLP模块的最佳嵌入尺寸参数取值很小。

3.3 启示

结合以前关于NeuMF模型的调参经验，原文提出的NeuMF实际上仅是一个如何使用神经网络来实现协同过滤方法的框架，我们可以结合具体的应用情景，**灵活调整输入与交互函数的建模方式**。比如，如若处理的数据规模太大，由于计算资源的限制将不能使用过大的嵌入尺寸来表示用户或物品的潜在特征时，实际上使用神经网络来学习相似性优于点积方法。或者说，当输入特征不仅有用户或物品ID，还有history embedding和side information等更丰富的信息时，相较于点积等线性模型，神经网络能够学习更复杂的非线性交互。又或者说，从工业实现的角度出发，在召回阶段考虑到时间效率，使用点积作为匹配函数更为合适；在精排阶段考虑到对时间效率的要求不高，如若需要追求更精确且拟合能力更强的模型，使用神经网络作为匹配函数则更为合适。

值得注意的是，关于MLP难以学习点积的结果，并不是质疑神经网络在推荐系统领域的应用价值。实际上，在使用深度学习方法解决其它领域特定问题的过程中，也常需要将MLP替换为更专门的结构，从而用于更好地完成特定模式的提取表征。例如，在图像分类任务中，使用CNN等结构能够更为有效地表征输入数据的空间结构；在机器翻译任务中，使用RNN等结构能够更好地考虑上下文情景。前述提及的所有专门网络结构，对于推进使用深度学习方法来解决各领域的关键问题至关重要，即使它们都可以由MLP进行近似。

1. Rendle S, Krichene W, Zhang L, et al. Neural collaborative filtering vs. matrix factorization revisited[C]//Proceedings of the 14th ACM Conference on Recommender Systems. 2020: 240-248. [↑](#footnote-ref-0)
2. https://www.zhihu.com/question/396722911/answer/1251909703 [↑](#footnote-ref-1)