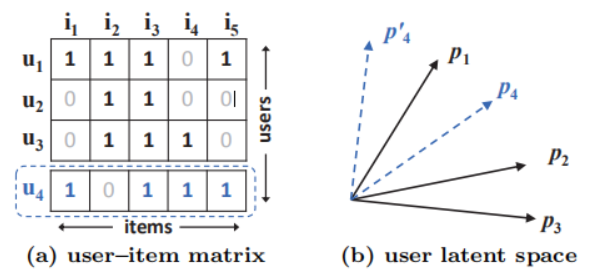
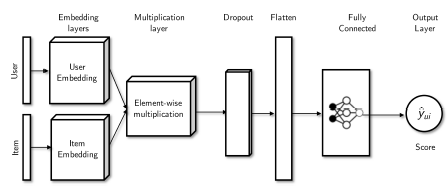
MF的局限性：

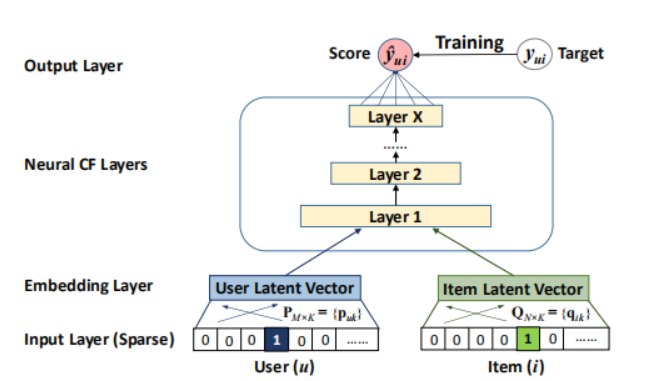


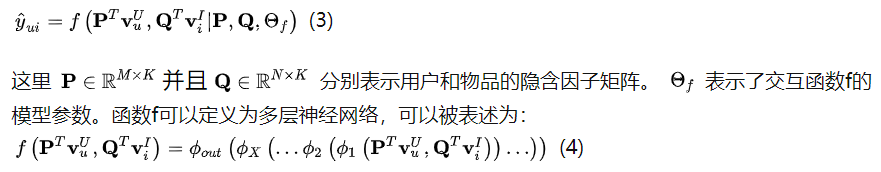
（a）中使用Jaccard作为相似性度量（交集/并集），（b）中使用cos作为相似性度量。

GMF网络（Loss：pointwise-MSE pairwise-交叉熵）：



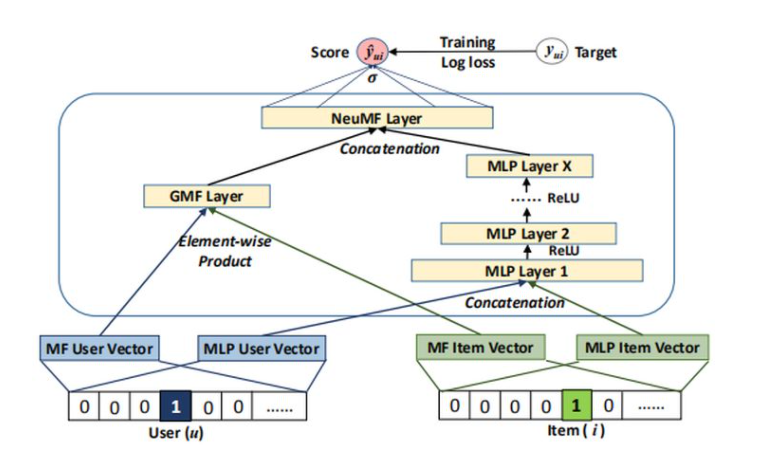
NCF结构：



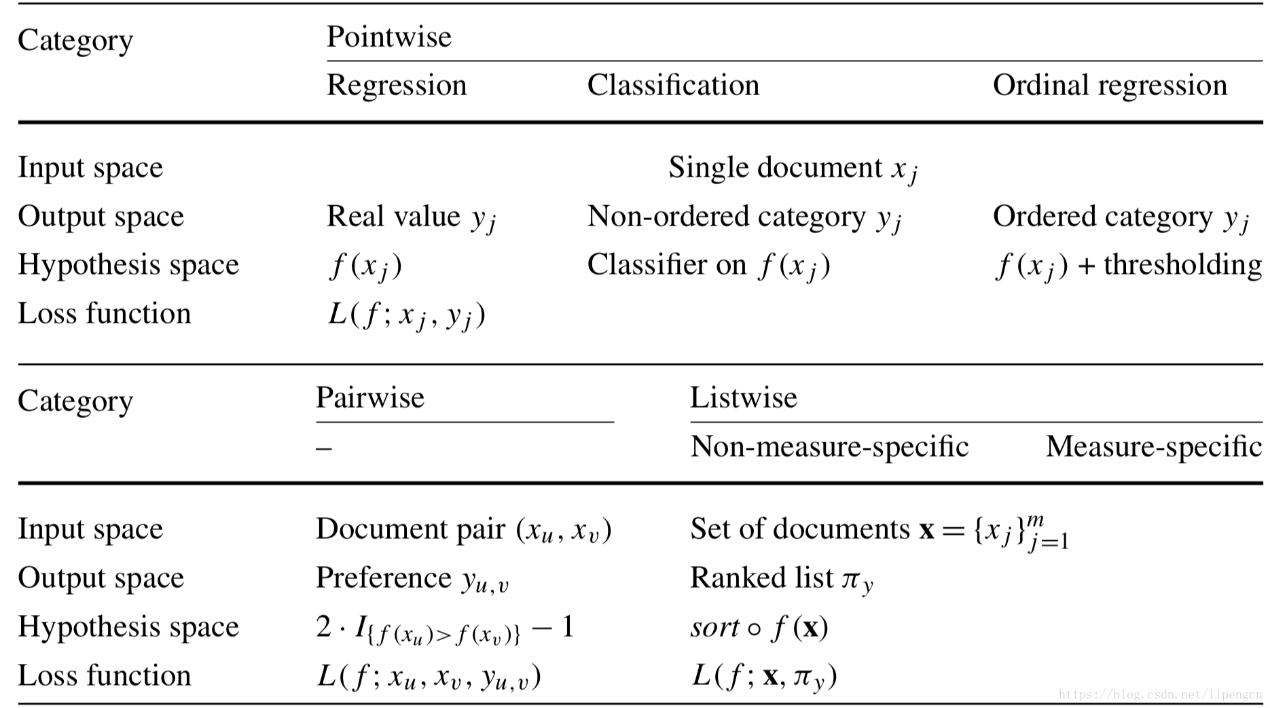


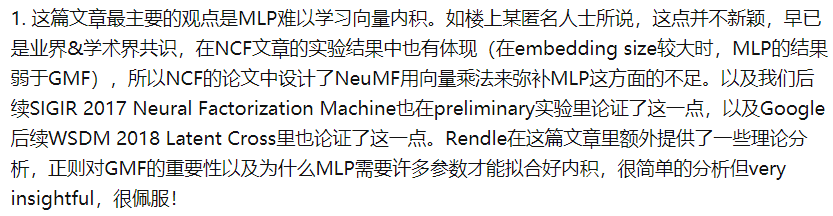
Q：为什么MF（即前文中提到的GMF）是特殊的NCF（给出映射）？

NeuMF模型（NCF+MLP）：



RS中的loss functions：





本文目标：神经网络不如点积效果好！

<https://zhuanlan.zhihu.com/p/144740834>

Related work

现代推荐系统都基于embedding，一个核心的操作是计算两个embedding的相似度打分；

直接做向量內积一个良好的基线；

主要对比的是NCF论文提出的NeuMF结构，提出相似度应该由神经网络这种万能拟合器去学习；

Approach

这篇文章以实验分析为主，方法介绍着重在符号定义和公式，篇幅不多；

关注点仅落在最后的相似度计算公式上，即给定两个向量如何给出一个标量打分，至于如何得出这两个向量不在本文讨论之列；

dot：基本的点击相似度;

MLP：将两个向量拼接到一起，灌MLP；

GMF：向量点积的每个维度要乘一个洗漱；

NeuMF：两个d维向量的前k维做GMF，后d-k维做MLP，再相加；

NeuMF pretrain：NeuMF的两个组件独立训练；

Experiments

还原实验

Movielens 1M 和 Pinterest，都是只有正样本的隐反馈数据；

每个用户的最后一个item被保留，和另外100个随机采样的item构成一个测试case；剩余的用户正样本加入到训练集中；

评价指标是top10 recall和NDCG；

训练方法：用logistic loss+L2正则，SGD(没有batching，没有momentum等等）+负采样；

尝试了各种embedding维度d；

除了d=192上面recall被NeuMF pretrain轻微超越，其他各个评估均是dot大幅胜出；

特别的，讨论了GMF为何不work，因为额外引入的系数没有加入到正则化项中，可能会使本来的正则化失效？(注：这个问号的细节在彩蛋里)

另外，也对原论文的超参调优方式提出疑问，应该用验证集而不是测试集来调优参数；

人造数据实验

为了说明MLP尽管是通用分类器，但可能无法拟合出dot的效果，构造了随机数据，用MLP学习dot；

在各种参数下，看起来确实都无法很好的拟合；

提到了一个很有意思的算法界：用MLP拟合d维空间二次多项式，需要O(d^4 / e^2)步才能收敛到误差e；

而从实验曲线观察，需要的样本数大约是O(d/e)^a，a在1~2之间；

彩蛋：Rendle是怎么做实验调参的？

论文的最后专门加了个附录详细交代实验是怎么做的；

loss = logloss + L2(P) + L2(Q);

用的sgd而不是mini-batch sgd，每个样本一更新；

因为要调超参，用切分测试集的方式又切了一遍训练集，hold-out了最后一个item作为调参验集；

作者的经验是，参数调好了，一般d越大越好，学习率越小越好，但是相应的训练时间会越长，希望大约的预算是epoch=256；

先固定正则系数为0，epoch=128，并且固定维度MovieLens=64，Pinterest=128，grid search了学习率=0.001，0.003，0.01和负采样系数4 8 16；

然后在得到的较优参数附近，固定d=128，epoch=256，搜索学习率、负采样系数和正则化系数；

没有提到MLP、GMF等模型的优化器；

读后记

某种意义上的现代推荐系统之父Rendle的文章，排版上还是延续了libFM的单列风格，13页洋洋洒洒；

这里需要吐槽一下百度和知乎的搜索团队，再表扬一下知乎的推荐团队，写这篇笔记的时候先在百度和知乎都搜了一下没搜到相关的解读，还窃喜了一下能抢个全网首发，写完了随手打开知乎，给我推荐的第一条就是前一天的一个回答；

论文是5.19发出，5.23在知乎有匿名用户提问，这周末陆续有答主开始认真回答，是个值得关注和讨论的问题啊；

NCF作者Xinagnan He有在知乎问题下面回答，值得一读，就不复制粘贴了；

He的回到中有提到四年前论文的成文过程，半年时间踩过的坑，其实也是近几年深度学习工具链发展对效率提升的一个缩影，在数据量上了之后，工业界在有的问题上确实是走在学术界前面的；

讨论：

1. MLP很难学习内积 WSDM2018 Latent cross
2. NeuMF和GMF不如MF。NCF优化log loss，MF优化BPR和eALS
3. DL像水一样具有高适应性，但是这也成为了被滥用的原因。
4. 工业界更关注模型的有效性，根据不同的场景选择最佳模型。



