**学习进度**

本周学习任务：

学习Part I：Advances in Collaborative Filtering

下周学习任务：

学习Part I：Item Recommendation from Implicit Feedback

**一、基准预测**

**1. 基本认知**

过去，主要通过考虑用户项目间的交互来理解评分，常忽略用户和项目本身的评分偏差。于是，基准预测（Baseline Predictors）作为一种以理解偏差为中心的非个性化推荐模型应运而生，它的评分预测过程完全由用户评分的慷慨度和项目的受欢迎程度解释。基准预测模型除了可以考虑评分偏差，对于缺乏足够交互数据的新用户或项目而言，还可以提供一个可信的基准预测值，从而有效缓解冷启动及数据稀疏性问题。

基准预测规则为：，表示全局评分均值，表示用户历史评分均值与之间的偏差。一般可通过解决最小二乘法问题，来估计可学习参数的值，具体公式如下所示。其中，第一项用来寻找与评分记录拟合得更好的和，第二项正则化项则用于避免过拟合现象。对于此类最小二乘法问题，可使用优化算法SGD学习收敛。



1. **考虑偏差的时序变化**

值得注意的是，用户平均偏差及项目平均偏差，有时候会随时间发生变化。引入时间效应的基准预测，通过将看做时间的函数，考虑了偏差的时效差异，预测规则为：。

在物品偏差函数中，通过特定取值范围的划分若干个时间段，每个时间段用一个专门的常数表示物品偏置。对于而言，关键在于权衡时间段的长度，从而保证每个时间段中都包含足够的评分记录。

在用户偏差函数中，存在线性及样条曲线两种实现方式，公式分别如下左右所示。其中，对于线性用户偏差函数而言：可学习参数有，超参的经验取值为0.4，表示当前时间与用户历史评分日期均值的差；对于样条曲线用户偏差函数而言：可学习参数有，超参用于将用户历史评分日期均匀的划分到个时间点中，超参控制曲线的平滑度（经验取值为0.3）。

；

1. **考虑偏差的短期瞬时变化**

特别的是，如若考虑用户评分偏差在某段时间内（比如在某一天中）发生的瞬时变化，则需在预测规则中额外引入，用于表示用户在某天内的评分偏差。同样，也可以引入，用于表示用户分季节的周期性评分偏差。

此外，物品偏差并不完全独立于用户，**用户也许会随时间改变对于物品偏差的看法**。此时，可以添加额外考虑短期瞬时变化的缩放因子，作为预测规则中物品偏差关于时间变化的置信权重。

至此，改进后的线性预测规则如下所示：



1. **讨论**

需要讨论的是，根据下图所示的实验结果，可以知道相较于线性预测规则Linear，样条曲线预测规则Spline增加了额外的灵活性，小幅提高了预测准确度。此外，在预测规则中额外考虑偏置在某段时间内发生的瞬时变化，能够相对更大幅度地提高准确度。然而，对于推荐系统而言，模型既需要凭借基准预测考虑用户的短期瞬时偏差，也需要凭借基准预测吸收这种偶然的短期波动，保持对于数据长期特征的记忆。

graphic

**二、矩阵分解模型**

在推荐系统中，矩阵分解模型（Matrix Factorization Models）作为一种潜在因素模型，它通过奇异值分解用户-项目交互矩阵的方式，能够更准确且灵活地应对数据稀疏性的情形。矩阵分解模型的常见算法，如下所示。

1. **奇异值分解（Singular Value Decomposition，SVD）**

SVD将用户和项目都映射到同一个维联合潜在因素空间中，用户-项目的交互矩阵被因式分解为潜在因素空间中向量间的内积。SVD的预测规则为：，为了估计可学习参数的取值，目标函数如下所示：



而在使用优化算法SGD迭代训练的过程中，需要循环遍历并预测训练集中的每个评级，计算对应的预测误差。继而根据如下公式，异步更新可学习参数，直至收敛。





1. **SVD++**

如若需要考虑用户点击、浏览或评价物品等隐式反馈记录，只需在SVD预测规则中追加表示特定隐式信息的因子向量即可，SVD++的预测规则为：



其目标函数及优化函数，参考前述公式，此处省略。值得注意的是，SVD++提供了在预测规则中，自由追加多种不同类型隐式反馈信息的灵活性。而且，不同类型隐式反馈信息的重要性，由算法自动学习获得。

1. **timeSVD++**

如若需要考虑用户偏好的时序变化，只需重构用户向量为关于时间的函数即可，重构后的用户向量为:。timeSVD++的预测规则为：



值得注意的是，由于项目本质上趋于静态，timeSVD++并未考虑项目的时序变化。此外，虽然timeSVD++的运行时间是SVD++的两倍，但是timeSVD++算法的收敛速度没有受到时序参数的影响。比较前述三种矩阵分解算法，相同点在于：它们的预测准确度都能随因子维数的增加而提高；不同点在于：从性能的角度出发，timeSVD++>>SVD++>SVD。由此可知，相较于考虑隐式反馈，捕捉时间效应带来的回报，某种程度上更为明显。

1. **邻域模型**
2. **基于相似度的邻域方法**

相较于矩阵分解模型，邻域方法的推荐结果更具解释性，而且还能够立即反映新评分。此类方法的关键在于，相似度的计算方式及预测规则的制定。若考虑使用基准预测消除评分偏差，同时再引入基于数据量的相似度置信权重，则改进后的度量物品间相似度的PC公式如下所示。其中，表示同时对物品和评分的用户数量，超参经验取值为100。



由此，改进后的基于物品预测规则，如下所示。其中，表示与目标物品i最相似的k个近邻物品集合。有时候在预测规则中取相似度的平方，效果更佳。



然而，如若近邻间的评分变化太大，此类方法往往效果不佳。而且，此类方法没有考虑到近邻间相互作用的同时，基于相似度的插值权重在衡量不同近邻评分的重要性时，还具有缺乏有界的局限性。此外，由于上述预测规则中的权重和为1，当缺乏交互数据时将发生过拟合现象。由此，可以妥协地引入惩罚因子用于控制邻域部分的预测置信度，即：近邻信息倘若越少，最终预测评分则越多地参考基准预测值。改进后的基于物品预测规则，如下所示：



1. **基于全局优化的邻域模型**

邻域模型的相似度计算，不依赖于人为偏好或先验知识，而是依赖于数据的全局最优化求解过程。邻域模型允许集成不同形式的显式或隐式输入，亦能方便地考虑输入的时序效应，具备更高度的可扩展性。同时，伴随着近邻数的适宜增加，一定程度上能不断提高模型性能。此外，邻域模型同样考虑到了过拟合问题，即：除非用户输入足够多的相关评级，否则模型将恢复到稳健的基准预测值。

基于物品的邻域预测规则，使用可学习的全局权重，代替基于相似度公式的可信插值系数。至此，额外引入隐式反馈因子的基于物品的邻域预测规则，如下所示。其中，超参用于控制规范化的程度，可学习参数有。



特别的，如若需要灵活地考虑近邻数量，则改进后的基于物品邻域预测规则，如下所示。其中，表示用户评价过的物品中，同物品最相关的个物品。然而，这种依赖相似度公式来考虑近邻数的灵活度，需要额外的计算代价。



领域模型在迭代训练及测试过程中，所必须的目标函数及优化算法，参考前述矩阵分解模型中的相关叙述，此处省略。

1. **引入因式分解的邻域模型**

参考矩阵分解思想，邻域模型亦可将全局权重因式分解，保持预测性能的同时（性能同SVD相当），显著降低模型计算的时空复杂度。

对于基于物品的邻域模型而言，可以将全局权重因式分解为目标物品与不同物品间的两个独立低维向量，不同物品是指目标用户评价过的其它物品。预测规则，如下所示。其中，隐式反馈因子，全局权重，小括号内部的项与目标物品无关，只依赖于目标用户。



对于基于用户的邻域模型而言，可以将全局权重因式分解为目标用户与不同用户间的两个独立低维向量，不同用户是指评价过目标物品的其他用户。预测规则，如下所示。其中，全局权重。



对于基于用户和物品的混合邻域模型而言，它通过组合训练前述两种模型的方式，提高模型的泛化性能。一种简单的追加类型组合预测规则，如下所示。显而易见的是，邻域模型也可以使用同样的方式和其它不同类型的模型集成。

**4. 考虑时序效应的邻域模型**

经验是，相比于设计更复杂的预测规则，考虑数据中的时序效应，能够对预测性能有更直观的影响。引入时序效应的基于物品的邻域预测规则，如下所示。其中，根据时间衰减物品和物品间的关联，可学习参数用于表示用户偏好的变化速率（如若从数据中观察到用户的偏好变化很快，即很大，则需要更多地削减历史交互信息对最终预测评分的影响）。然而，不恰当地处理数据的时序效应，或许会影响物品间的固有关系（物品间的相似度权重，某种程度上反映了物品的固有特点）。

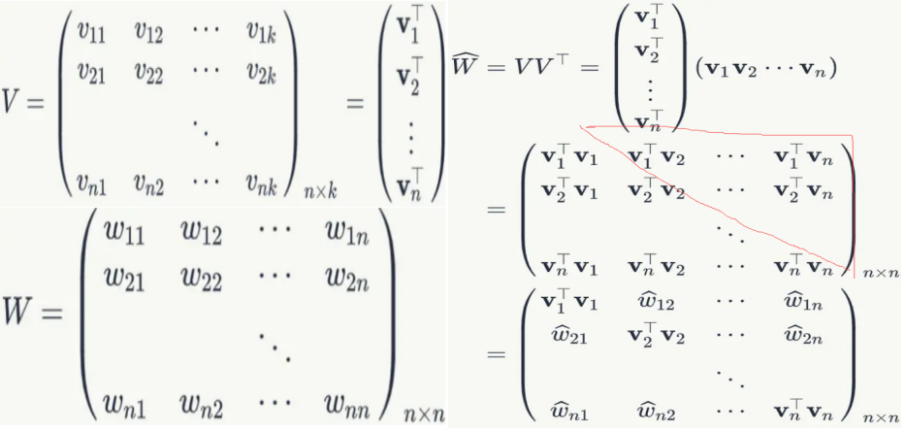


1. **因子分解机**

因子分解机（Factorization Machine，FM）作为一种自动组合二阶高维稀疏特征的机器学习模型，它主要通过矩阵分解二阶特征交叉关系权重为两个隐向量与内积的方式，更高效地捕捉了每个样本中特征分量间的关系。二阶FM模型的公式，如下所示。其中,表示特征向量的维数（特征分量数），隐向量的维数为。



因子分解机FM本质上，是将维的特征分量交叉项系数矩阵，转化成了维的隐向量矩阵与其自身转置的内积，即：（当且仅当足够大时，对称正定矩阵才能矩阵分解为）。具体推导过程，如下图所示。



相较于poly2模型，FM使用隐向量与的内积，代替了衡量二阶特征分量组合重要性的权重。使得FM模型在迭代训练的过程中，只要在样本中出现，模型就可以学习对应隐向量的嵌入表示。某种程度上，弥补了poly2模型关于参数的更新，只依赖于满足“”条件的稀缺样本数据的缺陷。

相较于传统因式分解模型，只能处理固定维度的离散特征。FM模型能够兼容任意实值的特征向量，即：无需特定的数据预处理或特征转换，就能够适应具有连续、类别及稀疏等特性的特征分量组合。FM模型能够处理的样本实例集，如下图所示。

graphic

另一方面，FM模型公式中的可以通过多项式化简的方式，将模型的复杂度从降到。简化结果如下所示，本质是将元素间的二次多项转化为元素间和的平方减去平方的和。



特别的，高阶FM模型公式，如下所示。从中可知，当时模型公式相对复杂，个特征分量间的相互关系也更难估计。



而FM模型在迭代训练的过程中，可用的优化算法有：随机梯度下降法SGD、交替最小二乘法ALS及马尔可夫链蒙特卡洛法MCMC等。在SGD优化算法中，需要强调的是关于隐向量的更新公式，如下所示。



参考资料：

https://blog.csdn.net/realnuannuan/article/details/93401214  
https://blog.csdn.net/lijingru1/article/details/88623136  
https://www.jianshu.com/p/f107724fda5d