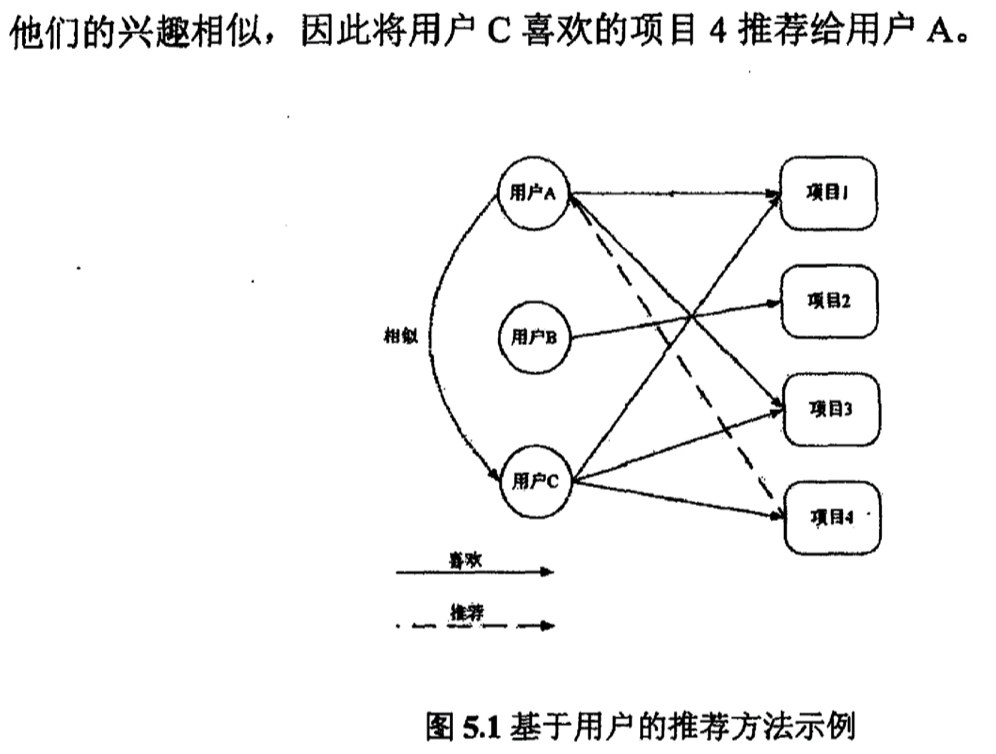
融合基于用户和基于项目方法的机器学习框架

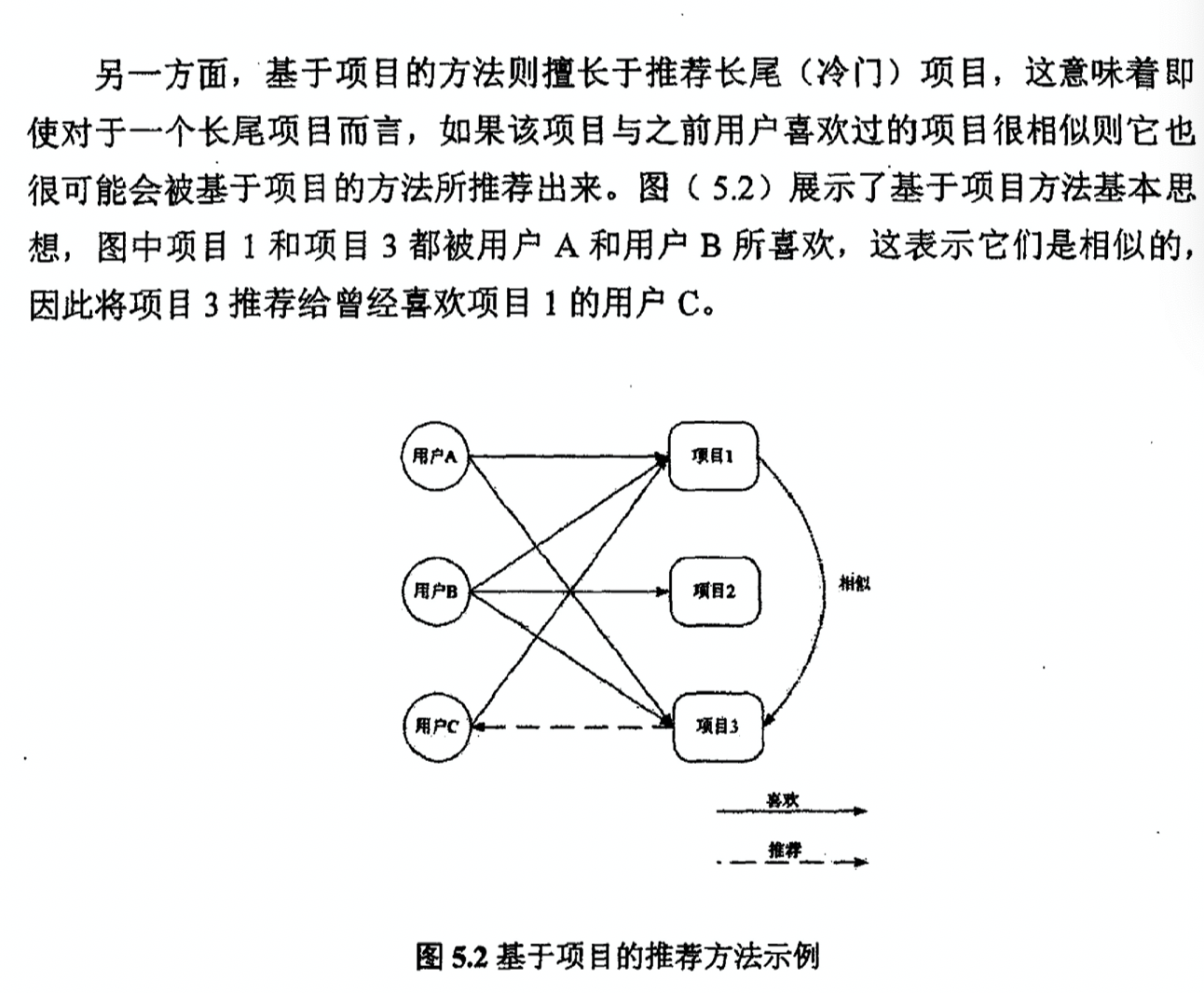
## 假设：

利用用户群历史标注数据（评分、观看、购买、点击）来预测用户的项目偏好程度。根据不同的假设，协同过滤算法可以分为两类：基于用户的，和基于项目的方法。

基于用户的方法假设：用户会喜欢与他具有相似的历史标注习惯的用户，所喜欢的项目。

基于项目的方法假设：用户会喜欢与他曾喜欢过的项目相类似的项目。





## 基于内存的实现方法：

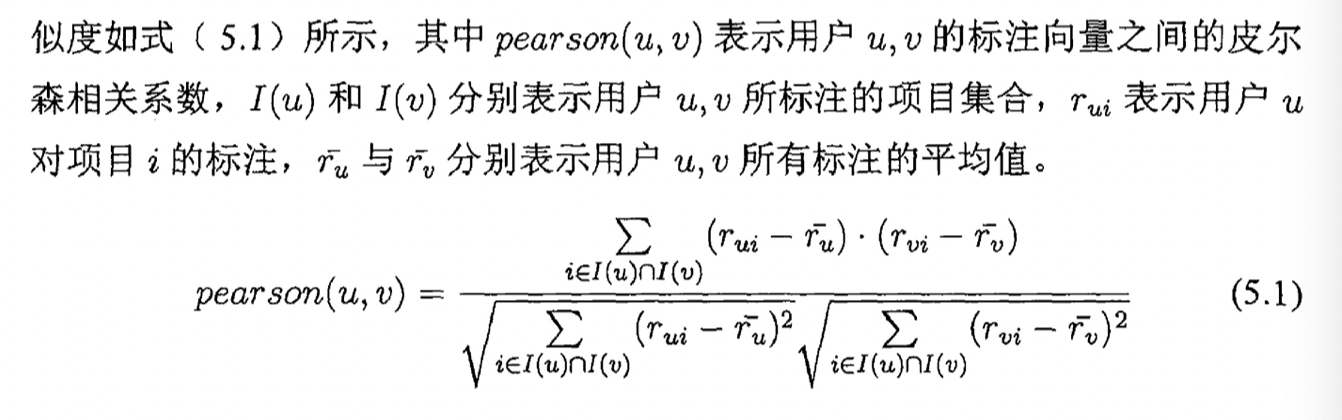
第一步，找出活跃标注 （active rating, 即待预测的用户项目对）的相似用户或相似项目集合；

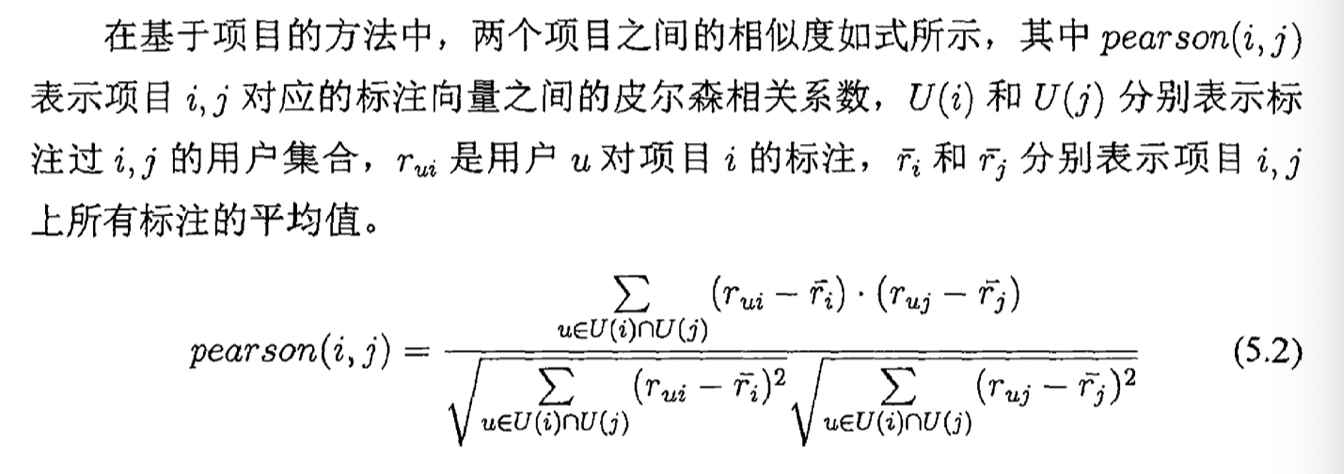
第二步，利用相似用户或项目的标注进行预测。

关于协同过滤算法的早期研究，主要集中在基于用户的方法。基于项目的方法能小幅度提高推荐精度，更好的可拓展性。

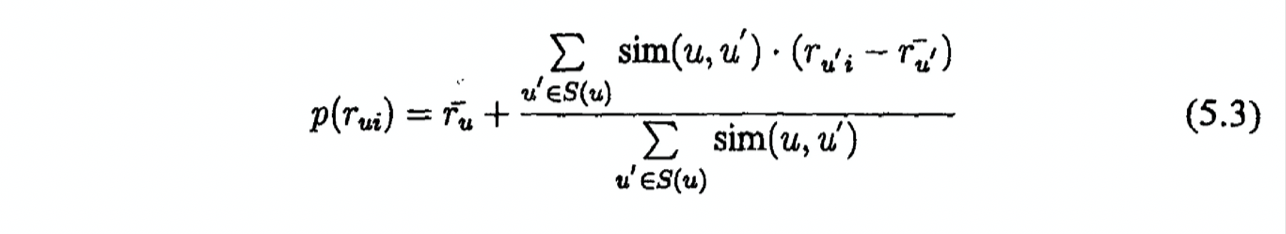
用户相似度计算：

皮尔森相关系数，以两个向量之间的先行相关性来衡量向量之间的相似度。

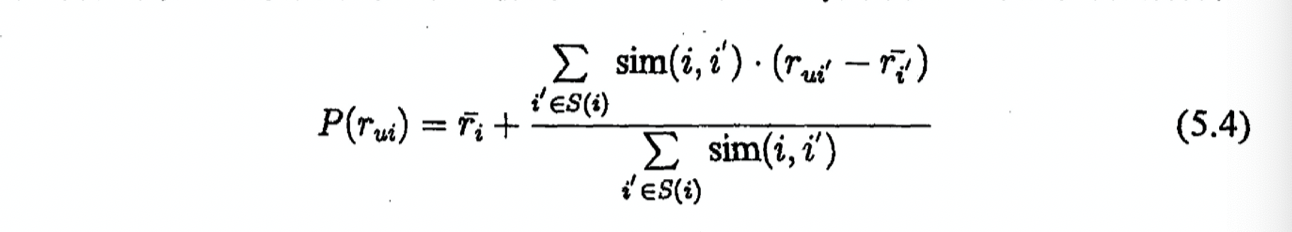




在最后的标注预测过程中，基于用户的方法，将相似的用户集中中的所有用户标注，用用户相似度进行加权平均，得到最终预测结果。



基于项目的方法将相似项目集合中的所有项目，用项目相似度进行加权平均得到最终的预测结果。



## 基于模型的实现方法

利用评价矩阵数据作为训练数据来训练相关的机器学习模型，然后利用模型进行推荐预测。

Billsus等人将协同过滤看成一个分类任务，采用矩阵奇异值分解来特征提取，采用神经网络来进行分类器的训练。

Breese等人则提出，协同过滤概率模型，并分别采用聚类模型和贝叶斯模型进行概率估计。隐语义模型（Latent semantic models）是一类概率生成模型，这类模型也可以被应用到协同过滤中，典型的代表有概率隐语义分析（probabilistic latent semantic analysispLSA）以及隐含狄利克雷分配（Latent Dirichlet Allocation, LDA)

按照不同的假设，基于模型的方法也可以分为基于用户的方法和基于项目的方法，例如在 pLSA 模型中，如果其中的隐变量表示用户群组，则该模型就是从用户的角度的出发来构建的；如果其中的隐变量表示项目类别，则该模型就是从项目的角度出发来构建的。