**算法说明**

**姓名：杨劲儒**

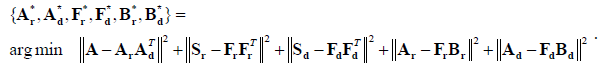
1. **简介**

本算法用于预测mirna与疾病之间的关系，借鉴了论文Predicting binary, discrete and continued lncRNA-disease associations via a unified framework based on graph regression预测lncRNA与疾病之间关系的方法。

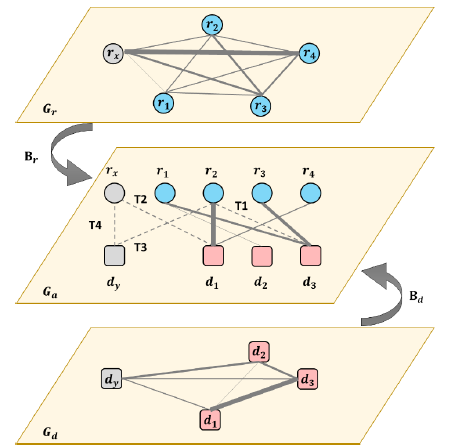
1. **算法原理**

**1.图回归**

图回归公式化后如下，其中A表示miRNA-疾病关联矩阵（即Ga），Sr表示miRNA相似矩阵（即Gr），Sd表示疾病相似矩阵（即Gd）。



三个矩阵对应的图如下：



**2.利用SVD计算Ar,Ad,Fr,Fd**

SVD是一种矩阵分解方法，可以降低矩阵的复杂度。它假定行代表的变量与第三方变量相关，列代表的变量也与这个第三方变量相关。排除掉相关程度较小的变量，即可化简矩阵。SVD分解后得到与行有关的矩阵U，系数矩阵S和与列有关的矩阵V。把S开方乘到两边即可得到我们要求的矩阵。系数矩阵S是一个对角矩阵，按照从大到小从左上角到右下角排列，每一个值代表其对应的属性在S,D关系间的重要程度，值越大，表示该属性越能反映S和D之间的联系，值越小，则越可能是噪音等误差，应当视情况舍去。

**3.利用PLSR计算Br,Bd**

PLSR是一种多因变量Y对多自变量X的回归建模方法.。我们的目标是求出Y=XB的矩阵B，B即为我们要求的矩阵。所使用的PLSR函数参考于<http://www.cnblogs.com/pegasus/p/3396085.html>需要注意的是，当残差矩阵不再减小时，就应当停止循环，以免受到噪音的影响。残差矩阵大小根据其2-范数的大小决定（此处应注意矩阵2-范数和向量2-范数的区别）。

**4.打分**

打分的公式如下，Fr表示miRNA之间的相似信息，Fd表示疾病之间的相似性信息，Θ=Br\*BdT为从Ar，Ad中提取得到的系数矩阵，包含了A的信息，用于沟通链接Fr和Fd。最终得到的左式即为所求的打分矩阵。

C:\Users\Blue\AppData\Roaming\Tencent\Users\706415254\QQ\WinTemp\RichOle\LQ37VO`4GO][5AO$86U%BMA.png