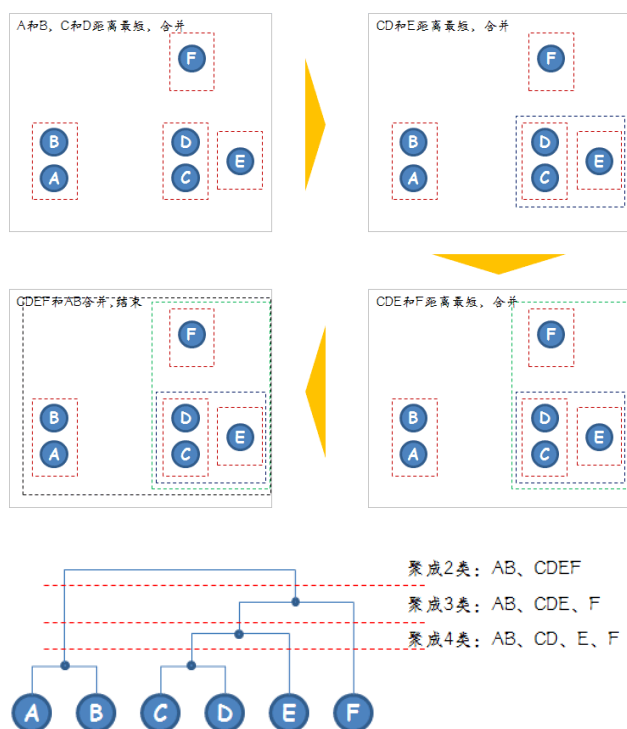


目前使用的聚类算法层次聚类其原理为

设有 n 个人，每个人测得 m 项指标，首先定义人和人之间的距离 (称为样本距离数) 和类与类之间的距离，初始将 n 个人看成 n 类，这时类间距离与样本距离等价。然后将距离最近的两类合并为新类，并计算新类与其他类的类间距离，再按最小距离准则并类。每次缩小一类，直到所有的样品都合并成为一类为止。



目前定义任何人之间的距离方法如下:

对第 i 个人赋予一个向量

$$u_i = (x_i, y_i, z_i)$$

x_i, y_i, z_i 分别为标准化后的 x 坐标、 y 坐标、运动方向

定义第 i 个人和第 j 个人之间的距离

$$dist(i, j) = 1 - \rho(i, j)^2$$

$\rho(i, j)$ 为第 i 个和第 j 个人的 spearman 相关系数。直观上容易理解，两个人相似度高时，此度量下的距离短，更倾向于分到同一类。