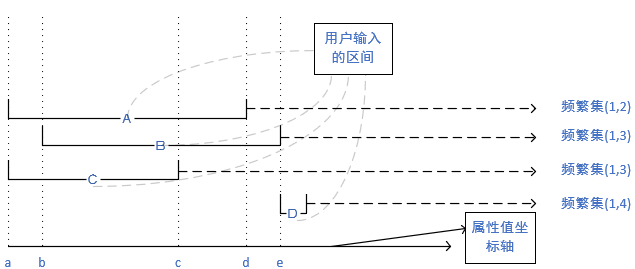
测试数据设计思路是从逆向设计，从需要的结果倒推回去得到频繁集粒度，最后再根据频繁集去生成原始的记录。

1. 需要的结果：

当前方法的创新点主要有两个，一个是那两个比例的提出，另一个是区间整合算法的提出，所以测试用例的设计也是围绕这两个方面展开。

1.1 区间整合算法

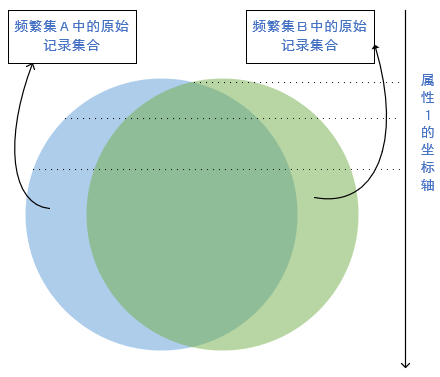
假设现在用户的输入如下：候选分割合并区间为[d,e]



当频繁集(1,3)的粒度为[c,e]时，我们方案的效果是最好的。所以将频繁集(1,3)的一个粒度设为[c,e]（因为是测试数据，所以展现一个，如果需要，其他区间也可以按照此方法设计）

1.2 展现那两个比例

影响部分占总体的比例的因素主要有两种，一种是噪音，另一种是其他频繁集的影响（如频繁集A为{闯红灯，违章停车}，频繁集B为{闯红灯，超速}，A中闯红灯有一个粒度为[3,5],B中闯红灯有一个粒度是[4，7]那么他们之间的比例肯定是会相互影响的，如下图所示）。



所以这里采用的做法是：（在频繁集生成原始记录这一阶段进行的）

1. 降低了噪音的比例。

2. 加大两个频繁集原始记录的重合度。