[positive-unlabeled (PU) learning](https://www.cnblogs.com/stAr-1/p/9494215.html)

PULearning的应用场景是，我们可以清晰地确定正样本，但是不能确定负样本，因为它有可能是正样本，只是我们还没有证明。这时我们可以把这部分不确定的样本称为无标签样本U，加上正样本P来建立模型。

问题可以转化为一个有约束条件的最优化问题：

在保证正例中错误率低于1-r的条件下，最小化无标签样本中U的正例数目。

建立PU分类器有两种方法：

（1）两步方法two- step approach

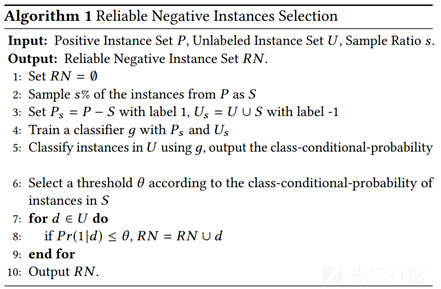
（2）直接方法direct approach

**two-step approach**:

第一阶段：从未标记实例中选择可靠的负例集RN，做法是：

在P中随机选取一部分正例S加入U中，这时两个数据集是P-S，我们叫做ps，和U+S，我们叫做us，用ps和us训练一个模型g。

然后用g对无标签样本U做分类，得到每个样本的概率，设定一个阈值a，如果样本概率低于a那么我们认为是一个可靠负例



第二阶段：利用正例P和可靠负例RN，训练一个传统的机器学习分类模型，用来预测新样本。

我们和本赛道单模第七名的选手进行了交流，他们在学习策略上使用了PU learning。Balabalabala……但我们在实现/学习时发现本题上PU learning的策略并不能带来好的效果，我们也向原作者反应了这一问题。看了他们的开源的代码，发现对方的实现上还是存在问题。（最后一句话加不加）