[Spaco原理及实现](http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=743927099)

[转至元数据结尾](http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=743927099#page-metadata-end)

* 由  [张宇轩](%20%20%20%20/display/~tianjinmouthzhangyx%0a)创建, 最后修改于[十一月 12, 2021](http://wiki.intra.xiaojukeji.com/pages/diffpagesbyversion.action?pageId=743927099&selectedPageVersions=1&selectedPageVersions=2)

[转至元数据起始](http://wiki.intra.xiaojukeji.com/pages/viewpage.action?pageId=743927099#page-metadata-start)

**一、背景**

机器学习中，根据学习任务的不同，分为监督学习、无监督学习、半监督学习(SSL)和强化学习.  
•有监督学习：从已标注数据中学习  
•无监督学习：从未标注数据中学习  
•半监督学习：从<少量>标注数据和<大量>未标注数据中学习

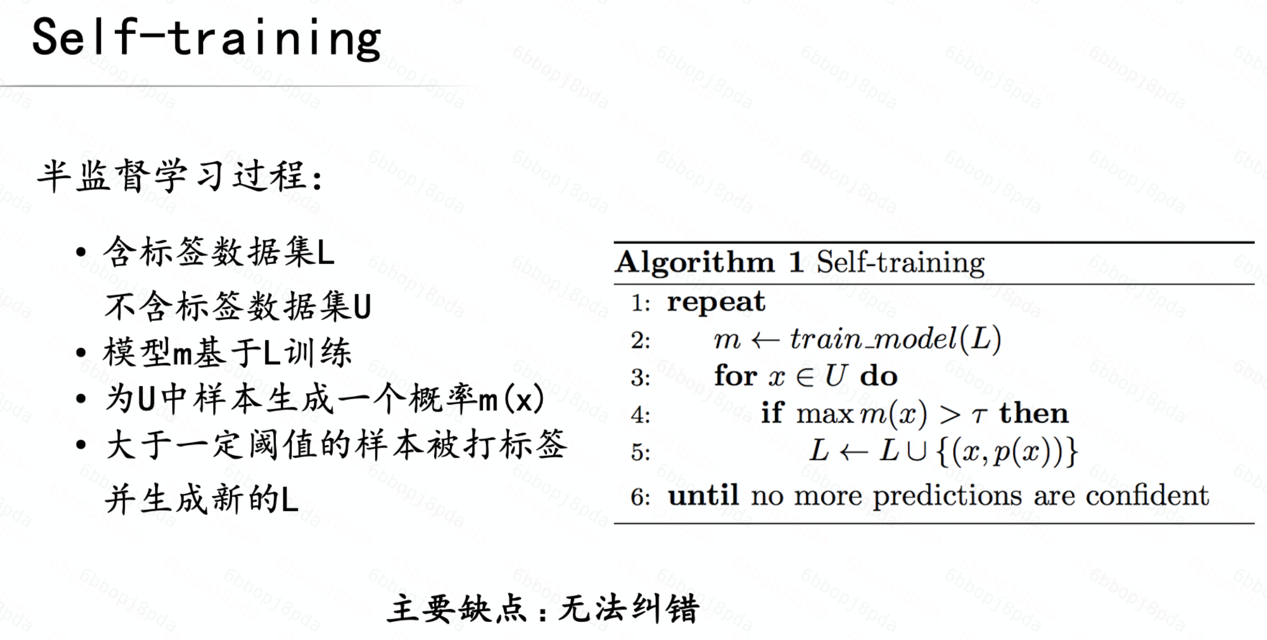
在我们的安全场景中，正样本极少（实际投诉量），还有很多数据是未知的，因此可以归类为半监督学习。

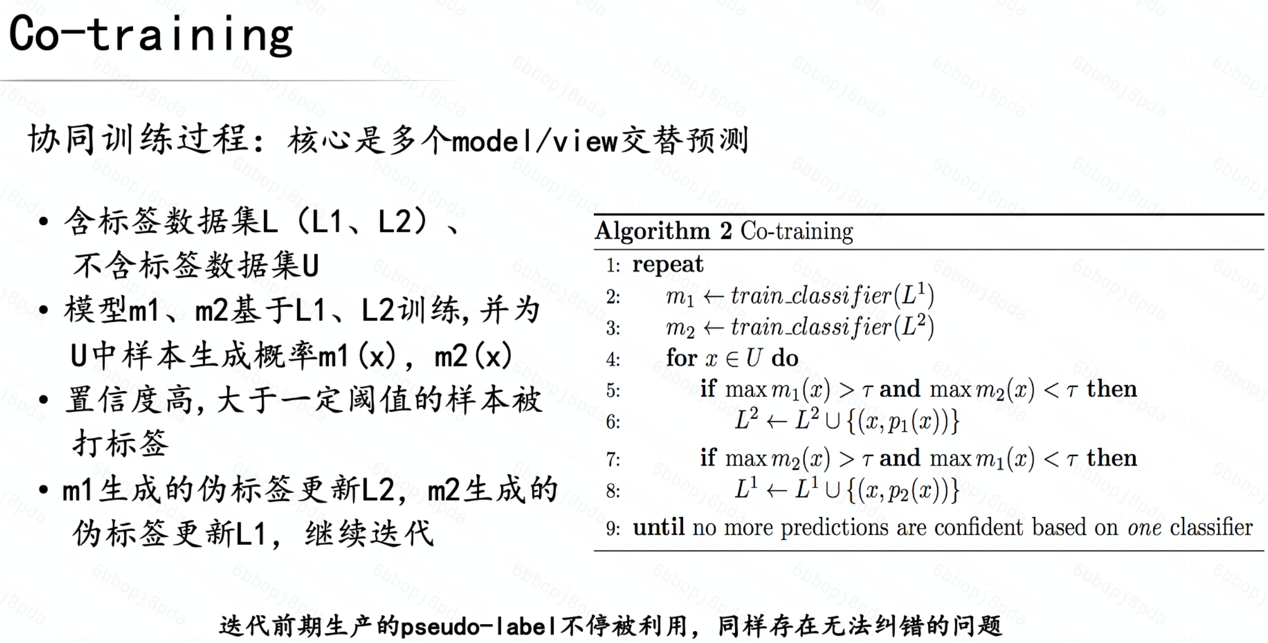
**二、Spaco 定义及核心原理**

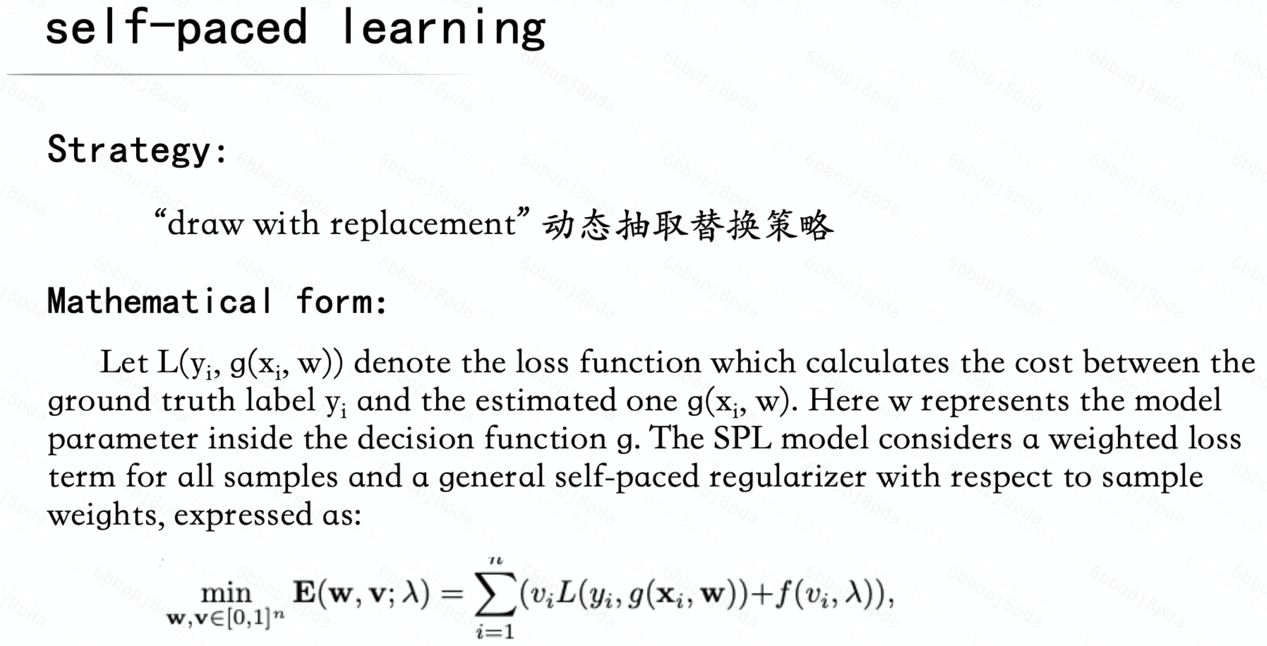
半监督学习算法中有一类特殊算法，能在未标记数据上生成代理标签（proxy-label），也就是去扩充数据集，并将其与已经标记好的数据共同使用。常用的有如下两种：

Self-training（自训练算法）、Co-training（协同训练算法），Spaco即为两者的结合： Self-Paced(自步训练) + Co-training(协同训练)

原理：







通过控制vi 取值控制样本对目标函数的贡献，同时可以控制每次训练加入的labeled samples. vi 取值与λ相关，在spaco中f (vi,λ)为 – λjvkj  .因此通过适度加大λ可以控制vi向量中1的个数，在迭代中起到步长的作用。

**三、训练流程及代码**

