

附录

APSO 欠采样算法

输入：原始数据集(imbalanced_data)

输出：平衡数据集(balanced_data)

```
1: cv_sets=crossValidate(imbalanced_data,cv_splits);
2: for j=1 to cv_splits do
3:   test_dataj=getTestData(cv_sets, j); //得到验证集
4:   train_dataj=getTestData(cv_sets, j); //得到训练集
5:   min_dataj=filterMinData(train_dataj); //从训练集中筛选出少数类数据
6:   maj_dataj=filterMajData(train_dataj); //从训练集中筛选出多数类数据
7:   //pso 优化样本子集得到种群信息
8:   particlesj=optimizeByAPSO(min_dataj,maj_dataj,test_dataj);
9:   //根据种群中每个粒子的最佳位置得到每个样本选择频率
10:   frequency_data=(frequency_data) U (getFrequency(particlesj));
11: end for
12: //根据样本选择频率从多数类中选出与少数类个数相同的子集并少数类合并得到平衡数据
13: balanced_data=getBalancedData (frequency_data,imbalanced_data);
14: return balanced_data;
```