实验设计

1. 实验数据集

（BONIVA、CIPRO、CYMBALTA、DEPOPROVERA、EFFEXORXR、LAMICTAL、LEVAQUIN、LEXAPRO、TOPAMAX、ZYRTEC）

1. 评价指标

AUC；

TP TN FP FN（AA，F-measure，g-means）

1. 对比实验

基分类器：

LRA，DT，PART，NB，KNN-1，KNN-5，SVM，ANN

1. 基分类器单独比较 8个

Self-training 8个

1. 半监督

比较方法 Co-training 8个

OS US SMOTE

③Imbalance 48个

Bagging Boosting RS

④Combine OS US SMOTE+ Co-training

16个 + Self-training

1. 实验流程

①Imbalance用默认

参数 ②unlable的标记比率：10%、20%、40%、80%

③迭代次数：100

④ P N

类别A

|  |  |
| --- | --- |
| N | P |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |

类别B

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **China** | | **German** | | **England** | | **Australia** | | **PAKDD** | |
| **N** | **P** | **N** | **P** | **N** | **P** | **N** | **P** | **N** | **P** |
| 2 | 1 | 2 | 1 | 3 | 1 | 1 | 1 | 4 | 1 |
| 4 | 2 | 4 | 2 | 6 | 2 | 2 | 2 | 8 | 2 |
| 6 | 3 | 6 | 3 | 9 | 3 | 3 | 3 | 12 | 3 |
| 8 | 4 | 8 | 4 | 12 | 4 | 4 | 4 | 16 | 4 |
| 10 | 5 | 10 | 5 | 15 | 5 | 5 | 5 | 20 | 5 |