**实验说明**

**实验前提：**

（1）已知某个项目的小版本数据集文件。

1. 运用各种缺陷预测模型(rf/lm/cart/svm)+不同的预测类型(density+num+prob)所得到的预测结果文件已知。
2. 已知N种工作量策略
3. 已知测试工作量覆盖的模块Top X (X==1代表全测,X=0.2代表只测20%)
4. 已知剩余缺陷率R1,R2

**实验结果:**

**版本1：按照原文作者的思路来实现，缺陷发现模型是GO模型。我画出了策略A1~A11,B1~B4,C1~C4，以及最优模型O的预测结果图。横坐标为“当前测试工作量/实际测试工作量×100%”，纵坐标为“model在分配策略S下能发现的缺陷总数 / 总的实际缺陷总数 × 100%"。**

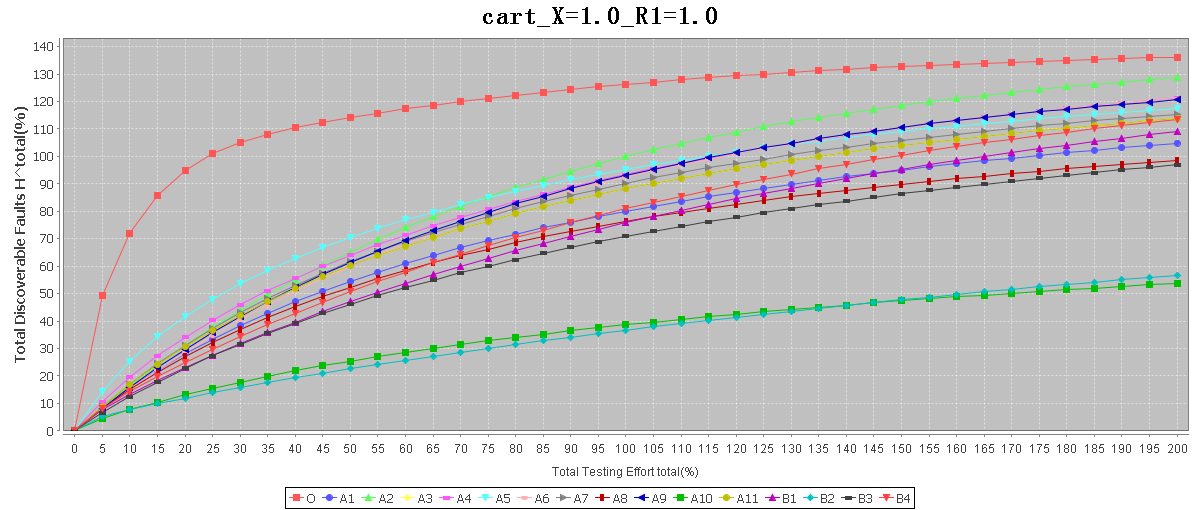
1. 版本1的结果都在这个目录下：

QQ截图20151128212558

1. 文件目录结构说明：

例：

1. GO/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.png 是预测结果图，如下：



其中，lang是项目名，density是预测类型，2\_by\_1 是指实验结果是基于缺陷预测时V2->v1的结果进行的，cart表示预测模型是cart,X=1.0代表全部模块都分配测试工作量，如果X=0.5代表只把测试工作量分给50%的模块，R1代表剩余缺陷率。

2. GO/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.xls是图中的每个点对应的具体坐标。

3. GO/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.doc是每个策略要达到实际bug时所需的test数目与实际数目的百分比。结果如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prediction type | Modeling technique | Required testing effort total(%) | | | | | | | | | | | | | | |
| A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 | A11 | B1 | B2 | B3 | B4 |
| density | cart | 175% | 100% | 135% | 117.5% | 115% | 120% | 130% | 212.5% | 117.5% | 450% | 135% | 165% | 685% | 220% | 150% |

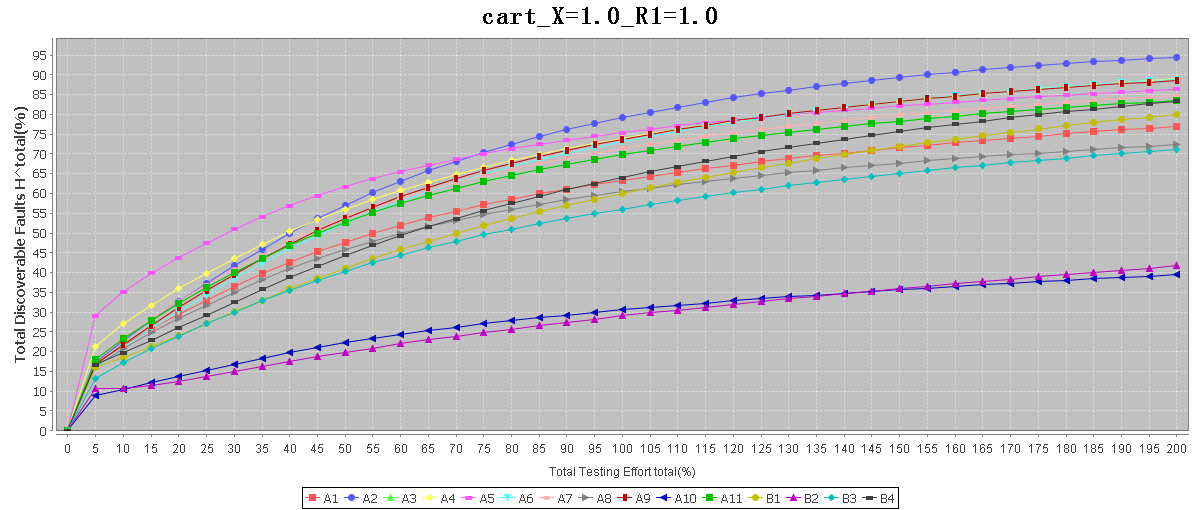
**版本2：与版本1不同的是，横坐标为“当前测试工作量/实际测试工作量×100%”，纵坐标改成了"model在分配策略S下能发现的缺陷总数 / 在最优分配策略O下能发现的缺陷总数 ×100%“。**

1. 版本2的结果都在这个目录下：

QQ截图20151202172605

1. 文件目录结构说明(同上)

GO\_2/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.png:



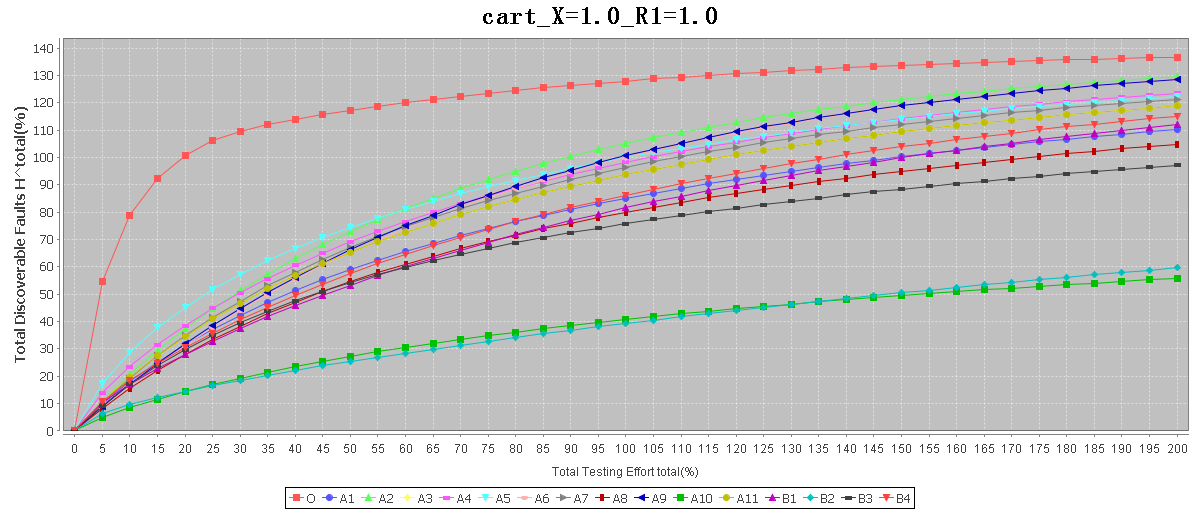
**版本3：缺陷发现模型仍然是Go模型，但是把方程中的Sloc换成了圈复杂度wmc。横坐标为“当前测试工作量/实际测试工作量×100%”，纵坐标为"model在分配策略S下能发现的缺陷总数 / 总的实际缺陷总数 × 100%“。**

（1）版本3的结果都在这个目录下：

QQ截图20151203135745

（2）文件目录结构说明(同上)

GO\_WMC/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.png:

****

GO\_WMC/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.doc:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prediction type | Modeling technique | Required testing effort total(%) | | | | | | | | | | | | | | |
| A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 | A11 | B1 | B2 | B3 | B4 |
| density | cart | 150% | 90% | 117.5% | 105% | 102.5% | 100% | 110% | 175% | 100% | 240% | 117.5% | 152.5% | 597.5% | 220% | 137.5% |

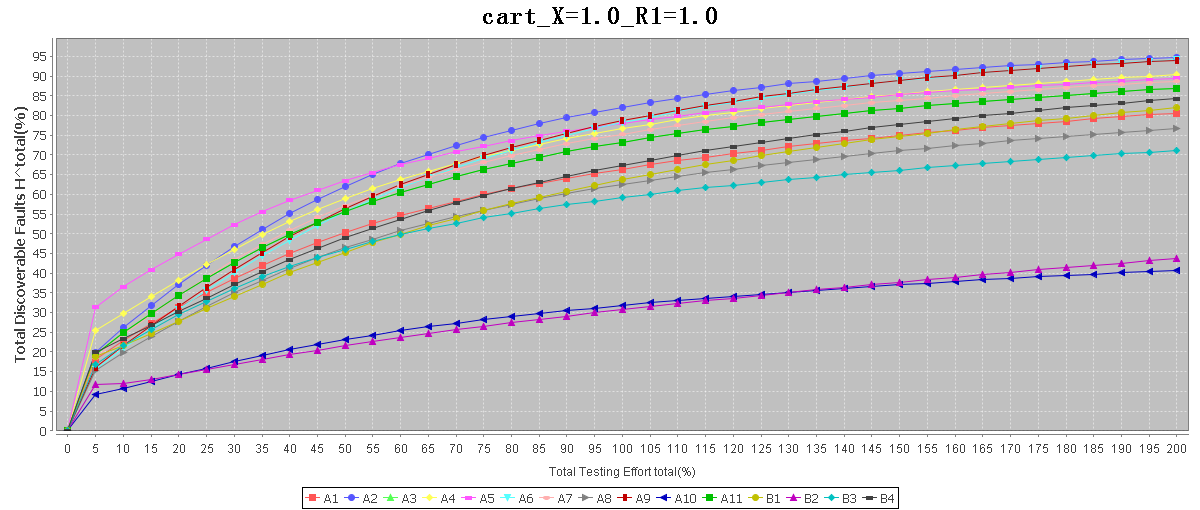
**版本4：与版本3不同的是，横坐标为“当前测试工作量/实际测试工作量×100%”，纵坐标改成了"model在分配策略S下能发现的缺陷总数 / 在最优分配策略O下能发现的缺陷总数 ×100%“。**

（1）版本4的结果都在这个目录下：

QQ截图20151203135808

（2）文件目录结构说明(同上)

GO\_WMC2/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.png:



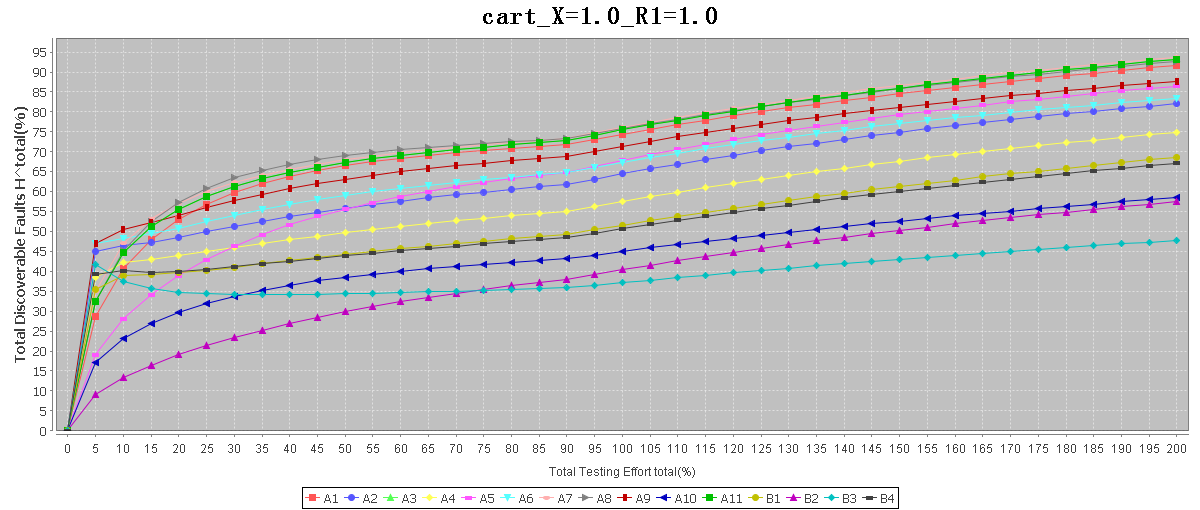
**版本5：缺陷发现模型仍然是Go模型，但是估计ai和b0的方式是通过v0来得到v1。横坐标为“当前测试工作量/实际测试工作量×100%”，纵坐标为"model在分配策略S下能发现的缺陷总数 / 在最优分配策略O下能发现的缺陷总数 ×100%“。**

（1）版本5的结果都在这个目录下：

QQ截图20151202173507

(2)文件目录结构说明(同上)

(3)样例：GO\_AIB0/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.png：

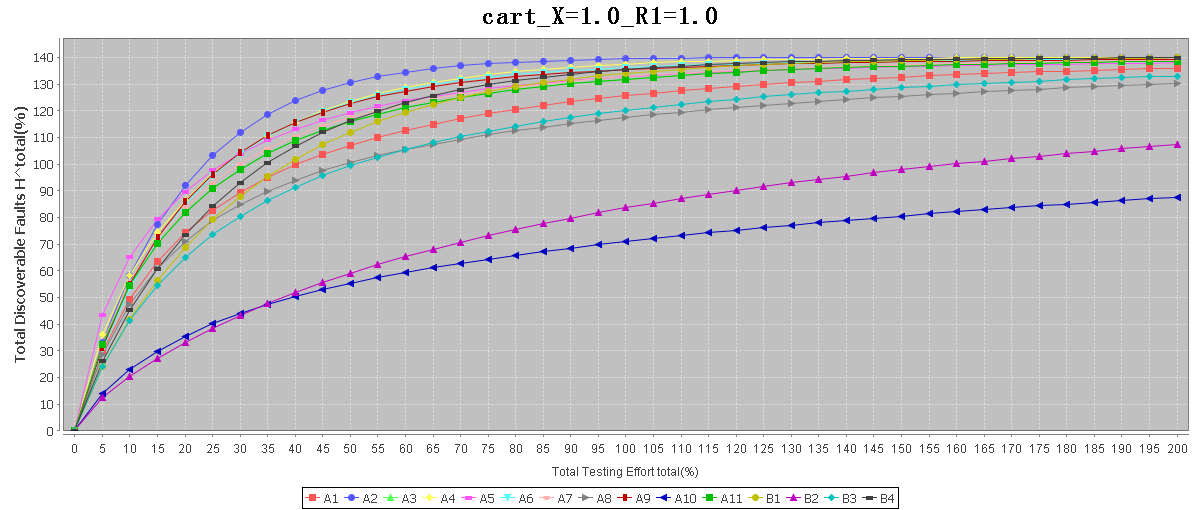


**版本6：缺陷发现模型是Rayleigh，Rayleigh无法得到最优工作量分配策略。横坐标为“当前测试工作量/实际测试工作量×100%”，纵坐标为“model在分配策略S下能发现的缺陷总数 / 总的实际缺陷总数 × 100%"。**

1. 版本6的结果都在这个目录下：

QQ截图20151202192643

1. 文件目录结构同版本1。
2. 样例：Rayleigh/lang/density/2\_by\_1/(lang.density.2\_by\_1)cart\_X=1.0\_R1=1.0.png:



存在的问题：

（1）因为DS模型无法计算b0（为了得到b0,要解方程ex + xex = c,c为常量，求x,无法得解）, 同样的原因，无法得到最优策略，所以未能进行实验。

（2）Rayleigh模型也是无法得到最优策略。

具体实验结果：https://github.com/faultPrediction/cefprat/tree/master/result