**Brain Storm Result Group05**

Our Group still wants to focus on the quality of Linux stable kernel code. There are some data we want to extract for indicators we designed in final project.

**The Properties we want to acquire:**

The number of fix tags, the author, commit date, fix data and numbers of commit lines.

We will extract the properties by using git command and regular expression.

**The Indication:**

1. Fix tags density of each version (belong to stability analysis)

We will try to get the fix tags density of each version and compare it with that of last version to get a change percent (tags density for this version/tags density for last version) and judge that with the version updating, the stability becomes better or not.

1. Better developer can produce better code (developer ranking and simple term analysis are involved)

We can extract the years that a developer has contributed for the kernel or the bug density of a developer’s commit to judge who is better. After that we will calculate the commit density (by count the author’s time for appearance) of those brilliant developers in each version to imply the code is stable or not.

1. The number of add line/the interval between every two subversion (stability analysis)

If the process of developing is regular and in an appropriate schedule, we may conclude that the code under that plan will be better. So, we use the numbers of add line of each subversion and divide it by the time interval it may take to see that if the indicator fluctuates in a limited range or has some rules.

1. All patch levels have the same distribution about bugs of sublevels

After getting numbers of bugs of sublevels, such as v4.4.1-90bugs, we make a histogram and infer that it conforms some distribution. Then we plot histogram of other patch levels to judge our model.

1. Bug density in 1000 lines (a kind of distance analysis?)

The number of fix tags / (number of lines/1000). The lower the indicator is the better the quality will be.

Absolutely, we will detect the data which may be regarded as outlier and then normalize it.

Additionally, we attempt to use multiple regression and take other advanced technologies into consider such as decision tree, CNN and Naive Bayes. We choose to use one of them and establish a model which can calculate the four indicators and produce a result.

Group members:

YuhaoZhang:320180940581

YumingDong:320180939701

HaoqiuYan:320180940440

YinruYe:320180940480

WanfengZhu:320180940691

QiyuanZhang:320180940541

YunHaoKang: 320180939851

JiachuanHe:320180939771