RQ1: Can we quantify interest of TD at the functional level? How much is the interest? (Version 5)

Yasutaka Kamei

Feb 13th, 2016

What's new in the version 5?

- Add histgrams

What's new in the version 4?

- We used the new dataset that solves one bug that is fixed by Everton.

Data Load

```
setwd("/Users/kamei/Research/techdebt/msr16_td_interest/")
source("./r_scripts/data_read.r")
```

Data Summary

- (Step 1) choose one of duplicated method and version name
- (Step 2) only use technical debt including metrics

```
nrow(data)
```

```
## [1] 837
```

```
apply(data.s1[,c("version_name","CountInput_v1","CountInput_v2")], 2, function(x){sum(x == -1) })
```

```
##  version_name CountInput_v1 CountInput_v2
##             8           221            94
```

Observation

- The number of methods that cannot be linked between Evernton's data and metrics data
  - 171 => 221 (introducing)
  - 101 => 94 (being found as last version)

(Step 3) use technical debt including non 0 for division

CountLine

```
summary(data.CountLine.all$Project)
```

```
##    apache-ant apache-jmeter         jruby
##            71           181           236
```

```
summary(data.CountLine.positive$Project)
```

```
##    apache-ant apache-jmeter         jruby
##            27            80            77
```

```
summary(data.CountLine.negative$Project)
```

```
##    apache-ant apache-jmeter         jruby
##            20            25            59
```

```
summary(data.CountLine.positive$Project) / summary(data.CountLine.all$Project) * 100
```

```
##    apache-ant apache-jmeter         jruby
##      38.02817      44.19890      32.62712
```

```
summary(data.CountLine.negative$Project) / summary(data.CountLine.all$Project) * 100
```

```
##    apache-ant apache-jmeter         jruby
##      28.16901      13.81215      25.00000
```

CountInput

```
summary(data.CountInput.all$Project)
```

```
##    apache-ant apache-jmeter         jruby
##            68           161           231
```

```
summary(data.CountInput.positive$Project)
```

```
##    apache-ant apache-jmeter         jruby
##            21            68            70
```

```
summary(data.CountInput.negative$Project)
```

```
##    apache-ant apache-jmeter         jruby
##            13            13            37
```

```
summary(data.CountInput.positive$Project) / summary(data.CountInput.all$Project) * 100
```

```
##    apache-ant apache-jmeter         jruby
##      30.88235      42.23602      30.30303
```

```
summary(data.CountInput.negative$Project) / summary(data.CountInput.all$Project) * 100
```

```
##    apache-ant apache-jmeter         jruby
##     19.117647      8.074534     16.017316
```

Observation

- The number of all methods is 837
  - (s1) 754
  - (s2) 488
- We use 71, 181, and 236 methods including technical debt.
  - The data set we used had 67 (ant), 169(jmeter) and 268(jruby) technical debt.
- 32.6%-44.2% of technical debt has positive interest.
- 13.8%-28.7% of technical debt has negative interest.

CountLine

```
# interest of CountLine (LOC)
fc <- factor(data.CountLine.all$Project)
interest <- data.CountLine.all$interest
tapply(interest, fc, summary)
```

```
## $`apache-ant`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -85.0000  -4.2910   0.0000  -0.8538   6.9050  75.0000
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -66.04    0.00    0.00   53.63   13.89 6667.00
##
## $jruby
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -95.8300  -0.4032   0.0000   6.2810   7.2440 362.5000
```

```
fc <- factor(data.CountLine.positive$Project)
interest <- data.CountLine.positive$interest
tapply(interest, fc, summary)
```

```
## $`apache-ant`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.370   6.155  11.110  18.400  25.400  75.000
```

```
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    1.562    6.920   18.010  126.300   50.000 6667.000
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.266   7.692  20.690  41.100  40.000 362.500
```

```
fc <- factor(data.CountLine.negative$Project)
interest <- data.CountLine.negative$interest
tapply(interest, fc, summary)
```

```
## $`apache-ant`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -85.000 -39.800 -23.160 -27.870  -6.561  -1.538
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -66.040 -17.240 -12.500 -15.740  -5.814  -2.000
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -95.830 -44.170 -16.670 -28.510  -9.167  -1.613
```

CountInput

```
# CountInput
fc <- factor(data.CountInput.all$Project)
interest <- data.CountInput.all$interest
tapply(interest, fc, summary)
```

```
## $`apache-ant`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -69.23    0.00    0.00   50.20   12.18 2700.00
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -66.67    0.00    0.00   23.03   20.00  900.00
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -68.75    0.00    0.00   21.70   14.29  450.00
```

```
fc <- factor(data.CountInput.positive$Project)
interest <- data.CountInput.positive$interest
tapply(interest, fc, summary)
```

```
## $`apache-ant`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    5.263   20.000   33.330  180.400  100.000 2700.000
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.556  12.380  25.000  60.470  50.000 900.000
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.143  17.500  33.330  87.880 100.000 450.000
```

```
fc <- factor(data.CountInput.negative$Project)
interest <- data.CountInput.negative$interest
tapply(interest, fc, summary)
```

```
## $`apache-ant`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -69.230 -50.000 -25.000 -28.910  -8.333  -8.333
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -66.67  -37.50  -29.63  -31.06  -18.18  -10.00
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -68.750 -50.000 -25.000 -30.810 -14.290  -6.667
```

Plot

```
library(reshape2)
library(ggplot2)

if(0){
idx <- data.CountLine.positive$Project == "apache-ant"
a1 <- data.frame(Interest=data.CountLine.positive[idx,"interest"])
g = ggplot(a1, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-ant.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountLine.positive$Project == "apache-jmeter"
a2 <- data.frame(Interest=data.CountLine.positive[idx,"interest"])
g = ggplot(a2, aes(x=Interest, y=..density.., fill=T, lims(x = c(0,400))))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jmeter.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountLine.positive$Project == "jruby"
a3 <- data.frame(Interest=data.CountLine.positive[idx,"interest"])
g = ggplot(a3, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
```

```r
ggsave(file = "/tex/figures/rq1-jruby.pdf", plot = g, width = 8.09, height = 5)
}
```