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/Downloads/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 Everton'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
```

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

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

```r
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

```r
# 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.
## -566.700   -4.487    0.000  -16.150    6.458   42.860
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -194.400    0.000    0.000    6.296   12.200   98.520
##
## $jruby
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -2300.0000    -0.4098     0.0000   -29.0900     6.7540    78.3800
```

```r
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.
##   1.351   5.798  10.000  13.980  20.240  42.860
```

```
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.538   6.471  15.260  22.260  33.330  98.520
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.250   7.143  17.140  21.740  28.570  78.380
```

```
#idx <- order(interest, decreasing = T)
#write.csv(file="a.csv", data.CountLine.positive[idx,])
#data.CountLine.positive$interest[idx]
#data.CountLine.positive$Remove_Comment[idx]

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.
## -566.700  -66.520  -30.210  -76.210   -7.028   -1.562
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -194.400  -20.830  -14.290  -25.640   -6.173   -2.041
##
## $jruby
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2300.000   -79.140   -20.000  -144.700   -10.090    -1.639
```

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.
## -225.0000    0.0000    0.0000   -0.7257   10.8300   96.4300
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -200.000    0.000    0.000    6.346   16.670   90.000
##
## $jruby
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -220.0000    0.0000    0.0000    0.1944   12.5000   81.8200
```

```r
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.
##    5.00   16.67   25.00   33.37   50.00   96.43
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.263  11.010  20.000  26.150  33.330  90.000
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.667  14.880  25.000  34.390  50.000  81.820
```

```r
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.
## -225.000 -100.000  -33.330  -57.700   -9.091   -9.091
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -200.00  -60.00  -42.11  -58.18  -22.22  -11.11
##
## $jruby
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -220.000 -100.000  -33.330  -63.850  -16.670   -7.143
```

Plot

```r
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(0,100))) + labs(x="Interest (%)")
g = g + geom_density(alpha = 0.5) + xlim(0, 100) + ylim(0,0.06) + 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(0,100)))) + labs(x="Interest (%)")
g = g + geom_density(alpha = 0.5) + xlim(0, 100) + ylim(0,0.06) + 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(0,100))) + labs(x="Interest (%)")
g = g + geom_density(alpha = 0.5) + xlim(0, 100) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jruby.pdf", plot = g, width = 8.09, height = 5)
}

if(0){
idx <- data.CountInput.positive$Project == "apache-ant"
a1 <- data.frame(Interest=data.CountInput.positive[idx,"interest"])
g = ggplot(a1, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,100))) + labs(x="Interest (%)")
g = g + geom_density(alpha = 0.5) + xlim(0, 100) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-ant-fanin.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountInput.positive$Project == "apache-jmeter"
a2 <- data.frame(Interest=data.CountInput.positive[idx,"interest"])
g = ggplot(a2, aes(x=Interest, y=..density.., fill=T, lims(x = c(0,100)))) + labs(x="Interest (%)")
g = g + geom_density(alpha = 0.5) + xlim(0, 100) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jmeter-fanin.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountInput.positive$Project == "jruby"
a3 <- data.frame(Interest=data.CountInput.positive[idx,"interest"])
g = ggplot(a3, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,100))) + labs(x="Interest (%)")
g = g + geom_density(alpha = 0.5) + xlim(0, 100) + ylim(0,0.04) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq1-jruby-fanin.pdf", plot = g, width = 8.09, height = 5)
}
```