

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 histograms

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] 332
```

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

```
## version_name CountInput_v1 CountInput_v2
##           6           119           46
```

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  
##           15           27           141
```

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

```
##      apache-ant apache-jmeter      jruby  
##           6           8           49
```

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

```
##      apache-ant apache-jmeter      jruby  
##           3           4           31
```

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

```
##      apache-ant apache-jmeter      jruby  
##      40.00000      29.62963      34.75177
```

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

```
##      apache-ant apache-jmeter      jruby  
##      20.00000      14.81481      21.98582
```

CountInput

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

```
##      apache-ant apache-jmeter      jruby  
##           14           22           139
```

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

```
##      apache-ant apache-jmeter      jruby  
##           2           6           34
```

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

```
##      apache-ant apache-jmeter      jruby  
##           4           3           18
```

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

```
##      apache-ant apache-jmeter      jruby
##      14.28571    27.27273    24.46043
```

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

```
##      apache-ant apache-jmeter      jruby
##      28.57143    13.63636    12.94964
```

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.
## -267.400   0.000   0.000 -31.470   6.389   21.430
##
## $`apache-jmeter`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -17.780   0.000   0.000   3.028   2.724   33.330
##
## $jruby
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -350.00   0.00   0.00  -5.46   7.50   78.38
```

```
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.
##   4.950   6.136   8.448  11.370  16.790   21.430
```

```
##
## $`apache-jmeter`
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.222   5.806   9.871  14.030  18.620  33.330
##
## $jruby
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.250   7.018  16.000  21.200  31.000  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.
## -267.40 -229.40 -191.30 -180.10 -136.40  -81.48
##
## $`apache-jmeter`
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -17.780  -9.247  -5.243  -7.621  -3.617  -2.222
##
## $jruby
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -350.000 -55.940 -20.000 -58.340 -11.270  -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.
## -225.000  -6.818   0.000 -22.910   0.000  57.140
##
## $`apache-jmeter`
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.000   0.000   0.000  -1.412   4.412  50.000
##
## $jruby
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -200.000   0.000   0.000   1.655   0.000  80.000
```

```
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.
##  25.00   33.04   41.07   41.07   49.11   57.14
##
## $`apache-jmeter`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.882   8.750  10.000  16.120  11.880  50.000
##
## $jruby
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.667  20.000  26.140  33.300  50.000  80.000
```

```
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.
## -225.000 -154.700  -84.380 -100.700  -30.400   -9.091
##
## $`apache-jmeter`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00  -58.33  -16.67  -42.59  -13.89  -11.11
##
## $jruby
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -200.000  -50.000  -30.950  -50.110  -12.950   -7.143
```

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(0,100))) + 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(0,0.04) + 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(0,100))) + 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)
}

```