RQ1: Non-SATD

Yasutaka Kamei

Apr 14th, 2016

Data Load

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

Data Summary

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

```
nrow(data)
```

```
## [1] 78705
```

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

```
##  version_name CountInput_v1 CountInput_v2
##             0         19594         15127
```

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
##          1357          4011          9983
```

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

```
##    apache-ant apache-jmeter         jruby
##           438           763          1686
```

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

```
##    apache-ant apache-jmeter         jruby
##           196           429          1827
```

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

```
##    apache-ant apache-jmeter         jruby
##      32.27708      19.02269      16.88871
```

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

```
##    apache-ant apache-jmeter         jruby
##      14.44363      10.69559      18.30111
```

CountInput

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

```
##    apache-ant apache-jmeter         jruby
##          1245          3691          9390
```

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

```
##    apache-ant apache-jmeter         jruby
##           278           845          3084
```

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

```
##    apache-ant apache-jmeter         jruby
##           144           298          2079
```

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

```
##    apache-ant apache-jmeter         jruby
##      22.32932      22.89352      32.84345
```

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

```
##    apache-ant apache-jmeter         jruby
##     11.566265      8.073693     22.140575
```

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.
##  -89.47    0.00    0.00   16.86   11.76 1000.00
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  -97.640    0.000    0.000    6.591    0.000 3060.000
##
## $jruby
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  -98.920    0.000    0.000    8.447    0.000 2000.000
```

```
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.
##    1.429   12.500   28.570   62.600   83.330 1000.000
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    1.205   11.110   25.000   50.630   50.000 3060.000
##
## $jruby
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    1.068   18.180   36.360   97.610  125.000 2000.000
```

```
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.
## -89.470 -28.140 -13.460 -23.150  -8.333  -1.370
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -97.64  -42.11  -23.08  -28.41  -10.00   -2.00
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -98.920 -66.670 -40.820 -43.920 -20.990  -1.235
```

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.
## -100.00    0.00    0.00   22.17    0.00 3800.00
##
## $`apache-jmeter`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00    0.00    0.00   11.99    0.00 1000.00
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00    0.00    0.00   38.41   33.33 8950.00
```

```
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.
##    2.985   25.000   50.000  119.700  100.000 3800.000
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    2.381   20.000   41.300   64.000   75.000 1000.000
##
## $jruby
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.80   33.33   60.00  147.80  133.30 8950.00
```

```
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.
## -100.000  -50.000  -33.330  -39.410  -20.000   -2.632
##
## $`apache-jmeter`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -100.000  -42.860  -27.620  -32.950  -20.000   -3.333
##
## $jruby
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -100.0000  -66.6700  -40.0000  -45.7600  -25.0000   -0.7092
```

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

if(0){
idx <- data.CountInput.all$Project == "apache-ant"
a1 <- data.frame(Interest=data.CountInput.all[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-fanin-non-SATD.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountInput.all$Project == "apache-jmeter"
a2 <- data.frame(Interest=data.CountInput.all[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-fanin-non-SATD.pdf", plot = g, width = 8.09, height = 5)

idx <- data.CountInput.all$Project == "jruby"
a3 <- data.frame(Interest=data.CountInput.all[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)
ggsave(file = "./tex/figures/rq1-jruby-fanin-non-SATD.pdf", plot = g, width = 8.09, height = 5)
}
```