

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")
```

```
## 8893 18152 21668 4156 35368
```

Data Summary

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

```
nrow(data)
```

```
## [1] 88237
```

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

```
## CountInput_v1 CountInput_v2
##              NA              NA
```

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)
```

```
##   ant  hadoop  jmeter  log4j  tomcat  NA's
##  5273  11247   8116   1387  17769    13
```

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

```
##   ant  hadoop  jmeter  log4j  tomcat
##   779   1232   1049   160   1196
```

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

```
##   ant  hadoop  jmeter  log4j  tomcat
##   359   544   1061    82    811
```

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

```
## Warning in summary(data.CountLine.positive$Project)/
## summary(data.CountLine.all$Project): longer object length is not a multiple
## of shorter object length
```

```
##      ant      hadoop      jmeter      log4j      tomcat      NA's
## 14.773374 10.954032 12.925086 11.535689 6.730823 5992.307692
```

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

```
## Warning in summary(data.CountLine.negative$Project)/
## summary(data.CountLine.all$Project): longer object length is not a multiple
## of shorter object length
```

```
##      ant      hadoop      jmeter      log4j      tomcat      NA's
## 6.808269 4.836845 13.072942 5.912040 4.564129 2761.538462
```

CountInput

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

```
##      ant hadoop jmeter log4j tomcat  NA's
## 4924 10712 7566 1225 16487 13
```

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

```
##      ant hadoop jmeter log4j tomcat
## 424 1884 1239 172 1811
```

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

```
##      ant hadoop jmeter log4j tomcat
## 252 744 498 89 823
```

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

```
## Warning in summary(data.CountInput.positive$Project)/
## summary(data.CountInput.all$Project): longer object length is not a
## multiple of shorter object length
```

```
##      ant      hadoop      jmeter      log4j      tomcat      NA's
## 8.610885 17.587752 16.375892 14.040816 10.984412 3261.538462
```

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

```
## Warning in summary(data.CountInput.negative$Project)/
## summary(data.CountInput.all$Project): longer object length is not a
## multiple of shorter object length
```

```
##      ant      hadoop      jmeter      log4j      tomcat      NA's
## 5.117790 6.945482 6.582078 7.265306 4.991812 1938.461538
```

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)
```

```
## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -88.240   0.000   0.000   6.671   0.000 1000.000
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -97.320   0.000   0.000   4.456   0.000 1350.000
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -96.510   0.000   0.000   4.054   0.000 7167.000
##
## $log4j
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -91.18    0.00    0.00   11.25    0.00 1233.00
##
## $tomcat
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -98.740   0.000   0.000   3.362   0.000 3200.000
```

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

```
## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.087  11.270  26.470  55.160  66.670 1000.000
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.9346  11.1100  25.4600  53.6000  56.8200 1350.0000
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.388  10.000  23.080  55.750  51.430 7167.000
##
## $log4j
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.124  15.940  36.550 114.900 100.000 1233.000
##
## $tomcat
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.581   7.692   19.020   68.570   50.000  3200.000
```

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

```
## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -88.2400 -29.5800 -14.8100 -21.7200 -7.8780  -0.8403
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -97.3200 -44.6200 -22.6200 -29.2600 -8.9920  -0.5405
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -96.5100 -26.9200 -21.4300 -24.1100 -11.1100  -0.6494
##
## $log4j
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -91.180 -50.000 -25.000 -33.880 -14.420  -1.923
##
## $tomcat
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -98.7400 -40.0000 -20.0000 -27.4700 -8.3330  -0.5291
```

CountInput

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

```
## $ant
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.000   0.000   0.000   5.284   0.000  3800.000
##
## $hadoop
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00   0.00   0.00   21.59   0.00  19850.00
##
## $jmeter
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00   0.00   0.00   9.04   0.00  1500.00
##
## $log4j
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.000   0.000   0.000   7.844   0.000   700.000
##
## $tomcat
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -100.00   0.00   0.00   10.86   0.00  15230.00
```

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

```
## $ant
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##      3.704   20.000   42.860   84.430  100.000 3800.000
##
## $hadoop
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##      0.98    20.00    40.00   135.90   100.00 19850.00
##
## $jmeter
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##      2.12    20.00    35.29   67.26   66.67 1500.00
##
## $log4j
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##      5.00    25.00    50.00   81.91  100.00   700.00
##
## $tomcat
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##      1.613   25.000   50.000   113.100 100.000 15230.000
```

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

```
## $ant
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##     -100.000  -50.000  -33.330  -38.810  -20.000   -3.125
##
## $hadoop
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##     -100.000  -50.000  -25.000  -33.280  -14.290   -1.613
##
## $jmeter
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##     -100.000  -35.120  -25.000  -30.010  -14.290   -2.632
##
## $log4j
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##     -100.000  -66.670  -44.440  -50.340  -25.000   -3.571
##
## $tomcat
##      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
##     -100.000  -42.860  -25.000  -31.430  -16.670   -1.099
```

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.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(0,400)))
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.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(0,400))))
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.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(0,400)))
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)
}

```