

Technical Report

Xunhui Zhang, Ayushi Rastogi, Yue Yu, Georgios Gousios

September 8, 2020

This is the appendix of TSE paper “pull-based development”.

1 Regression results for decision making with different contexts

1.1 The maturity of project

To find whether there is difference for the influence of factors on decision making for projects with different maturity. We split the data into three categories according to the time gap between project creation and data collection (project history in month), namely young projects ($proj_hist \leq 29$), semi-mature projects ($29 < proj_hist \leq 55$) and mature projects ($proj_hist > 55$). The result of regression models can be seen in Table 1.

Table 1: The result of regression models for projects with different maturity

	Young	Semi-mature	Mature
Num. obs	340,555	715,110	1,155,281
(Intercept)	1.23***	0.98***	1.2***
lifetime_minutes	-0.77*** (22.7%)	-0.63*** (17.3%)	-0.52*** (17.2%)
has_comments1	-0.88*** (10.3%)	-0.72*** (7%)	-1.05*** (13.7%)
prev_pullreqs	0.39*** (5.3%)	0.38*** (10.2%)	0.31*** (9.9%)
same_user0	2.1*** (32.1%)	1.99*** (38.4%)	1.75*** (40.9%)
ci_failed_perc	-0.6*** (2.2%)	-0.58*** (1.9%)	-0.25*** (0.2%)
open_pr_num	-0.34*** (3.7%)	-0.38*** (0.8%)	-0.2*** (0%)
core_member1	1.07*** (11.3%)	1.07*** (10.4%)	0.75*** (7.1%)
num_commits	0.32*** (7.2%)	0.3*** (6.8%)	0.27*** (5.1%)
project_age	0.03 (0%)	0.37*** (3.2%)	0.23*** (1.2%)
followers	0.2*** (1.6%)	0.14*** (1%)	0.17*** (1.6%)
hash_tag1	0.44*** (2%)	0.32*** (1.6%)	0.29*** (1.4%)
watchers	-0.12*** (0.1%)	0.02 (0%)	0.19*** (0.3%)
src_churn	0.15*** (0.5%)	0.09*** (0.1%)	0.02*** (0%)
team_size	0 (0%)	-0.02 (0%)	0.03*** (0%)
commits_on_	0.07*** (0.1%)	0.04*** (0%)	0.1*** (0.4%)
files_touched			
files_changed	-0.15*** (0.5%)	-0.14*** (0.5%)	-0.13*** (0.6%)
test_inclusion1	0.25*** (0.5%)	0.27*** (0.7%)	0.16*** (0.3%)

1.2 The popularity of project

In order to detect whether factors will have different influence on decision making for projects with different popularity. We split the data into three categories according to the star number of projects when collecting data, namely unpopular projects ($star_num \leq 1952$), semi-popular

projects ($1952 < star_num \leq 35153$) and popular projects ($star_num > 35153$). The result of regression models can be seen in Table 2.

Table 2: The result of regression models for projects with different popularity

	Unpopular	Semi-popular	Popular
Num.obs	208,700	415,911	1,507,258
(Intercept)	1.78***	1.27***	0.92***
lifetime_minutes	-0.81***(32.5%)	-0.68***(23.7%)	-0.53***(15.3%)
has_comments1	-0.88***(10.3%)	-0.79***(8.9%)	-0.98***(11.8%)
prev_pullreqs	0.31***(4%)	0.33***(8.4%)	0.34***(10.3%)
same_user0	1.68***(24.9%)	1.88***(32.6%)	1.85***(43.1%)
ci_failed_perc	-0.44***(1.2%)	-0.35***(0.7%)	-0.33***(0.6%)
open_pr_num	-0.5***(3.9%)	-0.25***(0.7%)	-0.19***(0%)
core_member1	0.97***(10.4%)	1.11***(13.4%)	0.8***(7.3%)
num_commits	0.3***(7.6%)	0.27***(6%)	0.29***(5.6%)
project_age	0.21***(1.9%)	0.16***(1.2%)	0.18***(0.9%)
followers	0.05***(0.1%)	0.16***(1.1%)	0.18***(1.9%)
hash_tag1	0.24***(0.6%)	0.29***(1%)	0.33***(1.8%)
watchers	0.28***(0.2%)	0.2***(0.2%)	0.22***(0.3%)
src_churn	0.16***(1%)	0.11***(0.3%)	0.03***(0%)
team_size	-0.11***(0%)	0.04***(0.1%)	0.01(0%)
commits_on_	0.16***(0.8%)	0.1***(0.3%)	0.07***(0.2%)
files_touched			
files_changed	-0.1***(0.2%)	-0.14***(0.5%)	-0.14***(0.6%)
test_inclusion1	0.28***(0.5%)	0.32***(0.9%)	0.17***(0.3%)

1.3 The programming language of project

We also split the dataset according to the programming language of project. There are 6 programming languages, namely Java, Javascript, Python, Ruby, Go, Scala. The result of each programming language is shown in Table 3.

Table 3: The result of regression models for projects with different programming language

	Java	JavaScript	Python	Ruby	Go	Scala
Num.obs	417,580	695,198	612,428	257,564	150,385	77,791
(Intercept)	1.64***	1.35***	1.6***	1.72***	1.41***	1.5***
lifetime_minutes	-0.47***(15.6%)	-0.65***(22.4%)	-0.61***(16.6%)	-0.62***(20.7%)	-0.6***(13.8%)	-0.33***(3.8%)
has_comments1	-0.88***(10.5%)	-0.91***(11%)	-1.14***(14.5%)	-0.95***(9.7%)	-0.66***(5.6%)	-0.83***(6.1%)
prev_pullreqs	0.36***(10.3%)	0.33***(10.5%)	0.35***(7.1%)	0.35***(10.4%)	0.4***(13.8%)	0.08***(6.1%)
same_user0	1.8***(47.6%)	1.78***(31.6%)	1.99***(39.7%)	1.91***(41.2%)	1.93***(40.4%)	1.51***(53.8%)
ci_failed_perc	-0.07***(0%)	-0.41***(1%)	-0.51***(1.3%)	-0.23***(0.3%)	-0.56***(1.9%)	-0.3***(1.5%)
open_pr_num	-0.24***(0.3%)	-0.34***(0.6%)	-0.2***(0.1%)	-0.13***(0.1%)	-0.02(0.1%)	-0.31***(0.9%)
core_member1	0.71***(6.8%)	0.96***(9.3%)	0.94***(9.2%)	0.88***(8.3%)	1.02***(10.1%)	0.6***(6.3%)
num_commits	0.22***(2.7%)	0.34***(8.2%)	0.26***(6%)	0.28***(4.1%)	0.38***(5.7%)	0.2***(8.9%)
project_age	0.11***(0.6%)	0.33***(2.3%)	0.22***(1.1%)	0.12***(0.4%)	0.05*(0%)	0.04(0.6%)
followers	0.13***(0.9%)	0.14***(1.2%)	0.16***(1.2%)	0.22***(2.6%)	0.24***(2.9%)	0.08***(0.5%)
hash_tag1	0.46***(2.4%)	0.23***(0.9%)	0.29***(1.3%)	0.24***(0.8%)	0.45***(3.2%)	0.61***(8.6%)
watchers	0.2***(0.3%)	0.04***(0%)	0.19***(0.2%)	0.14***(0.1%)	0.03(0%)	0.19***(0.2%)
src_churn	-0.02*(0.3%)	0.12***(0.4%)	0.06***(0.1%)	0.02(0%)	-0.07***(1.2%)	0.16***(1%)
team_size	0.07***(0.1%)	-0.02*(0%)	0.01(0%)	0(0%)	0.14***(0.3%)	0(0%)
commits_on_	0.12***(0.5%)	0.04***(0%)	0.08***(0.3%)	0.08***(0.2%)	0.04***(0%)	0.13***(1%)
files_touched						
files_changed	-0.13***(0.9%)	-0.12***(0.4%)	-0.11***(0.1%)	-0.15***(0.6%)	-0.14***(0.5%)	-0.13***(0.8%)
test_inclusion1	0.09***(0.1%)	0.14***(0.2%)	0.37***(1.3%)	0.19***(0.4%)	0.17***(0.4%)	-0.18.(0%)