Technical Report

March 11, 2022

This is the appendix of TSE paper "Pull Request Decision Explained: An Empirical Overall".

1 Results for decision making in special cases

Table 1 presents the results of how factors influence pull request decision when there exists contributor's comment or integrator's comment.

2 Results for factor *num_commits* at open and close time of pull requests

Table 2 shows the effect of the factor *num_commits* on pull request decisions at the time points of pull request open (submission) and close. From the results, we can see that the increase of *num_commits* decreases the merged probability of pull request at the time of pull request submission, while the opposite result is observed at the time of pull request close. Therefore, we think that the code modification during the process increases commit but promotes the pull request merge. Therefore, *num_commits* does not indicate the size of the patch in a strict sense and should depend on the specific moment.

3 Threshold sensitivity analysis

Figures 1 and Figure 2 show the differences between the model results for different choices of thresholds when studying the team size context and the

Table 1: presents the results of special cases, including *contrib_comment=1* and *inte_comment=1*.

| | Dependent variable: merged_or_not=1 | |
	contrib_comment=1	inte_comment=1
(Intercept)	11.0***	9.46***
same_user	0.50***[37.26%]	0.62***[18.83%]
lifetime_minutes	0.65***[19.81%]	0.61***[30.63%]
prior_review_num	1.52***[12.67%]	1.47***[12.70%]
core_member	1.30***[5.30%]	1.30***[5.66%]
num_commits	1.51***[10.76%]	1.54***[12.75%]
other_comment	1.13***[1.53%]	1.15***[2.35%]
ci_exists	1.10***[0.73%]	1.12***[1.13%]
hash_tag	1.12***[1.51%]	1.12***[1.81%]
files_added	0.90***[0.79%]	0.89***[1.03%]
prev_pullreqs	1.14***[0.71%]	1.17***[1.12%]
commits_on_files_touched	1.05***[0.21%]	1.04***[0.12%]
open_pr_num	0.94***[0.05%]	0.91***[0.12%]
account_creation_days	1.05***[0.24%]	1.05***[0.29%]
first_pr	0.93***[0.54%]	0.95***[0.32%]
test_churn	1.09***[0.40%]	1.09***[0.42%]
files_changed	0.96***[0.06%]	0.96***[0.07%]
project_age	1.10***[0.27%]	1.08***[0.19%]
reopen_or_not	0.97***[0.21%]	0.97***[0.18%]
contrib_open	1.03***[0.07%]	1.03***[0.09%]
stars	0.80***[0.60%]	0.81***[0.55%]
inte_open	1.14***[1.04%]	1.12***[0.88%]
description_length	1.00***[0.00%]	1.01***[0.02%]
pushed_delta	1.03***[0.08%]	1.03***[0.13%]
followers	1.02***[0.04%]	1.03***[0.05%]
contrib_cons	1.06***[0.20%]	1.05***[0.18%]
team_size	1.04***[0.03%]	1.06***[0.07%]
contrib_gender	0.98***[0.08%]	0.97***[0.11%]
files_deleted	0.96***[0.19%]	0.95***[0.26%]
pr_succ_rate	0.99***[0.01%]	0.99***[0.02%]
contrib_agree	0.96***[0.13%]	0.98***[0.04%]
contrib_extra	0.97***[0.08%]	0.98***[0.04%]
contrib_neur	1.00***[0.00%]	1.01***[0.00%]
inte_neur	1.05***[0.12%]	1.03***[0.00%]
num_comments	1.03***[0.09%]	1.17***[2.15%]
comment_conflict	1.00***[0.00%]	1.00***[0.00%]
Friday_effect	1.02***[0.03%]	1.01***[0.04%]
inte_agree	1.01***[0.00%]	1.00***[0.00%]
inte_extra	1.00***[0.00%]	1.01***[0.00%]
open_issue_num	0.98***[0.01%]	1.01***[0.00%]
sloc	1.02***[0.01%]	1.02***[0.01%]
test_inclusion	1.02***[0.03%]	1.02***[0.04%]
inte_cons	1.00***[0.00%]	1.02***[0.02%]
integrator_availability	0.99***[0.01%]	1.00***[0.00%]
src_churn	1.06***[0.21%]	1.06***[0.21%]
test_lines_per_kloc	1.01***[0.00%]	1.00***[0.00%]
perc_contrib_pos_emo	1.19***[3.83%]	-
perc_contrib_neg_emo	0.98***[0.05%]	-
perc_inte_pos_emo	-	1.24***[5.19%]
perc_inte_neg_emo	-	1.03***[0.10%]
Observations	770,280	823,711
AUC_train	0.837	0.836

Table 2: presents the results of num_commits at different time of pull request.

| | Dependent variable: merged_or_not=1 | | |
	num_commits(open)	num_commits(close)	num_commits(change)
(Intercept)	2.3418***	2.3398***	2.3446***
num__commits	−0.0505***	0.0003	0.0927***
Observations	3,170,221	3,170,221	3,170,221

project evolution context, respectively. The values in the figures indicate the Spearman correlation coefficients for the importance of different factors. The x and y coordinates indicate the thresholds. The red area in the lower-left corner shows the correlation of model results under different thresholds in the first interval. The red area in the upper-right corner shows the third, and the middle is the second interval.

From the results, we can see that, for different team sizes, when we choose 2-6 people as the threshold for small projects, the model results for small projects are strongly correlated ($\rho > 0.7$), and we find that the larger the threshold gap, the smaller the correlation of the results. Therefore, for the study of team size context, the threshold selection of 2-6 for distinguishing small projects satisfies the sensitivity detection. Similarly, we find that thresholds set to 8-12 for the division of large projects satisfy the sensitivity detection.

For the project evolution context, the results of threshold selection at different time points are different. We took six months as a span, selected three time points as the first interval threshold and the third interval threshold, respectively, and found that the results were all strongly correlated. Therefore, we considered 2015-12-01 to 2016-12-01 as the first and second interval division criteria and 2017-12-01 to 2018-12-01 as the second and third interval division criteria as satisfying the sensitivity detection.
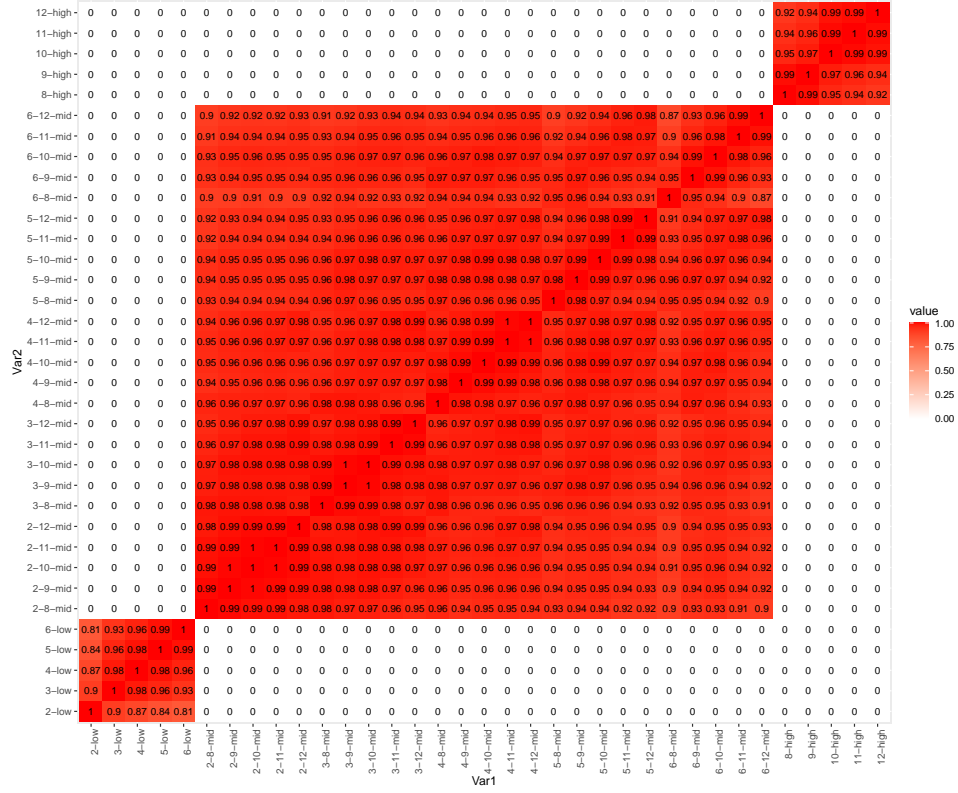
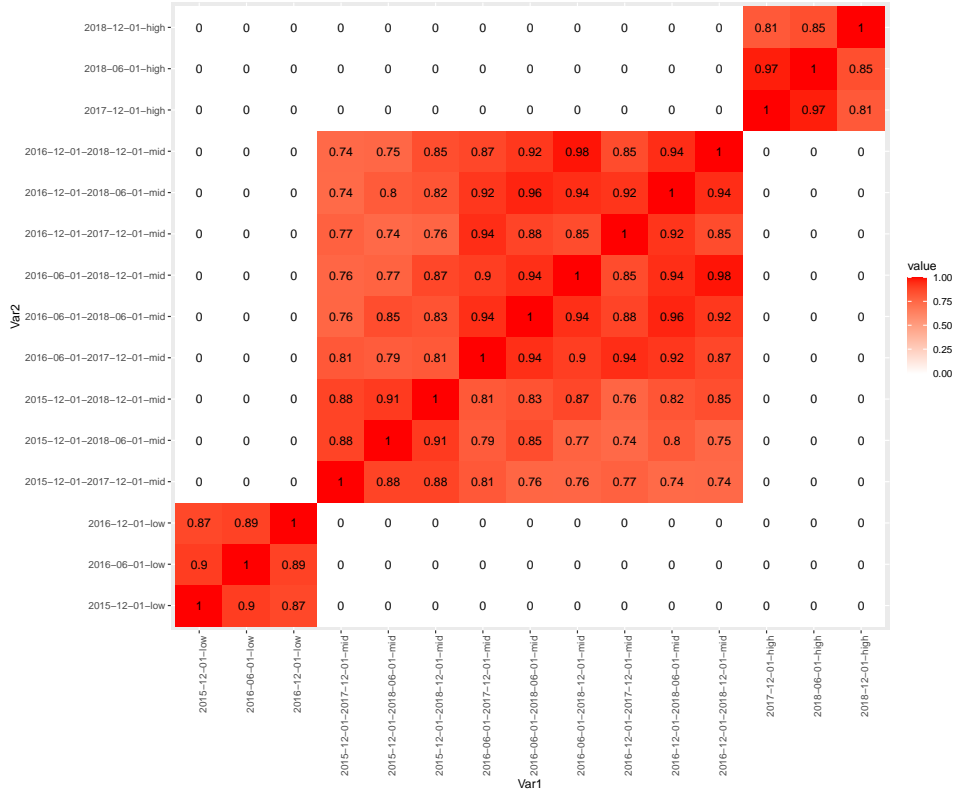Figure 1: The correlation between models with different team size thresholds

Figure 2: The correlation between models with different time thresholds