Applying the Submission Multiple Tier (SMT) Matrix to Detect Impact on Developer Interest on Open Source Project Survivability

Bee Chua

HAL Id: hal-01373057

https://hal.inria.fr/hal-01373057

Submitted on 28 Sep 2016

Applying the Submission Multiple Tier (SMT) Matrix to Detect Impact on Developer Interest on Open Source Project Survivability

Bee Bee Chua[1,2]

[1] University of Technology, Sydney, [2] University of New South Wales

Abstract. There is a significant relationship between project activity and developer interest on Open Source (OS) projects. Total project activity submission count number can be an indicator for gauging developer interest. The higher the project activity submission of a project is, the larger developer interest in a project. Our paper proposed that applying a Submission Multiple Tier (SMT) matrix can detect the impact of developer interest on project activity. Results showed more volume of OS projects with low project activity than high. Activity submission results also showed that developers are more likely to review than correct projects, with the first priority to find and fix bugs. Further research is needed to determine the impact of project activity type on developer motivation to contribute, participate and support OS projects.

1 INTRODUCTION

For any Open Source (OS) software to survive, appropriate OS infrastructure must be in place. These include the viability of project activities increases, high quality codes must produce, quality and qualifying peer developer networks expandable and OS models adoption are strategies for building an effective infrastructure OS for projects.

Developer support and collaboration is vitally important for maintaining an OS project's survivability, especially if the project is new. Project activity is crucial to survival. Making project activity viable is important as OS coordinators and sponsors can monitor project activity performance, including growth, reptutation and status.

This paper introduces the Submission Multiple Tier (SMT) matrix, which is based on a structured, hierachical tier for detecting submission pattern similarity over multiple project activities. The aim of the SMT is to allow possible detection of developer interest by activity type and project population. The SMT can assist OS coordinators and sponsors by: detecting projects with low survival difficulty based on the low submission counts from project activities; and identifying specific project activity interests of OS developers by the high submission counts.

The SMT was constructed to provide a summary of a total project number on each project acitvity by submission tier, so that OS spon sors and resource planning coordinators can provide better OS infrastructure support (for instance, developing a network strategy to increase developer interest on a particular project activity with extremely high number of unfixed bugs.

This paper is divided into five sections: section 2 discusses existing literature on OS infrastructure to support survivability, particularly project activity and developer interest; section 3 introduces the SMT and outlines the procedures for applying the SMT on a Sourceforge.net dataset; section 4 presents results; and section 5 conclusions and future work.

2 OPEN SOURCE INFRASTRUCTURE: PROJECT ACTIVITY AND DEVELOPER INTEREST

A number of studies have investigated OS variables for survival analysis, focusing on programming languages, licences, developer interest, operating systems and end user interest [1-10]. Studies [4-11] also used these variables to measure other project performance outcomes, such as successability, popularity, efficiency and effectiveness. They are essential variables for supporting OS infrastructure.

Good infrastructure is essential for OS project survival. The term 'infrastructure' is defined as "an underlying base or foundation especially for an organization or system" [11]. In this paper, infrastructure is classified as basic or advanced. Basic infrastructure is essential components needed to develop and support OS software (OSS), such as programming languages, licences, operating systems and developers, whereas advanced infrastructure refers to activities positively or negative influencing basic components of OS programs, projects and products. An example of an advanced OS infrastructure is project activity status, which influences developer interest and impacts project download. In addition to a high OSS adoption rate, project activity and high developer interest are vital infrastructure components to enable project survival.

Current OS literature discusses how project activity can be used to predict project success and popularity [3,4,9,10]; however, there is minimal work analysing this variable from a survivability perspective. Our research showed non-surviving projects were physically removed from an OS server, irrespective of project activity. This was in contrast to the study by Samoladas [12] that did not emphasise zero project activity for non-surviving projects, instead classifying project activity status as inactive. As a result their analysis on non-surviving projects showed an absence of zero project activity, which is a crucial variable to consider.

3 METHODS

The following procedures were followed to analyse project activity submissions and classify projects by survival status.

1. Sourceforge was chosen as it is the world's largest repository of OS projects, with over 100,000 projects and over a million registered users [13,14]. Sourceforge.net data dump files for 2010 and 2012 were downloaded ("stats_project_all"). Each file had 13 standard variables on download: project ID, number of developers, number of submissions opened and closed on bugs, support, patch, artefact and task. However only 10 standard variables were related to project activity and were selected for data review, creating 9 columns for the Multiple Submission Multiple Tier (SMT) Structure (Table 1).

Table 1 Submission Tier Matrix Structure.

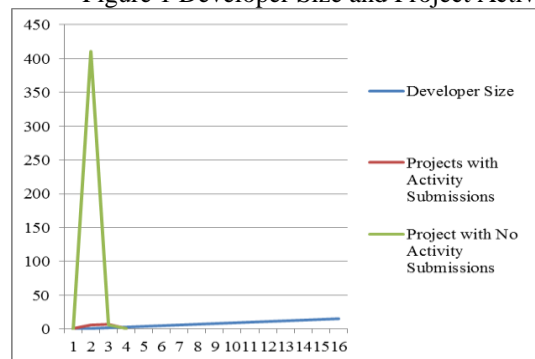| Variable Type	Activity	Description
Nominal	Bugs opened, support opened, patch opened, artefact opened, task opened	Number of submission opened
Nominal	Bugs closed, support closed, patch closed, artefact closed, task closed	Number of bugs opened
Nominal	Submission Tier 1 to 6	Submission Tier 1 to 6

2. Data were cleaned to remove blank submissions or projects with zero bug submissions opened, leaving 89,002 projects in 2010 and 89,916 in 2012. In 2010, submissions closed totalled 73,016 and for 2012, 66,777. Each project had up to five project activities (bug, support, patch, artefact and task). The project activity submission ranges varied.

3. The projects were then classified as surviving or non-surviving using the SMT. Project ID was used as the key identifier at both time points. If the same project number was present in 2010 and 2012 then it confirmed the project had survived (hosting) and had not been removed from the Sourceforge database. If a 2010 project ID had no corresponding 2012 proje ct ID and zero bug submissions opened (a value marked with '0') then we confirmed the project was deleted, i.e., no longer hosting. By 2010, 435 non-surviving projects had been deleted and archived in Sourceforge. These projects were examined further to determine the relationship between developer size and project activity. Table 2 shows non-surviving projects against project activity. Figure 1 shows effect of activity on developer size of non-surviving projects.

Table 2 Non-surviving Projects Against Project Activity.

Figure 1 Developer Size and Project Activity.

No. of Developers	Projects with Activity Submissions *	Projects With No Activity Submissions *
0	1	0
1	6	411
2	7	6
3	1	1
7	1	0
15	1	0

* Bug, support, artefact, patch, task.

4. The project total was counted for each activity case for 2010 and 2012, based on their submission number and per SMT tier in section 4.

4 RESULTS

Project Activities for Open Submissions Tiers

Table 3 illustrates the opened submissions pattern of multiple project activities, showing open submission tiers, total project count for each tier, project activities for 2010 and 2012 and project activity performance in terms of change between 2010 and 2012. Each project activity cell shows the project total number. The highest project population was bug submissions opened (26,453 projects in 2010 and 26,636 projects in 2012). Submission tier 1 had the lowest submission count (1–10 submissions), with bugs, support and patch being the three most active project activities in 2010 and 2012: bugs 26,453 (yr10) and 26,636 (yr12); support 17,053 (yr10) and 17,055 (yr12); patch 7402 (yr10) and 7112 (yr12). Projects in this tier could have a high source code quality and, as a result, low active project activity submission by developers for bugs, support, patch, artefact and task

Table 3 SMT Results for Open Submissions.

Submission Tier	Bug Submission Opened 2010	Bug Submission Opened 2012	Support Submission Opened 2010	Support Submission Opened 2012	Patch Submission Opened 2010	Patch Submission Opened 2012	Artefact Submission Opened 2010	Artefact Submission Opened 2012	Task Submission Opened 2010	Task Submission Opened 2012
Tier 1 1-10	26453	26636	17053	17055	7402	7112	7140	3049	1133	1335
Tier 2 11-100	6751	6719	5030	4983	4363	4636	4607	6156	1371	1338
Tier3 101-1000	1477	1495	1211	1205	1378	1413	1391	1813	1243	3771
Tier4 1001-10,000	128	132	99	105	167	174	168	344	374	372
Tier5 10,001-100,000	2	2	5	8	9	10	9	15	38	38
Total projects for each project activity	34811	34984	23398	23356	13319	13345	13315	11377	4159	6854

Project Activities Closed Submission Tiers

Table 4 displays six closed submissions tiers, one total project for each tier and ten other project activities for 2010 and 2012. The cells show total projects for each project activity closed submission. Closed bug submissions were the most active (19,258 closed in 2010 and 22,752 in 2012), with 1395 in 2010 and 1665 in 2012 for artefact submissions closed, and 449 and 454 respectively in 2010 and 2012 for task submission closed (see Table 4).

Table 4 SMT on Submissions Closed.

Submission Tier	Bug Submission Closed 2010	Bug Submission Closed 2012	Support Submission Closed 2010	Support Submission Closed 2012	Patch Submission Closed 2010	Patch Submission Closed 2012	Artefact Submission Closed 2010	Artefact Submission Closed 2012	Task Submission Closed 2010	Task Submission Closed 2012
Tier 1 1-10	19258	22752	9558	10398	3766	3049	1395	1665	449	454
Tier 2 11-100	6165	6215	3934	4050	3592	6156	2239	2067	750	677
Tier3 101-1000	24476	1723	1158	1161	1646	1813	1680	1603	854	807
Tier4 1001-10,000	171	177	135	138	264	344	432	288	366	1162
Tier5 10,001-100,000,0	1	5	3	5	10	15	18	18	18	35
Tier 6 100,001-100,0000	0	0	0	0	0	0	1	0	1	0
Total projects for each project activity	50071	30872	14788	15752	9278	11377	5765	5641	2438	3135

For tier 1, opened and closed submissions were similar and had the highest project populations. There were more opened than closed tier 1 bug submissions, suggesting developers were more interested in reporting bugs than correcting them. This could be due to developers finding it easier to report than correct during review, as no solution is required. The tier 2 project population ranged from 2000 to 6500 closed submissions: 6165 projects in 2010 and 6215 projects in 2012 for closed bug submissions; 3934 and 4050 in 2010 and 2012 respectively for closed support submissions; 3592 and 6156 for closed patch submissions; 2239 and 2067 for closed artefact submissions; and 750 projects in 2010 and 677 projects in 2012 for closed task submissions.

5 CONCLUSIONS AND FUTURE WORK

For the SMT matrix, three common survival patterns for project activity were found from the two submission patterns: 1) projects survive with a minimum of one project activity submission; 2) the most activity submissions – either open or closed – are bug submissions; and 3) more submissions can positively influence project survival from the support perspective. The submission patterns also revealed that many projects have very low project activity submissions and that developers review more than correct, motivated by project type rather than project activity. We plan to extend our work by investigating project submission based on project activity to confirm survivability to 5, 10 and 15 years. We will also validate the SMT on other OS repositories.

REFERENCES

1. A. Hars and S. Ou, "Working for free? – motivations of participating in open source projects," in Proceedings of the 34th Annual Hawaii International Conference on Systems Sciences, 2001.
2. A. Mockus, R. T. Fielding, and J. D. Herbsleb, "Two case studies of open source software development: Apache and Mozilla," ACM Transactions on Software Engineering and Methodology, Vol. 11, No. 3, pp. 309–346, 2002.
3. K. Crowston, J. Howison, and H. Annabi, "Information systems success in free and open source development: Theory and measures," Software Process and Practice, Vol. 11, No. 2, pp. 123–148, 2006.
4. V. Midha and P. Palvia, "Factors affecting the success of open source software," The Journal of Systems and Software, Vol. 85, No. 4, pp. 895–905, 2012.
5. S. Chen, "Determinants of survival of open source software: An empirical study," Academy of Information and Management Sciences Journal, Vol. 13, No. 2, pp. 119–128, 2010.
6. H.W. Lee, S.T. Kim, and S. Gupta, "Measuring open source software success," Journal of Omega, Vol. 37, No. 2, pp. 426–438, 2009.
7. N.J. Choi and S. Chengalur-Smith, "An exploratory study on the two new trends in open source software: End-users and service," in the Proceedings of the 42nd Hawaii International Conference on System Sciences, 2009.
8. J. Wang, "Survival factors for free open source software projects: A multi-stage perspective," European Management Journal,Vol. 30, No. 1, pp. 352–371, 2012.
9. C. Subramaniam, R. Sen, and M. L. Nelson, "Determinants of open source software project success: A longitudinal study," Journal of Decision Support Systems, Vol. 2, No. 46, pp. 576–585, 2009.
10. A.H. Ghapanchi and A. Aurum, "Competency rallying in electronic markets: Implications for open source project success," Journal of Electronic Markets, Vol. 22, No. 2, pp. 11–17, 2012.
11. http://en.wikipedia.org/wiki/SourceForge
12. I. Samoladas, L. Angelis and I. Stamelos , "Survival duration on the duration of open source projects " Journal of Software and Information Technology, Vol. 52, No.9, pp. 902-922, 2010
13. S. Christley and Madey, G. Analysis of Activity in the Open Source Software Development Community. In Proceedings of the 40th Hawaii International Conference on System Sciences.2007
14. http://www.sourceforge.net/