

与问题解决速度之间的相关性难以确定.

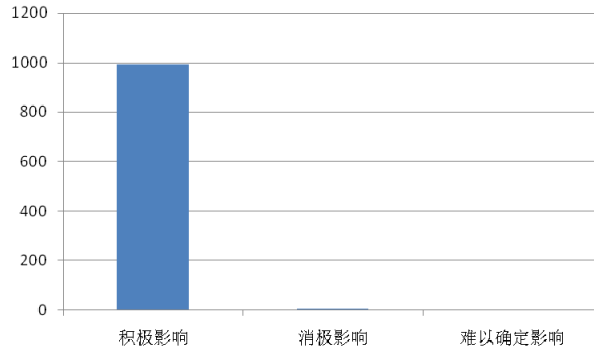


Fig.10 Relationship between problem comment speed and problem resolve speed

图 10 问题评论速度与问题解决速度的关系图

根据上面的实验,再结合实际的 GitHub 开发过程对实验 4 的分析,可以看出:在项目开发的过程中,当问题评论的速度增加时,表明参与开发的程序员对问题的关注度变高,从而在一定程度上使得问题解决速度加快.而对于还有存在消极影响或难以确定影响的 10 多个项目,本文也做了进一步的调研和分析,结果发现:这些项目也存在开发人员数目较少,且代码更新的速度也相当慢,从而使得问题评论速度与问题解决速度之间没有必然的积极影响.

实验 5. 解决问题总量 P_SOL_ALL 与代码变化速度 C_VEL

实验结果如图 11 所示.从图 11 可以看出:1 000 个项目中,超过 980 个项目的解决问题总量与代码变化速度存在消极影响,少于 20 个项目的问题评论速度与问题解决速度存在积极影响,只有 5 个项目中的问题评论速度与问题解决速度之间的相关性难以确定.

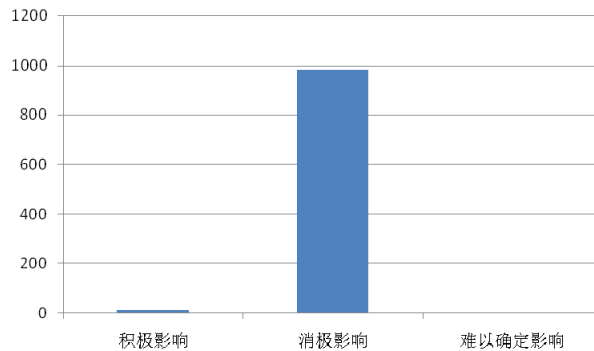


Fig.11 Relationship between problem resolve number and code change speed

图 11 解决问题总量与代码变化速度的关系图

根据上面的实验,再结合实际的 GitHub 开发过程对实验 5 的分析,可以看出:在项目开发的过程中,随着解决问题总量的增加,代码越来越趋于稳定,代码变化的速度也相应变慢.而对于还有存在消极影响或难以确定影响的 20 多个项目,本文也做了进一步的调研和分析,结果发现:这些项目也存在开发人员数目较少,且有些项目的代码基本上很长时间不再更新,从而使得解决问题总量与代码变化速度之间没有必然的相关性.

4.5 对研究问题的回答及对GitHub开源软件开发过程的几条建议

本文在此前提出了两个研究问题,在基于上述 5 个实验的基础上,这两个问题也有了相应的答案.影响 GitHub 开源软件开发过程中的主要因素之间存在一定的相关性,比如本文所提出的贡献者人数 N 与问题解决

速度 P_{SOL_VEL} 、问题增加速度 P_{ADD_VEL} 与请求合并增加速度 M_{ADD_VEL} 、问题评论速度 P_{COM_VEL} 与问题解决速度 P_{SOL_VEL} 之间就存在一定的积极影响,解决问题总量 P_{SOL_VEL} 与代码变化速度 C_{VEL} 则存在一定的消极影响。

利用这些结论,可以对基于 GitHub 开源软件的开发过程给出几条建议,这些建议可以给开源项目的管理者或是参与项目开发的程序员。

- (1) 在 GitHub 开源软件开发初期,适量地吸引更多的贡献者来参与.这样做的好处是:可以使得问题解决的速度在一定程度上加快,从而可能提高开发的效率;
- (2) 尽量提高提交问题的频率有助于请求合并增加的速度.这样做的好处是:随着提交的问题的频率变快,请求合并的速度也有可能加快,从而可能提高开发的效率;
- (3) 尽量提高问题评论的频率有助于问题尽快解决.这样做的好处是:随着对问题评论的频率变快,问题解决的时间间隔有可能会减少,从而可能提高开发的效率;
- (4) 尽量多解决问题,这样会尽早使得代码趋于稳定.这样做的好处是:随着解决问题的数量越来越多,代码中存在的问题可能会越来越少,从而让开发过程变得更有效率。

另外,本文也给出了一些实施的建议:对于吸引更多的贡献者来参与开发的问题,可以采取宣传、激励、寻求企业的支持等方法,并能经常进行积极地沟通与协作,使贡献者之间能够和谐地进行工作.对于经验非常丰富的贡献者,应该采取一定的奖励机制,使得项目能够更好更快地进行;有针对性地提高问题提交的频率,并且为贡献者创造不错的问题评论的氛围,让大家畅所欲言且能够集中焦点,这需要从管理学和心理学等多个角度来进行处理;对于解决问题的快慢,可以研究开源软件,尤其是基于 GitHub 的软件问题分配的各种策略,采用更好的分配策略,将问题分配给更合适的贡献者进行修改。

4.6 威胁实验有效性因素

GitHub 开源软件开发过程中影响因素的相关性分析虽然得到了多条很有意思的结论,但考虑到所分析的开源软件数目还不够多,且没有考虑更多的影响因素,比如贡献者的水平、评论中是否包含情感等因素,因此,后续还需要进行更多的实验和考虑更多的影响因素来做进一步的验证.另外,本文只考虑两种影响因素之间的相关性,缺乏对多种影响因素之间相关性的考察,且目前的分析还没有进一步量化,后续会在这些方面进行更深的研究.除此之外,本文所选择的实验程序还只是 1 000 个开源程序,与 GitHub 超过 1 200 万的项目相比依然是非常小的数目,因此也有可能存在一定的偶然性。

5 总结与展望

本文分析了 GitHub 开源软件开发过程中的影响因素,提出了问题解决速度、提交增加速度、问题增加速度、请求合并增加速度、问题评论速度、合并评论速度、解决问题总量和对代码变化速度等影响因素,并对这些因素之间存在的相关性进行了分析.通过对典型的 GitHub 开源软件进行实验,得出了对 GitHub 开源软件开发过程非常积极的结论.由于 GitHub 开源软件开发过程非常复杂,本文肯定还存在一些需要改进的地方.在后续的研究中,会考虑更多的影响因素和多种影响因素之间的相关性。

致谢 衷心感谢北京航空航天大学软件工程研究所开源软件库挖掘小组的成员在此前所做的工作,同时感谢在百忙之中对本文进行评阅的各位专家。

References:

- [1] DiBona C, Ockman S, Stone M. Open sources: Voices from the open source revolution. Contact Center Call Center & Ip Solutions, 1999.
- [2] Sen R, Singh SS, Borle S. Open source software success: Measure and analysis. Decision Support Systems, 2012,52(2):364–372. [doi: 10.1016/j.dss.2011.09.003]
- [3] Chacon S. Pro Git. 2nd ed., Berkeley: Apress, 2009.

- [4] Bird C, Rigby PC, Barr ET, Hamilton DJ, German DM, Devanbu P. The promises and perils of mining Git. In: Proc. of the 6th IEEE Int'l Working Conf. on Mining Software Repositories (MSR 2009). 2009,37(5):1-10. [doi: 10.1109/MSR.2009.5069475]
- [5] Zhou MH. Looking for micro-process in large-scale data. In: Proc. of the 2nd Int'l Workshop on Evidential Assessment of Software Technologies (EAST 2012). New York: ACM Press, 2012. 39-42. [doi: 10.1145/2372233.2372245]
- [6] Kalliamvakou E, Gousios G, Blincoe K, Singer L, German DM, Damian D. The promises and perils of mining GitHub. In: Proc. of the 11th Working Conf. on Mining Software Repositories (MSR 2014). New York: ACM Press, 2014. 92-101. [doi: 10.1145/2597073.2597074]
- [7] Mockus A, Fielding RT, Herbsleb JD. Two case studies of open source software development: Apache and Mozilla. ACM Trans. on Software Engineering Methodol, 2002,11(3):309-346. [doi: 10.1145/567793.567795]
- [8] Krishnamurthy S. Cave or community? An empirical examination of 100 mature open source projects. First Monday, 2002,7(6).
- [9] Hars A, Ou S. Working for free? Motivations of participating in open source projects. In: Proc. of the 34th Annual Hawaii Int'l Conf. on System Sciences (HICSS-34). 2001. [doi: 10.1109/HICSS.2001.927045]
- [10] Feller J, Fitzgerald B, Hissam SA, Huff KR. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. 2003. 3-21.
- [11] Sohn S, Mok M. A strategic analysis for successful open source software utilization based on a structural equation model. The Journal of Systems and Software, 2008,81(6):1014-1024. [doi: 10.1016/j.jss.2007.08.034]
- [12] Subramaniam C, Sen R, Nelson ML. Determinants of open source software project success: A longitudinal study. Decision Support Systems, 2009,46(2):576-585. [doi: 10.1016/j.dss.2008.10.005]
- [13] Katsamakos E, Georgantzas N. Why most open source development projects do not succeed? In: Proc. of the 1st Int'l Workshop on Emerging Trends in FLOSS Research and Development (FLOSS 2007). Washington: IEEE Computer Society, 2007. [doi: 10.1109/FLOSS.2007.15]
- [14] Dabbish L, Stuart C, Tsay J, Herbsleb J. Leveraging transparency. IEEE Software, 2013,30(1):37-43. [doi: 10.1109/MS.2012.172]
- [15] Peterson K. The GitHub open source development process. Technical Report, Mayo Clinic, 2013.
- [16] Gousios G, Pinzger M, van Deursen A. An exploratory study of the pull-based software development model. In: Proc. of the 36th Int'l Conf. on Software Engineering (ICSE 2014). New York: ACM Press, 2014. 345-355. <http://doi.acm.org/10.1145/2568225.2568260> [doi: 10.1145/2568225.2568260]
- [17] Zhou M, Mockus A. Who will stay in the FLOSS community? Modeling participant's initial behavior. IEEE Trans. on Software Engineering, 2015,41(1):82-99. [doi: 10.1109/TSE.2014.2349496]
- [18] Casebolt JR, Krein JL, Maclean AC, Delorey DP. Author entropy vs. file size in the GNOME suite of applications. In: Proc. of the 6th IEEE Int'l Working Conf. on Mining Software Repositories. 2009. 91-94. [doi: 10.1109/MSR.2009.5069484]
- [19] <https://github.com>



杨波(1981-),男,湖北荆州人,博士,副教授,CCF 专业会员,主要研究领域为软件测试,软件仓库挖掘。



吴际(1974-),男,博士,副教授,CCF 专业会员,主要研究领域为软件测试,软件可靠性分析。



于茜(1993-),女,硕士生,主要研究领域为软件测试。



刘超(1958-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为软件工程,软件测试。



张伟(1993-),男,硕士生,主要研究领域为软件过程方法与技术。