

What Make Long Term Contributors: Willingness and Opportunity in OSS Community



Minghui Zhou



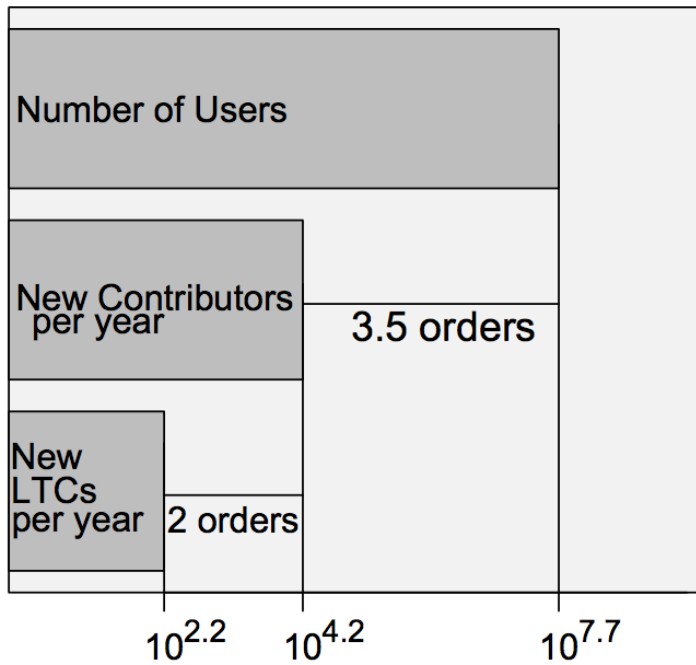
Audris Mockus

背景

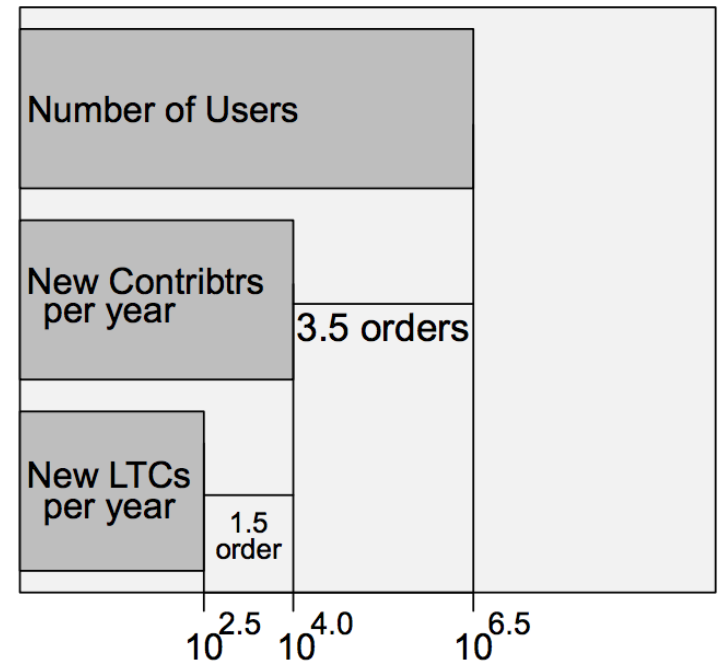
- Long-term contributor(LTC)はプロジェクトの成功に必要不可欠
- LTCでなければ重大なタスクを達成できにくい
 - 開発者は一人前になるのに少なくとも3年かかる
- 新人はほとんど将来的にLTCにならない

背景

Mozilla (average over 2000–2008)



Gnome (average over 1999–2007)



目的

- 参加初期の時点で将来Long-term contributor(LTC)になる開発者を予測したい
- 能力や意欲, 環境を数量化して測定し, 将来LTCになるかを推定するモデルを作る

アプローチ

- 未来のLTCに関係する要因を導き出す
 - 40 contributors (20 non-LTCs/20 LTCs)を調査
- 開発者にインタビュー
- Bugzillaの活動をメトリクスとして抽出
- ロジスティック回帰で分析

データセット

- Gnome(2000~2011/1)と
Mozilla(1999~2011/1)の不具合情報

Table I
PROJECTS

Project	Years	MLOC ⁷	Domain	Cntrbtrs
Gnome	10	7.9	UI	156,332
Evolution		.8	Calendar&Mailbox	21,041
Nautilus		.1	File manager	17,430
Epiphany		.1	Browser	3,716
Mozilla	12	20	UI	187,333
Firefox		5.3	Browser	47,690
Thunderbird		1.1	Mailbox	12,993
Calendar		.8	Calendar	4,130

結果

- 能力と意欲に関して
 - 不具合報告もしくはコメントを1度しかしていない開発者(One-Time-Contributor)が70%以上
 - Mozilla:70% Gnome:78%
 - OTCのreportの90%はFIXEDで終わり, そのうち35~47%はDUPLICATE
 - 対してLTCのreportの50%がFIXEDで終わり, そのうち15%がDUPLICATE

考察

Table II
QUARTILES OF NUMBER OF COMMENTS

Resolution	Gnome				Mozilla			
	25%	50%	75%	100%	25%	50%	75%	100%
DUPLICATE	1	1	1	283	2	3	5	722
INCOMPLETE	1	2	2	200	2	3	5	168
INVALID	1	2	4	247	2	3	6	384
WONTFIX	1	2	5	819	2	5	9	760
FIXED	2	3	6	1578	3	7	14	756

- 開発者の意欲や能力はタスクの種類や数で測定できる

結果

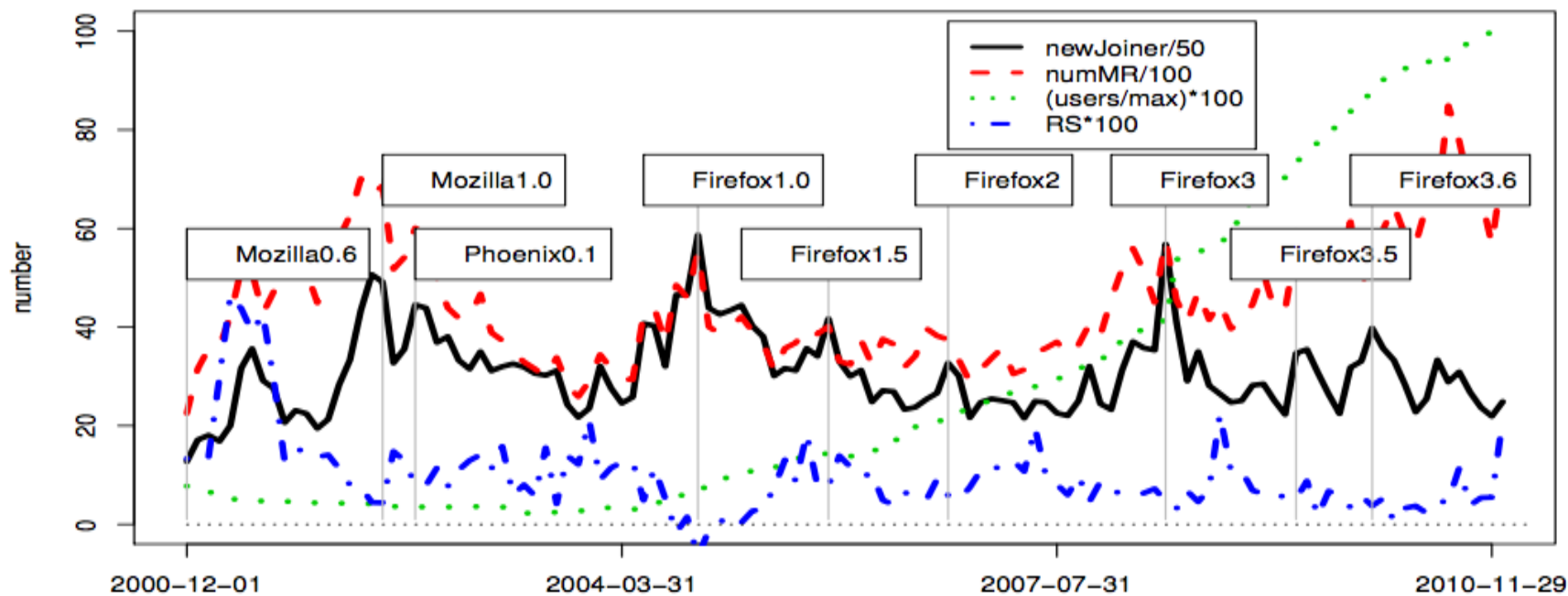


Figure 1. Macro-climate in Mozilla

結果

- ❖ Macro-climate: *popularity*:
 - ❖ “GNOME is something which you can show to your friends and family members”
- ❖ Micro-climate: *attention, number of peers, performance of peers*
 - ❖ “With bugzilla, ... the feedback from the developers shows that they care, and appreciate the effort I made, and actively work to solve the bug in a way that I can see progress.”
 - ❖ “As I met a lot of nice people at GUADECs who became friends there was also a personal component involved in the motivation.”
 - ❖ “I learned a lot from this leading open source project while working with other contributors”

考察

- 環境に関して
 - 以下を測定できる
 - Macro-climate (プロジェクト全体)
 - プロジェクトの人気
 - プロジェクトのSociality
 - Micro-climate (個々人にユニーク)
 - 同僚の数
 - 同僚の生産性
 - 同僚間のネットワーク
 - 同僚からの関心

メトリクス

Table III
PREDICTORS FOR PARTICIPANT i

Dimension	Predictor	Description
Willingness (and capacity)	nCmt	Logarithm of the number of comments +1
	GotFix	At least one of the issues reported by i was fixed
	withBB FNotRep	First report by i uses a crash reporting tool i starts participation with a comment
Macro- climate	nUsr	Number of product users when i joins
	RS	Project's relative sociality
Micro- climate	nPeer	Logarithm of i 's peers' group size: $\ln \ \cup_{p \in Peers(i)} Peers(p)\ + 1$, where $Peers(p)$ is the peer group for p
	pShared	Logarithm of the social clustering of i 's peer group $\ln \frac{\sum_{p \in Peers(i)} \ Peers(p)\ }{\ \cup_{p \in Peers(i)} Peers(p)\ }$
	PeerPerf	Logarithm of the minimum productivity (issues/month) of the peers $\ln \min_{p \in Peers(i)} nmr_p + 1$
	LckAttn	The longest duration between the i 's ac- tion until the response is less than 1 hour

分析結果

Table IV
MODEL FOR GNOME (125,665 OBSERVATIONS)

	Est	Std.Err.	x	x_{alt}	$\frac{Odds(x_{alt})}{Odds(x)}$
(Intcpt)	-4.79	0.193			
nUsr	-1.95	0.0908	0.528	0.87	51/100
RS	-0.981	0.0588	-0.794	-0.468	73/100
GotFix	0.829	0.0354	F	T	229/100
withBB	-1.08	0.0556	T	F	295/100
nCmt	0.719	0.0314	ln 2	ln 4	165/100
nPeer	-0.0543	0.00673	ln 1779	ln 5405	94/100
pShared	2.00	0.182	ln 1.06	ln 1.17	122/100
LckAttn	-0.501	0.0778	F	T	61/100
PeerPerf	0.218	0.00496	ln 15	ln 318	195/100

精度77%

分析結果

Table V
MODEL FOR MOZILLA (130,471 OBSERVATIONS)

	Est	Std.Err.	x	x_{alt}	$\frac{Odds(x_{alt})}{Odds(x)}$
(Intcpt)	-7.49	0.419			
nUsr	-0.601	0.15	0.308	0.57	85/100
RS	0.701	0.293	0.0738	0.173	107/100
GotFix	0.74	0.0831	F	T	210/100
FNotRep	0.507	0.0821	F	T	166/100
nCmt	0.819	0.0409	ln 2	ln 5	212/100
nPeer	0.142	0.0205	ln 1650	ln 4266	114/100
pShared	2.35	0.135	ln 1.12	ln 1.32	148/100
LckAttn	-0.325	0.124	F	T	72/100
PeerPerf	0.0649	0.0131	ln 15	ln 249	120/100

精度81%

LTCになる確率の変化

	Gnome	Mozilla
レスポンスが遅い	-39%	-28%
同僚の労働力を上げる	+95%	+20%
連携を密にする	+22.15%	+48%
ユーザーの数が増える	-49%	-15%
同僚グループの規模が大きくなる	-6%	+14%

考察

- 能力や意欲, 環境をメトリクスとして将来 Long-term contributor(LTC)になる開発者を予測できる

まとめ

- 参加初期の時点で将来Long-term contributor(LTC)になる開発者を予測したい
- 能力や意欲, 環境をメトリクスとし, ロジスティック回帰モデルを作成

所感

- Long-term contributor(LTC)はコミッターの位置づけに似ている
- プロダクトメトリクスもコミッター予測の指標となりうるかも