

Open Borders? Immigration in Open Source Projects

Christian Bird, Alex Gourley, Prem Devanbu
Anand Swaminathan, Greta Hsu

60206052

岡坂亮汰

背景

- OSSプロジェクトはボランティアのチームによって構成
- その中には中心となる開発者チームが存在する
- 開発者に選ばれる要因とは？

目的

開発者になる可能性が**コミュニティの在任期間**によって
どのように変化するかを考察し、また**社会的地位**や**スキル**
などの要素の重要性を定量的に評価

- プロジェクトへの滞在期間
- スキルレベル
- 個人の評判

仮説

- **仮説 1** : 開発者の地位を獲得する可能性は、任期とともに上昇し、ある時点でピークに達し、その後減少する
- **仮説 2** : パッチの提出などのスキルレベルは開発者になる可能性を高める
- **仮説 3** : 社会的地位(個人の評判)は開発者の地位達成にプラスの影響を与える

データの抽出

- ・ **ソースコードリポジトリ情報**(誰がどのファイルをいつ変更したか)と**アーカイブ情報**(電子メールを誰が誰に返信したか)を電子メールで収集
- ・ 電子メールから**パッチ提出**を抽出

データの整形

Variable	Mean	Std. Dev	Min	Max
(Postgres) $n = 169,118$ $c = 3,283$ $i = 19$				
patch	.04	.20	0	1
norm_indegree	.0024	.0065	0	.18
success_pct	1.27	9.78	0	100
devs_cum	13.92	4.14	1	20
sent_cum	13.12	63.55	1	2370
(Apache) $n = 50,170$ $c = 1,445$ $i = 30$				
patch	.10	.30	0	1
indegree	.0026	.0049	0	.087
success_pct	1.13	8.98	0	100
sent_cum	7.34	18.18	1	273
devs_cum	46.24	9.76	3	57
(Python) $n = 45,216$ $c = 1,320$ $i = 62$				
patch	.04	.20	0	1
norm_indegree	.0098	.032	0	.70
success_pct	.62	6.98	0	100
devs_cum	58.69	14.21	0	79
sent_cum	13.87	54.44	1	1648

n : レコード数

c : 候補数

i : 開発者になった人数

分析

- **生存時間分析**を用いる
- **過去のパッチ受付履歴**や**ソーシャルネットワークの状況**などの予測変数が非開発者に与える影響を調べる

生存時間分析

- ・ 死亡率や雇用期間などの時間に依存する事象をモデル化できる分析手法
- ・ ハザード比を算出し、事象が起こる確率を求める

予測変数を用いて開発者になる確率を算出し、分析する

使用する予測変数

- **tp1~tp6** : 参加してから開発者になるまでの期間
- **indegree** : 個人に返信されたメッセージの合計数
- **patch** : 個人が以前にパッチを提出したかどうか
- **acceptance_rate** : パッチ提出が受け入れられた割合
- **sent_cum** : 個人が送信したメッセージの合計数
- **devs_cum** : コミュニティ内の開発者の総数
- **time_trend** : 個人の在籍期間

分析結果

1：開発者になるまでの期間

Variable	Apache Model			Postgres Model			Python Model		
	Coef.	z	P > Z	Coef.	z	P > Z	Coef.	z	P > Z
tp1 (0-6 months)	-6.69	-2.97	0.003	-20.63	-10.62	0.000	1.10	0.67	0.5
tp2 (6-12)	-6.09	-2.66	0.008	-5.20	-2.32	0.021	1.18	0.69	0.492
tp3 (12-24)	-6.48	-2.7	0.007	-6.51	-2.9	0.004	0.098	0.06	0.955
tp4 (24-36)	-7.98	-2.93	0.003	-5.77	-3.00	0.003	-0.23	-0.13	0.9
tp5 (36-48)	-6.79	-2.53	0.011	-5.98	-2.88	0.004	0.10	0.06	0.96
tp6 (48-60)	-8.16	-2.60	0.009	-8.14	-2.88	0.004	0.91	0.47	0.64
indegree*	61.21	2.78	0.005	72.2	5.46	0.000	5.49	2.92	0.004
patch*	2.71	5.09	0.000	0.98	1.42	0.154	2.09	5.95	0.000
acceptance_rate*	0.15	2.03	0.043	0.0039	0.45	0.656	0.0062	0.10	0.918
sent_cum*	0.014	4.20	0.000	.00158	1.16	0.246	0.0041	1.86	0.062
devs_cum	0.14	1.84	0.066	1.09	2.17	0.03	0.14	2.81	0.005
time_trend	-0.91	-1.90	0.057	-2.06	-1.93	0.054	-1.46	-3.19	0.001

分析結果

1：開発者になるまでの期間

ApacheとPostgreではハザード比が増加し、その後減少するという仮説を支持しているが、Pythonではそうはなっていない。

→ PythonはApacheとPostgreに比べて4年遅く登場したため、非開発者に対するコミュニティの反応は進化途中である。

開発者の地位を獲得する可能性は、任期とともに上昇し、ある時点でピークに達し、その後減少するという仮説1を支持

分析結果

2：パッチ提出

Variable	Apache Model			Postgres Model			Python Model		
	Coef.	z	P > Z	Coef.	z	P > Z	Coef.	z	P > Z
tp1 (0-6 months)	-6.69	-2.97	0.003	-20.63	-10.62	0.000	1.10	0.67	0.5
tp2 (6-12)	-6.09	-2.66	0.008	-5.20	-2.32	0.021	1.18	0.69	0.492
tp3 (12-24)	-6.48	-2.7	0.007	-6.51	-2.9	0.004	0.098	0.06	0.955
tp4 (24-36)	-7.98	-2.93	0.003	-5.77	-3.00	0.003	-0.23	-0.13	0.9
tp5 (36-48)	-6.79	-2.53	0.011	-5.98	-2.88	0.004	0.10	0.06	0.96
tp6 (48-60)	-8.16	-2.60	0.009	-8.14	-2.88	0.004	0.91	0.47	0.64
indegree*	61.21	2.78	0.005	72.2	5.46	0.000	5.49	2.92	0.004
patch*	2.71	5.09	0.000	0.98	1.42	0.154	2.09	5.95	0.000
acceptance_rate*	0.15	2.03	0.043	0.0039	0.45	0.656	0.0062	0.10	0.918
sent_cum*	0.014	4.20	0.000	.00158	1.16	0.246	0.0041	1.86	0.062
devs_cum	0.14	1.84	0.066	1.09	2.17	0.03	0.14	2.81	0.005
time_trend	-0.91	-1.90	0.057	-2.06	-1.93	0.054	-1.46	-3.19	0.001

分析結果

2：パッチ提出

ApacheとPythonでは非常に強い影響を与えるが、
Postgresにおいては影響は少ない。

→ Postgreでは非開発者が、ApacheやPythonよりも多くパッチ
を提出しており、開発者の地位を獲得する上で有利にはならない

**パッチの提出などのスキルレベルは開発者になる可能性を高める
という仮説2を支持**

分析結果

3：開発者の総数

Variable	Apache Model			Postgres Model			Python Model		
	Coef.	z	P > Z	Coef.	z	P > Z	Coef.	z	P > Z
tp1 (0-6 months)	-6.69	-2.97	0.003	-20.63	-10.62	0.000	1.10	0.67	0.5
tp2 (6-12)	-6.09	-2.66	0.008	-5.20	-2.32	0.021	1.18	0.69	0.492
tp3 (12-24)	-6.48	-2.7	0.007	-6.51	-2.9	0.004	0.098	0.06	0.955
tp4 (24-36)	-7.98	-2.93	0.003	-5.77	-3.00	0.003	-0.23	-0.13	0.9
tp5 (36-48)	-6.79	-2.53	0.011	-5.98	-2.88	0.004	0.10	0.06	0.96
tp6 (48-60)	-8.16	-2.60	0.009	-8.14	-2.88	0.004	0.91	0.47	0.64
indegree*	61.21	2.78	0.005	72.2	5.46	0.000	5.49	2.92	0.004
patch*	2.71	5.09	0.000	0.98	1.42	0.154	2.09	5.95	0.000
acceptance_rate*	0.15	2.03	0.043	0.0039	0.45	0.656	0.0062	0.10	0.918
sent_cum*	0.014	4.20	0.000	.00158	1.16	0.246	0.0041	1.86	0.062
devs_cum	0.14	1.84	0.066	1.09	2.17	0.03	0.14	2.81	0.005
time_trend	-0.91	-1.90	0.057	-2.06	-1.93	0.054	-1.46	-3.19	0.001

分析結果

4：個人の在籍期間

Variable	Apache Model			Postgres Model			Python Model		
	Coef.	z	P > Z	Coef.	z	P > Z	Coef.	z	P > Z
tp1 (0-6 months)	-6.69	-2.97	0.003	-20.63	-10.62	0.000	1.10	0.67	0.5
tp2 (6-12)	-6.09	-2.66	0.008	-5.20	-2.32	0.021	1.18	0.69	0.492
tp3 (12-24)	-6.48	-2.7	0.007	-6.51	-2.9	0.004	0.098	0.06	0.955
tp4 (24-36)	-7.98	-2.93	0.003	-5.77	-3.00	0.003	-0.23	-0.13	0.9
tp5 (36-48)	-6.79	-2.53	0.011	-5.98	-2.88	0.004	0.10	0.06	0.96
tp6 (48-60)	-8.16	-2.60	0.009	-8.14	-2.88	0.004	0.91	0.47	0.64
indegree*	61.21	2.78	0.005	72.2	5.46	0.000	5.49	2.92	0.004
patch*	2.71	5.09	0.000	0.98	1.42	0.154	2.09	5.95	0.000
acceptance_rate*	0.15	2.03	0.043	0.0039	0.45	0.656	0.0062	0.10	0.918
sent_cum*	0.014	4.20	0.000	.00158	1.16	0.246	0.0041	1.86	0.062
devs_cum	0.14	1.84	0.066	1.09	2.17	0.03	0.14	2.81	0.005
time_trend	-0.91	-1.90	0.057	-2.06	-1.93	0.054	-1.46	-3.19	0.001

分析結果

5：個人に送信されたメッセージの合計数

Variable	Apache Model			Postgres Model			Python Model		
	Coef.	z	P > Z	Coef.	z	P > Z	Coef.	z	P > Z
tp1 (0-6 months)	-6.69	-2.97	0.003	-20.63	-10.62	0.000	1.10	0.67	0.5
tp2 (6-12)	-6.09	-2.66	0.008	-5.20	-2.32	0.021	1.18	0.69	0.492
tp3 (12-24)	-6.48	-2.7	0.007	-6.51	-2.9	0.004	0.098	0.06	0.955
tp4 (24-36)	-7.98	-2.93	0.003	-5.77	-3.00	0.003	-0.23	-0.13	0.9
tp5 (36-48)	-6.79	-2.53	0.011	-5.98	-2.88	0.004	0.10	0.06	0.96
tp6 (48-60)	-8.16	-2.60	0.009	-8.14	-2.88	0.004	0.91	0.47	0.64
indegree*	61.21	2.78	0.005	72.2	5.46	0.000	5.49	2.92	0.004
patch*	2.71	5.09	0.000	0.98	1.42	0.154	2.09	5.95	0.000
acceptance_rate*	0.15	2.03	0.043	0.0039	0.45	0.656	0.0062	0.10	0.918
sent_cum*	0.014	4.20	0.000	.00158	1.16	0.246	0.0041	1.86	0.062
devs_cum	0.14	1.84	0.066	1.09	2.17	0.03	0.14	2.81	0.005
time_trend	-0.91	-1.90	0.057	-2.06	-1.93	0.054	-1.46	-3.19	0.001

分析結果

5：個人に送信されたメッセージの合計数

- コミュニティ内の個人に対する反応の広がり の尺度

**3つのプロジェクトで影響があるという結果から、
個人の評判は開発者になるために重要**

まとめ

- 開発者になる可能性は滞在期間によって異なる
- パッチ提出などの技術は開発者になる機会に影響を与える
- 個人の評判は開発者になることに影響する