

研究報告

山崎大輝

既存研究

- ▶ 先週輪講発表したもの
+
- ▶ GNOMEを対象に
OSSecosystemsにおける開発者の役割進化に影響を
与える要因の調査
- ▶ development core developers,
collaborative core developers,
project leader を予測

Developer Role Evolution in Open Source Software Ecosystem: An Explanatory Study on GNOME

Can Cheng, Bing Li, Zeng-Yang Li,
Yu-Qi Zhao, Feng-Ling Liao

Cheng C, Li B, Li ZY et al. Developer role evolution in open source software ecosystem: An explanatory study on GNOME. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 32(2): 396-414 Mar. 2017. DOI 10.1007/s11390-017-1728-9

やってること

- ▶ 開発者の役割進化について調査
 - ◆ どの要因が役割進化に対し効果的か？
 - ◆ 要因で予測モデルを作るとどうなるか？

RQ1~3

- ▶ Do **subjective willingness** and **project environment** indicators affect that co-developers evolve into developmental core developers(collaborative core developers,project leaders) in the GNOME ecosystem?
- ▶ 要は 色々な開発者になる際に意欲やプロジェクトの環境がどのような影響を与えているか？

subjective willingness

Table 1. Indicators of Subjective Willingness

Indicator	Calculation Method
TCI (total commit intensity)	Let x = total commit intensity of a developer in his/her first year in the OSS ecosystem, then $TCI = \ln(x + 1)$
TFI (total fix intensity)	Let x = total fix intensity of a developer in his/her first year in the OSS ecosystem, then $TFI = \ln(x + 1)$
TNP (total number of projects a developer joined)	Let x = total number of projects that a developer joined in his/her first year in the OSS ecosystem, then $TNP = \ln(x + 1)$
MFI (maximum fix intensity in a single project)	Let x = maximum fix intensity of a developer in a single project in his/her first year in the OSS ecosystem, then $MFI = \ln(x + 1)$

Project Environment

Table 2. Indicators of Project Environment

Indicator	Calculation Method
DMP (duration of the main project)	Let x = duration of the main project (from the set-up time of the project) of the developer, then $DMP = \ln(x + 1)$
AFIP (average fix intensity of peers)	Let x = average fix intensity of peers in the main project of a developer in his/her first year in the OSS ecosystem, then $AFIP = \ln(x + 1)$
MFIP (minimum fix intensity of peers)	Let x = minimum development intensity of peers in the main project of a developer in his/her first year in the OSS ecosystem, then $MFIP = \ln(x + 1)$
NP (number of peers)	NP = number of developers who collaborated in the main project of a developer in the first year after joining the ecosystem

Project Environment

NM (number of members)

Number of developers who worked in the main project of a developer before the developer joins the ecosystem

SCCP
(social clustering coefficient of peers)

SCCP = social clustering coefficient of peers in the main project of a developer (this indicator is calculated in the cooperation network of peers right before the developer joins the main project, and the network is constructed by the peers' cooperative connections)

TIC (total intensity of commits of the main project)

Let x = total intensity of commits of peers in a developer's main project in his/her first year in the OSS ecosystem, then
$$\text{TIC} = \ln(x + 1)$$

Control Variables

Table 3. Control Variables

Variable	Calculation Method
ADC (average degree centrality of peers)	Let x_i = degree centrality of peers of developer i before the developer joins the ecosystem, n = number of peers, and $y_i = \frac{\sum_{i=1}^n x_i}{n}$, then $ADC = \ln(y_i + 1)$
ACC (average closeness centrality of peers)	Let x_i = closeness centrality of peers of developer i before the developer joins the ecosystem, and n = number of peers, then $ACC = \frac{\sum_{i=1}^n x_i}{n}$
ABC (average betweenness centrality of peers)	Let x_i = betweenness centrality of peers of developer i before the developer joins the ecosystem, n = number of peers, and $y_i = \frac{\sum_{i=1}^n x_i}{n}$, then $ABC = \ln(y_i + 1)$

VIF

- ▶ 独立変数間の多重共線性を検出するための指標の1つ。

独立変数間の相関係数行列の逆行列の対角要素であり、値が大きい場合はその変数を分析から除いた方がよいと考えられる。

10を基準とすることが多い。(今回は5)

- ▶ 例えば[今月の降水量]と[今月の雨の日数]はVIFが高くなる

Table 4. Remaining Indicators after
Multicollinearity Examination

Indicator	VIF
NP	1.993
SCCP	1.602
ACC	1.839
TNP	1.420
MFI	1.759
MFIP	1.905
DMP	1.973
AFIP	1.480
ABC	1.843

RQ1

- TNP,MFI,SCCP,AFIP affect

Developmental Core Developers

	<i>B</i>	S.E.	Walds	Sig.	Exp(<i>B</i>)
TNP	0.859	0.155	30.819	<0.001***	2.361
MFI	0.899	0.073	152.253	<0.001***	2.458
ABC	0.046	0.044	1.075	0.300	1.047
NP	−0.015	0.008	3.685	0.055	0.985
SCCP	−0.849	0.396	4.596	0.032**	0.428
ACC	−0.532	0.621	0.735	0.391	0.587
MFIP	0.161	0.088	3.358	0.067	1.175
DMP	0.005	0.049	0.009	0.923	1.005
AFIP	−0.210	0.087	5.878	0.015**	0.811
Constant	−3.857	0.489	62.102	0.000	0.021

RQ2

- ▶ TNP,MFI,ABC affect

for Collaborative Core Developers

	<i>B</i>	S.E.	Walds	Sig.	Exp(<i>B</i>)
TNP	0.811	0.124	43.073	<0.001***	2.250
MFI	0.140	0.045	9.825	0.002***	1.151
ABC	0.134	0.041	10.497	<0.001***	1.144
NP	0.002	0.005	0.090	0.764	1.002
SCCP	−0.134	0.310	0.188	0.665	0.874
ACC	−0.307	0.583	0.278	0.598	0.735
MFIP	−0.118	0.068	3.018	0.082	0.889
DMP	0.056	0.040	1.915	0.166	1.058
AFIP	−0.037	0.053	0.495	0.482	0.964
Constant	−2.500	0.412	36.865	0.000	0.082

RQ3

- ▶ TNP,MFI,SCCP affect

for Project Leaders

	<i>B</i>	S.E.	Walds	Sig.	Exp(<i>B</i>)
TNP	1.035	0.165	39.227	<0.001***	2.815
MFI	0.371	0.075	24.713	<0.001***	1.449
ABC	0.088	0.081	1.162	0.281	1.092
NP	−0.007	0.009	0.629	0.428	0.993
SCCP	−0.942	0.492	3.674	0.050**	0.390
ACC	−1.784	1.053	2.871	0.090	0.168
MFIP	−0.048	0.102	0.218	0.641	0.953
DMP	−0.103	0.063	2.665	0.103	0.902
AFIP	0.006	0.105	0.003	0.953	1.006
Constant	−3.311	0.761	18.927	0.000	0.036

予測モデル

Table 12. Results Using All Indicators

	Precision	Recall	<i>F</i> -Value
Developmental core developers	0.804	0.739	0.770
Collaborative core developers	0.653	0.405	0.500
Project leaders	0.613	0.259	0.365

significant indecatorsで10fold cross validation

Table 13. Effect of Significant Indicators

	Precision	Recall	AUC	<i>F</i> -Value	Indicator
Developmental core developers	0.807	0.727	0.896	0.765	TNP+MFI+SCCP+AFIP
Collaborative core developers	0.654	0.399	0.685	0.496	MFI+TNP+ABC

Note: AUC (area under receiver operating characteristic (ROC) curve) is used to evaluate the performance of a binary classifier. If the AUC value is close to 0.5, it means that the classification model is close to a random prediction.

TODO

- ▶ 既存研究・参考文献探しと整理
- ▶ 既存研究のデータセットで実験
 - ◆ GNOME DB , Apache DB and script
 - Apacheのデータセット整形のためコードリレーディング中
 - (GNOMEの方も実験してみたい)