

Who is going to Mentor Newcomers in Open Source Projects?



Gerardo Canfora



Massimiliano Di
Penta



Rocco Oliveto



Sebastiano
Panichella



UNIVERSITÀ
DEGLI STUDI
DEL MOLISE

@FSE2012

背景

- プロジェクトに新しく参加したnewcomerは技術スキルおよび組織管理の理解が必要
- newcomerへの指導の必要性は高まっている
[1]

[1] Barthélémy Dagenais, Harold Ossher, Rachel K. E. Bellamy, Martin P. Robillard, and Jacqueline P. de Vries. Moving into a new software project landscape. ICSE'10

目的・アプローチ

- 目的
 - newcomerが適切な指導を受けられるようにする
- アプローチ
 - mentor候補者を特定し, newcomerに適切なmentorを推薦する手法を構築する

優れたmentorの特徴

- newcomerが関心を持つトピックに関する専門知識が豊富である
- 他の開発者をサポートできる十分な能力がある

優れたmentorの特徴

- newcomerが関心を持つトピックに関する専門知識が豊富である

SVN

CVS

GIT



- 他の開発者をサポートできる十分な能力がある



提案手法

- YODA(Young and newcOmer Developer Assistant)
 - OSSプロジェクトにおけるmentorを定義するためのアプローチ

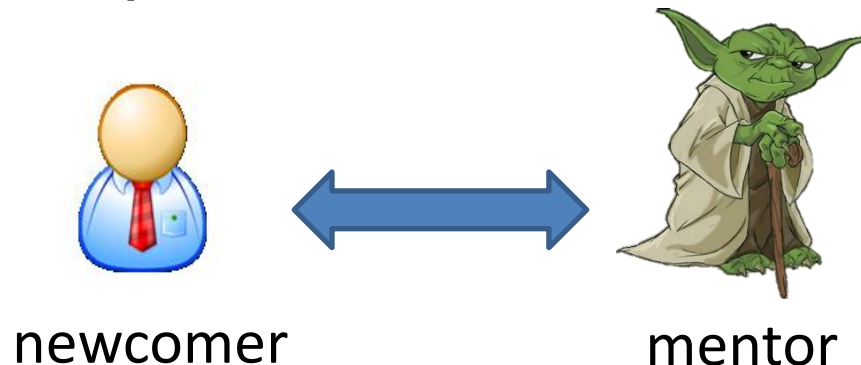


YODA

1. 過去の開発履歴からmentorを特定



2. mentorを推薦

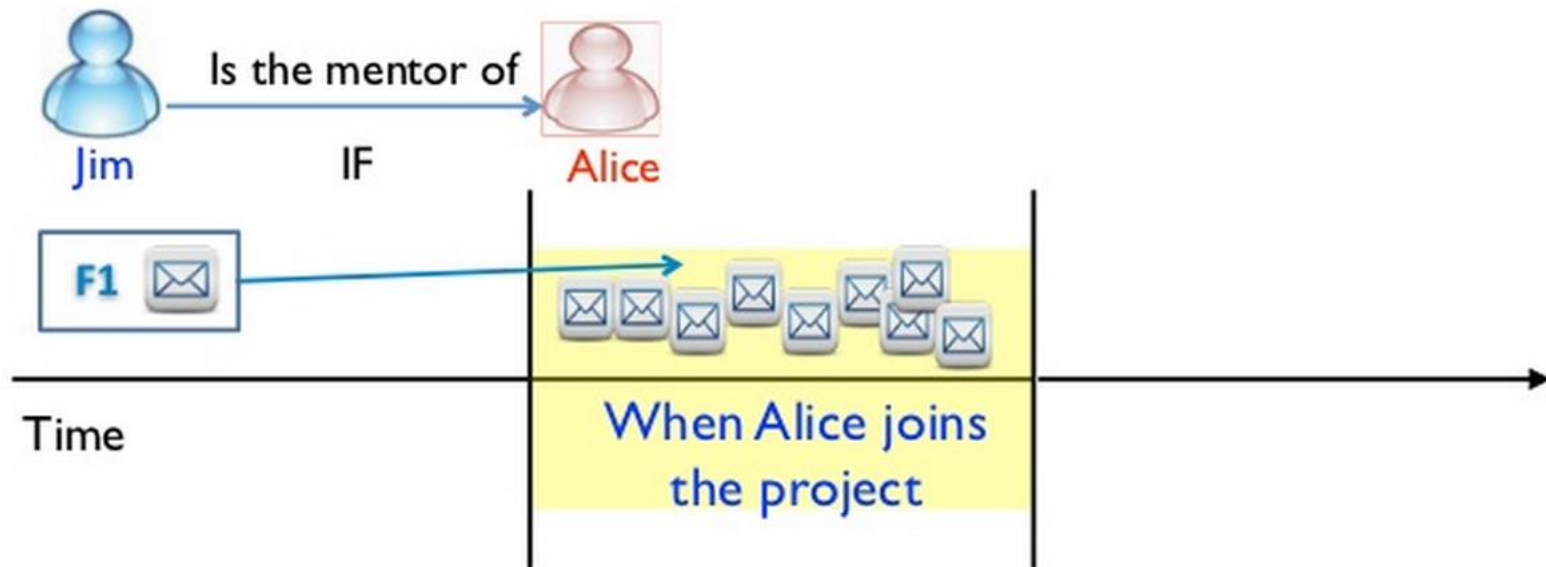


mentor適性

1. newcomerの参加初期にnewcomerと多くのメールのやりとりをしている
2. newcomerよりメールの送信数が多い
3. newcomerより経験が長い
4. newcomerが参加してすぐメールのやりとりが始まっている
5. コミットを活発におこなっている

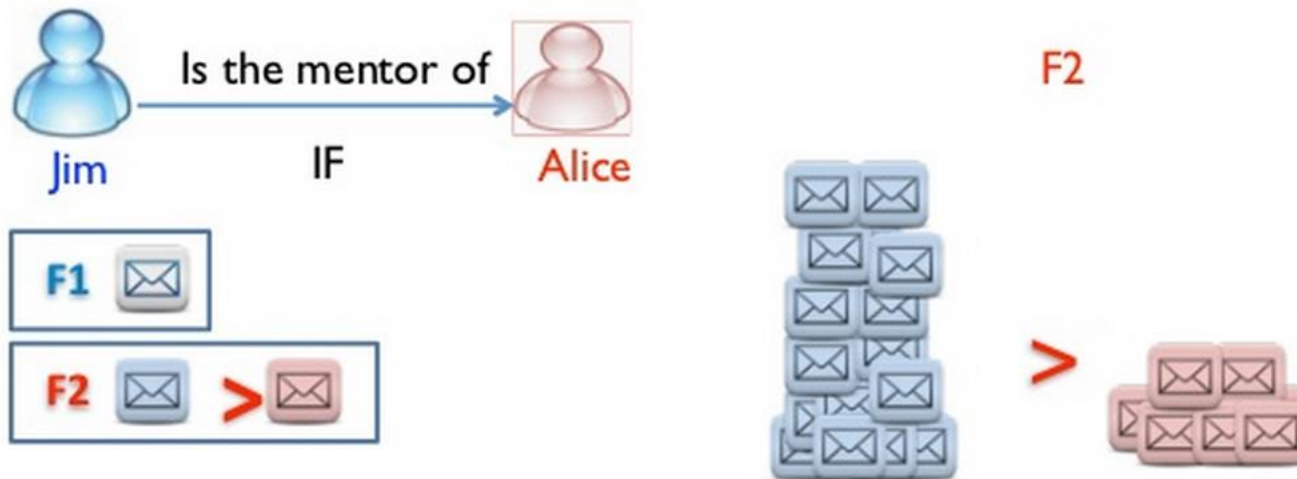
mentor指標

- F1: newcomerの参加初期における,
newcomerとのメールの割合



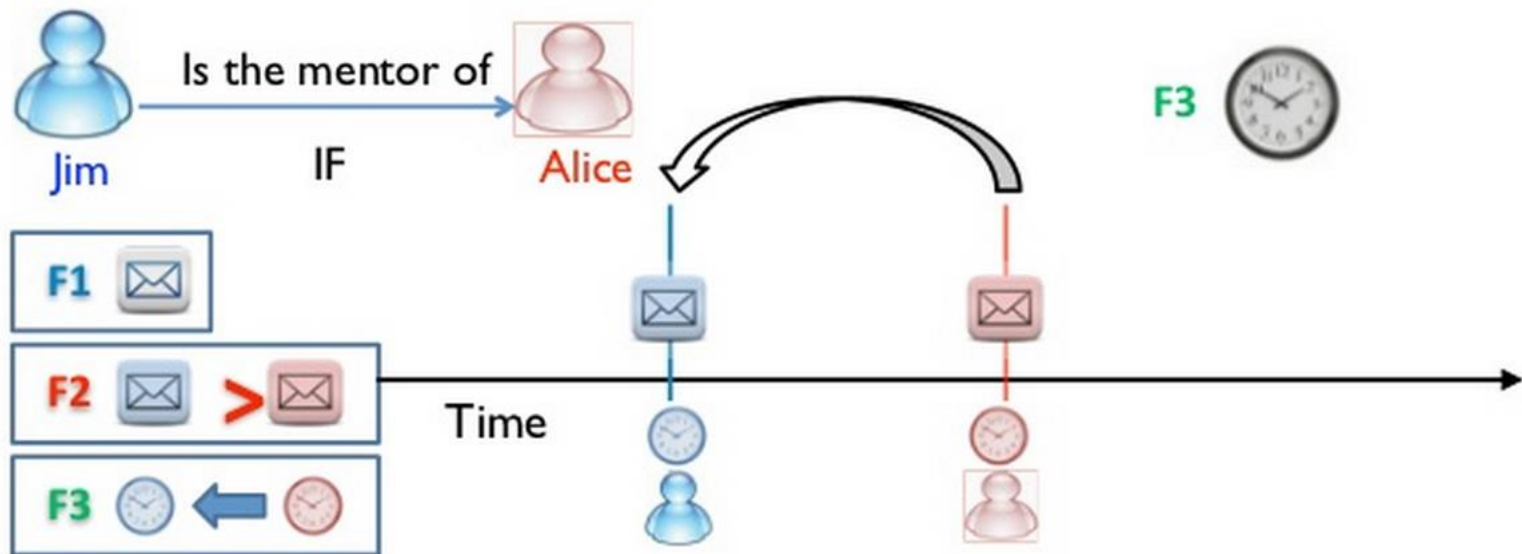
mentor指標

- F2: newcomerとのメール送信数の差



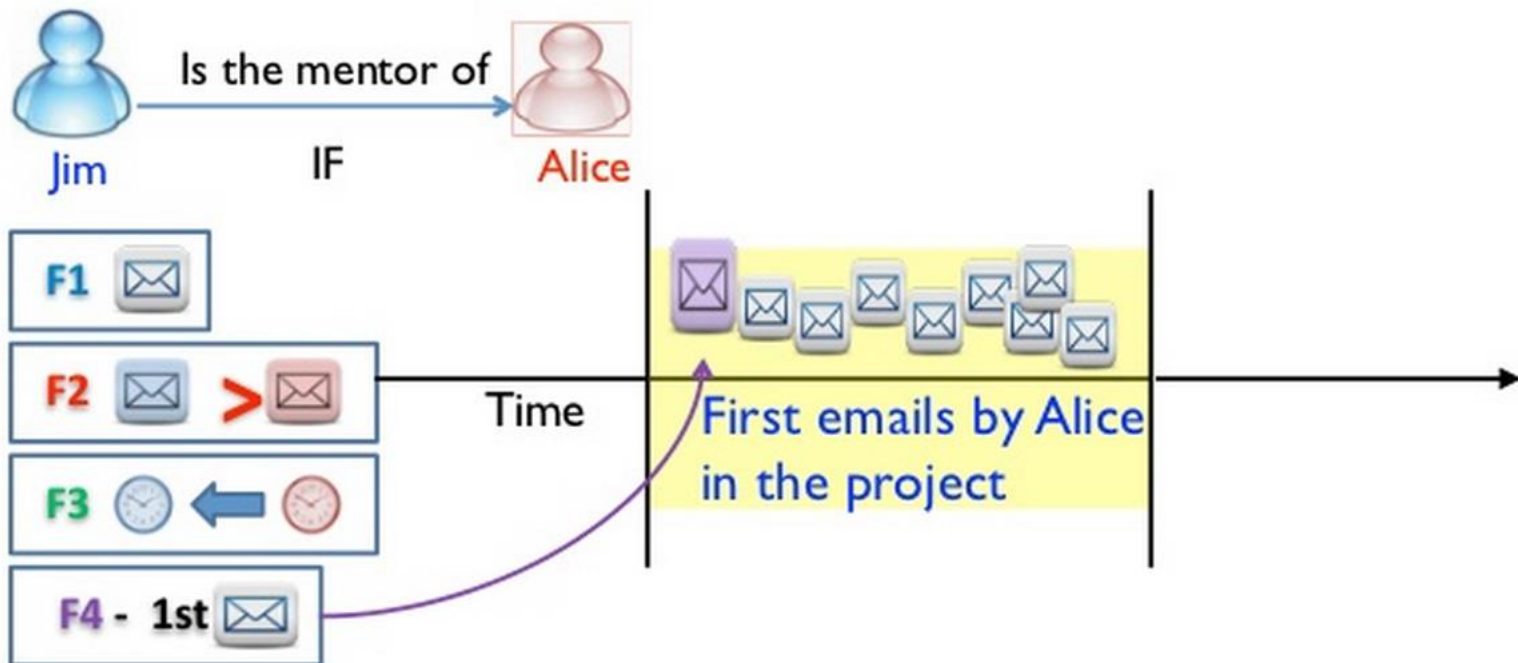
mentor指標

- F3: newcomerとの初メール送信日の差



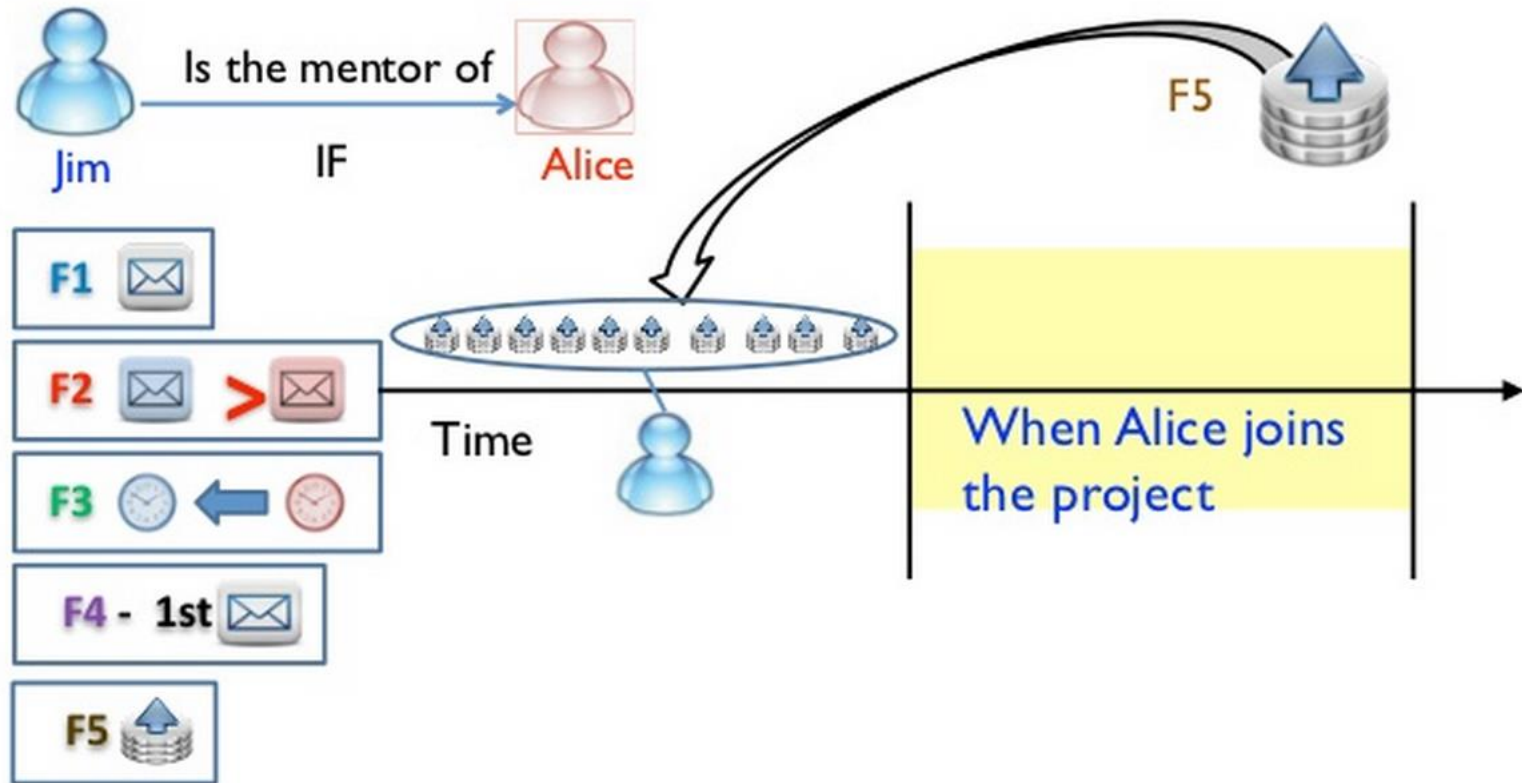
mentor指標

- F4: newcomerの初メール送信日からmentorとメールをするまでの期間



mentor指標

- F5: newcomer参加以前のコミットの数



スコアリング

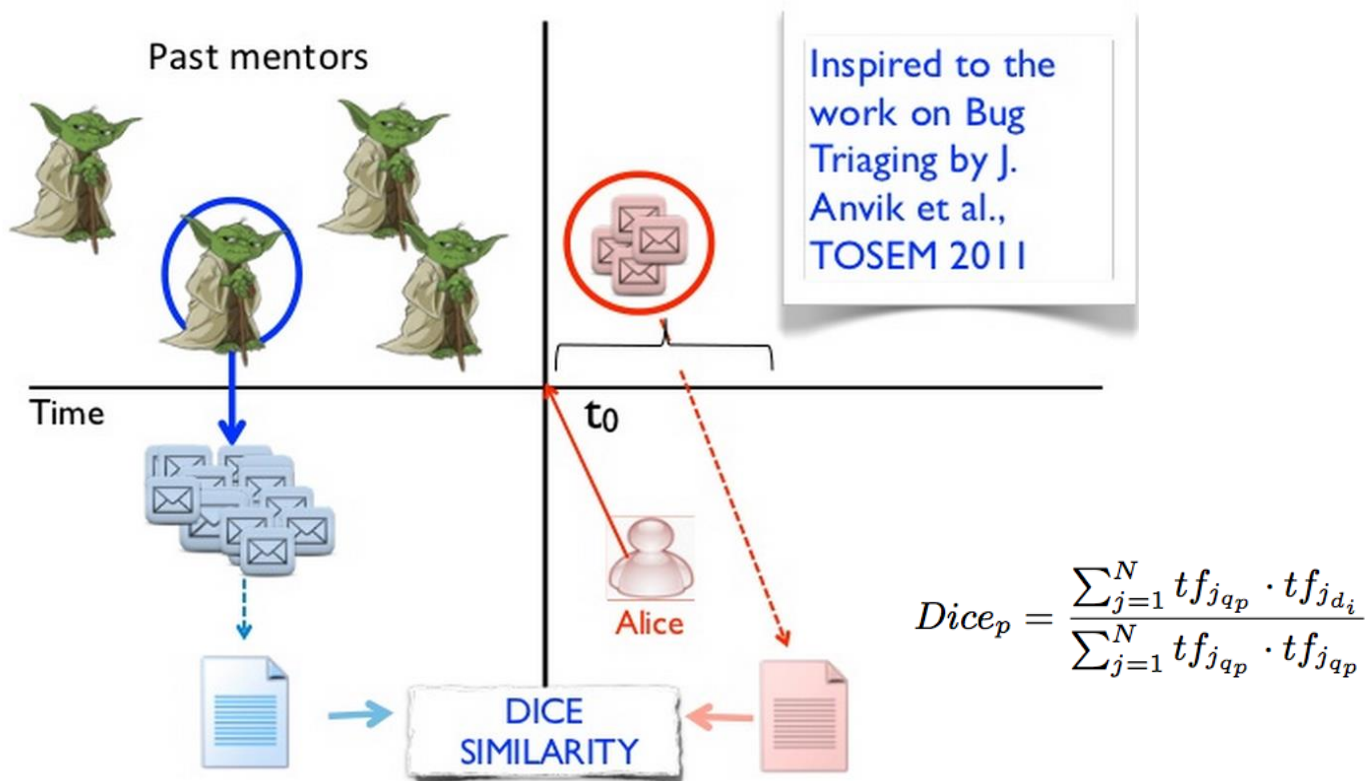
- mentorとしてのスコアを出してランク付け

$$\sum_{i=1}^5 w_i f_i$$

mentor -newcomer	score
Jim-Alice	2.5
Jim-Tom	2.5
David-Michael	2.0

mentorの推薦

- バグトリージと同様の手法で適当なmentorを決定



実験

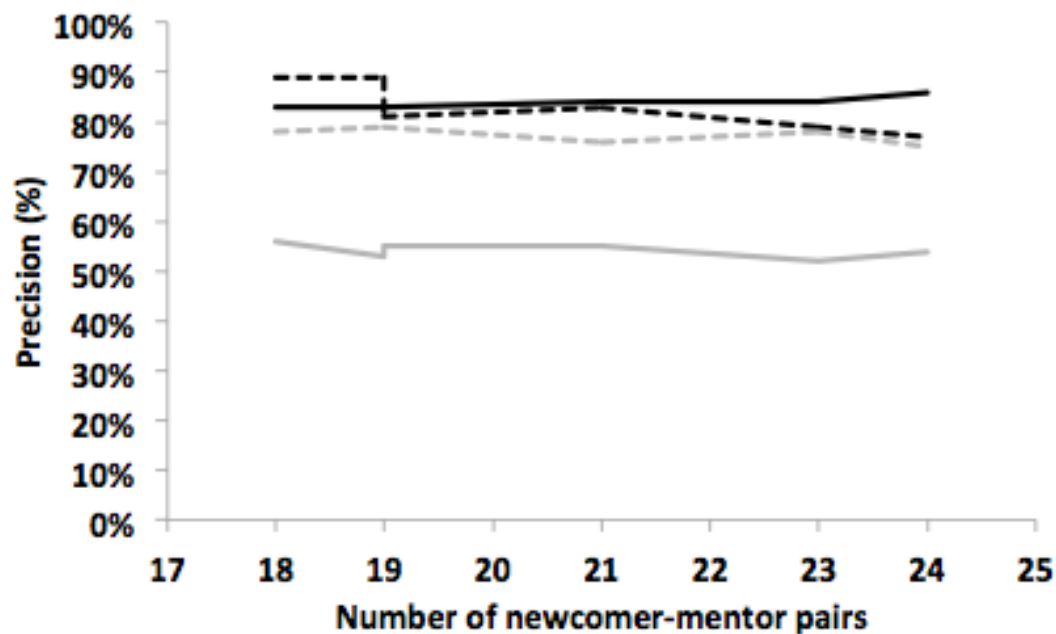
- データセット

	Apache	FreeBSD	PostgreSQL	Python	Samba
Period (Training set)	08/2001-03/2002	11/1998-02/2000	10/1998-05/2001	05/2000-05/2001	04/1998-09/2000
Period (Test set)	04/2002-12/2008	03/2000-10/2008	06/2001-03/2008	06/2001-12/2008	10/2000-12/2008
# of Mentors (Training set)	19	65	10	28	17
# of Newcomers (Training set)	13	33	8	32	33
# of Newcomers (Test set)	13	33	7	31	33

リサーチクエストション

- RQ1
 - どれほどmentorを特定できるか？
- RQ2
 - mentorの推薦がどの程度可能であるのか？

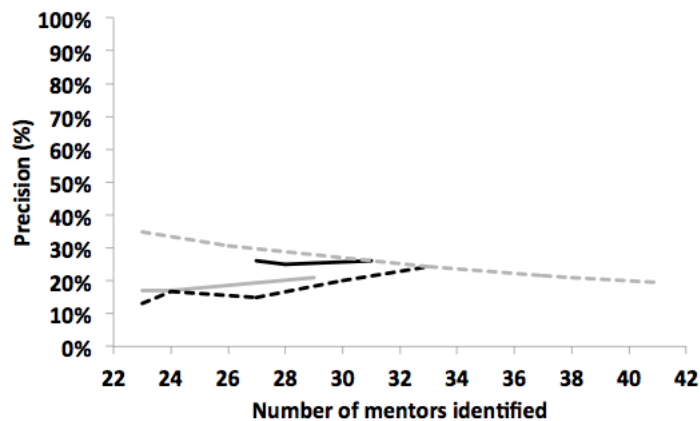
結果



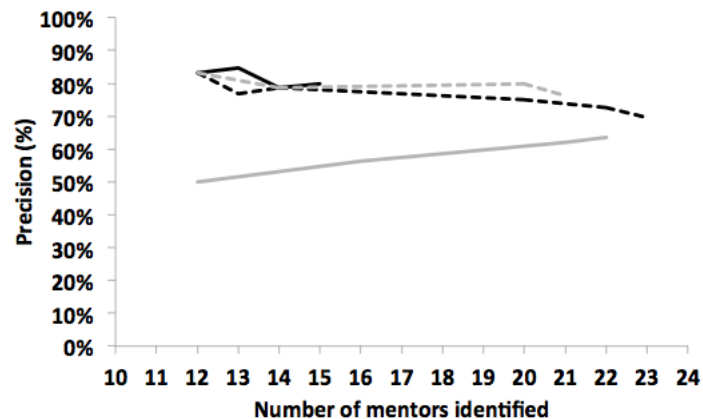
(a) Apache httpd

—f1 —f5 (baseline) - - $0.5*f1 + 0.25*f2 + 0.25*f3$ - - $0.5*f1 + 0.25*f2 + 0.25*f4$

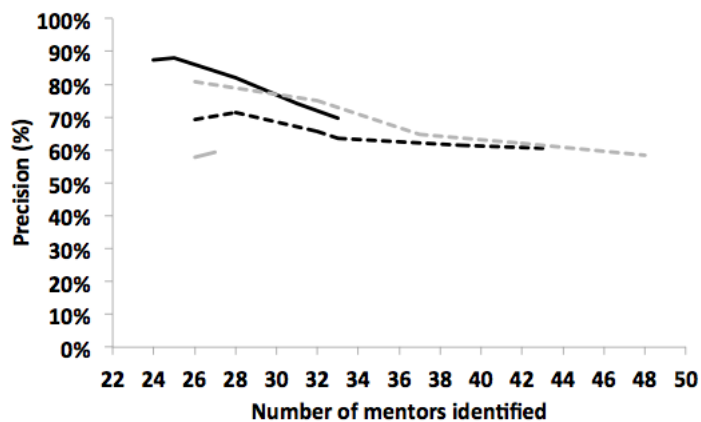
結果



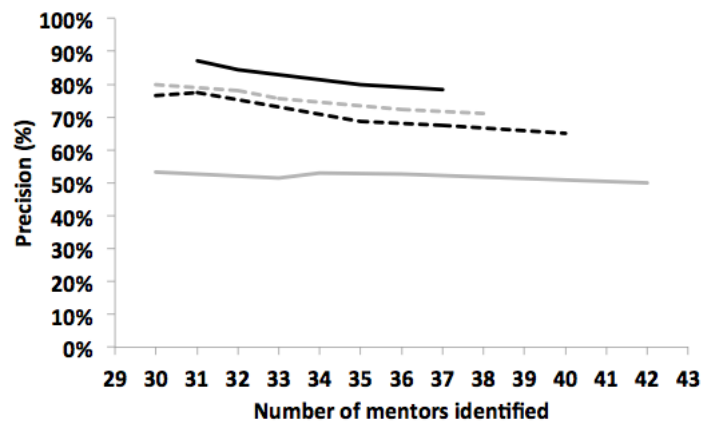
(b) FreeBSD



(c) PostgreSQL



(d) Python



(e) Samba

—f1 —f5 (baseline) - - $0.5*f1+0.25*f2+0.25*f3$ - - $0.5*f1+0.25*f2+0.25*f4$

結果

Table 2: Precision (%) and number of newcomer-mentor pairs identified for different values of λ .

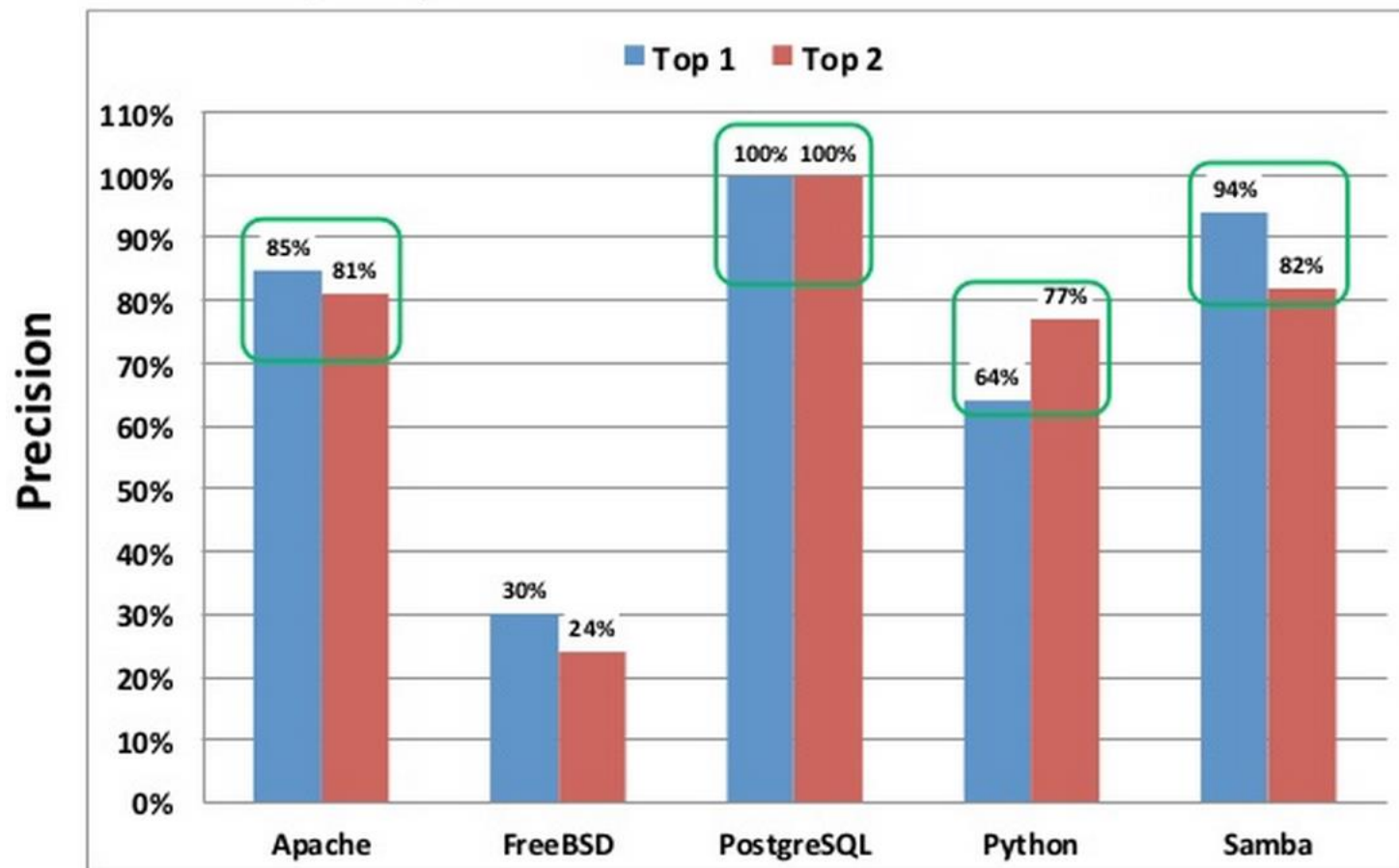
System	Formula	$\lambda = 100$		$\lambda = 90$		$\lambda = 80$		$\lambda = 70$		$\lambda = 60$		$\lambda = 50$	
Apache httpd	f_1	83%	18	83%	18	83%	18	84%	19	84%	19	86%	21
	$f_5(baseline)$	56%	18	53%	19	55%	20	55%	20	52%	21	54%	26
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_3$	83%	18	85%	20	82%	22	83%	24	80%	25	77%	26
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_4$	78%	18	75%	20	77%	22	77%	22	77%	22	78%	23
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_3$	89%	18	89%	18	81%	21	83%	23	79%	24	77%	26
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_4$	78%	18	79%	19	79%	19	76%	21	78%	23	75%	24
FreeBSD	f_1	26%	27	26%	27	26%	27	26%	27	25%	28	26%	31
	$f_5(baseline)$	17%	23	17%	23	17%	23	17%	24	17%	24	21%	29
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_3$	17%	23	17%	24	17%	24	17%	24	19%	26	23%	31
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_4$	30%	23	25%	28	20%	35	20%	35	21%	39	20%	40
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_3$	13%	23	13%	23	17%	24	15%	27	20%	30	24%	33
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_4$	35%	23	31%	26	24%	33	22%	37	21%	38	20%	41
PostgreSQL	f_1	83%	12	85%	13	79%	14	79%	14	80%	15	80%	15
	$f_5(baseline)$	50%	12	50%	12	50%	12	56%	16	62%	21	64%	22
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_3$	83%	12	76%	17	75%	20	65%	23	62%	26	55%	29
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_4$	83%	12	77%	13	73%	15	76%	21	73%	22	70%	23
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_3$	83%	12	77%	13	79%	14	75%	20	73%	22	70%	23
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_4$	83%	12	79%	14	79%	14	79%	14	80%	20	76%	21
Python	f_1	88%	24	88%	24	88%	25	82%	28	74%	31	70%	33
	$f_5(baseline)$	58%	26	58%	26	59%	27	59%	27	59%	27	59%	27
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_3$	62%	26	68%	31	66%	32	64%	36	63%	40	61%	41
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_4$	73%	26	75%	28	74%	34	63%	40	60%	45	57%	47
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_3$	69%	26	71%	28	66%	32	64%	33	62%	39	60%	43
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_4$	81%	26	81%	26	75%	32	65%	37	62%	42	58%	48
Samba	f_1	87%	31	84%	32	84%	32	80%	35	80%	35	78%	37
	$f_5(baseline)$	53%	30	52%	33	52%	33	53%	34	53%	36	50%	42
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_3$	77%	30	72%	32	69%	35	69%	36	66%	38	64%	42
	$0.33 \cdot f_1 + 0.33 \cdot f_2 + 0.33 \cdot f_4$	73%	30	70%	33	68%	34	67%	36	63%	40	63%	41
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_3$	77%	30	77%	31	69%	35	69%	35	68%	37	65%	40
	$0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_4$	80%	30	78%	32	78%	32	76%	33	72%	36	71%	38

結果

- RQ1
 - どれほどmentorを特定できるか？
- これらの組み合わせなら70%以上のprecisionで特定可能である

$$\begin{aligned} &0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_3 \\ &0.5 \cdot f_1 + 0.25 \cdot f_2 + 0.25 \cdot f_4 \end{aligned}$$

結果



結果

- RQ2
 - mentorの推薦がどの程度可能であるのか？
- FreeBSDを除いて, Top1なら64%以上, Top2なら77%以上のprecisionで正しく推薦ができている

アンケート

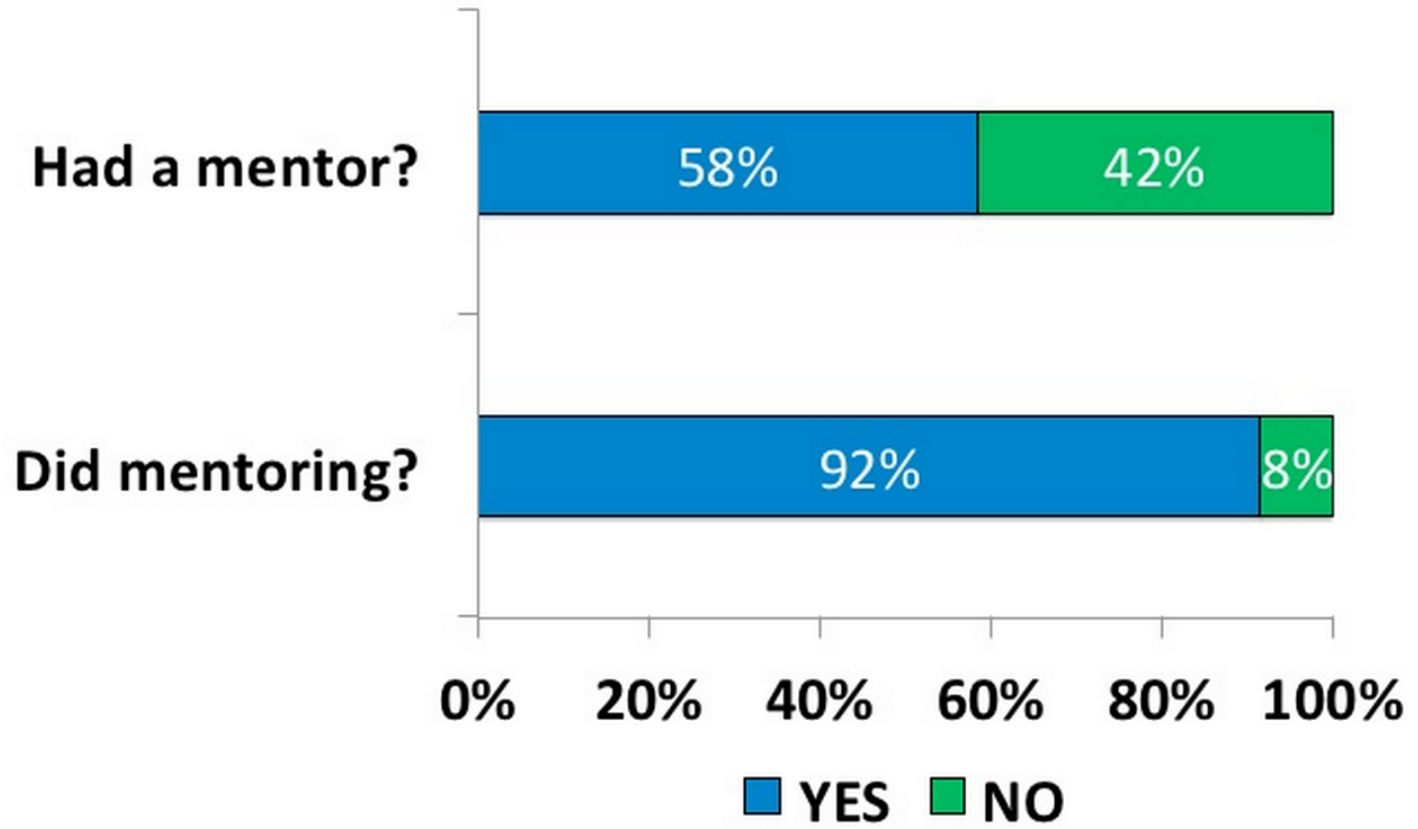
- mentor-newcomer関係の開発者にアンケート

Samba	 37
FreeBSD	 37
Postgre-SQL	 15
Python	 23
Apache	 23

Samba	
FreeBSD	
Postgre-SQL	
Python	
Apache	-

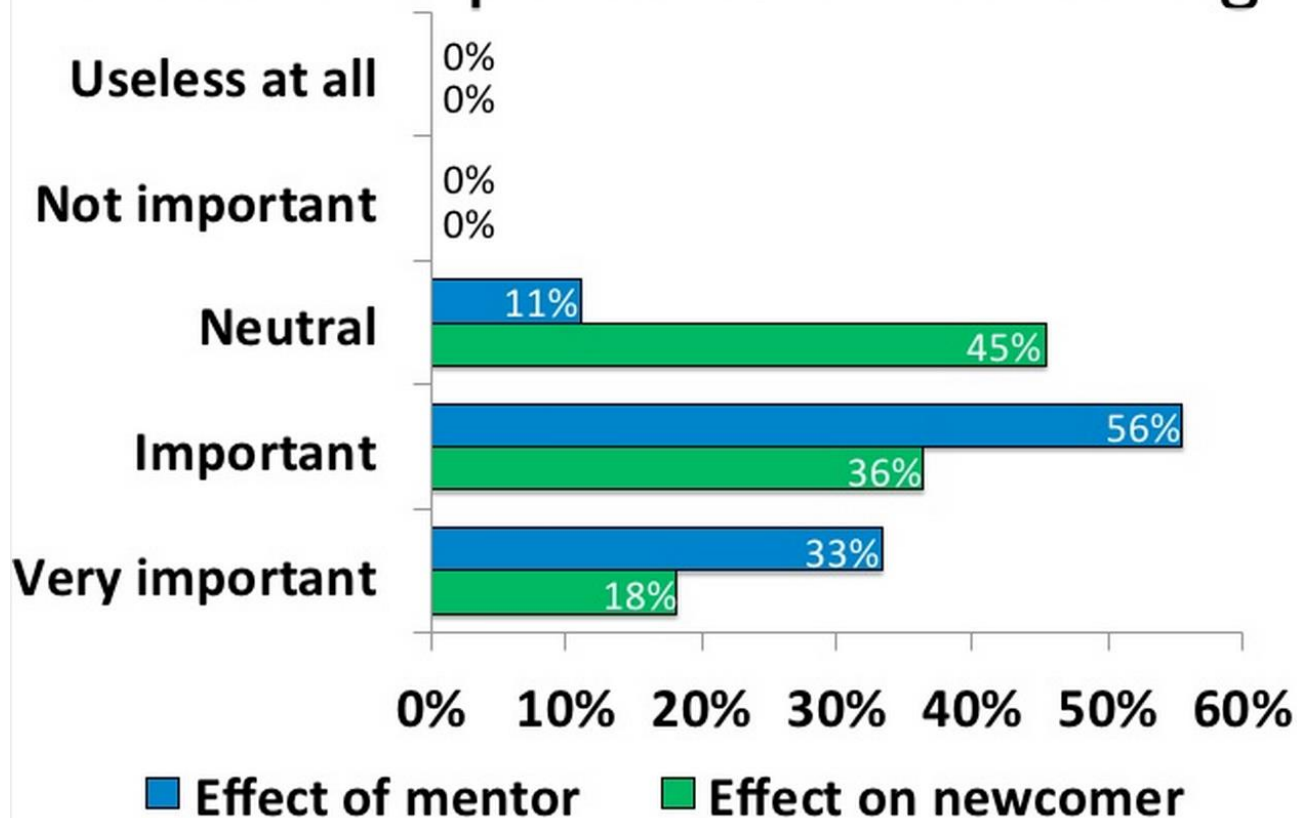
アンケート

Done/received mentoring?



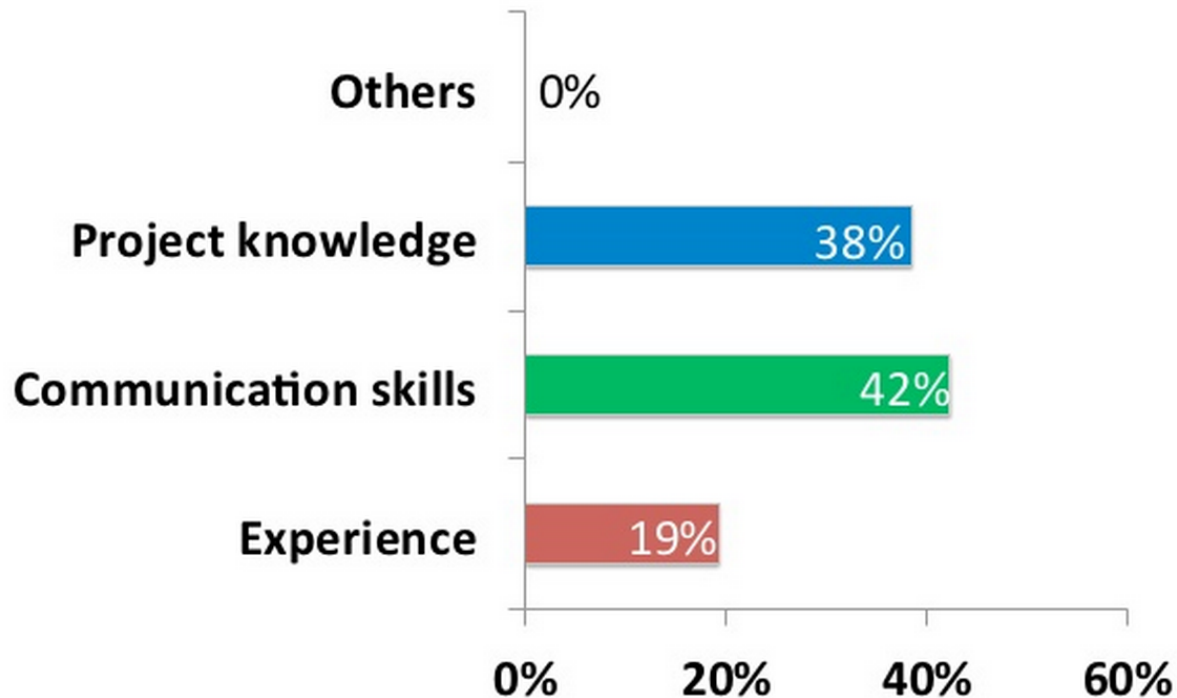
アンケート

Perceived importance of mentoring



アンケート

What makes a good mentor



まとめ

- OSSプロジェクトにおけるmentorを定義するためのアプローチYODAを提案した
- FreeBSDを除いて, Top1なら64%以上, Top2なら77%のprecisionで正しく推薦ができた

所感

- mentorの役割って結構大事かも...
 - Good Committer
 - Long Term Contributor
 - Project Evolution
- eclipse pluginとして実装したYODAのデモ動画 (ICSE2013)
 - <https://www.youtube.com/watch?v=4yrbYT-LAXA>