Supplementary Materials for FCFLA

Anonymous authors

November 29, 2023

This is the online supplement for the paper submitted to ISSTA-2024.

S-1 Parameter settings

The parameter settings for FCFLA are shown in Table S-1. The α is set to 0.5 to balance the weight of the conditional and posterior probabilities. The threshold ρ is set to 0 in our experiments, and it is an observed value that may need to be adjusted if FCFLA is used in other data sets. The υ is a tiny constant that is set to 0.001.

Table S-1: Parameter settings of FCFLA in our experiments

Parameter	Value
α	0.5
ρ	0
υ	0.001

S-2 Ablation and sensitivity analysis

To validate the effectiveness of the proposed block-level-based strategies, we generated several different variants of FCFLA for performance comparison on six systems. These variants are, without Evidence 1 (`WE1`), without Evidence 2 (`WE2`), without Evidence 3 (`WE3`), with Evidence 1 alone (`WF_E1`), with Evidence 2 alone (`WF_E2`), with Evidence 3 alone (`WF_E3`), and `WFR` (i.e., using only the procedural spectrum ranking). In fact, `WF_E1`, `WF_E2`, and `WF_E3` represent without fusion module of D-S evidence theory.

The result of comparing with the above variants on $RANK$ and $EXAM$ is presented in Fig. S-1. It can be seen that each part is helpful for FCFLA. For example, if Evidence 1 is absent, both $RANK$ and $EXAM$ for `WE1` significantly increase compared to FCFLA. Similarly, without Evidence 3, the fault localization performance of FCFLA is more affected, which demonstrates the effectiveness of utilizing the relevance of each block to the test result set for detecting faults in the block. For the fusion part, no matter which evidence is used alone to detect suspicious blocks, it is worse, and this also proves that the three pieces of evidence we used can be well fused based on the D-S evidence theory. It is worth noting that Evidence 2 has less impact on the performance improvement of FCFLA compared to Evidence 1 and Evidence 3. We analyzed the dataset and found that when

the coverage of the test set is low, the information from unit tests plays a limited role in suspicious block detection. When the coverage of the test set is high, Evidence 2 can significantly improve suspicious block detection. Notably, FCFLA is not dependent on just one part. For example, the $RANK$ of `WFR` is better in `Ochiai` than `WE1`, but the $EXAM$ of `WFR` is worse in `Dstar`. When all the parts contribute collectively, i.e., FCFLA, overall excellence can be achieved.

In addition, there is a critical variable in FCFLA, i.e., α. To determine the effect of different α on the fault localization performance, we set α with 0 to 1 to observe the changes in $RANK$ and $EXAM$, respectively. The corresponding results are shown in Fig. S-2. From Fig. S-2, when $\alpha \geq 0.5$, i.e., a greater weight for the posterior probability, FCFLA can achieve better performance. However, FCFLA reaches optimal performance only when $\alpha = 0.5$. This also proves that it is necessary to consider the conditional and posterior probabilities jointly.

In general, the proposed strategies are effective, especially the fusion of multi-source information, which can enable block-level-based fault location of SPLs.

S-3 Setting the validity of υ

In FCFLA, the purpose of setting υ is simply to make a difference in the scores of the suspicious statements being executed. In this research question, a variant, which sets υ=0, is generated to verify its validity. In addition, we also set υ=-0.001 to compare the performance of υ as 0.001. υ=0 means that all executed statements in a block have the same suspicion score, while υ=-0.001 indicates that an earlier executed statement is more likely to be ranked at the top.

The average $RANK$ and $EXAM$ on 338 cases are compared with the original FCFLA (υ=0.001) on `Barinel`, `Dstar`, `Ochiai`, `Op2`, and `Tarantula`, respectively. The result is shown as Fig. S-3 When α is set to 0, i.e., the suspicious scores of all statements are the same in the block, the performance of FCFLA decreases in all ranking metrics. And when υ=-0.001, FCFLA's performance is still improved compared to υ=0, which means that it is meaningful to keep the suspicious scores between statements in a block slightly different. In addition, υ=0.001 is better than υ=-0.001 in performance, especially in the $EXAM$ metric.

Overall, it is effective to increment or decrease the suspicious values of statements in a block in a sequential manner. Incrementing is more effective in improving FCFLA than decrementing.

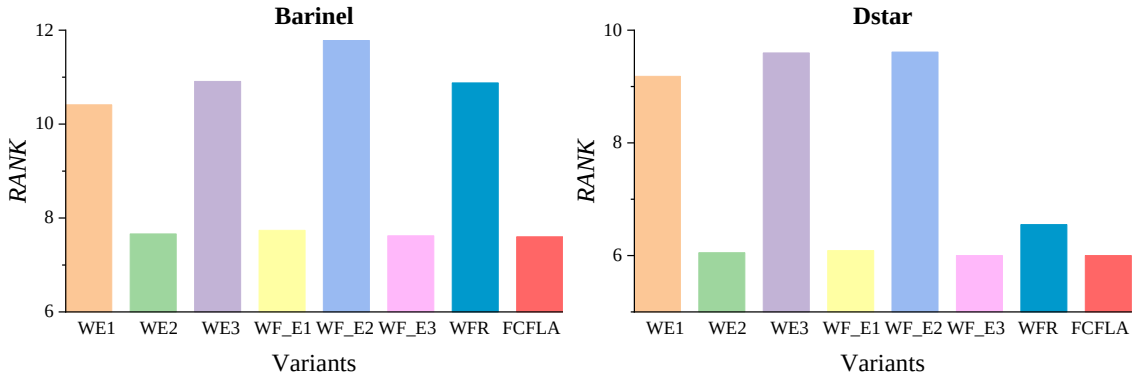
S-4 Supplementary results

In this section, we provide some supplementary results in the form of either tables or figures.

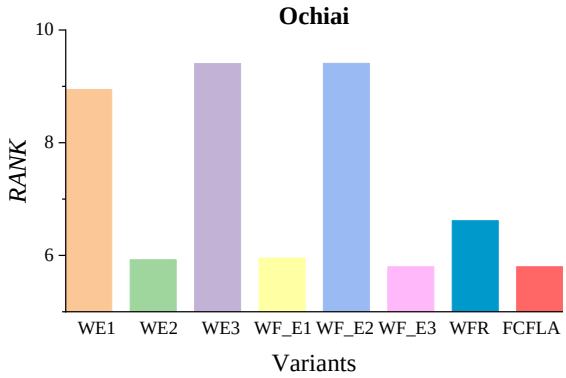
Table S-2 shows the comparison of *Rank* and $EXAM$ of FCFLA with VarCop, S-SBFL, SBFL, and FB on 30 ranking metrics.

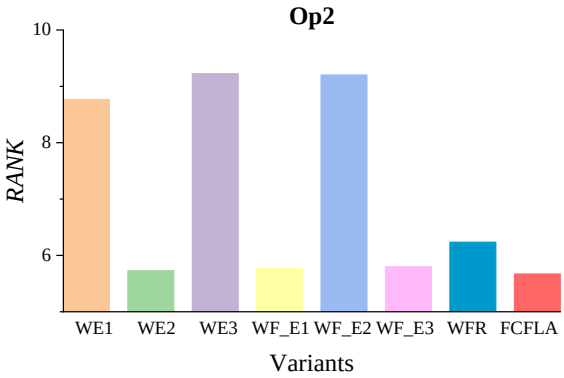
Table S-2: Comparison of *Rank* and *EXAM* of four baselines and our method on test datasets

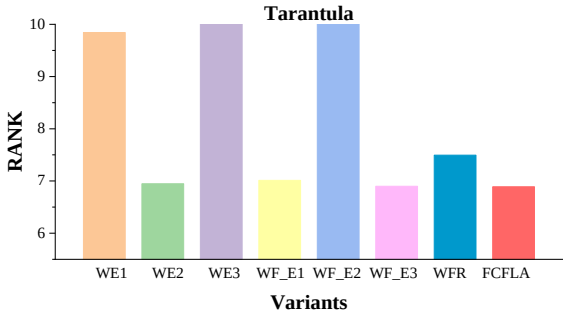
No	Ranking Metric	*Rank*					*EXAM*				
		FCFLA	VarCop	S-SBFL	SBFL	FB	FCFLA	VarCop	S-SBFL	SBFL	FB
1	Barinel	**7.61**	7.83	9.88	11.48	136.27	**1.76**	2.11	2.87	3.15	21.79
2	Dstar	**6.03**	6.16	7.20	8.09	108.78	**1.41**	1.77	1.94	2.02	15.88
3	Ochiai	**5.88**	6.19	7.25	8.14	109.91	**1.41**	1.77	1.95	2.03	16.10
4	Op2	**5.67**	5.86	6.07	6.74	106.99	**1.38**	1.71	1.75	1.80	15.36
5	Tarantula	**6.89**	6.96	9.88	11.48	136.27	**1.65**	1.98	2.87	3.15	21.79
6	Kulczynski2	**5.44**	5.61	3.36	7.08	108.23	**1.36**	1.67	1.77	1.83	15.59
7	M2	**5.79**	5.94	6.22	6.82	108.43	**1.38**	1.71	1.76	1.81	15.77
8	Harmonic Mean	**5.70**	5.95	6.52	7.28	149.70	**1.38**	1.72	1.80	1.86	21.37
9	Zoltar	**5.63**	6.00	6.12	6.78	107.57	**1.32**	1.68	1.75	1.80	15.45
10	Geometric Mean	**5.75**	6.05	7.37	8.29	149.70	**1.40**	1.76	1.99	2.09	21.37
11	Ample2	**5.95**	6.15	6.16	6.86	149.58	**1.42**	1.75	1.77	1.82	21.30
12	Rogot2	**5.95**	6.22	6.52	7.28	133.66	**1.49**	1.80	1.80	1.86	22.24
13	Sorensen Dice	**6.23**	6.50	8.79	10.17	115.72	**1.48**	1.84	2.41	2.62	17.29
14	Goodman	**6.23**	6.50	8.79	10.17	115.72	**1.48**	1.84	2.41	2.62	17.29
15	Jaccard	**6.37**	6.63	8.79	10.17	115.72	**1.48**	1.83	2.41	2.62	17.29
16	Dice	**6.37**	6.63	8.79	10.17	115.72	**1.48**	1.83	2.41	2.62	17.29
17	Anderberg	**6.39**	6.68	8.79	10.17	115.72	**1.48**	1.84	2.41	2.62	17.29
18	Cohen	**6.59**	6.81	8.93	10.33	152.04	**1.51**	1.87	2.47	2.70	21.61
19	Fleiss	6.95	**6.82**	12.24	52.03	145.70	**1.82**	2.09	3.51	9.13	21.65
20	Simple Matching	**6.82**	6.88	28.00	242.70	158.19	**1.82**	2.11	6.67	30.68	21.96
21	Humman	**6.82**	6.88	28.00	242.70	158.19	**1.82**	2.11	6.67	30.68	21.96
22	Wong2	**6.82**	6.88	28.00	242.70	158.19	**1.82**	2.11	6.67	30.68	21.96
23	Hamming	**6.82**	6.88	28.00	242.70	158.19	**1.82**	2.11	6.67	30.68	21.96
24	Sokal	**6.84**	6.91	28.00	242.70	158.19	**1.85**	2.15	6.67	30.68	21.96
25	Euclid	**6.82**	6.96	28.00	242.70	158.19	**1.85**	2.17	6.67	30.68	21.96
26	Rogers Tanimoto	**6.85**	7.05	28.00	242.70	158.19	**1.83**	2.20	6.67	30.68	21.96
27	Scott	**7.19**	7.38	13.22	50.86	147.65	**1.83**	2.17	3.76	8.79	22.23
28	Rogot1	**7.19**	7.38	13.22	50.86	147.65	**1.83**	2.17	3.76	8.79	22.23
29	Russell Rao	**12.13**	14.06	17.87	24.00	309.62	**2.43**	3.58	5.05	6.39	39.27
30	Wong1	**12.13**	14.06	17.87	24.00	309.62	**2.43**	3.58	5.05	6.39	39.27
	Average	**6.795**	7.094	13.625	68.605	146.777	**1.647**	2.034	3.545	9.842	21.015

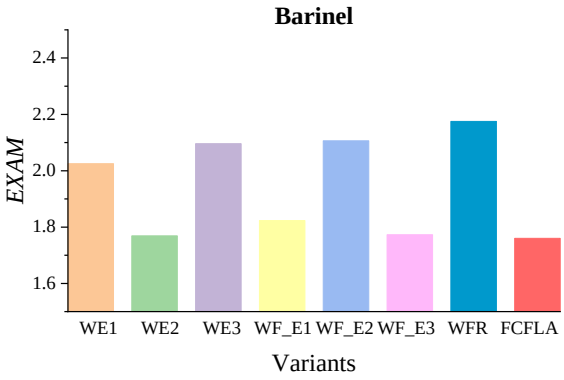

Barinel
RANK
12
10
8
6
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants

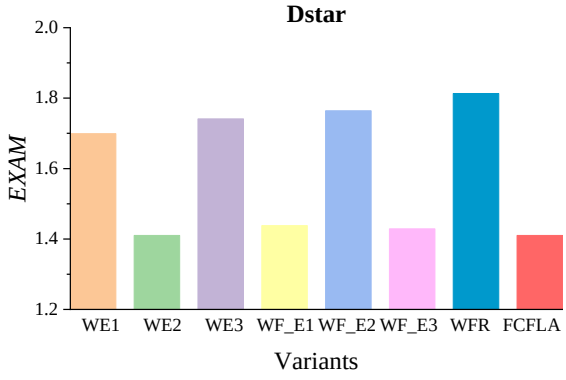
Dstar
RANK
10
8
6
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants

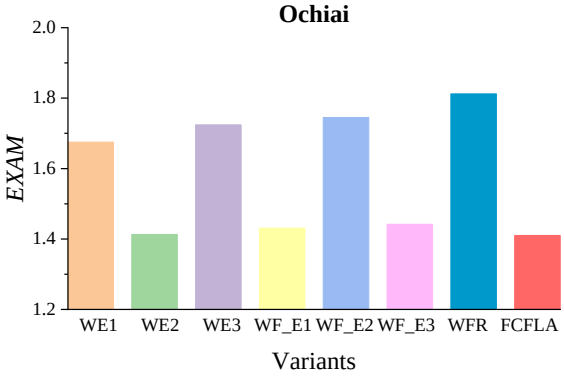

Ochiai
RANK
10
8
6
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants

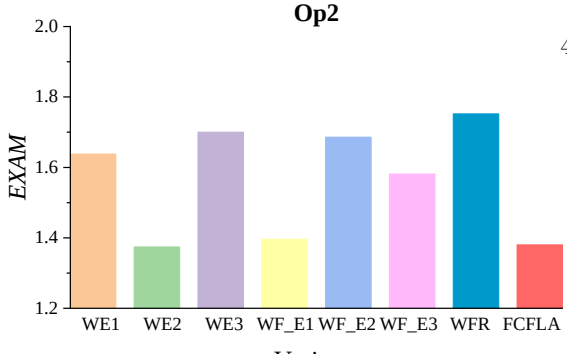

Op2
RANK
10
8
6
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants

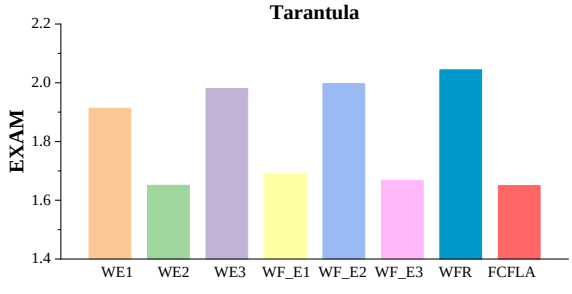

Tarantula
RANK
10
9
8
7
6
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants


Barinel
EXAM
2.4
2.2
2.0
1.8
1.6
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants


Dstar
EXAM
2.0
1.8
1.6
1.4
1.2
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants


Ochiai
EXAM
2.0
1.8
1.6
1.4
1.2
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA
Variants


Op2
EXAM
2.0
1.8
1.6
1.4
1.2
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA


Tarantula
EXAM
2.2
2.0
1.8
1.6
1.4
WE1 WE2 WE3 WF_E1 WF_E2 WF_E3 WFR FCFLA

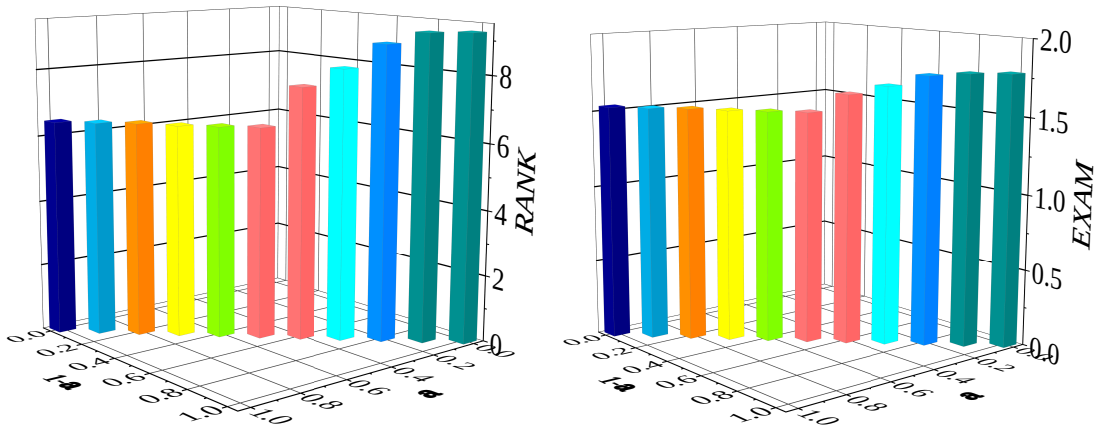


Figure S-2: The effect of setting different Θ on $RANK$ and $EXAM$ respectively.

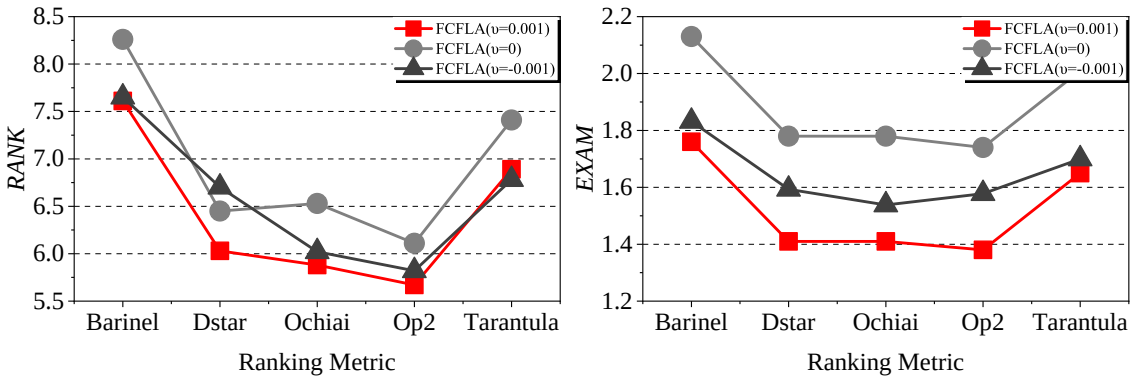


Figure S-3: The effect of setting different υ on $RANK$ and $EXAM$ respectively.