

An Empirical Study on Noisy Label Learning for Program Understanding

Wenhan Wang
Nanyang Technological University
Singapore
wwhjacob@hotmail.com

Yanzhou Li
Nanyang Technological University
Singapore
yanzhou001@e.ntu.edu.sg

Anran Li*
Nanyang Technological University
Singapore
anran.li@ntu.edu.sg

Jian Zhang
Nanyang Technological University
Singapore
jian_zhang@ntu.edu.sg

Wei Ma
Nanyang Technological University
Singapore
ma_wei@ntu.edu.sg

Yang Liu*
Nanyang Technological University
Singapore
yangliu@ntu.edu.sg

ICSE '24: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering May 2024,
Article No.: 227, Pages 1-12

発表者：野口

背景 | プログラム理解タスク

ソースコードに関する様々なタスク

- 例) プログラム分類, 脆弱性予測, コード要約

ディープラーニング(DL)モデルで学習

- 大規模データセットで0から訓練
- 事前学習済みモデルをファインチューニング

⇒ 訓練データにノイズが多いとモデルの性能を低下させてしまう

—— プログラム理解タスク ——

高品質, ノイズの少ないデータセットを収集するのは非常に困難^[28]

[28] Lili Mou, Ge Li, Lu Zhang, Tao Wang, and Zhi Jin. 2016. Convolutional neural networks over tree structures for programming language processing. In Proceedings of the AAAI conference on artificial intelligence, Vol. 30.

ノイズラベル学習 (NLL)

不正確なラベルでの影響を最小限にする機械学習手法

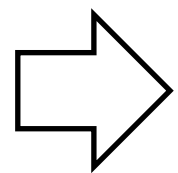
- DLモデルのロバスト性を向上させ、誤ラベルを検出できる

ロバスト性

モデルが、ノイズの多いデータに対して、
どれだけ影響を受けにくいかを表す性質

NLLのプログラム理解における課題点

- 分類タスクのみで、生成タスクは調査されていない
- 事前学習済みモデルは検証されていない



プログラム理解についてNLLの体系的な研究が必要

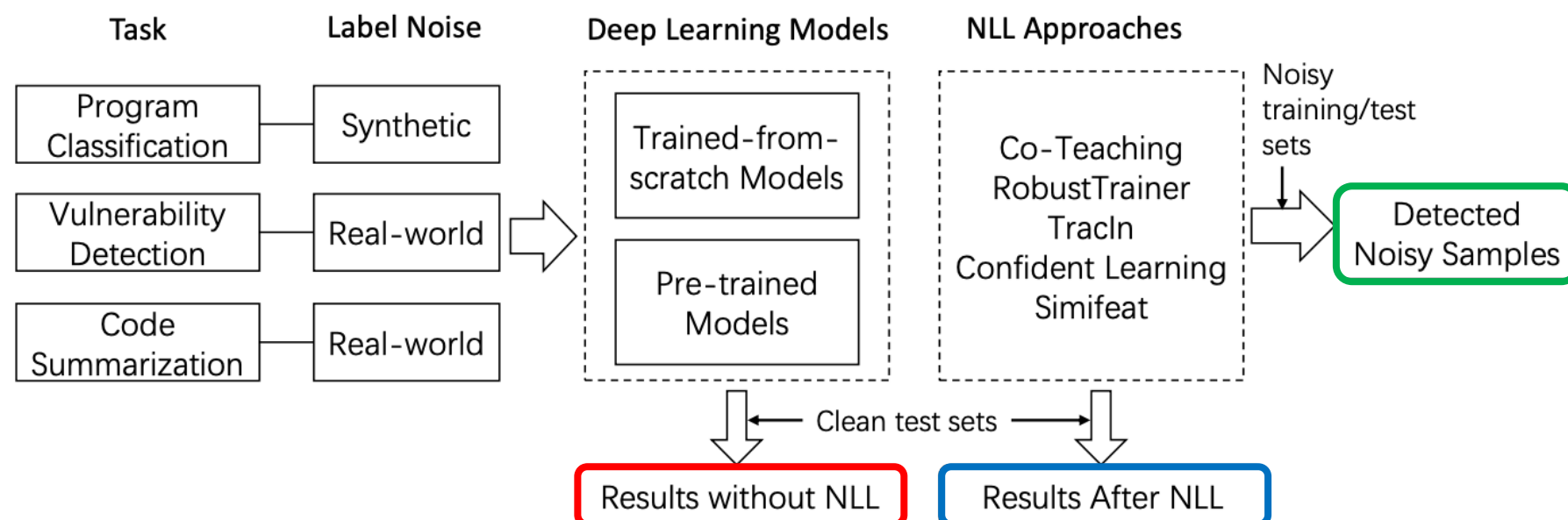
目的とリサーチクエスチョン

目的

- ノイズラベル学習がプログラム理解に与える影響を明らかにする

リサーチクエスチョン (RQ)

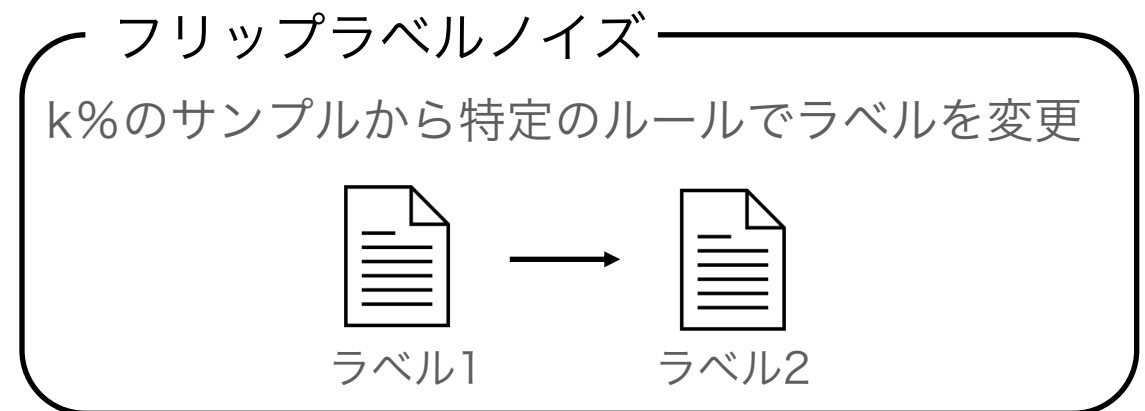
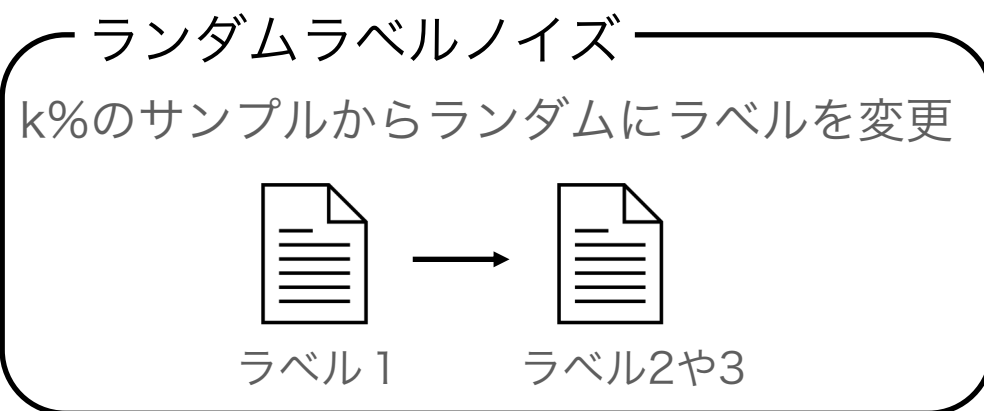
- RQ1** : NLLを使用しない場合の, DLモデルの性能は？
- RQ2** : NLL使用でプログラム分類におけるノイズへの影響は？
- RQ3** : NLLのノイズがあるプログラム理解タスクでの機能は？



データセット | 分類・検出タスク

プログラム分類タスク

- CodeNet^[33]データセットを使用
 - 250のクラスを含むJavaデータセットを選択
 - ランダムラベルノイズとフリップラベルノイズを注入



脆弱性検出タスク

- Devign^[49]データセットを使用
 - 2つのOSSプロジェクトからC言語のコードスニペットを収集
 - 46%が脆弱とラベル付け

データセット | コード要約タスク

TLCデータセットを使用

- 約41%がノイズデータである
 - ルールベースに従っており，人間の開発者の基準を反映していない

⇒ データセットの人間評価を実施

データセットを人間評価

- 対象者：博士課程学生，ポスドク6人
- 評価方法：0から2までのスコアで評価→平均スコア1.42

```
Code:
public static boolean startsWith(String s1,
String s2){
    if (s1 == null || s2 == null) {
        return false;
    }
    return s1.startsWith(s2);
}
```

Summary:
Check if one string starts with another string.

スコア 2

要約が機能を記述

```
Code:
public S2Cap complement(){
    double cHeight=isFull() ? -1 : 2 -
Math.max(height,0.0);
    return S2Cap.fromAxisHeight(S2Point.neg(axis),cHeight);
}
```

Summary:
Return the complement of the interior of the cap. A cap and its complement have the same boundary but do not share any interior points. The complement operator is not a bijection, since the complement of a singleton cap (containing a single point) is the same as the complement of an empty cap.

スコア 1

要約が部分的に機能を記述

```
Code:
protected void engineSetPadding(String padding)
throws NoSuchPaddingException {
    if (!padding.equalsIgnoreCase("NoPadding")) {
        throw new NoSuchPaddingException("Padding "
+ padding + " unknown.");
    }
}
```

Summary:
should never be called.

スコア 0

要約が機能を記述していない

DLモデルとノイズ学習アプローチ

DLモデル

Type	Model	Parameters
Small trained-from-scratch	LSTM	7M
	GIN	7M
	Transformer (summarization)	30M
Large pre-trained	CodeBERT	125M
	GraphCodeBERT	125M
	UnixCoder	125M
	PLBART	140M

 : 分類タスク向けモデル

 : 要約タスク向けモデル

ノイズラベル学習アプローチ

- TracIn^[32], Co-teaching^[14], Rubust Trainer^[25],
Confiden Learning^[30], Simifeat^[50]

[32] Garima Pruthi, Frederick Liu, Satyen Kale, and Mukund Sundararajan. 2020. Estimating training data influence by tracing gradient descent.

[14] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels.

[25] Zhong Li, Minxue Pan, Yu Pei, Tian Zhang, Linzhang Wang, and Xuandong Li. 2022. Robust Learning of Deep Predictive Models from Noisy and Imbalanced Software Engineering Datasets.

[30] Curtis Northcutt, Lu Jiang, and Isaac Chuang. 2021. Confident learning: Estimating uncertainty in dataset labels. Journal of Artificial Intelligence Research 70 (2021), 1373–1411.

[50] Zhaowei Zhu, Zihao Dong, and Yang Liu. 2022. Detecting corrupted labels without training a model to predict. In International Conference on Machine Learning. PMLR, 27412–27427.

RQ1

NLLを使用しない場合，異なるラベルノイズでのDLモデルの性能は？

実験手順

1. モデルをノイズのないデータセットで実行
2. 訓練データに異なるパターンのノイズをいれ再度モデルを実行
3. ランダムノイズ(10-50%)， フリップノイズ(10-40%)で検証

	Model	No noise	10%random	20% random	30%random	50% random	10% flip	20%flip	30% flip	40% flip
小型モデル	LSTM	93.04	86.68(6.36)	82.88 (10.16)	78.69(14.35)	56.57 (36.47)	83.64(9.40)	77.59 (15.45)	68.22(24.82)	56.32 (36.72)
	GIN [42]	93.23	89.16(4.07)	86.72 (6.51)	83.64(9.59)	78.73 (14.50)	85.17(8.06)	81.51 (11.72)	73.81(19.42)	61.42 (31.81)
大型モデル	CodeBERT [11]	98.05	97.57(0.48)	97.26 (0.79)	96.87(1.18)	96.05 (2.00)	97.43(0.62)	96.05 (2.00)	91.22(6.83)	85.58 (12.47)
	GraphCodeBERT [13]	98.26	97.62(0.64)	97.60 (0.66)	97.15(1.11)	96.48 (1.78)	97.77(0.49)	96.41 (1.85)	92.87(5.39)	87.85 (10.41)
	UnixCoder [12]	98.39	97.86(0.53)	97.29 (1.10)	97.18(1.21)	96.67 (1.72)	97.75(0.64)	96.52 (1.87)	93.06(5.33)	90.21 (8.18)

ランダムノイズ

小型モデル：ノイズの割合が高いほど
精度は低下

大型モデル：ノイズの割合が高くなっても
高精度を維持

フリップノイズ

小型モデル：ノイズの割合が高いほど
精度は低下

大型モデル：40%のフリップノイズでは
精度が10%ほど低下

RQ2

NLL使用でプログラム分類におけるノイズへの影響は？

分類精度と誤ラベル検出を評価

- 一部抜粋（50%のランダムノイズ）

モデル		NLLなし	NLLあり (Co-teaching)
小型モデル	LSTM	56.57	77.85
	GIN	78.73	86.42
大型モデル	CodeBERT	96.05	96.43

分類精度

モデル	訓練データ (Train)	テストデータ (Test)
LSTM	0.857	0.858
GIN	0.846	0.868
CodeBERT	0.983	0.945

F1-score

RQ3 NLLのノイズがあるプログラム理解タスクでの機能は？

脆弱性検出

実験設定

- 70の脆弱性ラベル付きサンプルを作成^[8]（20%が誤ラベル）
- 70のうち60はDevignの訓練セット
- TracInの評価は推定ノイズ率（20%, 50%）を採用
 - Devignの事前知識が正確でない可能性があるため

結果（F1-score）

モデル	TracIn20%	TracIn50%	Confident Learning	Baseline class
LSTM	0.347	0.390	0.631	0.583
CodeBERT	0.521	0.439	0.800	0.750
GraphCodeBERT	0.695	0.536	0.736	0.782
UnixCoder	0.695	0.536	0.750	0.769

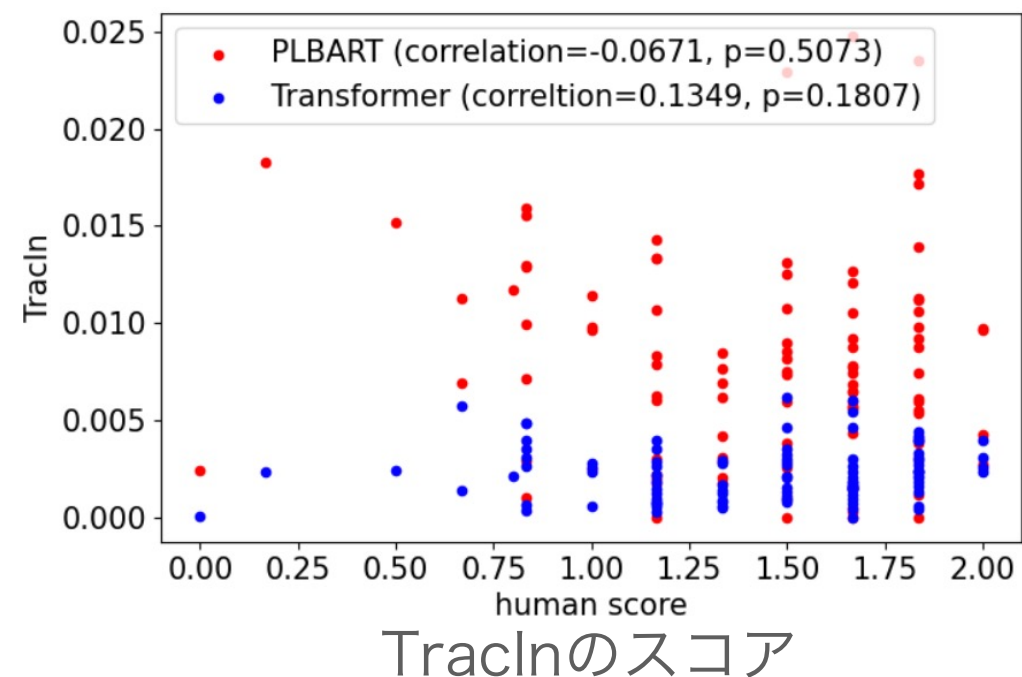
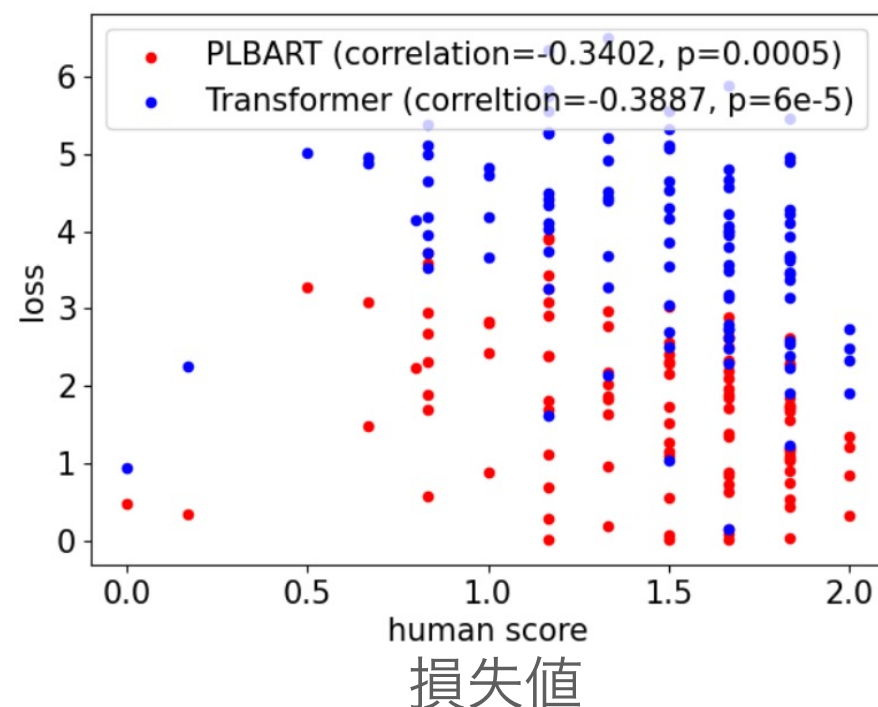
RQ3 NLLのノイズがあるプログラム理解タスクでの機能は？

コード要約

実験設定

- NLLが付与したスコアと人間評価スコアの分布を比較
- 損失値を使用
 - 損失値が高い：ノイズが多いか低品質な要約 損失値が低い：高品質な要約
- スピアマンの順位相関係数を採用し、スコアの関連性を測定

結果



まとめ

目的

- ノイズラベル学習（NLL）がプログラム理解タスクに与える影響を調査

結果

- NLLを行うことでノイズラベルを正確に分類, 検出可能

今後の課題

- コード生成やプログラム修正分野でのNLLの検証

補足 | RQ2の分類精度

Model	Approach	10%random	20% random	30%random	50% random	10%flip	20% flip	30%flip	40% flip
LSTM	Co-teaching	90.26(3.58)	88.25 (5.37)	83.37(4.68)	77.85 (21.28)	87.35(3.71)	84.27 (6.68)	76.67(8.45)	68.25 (11.93)
	TracIn	90.17(3.49)	89.46 (6.58)	84.05(5.36)	76.81 (20.24)	88.81(5.17)	88.62 (11.03)	73.30(5.08)	69.26 (12.94)
	Confident Learning	90.59(3.91)	86.96 (4.08)	81.72(3.03)	65.12 (8.55)	87.49(3.85)	86.41 (8.82)	75.26(7.04)	60.28 (3.96)
	RobustTrainer	91.37(4.69)	91.29 (8.41)	83.98(5.29)	75.64 (19.07)	88.96(5.32)	88.47 (10.88)	75.41(7.19)	61.79 (5.47)
	Baseline-class	91.82(5.14)	90.23 (7.35)	85.11(6.42)	73.13 (16.56)	88.58(4.94)	85.43 (7.84)	74.64(6.42)	57.69 (1.37)
GIN [42]	Co-teaching	91.28(2.12)	90.73 (4.01)	87.18(3.54)	86.42 (7.69)	89.25(4.08)	88.41 (6.90)	81.02(7.21)	74.03 (12.61)
	TracIn	90.74(1.58)	90.35 (3.63)	87.39(3.75)	83.26 (4.53)	89.79(4.62)	88.47 (6.96)	78.88(5.07)	75.78 (14.36)
	Confident Learning	92.08(2.92)	88.95 (2.23)	86.94(3.30)	80.78 (2.05)	90.51(5.34)	87.68 (6.17)	80.01(6.20)	67.13 (5.71)
	RobustTrainer	89.65(0.49)	90.51 (3.79)	84.75(1.11)	82.86 (4.13)	89.48(4.31)	85.44 (3.93)	76.37(2.56)	59.51 (-1.91)
	Baseline-class	92.21(3.05)	90.95 (4.23)	87.02(3.38)	85.63 (6.90)	90.55(5.38)	87.48 (5.97)	77.00(3.19)	65.23 (3.81)
CodeBERT [11]	Co-teaching	97.58(0.01)	97.40 (0.14)	97.05(0.28)	96.43 (0.38)	97.52(0.09)	96.95 (0.90)	91.29(0.07)	86.28 (0.70)
	TracIn	97.85(0.28)	97.80 (0.54)	97.44(0.47)	96.83 (0.78)	97.90(0.47)	97.78 (1.73)	96.70(5.48)	97.06 (11.48)
	Confident Learning	98.05(0.48)	97.81 (0.55)	97.47(0.60)	96.69 (0.64)	98.02(0.59)	97.71 (1.66)	96.79(5.57)	94.27 (8.69)
	Simifeat	97.60(0.03)	96.76 (-0.50)	96.66(-0.21)	96.21 (0.16)	97.49(0.06)	96.79 (0.74)	91.12(-0.10)	89.83 (4.25)
	RobustTrainer	97.65(0.08)	97.05 (-0.21)	96.92(0.05)	97.01 (0.96)	97.71(0.28)	96.61 (0.56)	93.19(1.97)	84.87 (-0.71)
	Baseline-class	97.91(0.34)	97.81 (0.55)	97.76(0.89)	97.23 (1.18)	97.94(0.51)	97.25 (1.30)	96.42(3.20)	91.79 (6.21)
GraphCodeBERT [13]	Co-teaching	97.64(0.02)	97.65 (0.05)	97.35(0.20)	96.91 (0.43)	97.77(0.00)	97.66 (1.25)	92.87(0.00)	90.19 (2.34)
	TracIn	98.16(0.54)	98.14 (0.54)	97.30(0.15)	97.16 (0.68)	98.00(0.23)	97.82 (1.41)	97.44(4.57)	97.21 (9.36)
	Confident Learning	98.24(0.62)	97.99 (0.39)	97.76(0.61)	97.13 (0.65)	98.26(0.49)	98.05 (1.64)	97.10(4.23)	94.63 (6.78)
	Simifeat	97.67(0.05)	97.22 (-0.38)	97.13(-0.02)	96.54 (0.06)	97.85(0.08)	97.35 (0.94)	94.12(1.25)	90.50 (2.65)
	RobustTrainer	97.96(0.34)	97.59 (-0.01)	97.51(0.36)	97.66 (1.18)	97.96(0.19)	96.78 (0.37)	93.03(0.16)	88.13 (0.28)
	Baseline-class	98.12(0.50)	98.06 (0.46)	97.94(0.79)	97.47 (0.99)	98.02(0.25)	97.72 (1.31)	95.89(3.02)	92.73 (4.88)
UnixCoder [12]	Co-teaching	97.99(0.13)	97.62 (0.33)	97.26(0.08)	97.05 (0.38)	97.89(0.14)	97.39 (0.87)	94.05(0.99)	92.41 (2.20)
	TracIn	98.25(0.39)	98.25 (0.96)	97.59(0.33)	97.49 (0.82)	98.06(0.31)	98.18 (1.66)	97.35(4.29)	97.31 (7.10)
	Confident Learning	98.32(0.46)	98.06 (0.77)	97.85(0.67)	97.37 (0.70)	98.29(0.54)	98.20 (1.68)	97.37(4.31)	95.19 (4.98)
	Simifeat	97.86(0.00)	97.25 (-0.04)	97.22(0.04)	96.55 (-0.12)	97.81(0.06)	97.31 (0.79)	93.01(-0.05)	92.69 (2.48)
	RobustTrainer	97.95(0.09)	97.48 (0.19)	97.07(-0.11)	97.06 (0.39)	97.85(0.10)	97.42 (0.90)	94.18(1.12)	91.59 (1.38)
	Baseline-class	98.26(0.40)	98.02 (0.73)	97.71(0.53)	97.81 (1.14)	98.11(0.36)	98.03 (1.51)	95.22(2.16)	93.67 (3.46)

補足 | RQ2の誤ラベル検出 (F1)

Model	Approach	10% random			20% random			30% random			50% random			10% flip			20% flip			30% flip			40% flip		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
LSTM	TracIn	74.17	74.17	74.17	82.51	82.51	82.51	80.05	80.05	80.05	85.70	85.70	85.70	72.96	72.96	72.96	84.27	84.27	84.27	71.82	71.82	71.82	69.11	69.11	69.11
	Confident Learning	84.78	79.33	81.96	90.95	84.58	87.65	88.57	76.31	81.99	82.18	82.34	82.26	85.23	72.81	78.53	77.50	87.72	82.30	87.15	68.0	76.39	59.98	65.94	62.82
	Baseline-class	76.37	94.11	84.32	72.65	99.96	84.15	71.38	98.53	82.79	74.04	99.83	85.02	64.25	97.72	77.53	60.74	91.43	72.99	65.48	94.74	77.44	56.10	70.55	62.50
GIN [42]	TracIn	74.00	74.00	74.00	84.42	84.42	84.42	80.20	80.20	80.20	84.56	84.56	84.56	75.11	75.11	75.11	73.57	73.57	73.57	76.21	76.21	76.21	69.71	69.71	69.71
	Confident Learning	92.21	80.27	85.83	90.65	85.72	88.12	92.53	81.90	86.89	91.66	84.24	87.10	91.18	78.82	84.55	77.03	89.09	82.62	75.11	85.27	79.87	59.86	65.11	62.38
	Baseline-class	82.68	94.16	88.05	75.50	99.98	86.03	76.98	96.67	88.62	85.71	99.94	91.36	78.15	69.47	73.55	62.54	91.58	74.33	66.91	80.91	73.25	56.43	68.41	61.85
CodeBERT [11]	TracIn	80.53	80.53	80.53	94.21	94.21	94.21	91.13	91.13	91.13	98.37	98.37	98.37	84.87	84.87	84.87	93.69	93.69	93.69	94.01	94.01	94.01	94.35	94.35	94.35
	Confident Learning	98.99	95.60	97.26	99.04	93.97	96.43	99.22	94.46	96.78	99.18	94.61	96.84	96.78	96.93	96.85	96.80	96.56	96.68	93.34	93.82	93.58	91.79	89.74	90.75
	Simifeat	67.42	61.09	64.10	61.10	78.56	68.74	60.94	68.25	64.39	82.75	60.20	69.70	63.77	55.51	59.35	56.67	70.10	62.67	64.87	73.64	68.98	65.23	51.94	57.83
	Baseline-class	88.41	100.0	93.85	92.73	100.00	96.23	95.34	99.99	97.61	96.88	99.99	98.41	88.36	99.73	93.70	89.91	98.57	94.04	83.99	97.88	90.41	82.37	95.30	88.37
GraphCodeBERT [13]	TracIn	85.47	85.47	85.47	96.41	96.41	96.41	93.02	93.02	93.02	99.17	99.17	99.17	84.31	84.31	84.31	95.82	95.82	95.82	94.97	94.97	94.97	93.08	93.08	93.08
	Confident Learning	98.98	94.71	96.79	99.16	95.04	97.06	99.17	94.94	97.01	99.35	93.53	96.36	97.06	97.08	97.07	96.92	97.04	96.98	95.22	92.87	94.03	92.22	89.41	90.79
	Simifeat	66.39	59.51	62.76	57.36	85.26	68.58	67.77	60.84	64.12	79.31	73.06	76.06	63.31	69.33	66.18	62.18	59.67	60.90	73.66	61.79	67.21	69.37	61.26	65.06
	Baseline-class	87.29	99.98	93.20	94.47	99.98	97.15	95.43	99.99	97.66	97.29	99.99	98.62	89.82	99.82	94.55	90.23	98.57	94.21	87.23	97.98	92.29	84.11	94.49	89.00
UnixCoder [12]	TracIn	90.24	90.24	90.24	97.43	97.43	97.43	92.95	92.95	92.95	99.31	99.31	99.31	83.71	83.71	83.71	96.62	96.62	96.62	93.15	93.15	93.15	95.83	95.83	95.83
	Confident Learning	96.96	92.87	94.87	98.57	89.23	93.67	99.80	92.75	96.15	99.30	92.07	95.55	97.91	94.02	95.93	95.14	93.90	94.51	97.87	91.53	94.59	93.69	92.66	93.17
	Simifeat	63.78	70.87	67.14	68.56	81.67	74.54	62.30	66.10	64.14	85.72	65.20	74.06	65.54	46.80	54.61	60.19	72.38	65.72	56.42	61.12	58.67	70.50	47.78	56.96
	Baseline-class	93.48	98.13	95.74	94.03	99.98	96.91	92.23	97.87	94.96	96.72	99.99	98.32	92.35	98.93	95.53	87.86	100.00	93.54	90.61	98.16	94.23	91.15	95.98	93.50

訓練データ

Model	Approach	10% random			20% random			30% random			50% random			10% flip			20% flip			30% flip			40% flip		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
LSTM	TracIn	67.31	67.31	67.31	76.15	76.15	76.15	81.02	81.02	81.02	85.82	85.82	85.82	80.83	80.83	80.83	79.91	79.91	79.91	67.95	67.95	67.95	71.04	71.04	71.04
	Confident Learning	81.09	78.33	79.69	80.37	80.67	80.52	78.24	73.84	75.98	73.42	84.48	78.56	74.75	80.51	77.52	62.62	84.00	71.75	75.31	69.4	72.23	55.48	65.17	59.93
	Baseline-class	70.42	96.20	81.32	60.54	99.93	75.40	69.99	97.16	81.36	67.38	99.73	80.43	60.33	97.16	74.44	48.77	89.83	63.82	52.62	88.24	65.93	52.35	71.53	60.46
GIN [42]	TracIn	71.06	71.06	71.06	78.63	78.63	78.63	79.90	79.90	79.90	86.85	86.85	86.85	71.95	71.95	71.95	67.30	67.30	67.30	71.46	71.46	71.46	67.10	67.10	67.10
	Confident Learning	85.72	78.56	81.98	83.50	83.33	83.42	89.20	74.09	80.95	91.08	83.49	87.12	86.16	81.2	83.61	68.76	86.07	76.45	81.64	73.18	77.18	57.09	65.28	60.91
	Baseline-class	80.71	91.93	85.95	65.57	99.93	79.19	79.16	89.20	83.88	83.23	99.92	90.81	78.99	84.98	81.88	54.43	91.03	68.13	71.43	74.61	72.98	54.24	70.27	61.26
CodeBERT [11]	TracIn	73.23	73.23	73.23	79.83	79.83	79.83	85.76	85.76	85.76	94.56	94.56	94.56	83.60	83.60	83.60	79.90	79.90	79.90	87.52	87.52	87.52	91.90	91.90	91.90
	Confident Learning	99.76	91.47	95.43	96.69	92.47	94.53	98.57	91.27	94.78	98.89	92.76	95.73	98.06	85.27	91.22	95.40	96.10	95.75	99.09	81.48	89.43	82.29	79.80	81.03
	Simifeat	54.22	58.02	56.05	51.24	76.60	61.40	65.16	61.04	63.04	81.62	55.23	65.88	56.52	73.88	64.04	46.95	69.03	55.89	80.28	57.34	66.9	60.91	53.98	57.24
	Baseline-class	89.51	95.80	92.55	89.15	100.00	94.27	88.75	100.0	94.04	95.88	100.00	97.89	79.32	91.89	85.14	88.43	99.90	93.82	82.06	97.87	89.27	73.16	83.18	77.85
GraphCodeBERT [13]	TracIn	76.31	76.31	76.31	81.17	81.17	81.17	83.88	83.88	83.88	95.24	95.24	95.24	84.17	84.17	84.17	81.20	81.20	81.20	89.46	89.46	89.46	93.20	93.20	93.20
	Confident Learning	99.10	88.36	93.42	98.53	91.70	94.99	99.98	90.45	94.98	99.24	92.55	95.77	98.24	80.78	88.66	95.98	95.50	95.74	99.15	84.46	91.22	89.74	87.75	88.73
	Simifeat	57.69	64.0	60.68	46.87	68.13	55.53	55.04	67.93	60.81	77.84	47.69	59.14	83.6	59.83	69.74	44.32	64.33	52.48	54.12	82.76	65.44	59.50	51.85	55.41
	Baseline-class	89.17	97.33	93.07	91.60	99.97	95.60	93.57	98.67	96.05	96.47	99.99	98.20	90.11	94.11	92.07	89.11	99.03	93.81	88.78	99.31	93.75	82.90	91.87	87.15
UnixCoder [12]	TracIn	82.97	82.97	82.97	90.00	90.00	90.00	87.77	87.77	87.77	97.39	97.39	97.39	80.06	80.06	80.06	89.80	89.80	89.80	92.37	92.37	92.37	96.03	96.03	96.03
	Confident Learning	96.78	91.6	94.12	98.13	90.97	94.41	98.21	91.9	94.95	98.67	90.81	94.58	93.98	91.89	92.92	95.79	94.57	95.17	97.56	88.16	92.62	92.83	92.17	92.50
	Simifeat	49.88	64.78	56.36	58.50	84.30	69.07	52.88	64.75	58.21	88.85	65.81	75.61	55.56	72.16	62.78	52.74	74.27	61.68	67.15	57.36	61.87	52.67	55.85	54.21
	Baseline-class	88.95	96.04	92.36	92.45	99.97	96.06	94.83	99.3	97.01	95.45	100.00	97.58	88.41	96.27	92.17	90.10	100.00	94.79	88.75	97.96	93.13	91.42	97.98	94.59

テストデータ