

Tracing with Less Data: Active Learning for Classification-Based Traceability Link Recovery

Chris Mills, Javier Escobar-Avila, Aditya Bhattacharya, Grigoriy Kondyukov, Shayok Chakraborty, Sonia Haiduc

Department of Computer Science

Florida State University

Tallahassee, USA

{cmills, escobara, abhattac, kondyuko, shayok, shaiduc}@cs.fsu.edu

背景

- ソフトウェアのトレーサビリティのレベルが高いシステムでは、要素（プロダクトコード、テストコード、ドキュメント）がどのように対応しているかを安易に理解できる
- トレーサビリティが高いシステムは良いコードと、バグが少ないシステムであるとするされている
- しかし、トレーサビリティの管理はプロジェクトにとって優先事項の低い作業である

TLR (Traceability Link Recovery)

- ソフトウェア要素間のリンクを対応付けることを**TLR**という
 - TLRは要件やコード，テストが作成された後，時間がたって行われることが多いため非常にコストのかかる作業
- TLRの半自動化アプローチ
 - 教師あり学習を用いたアプローチ (**TRAIL**) [9]が提案されている
 - 学習モデル構築のためには，ラベル付けされた大量の学習データが必要
 - 実際，殆どのプロジェクトではトレーサビリティデータを収集していないため使うことができない

目的とアプローチ

- 目的
 - 多くのプロジェクトで利用できるよう、少ないデータ量でTraceability Link Recoveryを行える自動化手法の構築を行う
- アプローチ（提案手法：ALCATRAL）
 - 既存手法TRAILに**アクティブラーニング**を用いた学習方法を組み込むことで、予測精度を維持しつつデータ量を削減する

アクティブラーニング

- ・ 少ないラベル付けで高い認識率を実現する手法
- ・ モデルに影響を与えるデータにのみラベル付けを行う

ALCATRALの概要

1. トレーサビリティリンクの特徴量（131種類）の抽出

- IR-ranking Features
- Query Quality Features
- Artifact Statics Feature

2. 相関を用いた特徴量の選択

- 良い特徴量はTargetの値と相関を持っていると仮定して，相関の低いものを削減

3. SMOTEによるデータ不均衡の対応

4. 分類器の構築（RandomForest）

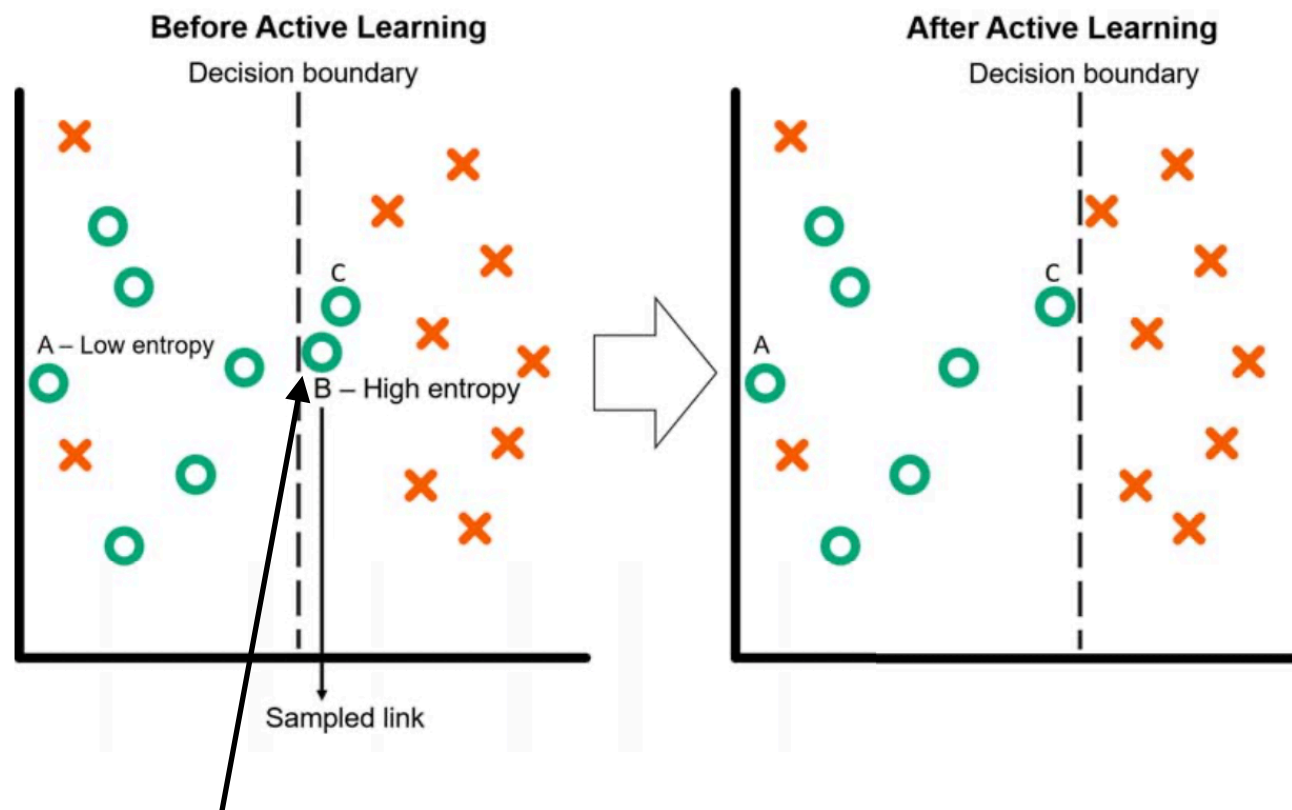
5. アクティブラーニング

TRAIL
既存手法

ALCATRAL

アクティブラーニング

- アクティブラーニング：
予測モデルが不確かな予測をしたデータに対してラベルを与える



$$E(x) = - \sum_{i \in \{0,1\}} P_i(x) \log P_i(x)$$

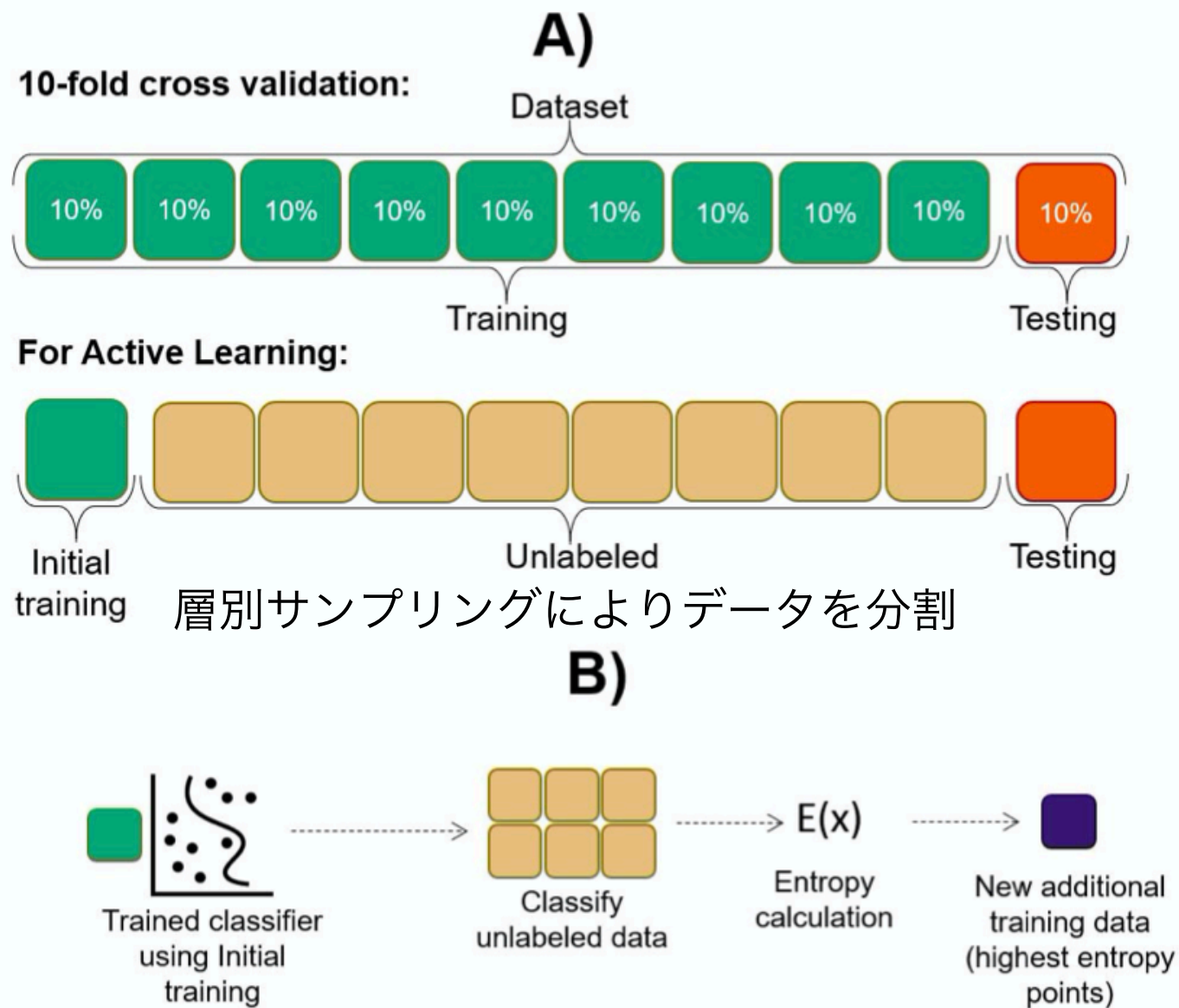
不確実性の尺度としてエントロピー
を用いる

境界線の周辺にある不確かなデータを
取り出してラベルを与える

リサーチクエスチョン

1. アクティブラーニングを一度適用するだけで、既存手法TRAILと同様の精度を保ちつつデータ量を削減できるか？
2. アクティブラーニングを繰り返し適用すると提案手法ALCATRALの精度は向上するか？

実験方法



もっとも不確かな予測をしたデータを学習データとして追加する

Fig. 2. Data splitting and processing

毎回学習データを作成

データセット

TABLE I
DATASETS USED IN THE EVALUATION

System	Invalid Links	Valid Links	Artifact Types [†]
eAnci	7091	554 (7.25%)	UC, CC
EasyClinic	1317	93 (6.60%)	UC, CC
EasyClinic	871	69 (7.34%)	ID, CC
EasyClinic	1177	83 (6.59%)	ID, TC
EasyClinic	574	26 (4.33%)	ID, UC
EasyClinic	2757	204 (6.89%)	TC, CC
EasyClinic	1827	63 (3.33%)	TC, UC
eTour	6363	365 (5.43%)	UC, CC
iTrust	1493	58 (3.74%)	UC, CC
MODIS	890	41 (4.40%)	HighR, LowR
SMOS	5656	1044 (15.58%)	UC, CC
Total	30016	2600 (7.97%)	

[†] HighR = High-level Requirements, LowR = Low-level Requirements, UC = Use Cases, CC = Code Classes, ID = Interaction Diagrams, TC = Test Cases.

RQ1 結果

※F値による評価

TABLE II
PERFORMANCE OF ALCATRAL BY AMOUNT OF TRAINING DATA ADDED USING ACTIVE LEARNING AND RANDOM SELECTION

Initial				Size of the training set														TRAIL baseline as in [9]
10%		20%		30%		40%		50%		60%		70%		80%		90%		
Dataset		ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	
eAnci	52.2	73.1*	61.4	77.4*	65.9	<u>77.4*</u> [‡]	69.0	77.0*	71.7	76.4*	73.6	75.9*	75.1	75.2*	76.7	74.6*	78.0	77.9
eC-CC/UC	45.8	62.4*	53.9	65.5*	57.8	66.5*	60.1	67.1* [†]	62.4	<u>67.8*</u> [†]	63.3	67.6* [†]	65.0	67.3* [†]	65.9	67.4 [†]	67.6	67.5
eC-ID/CC	42.1	64.0*	51.6	69.4*	57.9	71.3*	62.7	71.8* [‡]	64.6	<u>72.3*</u> [†]	67.5	71.7* [‡]	69.4	72.4* [†]	70.6	<u>72.5[†]</u>	72.1	72.8
eC-ID/TC	63.2	90.1*	75.6	93.1* [†]	81.4	93.6* [†]	85.1	93.6* [†]	87.6	93.3* [†]	89.4	93.7* [†]	91.3	<u>93.8*</u> [‡]	92.4	<u>93.6[†]</u>	93.4	93.4
eC-ID/UC	25.3	50.9*	29.2	52.3*	39.9	54.3*	43.2	54.7* [‡]	47.1	55.4* [‡]	49.0	55.5* [‡]	51.6	<u>55.6[†]</u>	53.9	<u>55.7*</u> [‡]	57.8	57.6
eC-TC/CC	73.8	94.7*	86.4	97.7* [‡]	91.5	98.0*	93.7	98.0*	95.2	97.9*	96.1	98.0*	96.8	<u>98.1*</u>	97.2	<u>98.0*</u>	97.4	97.5
eC-TC/UC	56.2	91.3*	69.4	93.0*	77.1	93.3*	81.9	<u>93.4*</u>	85.8	92.9*	88.1	93.1*	90.9	<u>93.0</u>	92.9	92.9*	94.7	94.5
eTour	42.5	55.0*	49.6	57.0*	52.1	57.5*	54.4	<u>57.9*</u>	56.3	58.2*	57.1	58.3	58.5	58.4*	59.7	<u>58.5*</u>	60.7	60.8
iTrust	30.9	62.5* [‡]	47.9	64.1*	53.7	<u>64.5*</u>	57.7	63.7*	59.3	63.3*	60.0	62.7*	60.5	63.0*	61.1	62.3	61.4	60.9
MODIS	32.1	52.0*	42.3	57.2*	48.7	<u>58.6*</u>	53.1	59.8*	56.0	61.2*	59.2	61.6*	60.3	62.6	62.4	<u>62.7*</u>	63.8	64.3
SMOS	47.6	60.2*	56.7	67.0*	62.3	71.4*	66.6	73.8*	70.4	75.2*	73.5	75.9	76.0	<u>76.5*</u>	77.9	<u>76.3*</u>	79.8	79.7
Average	46.5	68.8*	56.7	72.2*	62.6	73.3*	66.1	73.7*	68.7	74.0*	70.6	74.0*	72.3	74.2*	73.7	74.0*	75.1	75.2

*=there is a significant statistical difference between ALCATRAL and TRAIL trained with the same amount of data. †=there is no significant statistical difference between ALCATRAL and the TRAIL baseline [9]. ‡=there is a significant statistical different between ALCATRAL and the TRAIL baseline [9], but the effect size is small or negligible. **Bolded** values show the lowest percentage of training data with which ALCATRAL achieves results that are comparable to the TRAIL baseline [9] (no statistically significant difference or a low or negligible effect size). Underlined values indicate the best result overall for a dataset (for a line).

- ALCATRALは学習データの30%のみを使用して、
TRAILとの誤差3%未満の精度を発揮する

RQ2結果

※F値による評価

TABLE III
PERFORMANCE OF AL-TRAIL WITH ITERATIVE TRAINING

Dataset	Initial 10%	Size of the training set																		TRAIL baseline as in [9]
		20%		30%		40%		50%		60%		70%		80%		90%				
		ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL	ACL	TRL			
eAnci	52.2	72.8*	61.4	78.7*	65.9	<u>78.8*</u>	69.0	78.8*	71.7	78.6*	73.6	78.3*	75.1	78.2* [‡]	76.7	78.1 [‡]	78.0	77.9		
eC-CC/UC	45.8	62.4*	53.9	66.1*	57.8	<u>67.7*</u> [‡]	60.1	67.3* [†]	62.4	67.6* [†]	63.3	67.4* [†]	65.0	67.1* [‡]	65.9	67.5 [†]	67.6	67.5		
eC-ID/CC	42.1	64.3*	51.6	70.5*	57.9	<u>72.4*</u> [†]	62.7	72.3* [†]	64.6	72.4* [†]	67.5	72.5* [†]	69.4	<u>72.6*</u> [†]	70.6	72.6 [†]	72.1	72.8		
eC-ID/TC	63.2	90.0*	75.6	<u>94.1*</u>	81.4	93.6* [‡]	85.1	93.6* [†]	87.6	93.6* [†]	89.4	93.5* [†]	91.3	<u>93.4*</u> [†]	92.4	93.5 [†]	93.4	93.4		
eC-ID/UC	25.3	52.0*	29.2	54.9*	39.9	53.7*	43.2	<u>55.5*</u> [‡]	47.1	56.6* [†]	49.0	57.3* [†]	51.6	56.7* [†]	53.9	57.2 [†]	57.8	57.6		
eC-TC/CC	73.8	95.2*	86.4	97.6* [‡]	91.5	<u>97.7*</u>	93.7	97.6* [†]	95.2	97.6* [†]	96.1	97.5* [†]	96.8	97.5* [†]	97.2	97.4 [†]	97.4	97.5		
eC-TC/UC	56.2	91.8*	69.4	94.9* [‡]	77.1	<u>95.3*</u>	81.9	95.2* [‡]	85.8	94.5* [†]	88.1	94.7* [†]	90.9	94.8* [‡]	92.9	94.6 [†]	94.7	94.5		
eTour	42.5	56.1*	49.6	58.6*	52.1	59.7*	54.4	60.1*	56.3	60.5* [†]	57.1	60.7* [†]	58.5	<u>61.1*</u> [†]	59.7	60.6 [†]	60.7	60.8		
iTrust	30.9	62.2* [‡]	47.9	63.1*	53.7	<u>63.4*</u>	57.7	62.1*	59.3	61.8* [‡]	60.0	60.7 [†]	60.5	<u>61.2[†]</u>	61.1	61.3 [†]	61.4	60.9		
MODIS	32.1	52.2*	42.3	59.5*	48.7	61.6*	53.1	62.7* [‡]	56.0	63.3* [‡]	59.2	<u>63.6*</u> [†]	60.3	64.0* [†]	62.4	64.1 [†]	63.8	64.3		
SMOS	47.6	62.9*	56.7	75.3*	62.3	79.9* [‡]	66.6	<u>80.0*</u> [‡]	70.4	79.4* [‡]	73.5	79.3*	76.0	79.6* [†]	77.9	79.7 [†]	79.8	79.7		
Average	46.5	69.3	56.7	73.9	62.6	74.9	66.1	75.0	68.7	75.1	70.6	75.0	72.3	75.1	73.7	75.2	75.1	75.2		

*=there is a significant statistical difference between ALCATRAL and TRAIL trained with the same amount of data. †=there is **no** significant statistical difference between ALCATRAL and the TRAIL baseline [9]. ‡=there is a significant statistical different between ALCATRAL and the TRAIL baseline [9], but the effect size is small or negligible. **Bolded** values show the lowest percentage of training data with which ALCATRAL achieves results that are comparable to the TRAIL baseline [9] (no statistically significant difference or a low or negligible effect size). Underlined values indicate the best result overall for a dataset (for a line).

• ALCATRALは同程度の学習データで学習したTRAILよりも
高い精度を出す。

• 70%のデータを用いると，100%のデータ量のTRAILと同等

まとめと所感

- まとめ

- TLRの自動識別タスクにおいてアクティブラーニングの有効性を示した
 - ・ 少ない精度の低下で、最大70%のデータ量の削減可能であることを示した

- 所感

- 既存手法を前提とする部分が多すぎて、前提部分の理解が難しい（背景とか特徴量の選択. . .）
- アクティブラーニングは他の識別タスクでも利用できそう